OF 77-549 HAS BEEN SUPERSEDED BY

PUBLICATION IN: Geology, v. 5, no. 8,

p. 464-466, 1977.

B.Weld
7/77

Date 6/23/77

t,

ons

tterer ______ for the

pen files:

TITLE: Application of linear statistical models of earthquake magnitude versus fault length in estimating maximum expectable earthquakes

AUTHOR(S): Robert K. Mark

CONTENTS: 15 p., ---- pls., 3 figs., ______ tables.

Map scale: ______

Depositories:

USGS Library, Room 4A100, 12201 Sunrise Valley Dr., Reston, VA 22092
USGS Library, 1526 Cole Blvd. at West Colfax Ave., Golden, CO
(Mail address: Stop 914, Box 25046, Federal Center, Denver, CO 80225)
USGS Library, 345 Middlefield Rd., Menlo Park, CA 94025
USGS, Room 504, Custom House, San Francisco, CA 94111
USGS, Room 7638, Federal Bldg., Los Angeles, CA 90012

Release date: JUNE 1977

Area: --

Report No. 77-549

(*) Asterisks indicate depositories holding reproducibles.

U.S. GEOLOGICAL SURVEY
Reston, VA 22092

Memorandum

Date 6/23/77

To: Branch of Plans and Program Management, Publications Division -

From: Chief, Office of Scientific Publications

Subject: New USGS open-file report

The following report was authorized by W. P. Ketterer for the Director on 6/22/77 for release in the open files:

TITLE: Application of linear statistical models of earthquake magnitude versus fault length in estimating maximum expectable earthquakes

AUTHOR(S): Robert K. Mark

CONTENTS: 15 p., ___ pls., 3 figs., ___ tables.

Map scale: ___

Depositories:

USGS Library, Room 4A100, 12201 Sunrise Valley Dr., Reston, VA 22092
USGS Library, 1526 Cole Blvd. at West Colfax Ave., Golden, CO
(Mail address: Stop 914, Box 25046, Federal Center, Denver, CO 80225)
USGS Library, 345 Middlefield Rd., Menlo Park, CA 94025
USGS, Room 504, Custom House, San Francisco, CA 94111
USGS, Room 7638, Federal Bldg., Los Angeles, CA 90012

Release date: JUNE 1977

Area: --

Report No. 77-549

(*) Asterisks indicate depositories holding reproducibles.

Application of linear statistical models of earthquake magnitude versus fault length in estimating maximum expectable earthquakes

Robert K. Mark

U.S. Geological Survey, Menlo Park, Calif.

Abstract

Correlation or linear regression estimates of earthquake magnitude from historic magnitude and length of surface rupture data should be based upon the correct regression. For example, the regression of magnitude on log(length) can be used to estimate magnitude, but the regression of log(length) on magnitude cannot. Regression estimates are most probable values and estimates of maximum values require consideration of one-sided confidence limits.

U.S. GEOLOGICAL SURVEY OPEN-FILE REPORT 77-549

JUNE, 1977

Scheduled for Publication in Geology, August, 1977

In estimating maximum expectable earthquakes, it is common practice to assume a maximum length of surface rupture (typically one half the fault length) and use 'lines of best fit' to estimate maximum magnitude from graphs comparing historic earthquake magnitudes and associated surface fault-rupture lengths. This note discusses the interpretation and use of linear regression or correlation models for making statistical inferences from data on historical events. For example, Bonilla and Buchanan (1970) report length of surface rupture (L) and Richter magnitude (M) for those earthquakes for which these data were available and presented "best fit" equations of the form log(L)=a+bM, that is, the linear regressions of log(length) on magnitude (fig. 1, line AA'). Other authors (eg. Tocher, 1958; Iida, 1965) have calculated regressions of magnitude on log(length) (fig. 1, lines BB' and CC').

I will argue that all these regression lines have been used incorrectly to estimate maximum earthquake magnitudes from maximum fault rupture lengths. That is, the wrong regression line (log(length) on magnitude) has been used to estimate magnitude from maximum rupture length, or regression estimates have been interpreted as maximum rather than most likely magnitudes (e.g. Greene and others, 1973; Wentworth and others, 1973; Wesson and others, 1974 and 1975).

A correlation model

Many models can be used to draw statistical inferences from the data on magnitude and length of rupture. A transformation to log(length) is used because it tends to normalize the data and to enhance the linear relationship. For the purpose of this discussion, a correlation model is postulated in which it is assumed that the magnitude - log(length) data are randomly drawn from the population of earthquakes with associated surface rupture and that such a population has a bivariate normal distribution (fig. 2). As indicated below, these assumptions are more restrictive than necessary. As indicated in figure 2, the regression line of Y on X, or $Y=\alpha+\beta X$, passes through the most probable value of Y for each X and is the appropriate line to estimate Y given X. The other regression line, the regression of X on Y, passes through the most probable value of X for each Y and will not provide an unbiased estimate of Y given X.

Thus, the line of Bonilla and Buchanan in figure 1 is not the correct regression line for estimating earthquake magnitude from fault length. Rather, the appropriate regression of magnitude on log(length), calculated using their strike-slip fault data, is line DD' (fig. 1). It is similar to the equivalent regression lines of the other authors.

Estimation of maximum earthquake magnitudes

The regression lines of magnitude on log(length) can be used to estimate the most likely magnitude for a given maximum rupture. It must be stressed that such an estimate is <u>not</u> a maximum magnitude, but rather the magnitude that could be expected to be exceeded in 50% of the earthquakes associated with that rupture length.

It is possible to use the statistical model to estimate the magnitude, as a function of length, that could be expected to be exceeded in a given fraction $1-\alpha$ of surface-rupture occurrences, using a one-sided confidence limit (Wonnacott and Wonnacott, 1972, p. 280):

$$M_\alpha(L) = M(L) + t_{1-\alpha}\; S\left(\frac{1}{n} + 1 + \frac{(LOG(L) - \overline{LOG(L)})^2}{\sum_{i=1}^{n}(LOG(L_i) - \overline{LOG(L)})^2}\right)^{1/2}$$

where M(L) is the regression value, $t_{1-\alpha}$ is the critical value of the t distribution with (n-2) degrees of freedom, and s is the standard error of the regression. That is, the curve $M_\alpha(L)$ is the locus of points such that for a particular L, $1-\alpha$ is the probability that the magnitude will exceed M_α. Note that the regression line M(L) is equivalent to $M_{0.5}(L)$.

As an example, Bonilla and Buchanan (1970) report data on strike-slip faults (n=20) and calculate the regression line (L in meters)

$\log(L)=1.915+0.389M$ $\quad r=.70 \quad s=.52$

The regression of M on log(L) yields

$M=1.235+1.243\log(L)$ $\quad r=.70 \quad s=.93$

These lines are plotted in figure 3, along with the data points. Also plotted are the curves $M_{0.75}$ and $M_{0.95}$ for the regression of M on log(L). A magnitude value from the regression line M(L) can be referred to as the most likely magnitude for a given rupture length, and a value from $M_{\alpha}(L)$ as a maximum expectable earthquake magnitude at exceedance probability $1-\alpha$.

The line EE' on figure 3 connects the points that form the right-side envelope of the data. This field lies entirely to the left of $M_{0.95}$, and on the basis of the model, there are potential events larger than EE' that have probabilities in excess of 5%.

The preceding numerical results are somewhat model dependent, in that they depend on the population distribution and sample selection, but the general implications have wide application. Estimates of most likely earthquake magnitudes for a given value of an 'independent variable' (such as rupture length or fault displacement) must be based on the correct regression, and estimates of 'maximum magnitude' require consideration of the distribution about the regression line and the application of one-sided confidence limits.

These results can also be derived from a less restrictive linear regression model in which log(L) is treated as an independent variable and M is assumed to be normally distributed about the regression line (M on log(L)) with variance independent of L (Hays, 1973, ch. 15). If the data warrant, these models could be expanded to include additional 'independent variables' such as tectonic setting and hypocentral depth. A statistical approach is also needed to estimate the maximum surface rupture (at some exceedance probability) for a given total fault length.

REFERENCES

Bonilla, M. G., and Buchanan, J. M., 1970, Interim report on world wide historic surface faulting: U. S. Geol. Survey open-file rept., 32 p.

Greene, W. H., Lee, W. H. K., McColloch, D. S., and Brabb, E. E., 1973, Faults and earthquakes in the Monterey Bay region, California: U. S. Geol. Survey text to accompany map MF-518, 14 p.

Hays, W. L., 1973, Statistics for the social sciences: New York, Holt, Rinehart, and Winston, 954 p.

Iida, Numizi, 1965, Earthquake magnitude, earthquake fault and source dimensions: Nogoya Univ. Jour. Earth Sci., v. 13, p. 115-132.

Tocher, Don, 1958, Earthquake energy and ground breakage: Seismol. Soc. America Bull., v. 48, p. 147-153.

Wonnacott, Thomas H., and Wonnacott, Ronald J., 1972, Introductory statistics for business and economics: New

York, Wiley, 622 p.

Wentworth, C. M., Bonilla, M. G.. and Buchanan, J. M., 1973, Seismic environment of the Burro Flats site, Ventura County, California: U. S. Geol. Survey open-file rept., 35 p.

Wesson, R. L., Helley, E. J.. Lajoie, K. R., and Wentworth, C. M., 1975, Faults and future earthquakes, _in_ R. D. Borcherdt, ed., Studies for seismic zonation of the San Francisco Bay region: U. S. Geol. Survey Prof. Paper 941-A, p. A5-A30.

Wesson, R. L., Page, R. A., Boore, D. M.. and Yerkes, R. F., 1974, Expectable earthquakes in the Van Norman Reservoirs area: U. S. Geol. Survey Circ. 691-B, 9 p.

ACKNOWLEDGMENTS

Reviewed by D. R. Dawdy, D. G. Herd, R. A. Page, and D. M. Perkins.

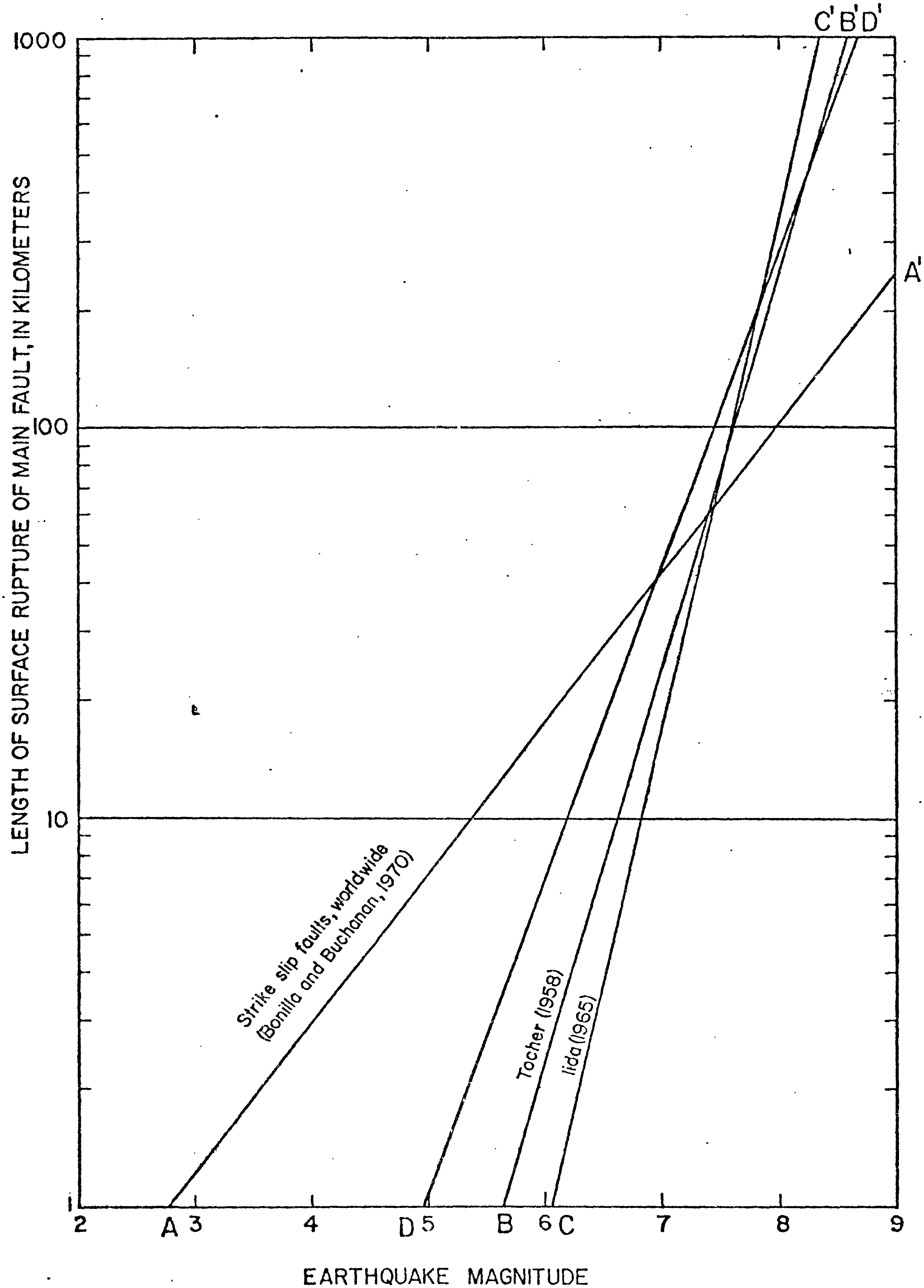
FIGURE CAPTIONS

Figure 1. Length of observed surface rupture in relation to earthquake magnitude. Line AA' is a regression line of log(length) on magnitude. Lines BB', CC', and DD' are regression lines of magnitude on log(length). Lines AA' and DD' are based on the same data.

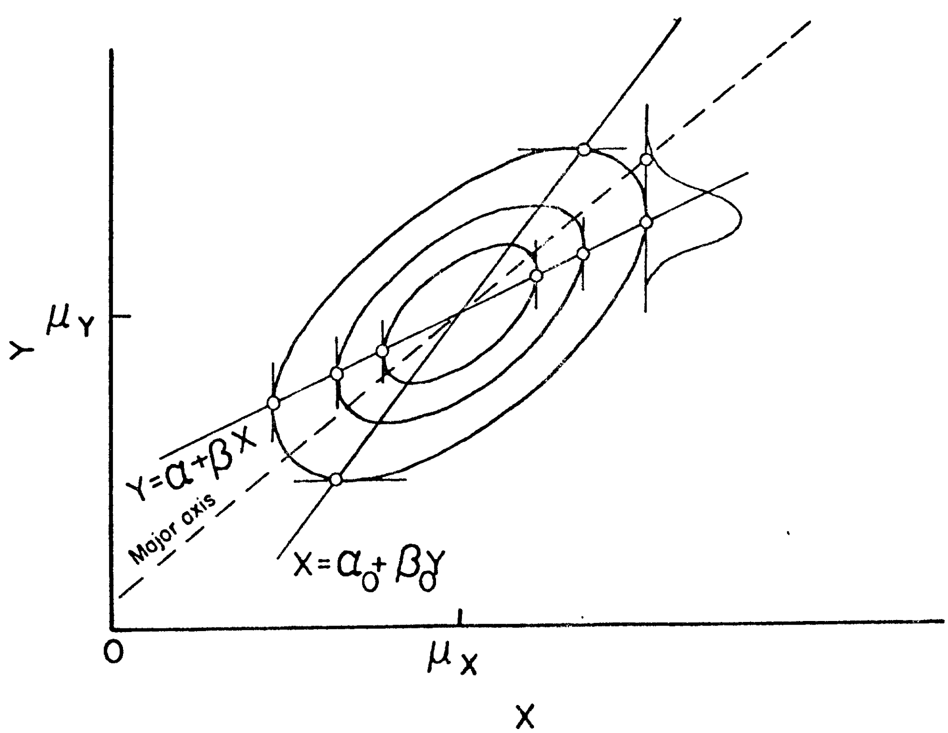
Figure 2. The two regression lines in a bivariate normal population. The contours are equal probability density. Modified from Wonnacott and Wonnacott (1972).

Figure 3. Length of observed surface rupture in relation to earthquake magnitude for the strike-slip fault data of Bonilla and Buchanan(1970). Line AA' is the regression line of log(length) on magnitude and could be used to estimate the most likely rupture length associated with a given magnitude earthquake. Line BB' is the regression line of magnitude on log(length) and could be used to estimate the most likely earthquake magnitude associated with a given

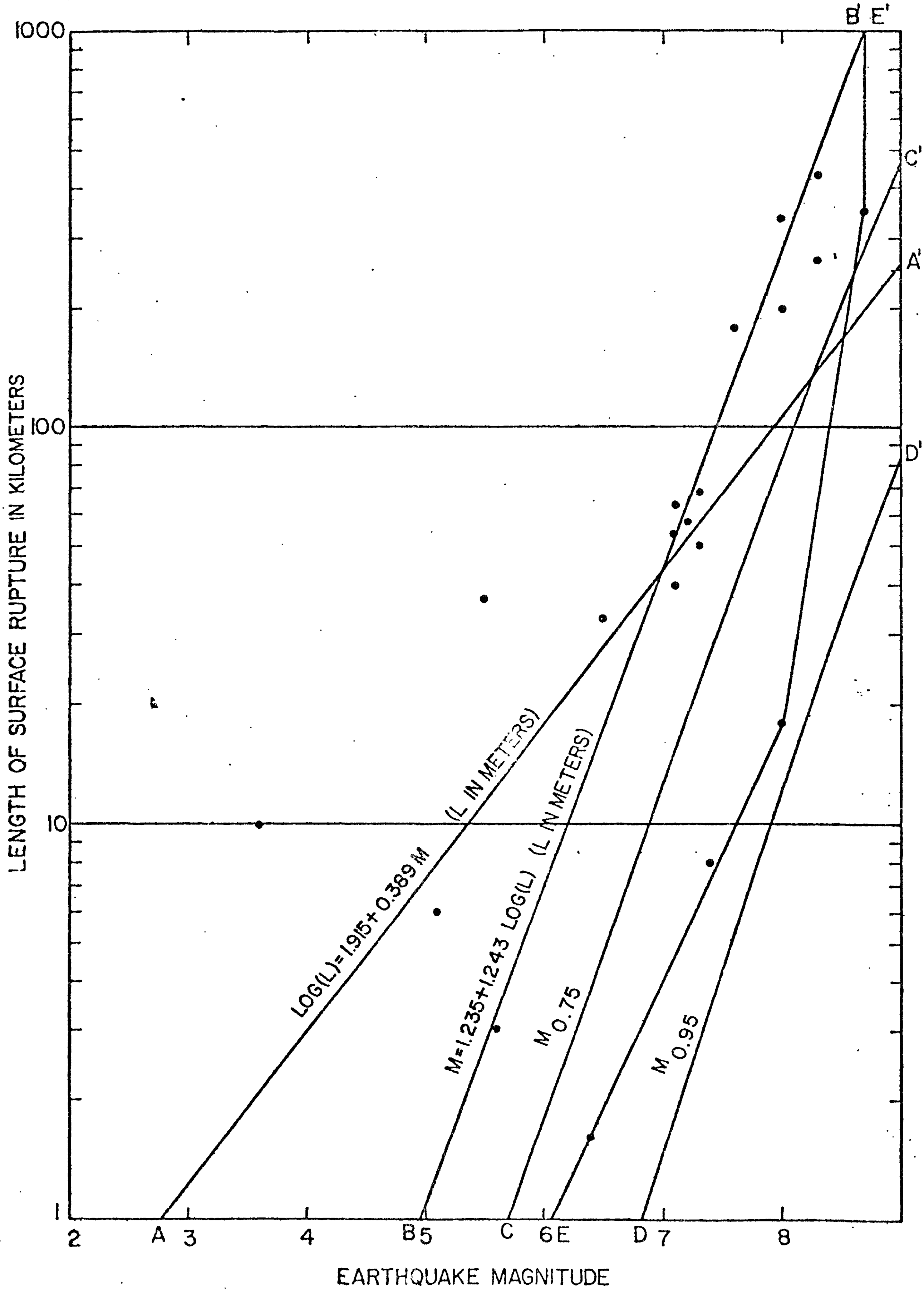
length of surface rupture. On the basis of the correlation model, half the earthquakes associated with a given length of surface rupture could be expected to be larger than BB'. The magnitudes given by curve DD' could be expected to exceed 95% of the earthquakes associated with a given length of surface rupture. The broken line EE' is the right-side envelope of observed data.



C' B' D'
1000
100
10
1
A'
LENGTH OF SURFACE RUPTURE OF MAIN FAULT, IN KILOMETERS
Strike slip faults, worldwide
(Bonilla and Buchanan, 1970)
Tocher (1958)
Iida (1965)
2
A 3
4
D 5
B 6 C
7
8
9
EARTHQUAKE MAGNITUDE



μ_Y
Y
$Y = \alpha + \beta X$
Major axis
$X = \alpha_0 + \beta_0 Y$
0
μ_X
X



LENGTH OF SURFACE RUPTURE IN KILOMETERS
EARTHQUAKE MAGNITUDE
1000
100
10
1
2
3
4
5
6
7
8
A
B
C
D
E
A'
B'
C'
D'
E'
LOG(L)=1.915+0.389 M (L IN METERS)
M=1.235+1.243 LOG(L) (L IN METERS)
M 0.75
M 0.95