7th Canadian Conference on Earthquake Engineering / Montreal / 1995
7ième Conférence canadienne sur le génie paraséismique / Montréal / 1995

A Seismicity Model for Eastern Canada

L.E. Chouinard and M. Diwani[1]

ABSTRACT

Seismicity in intra-plate regions, such as Eastern Canada, is usually estimated for seismic sources specified by experts drawing upon a general knowledge of historical seismicity and seismo-tectonic characteristics. The usual assumptions associated with these sources are that main earthquakes can be modelled as an homogeneous Poisson process in space and in time, and that the magnitude of events is exponentially distributed. A model is presented which relaxes all the above assumptions. The model can be used for exploratory data analysis with only historic earthquake data, or in a more conventional way with seismic sources defined by an expert.

INTRODUCTION

The calculation of seismic hazard at a given site requires knowledge of the recurrence laws of the nearby earthquakes sources. The statistical estimation of such laws from past activity presents however several problems:

1. Earthquakes are typically reported on different magnitude scales and a conversion to a single scale is necessary;
2. Seismic data display a considerable amount of clustering which is contrary to the common assumption of Poisson events;
3. The reporting of historical events is incomplete, equally for low magnitude and early periods;
4. The short historical data often do not support the hypothesis of homogeneous seismicity within large geographical regions;
5. The reported epicentral locations and magnitudes are subject to estimation errors.

In this paper, a method of recurrence law estimation is presented which accounts for catalog incompleteness, regards seismicity as spatially varying, and makes appropriate corrections for reporting errors. Procedures for magnitude conversion and earthquake clustering has been discussed elsewhere (Van Dyck, 1986, Van Dyck and Veneziano, 1987). It is assumed here that such procedures have been applied to the historical catalog and that dependent events, including foreshocks and aftershocks, have been removed. Therefore, earthquake size is measured in a

---

[1]Assistant Professor and Graduate Student, Department of Civil Engineering and Applied Mechanics, McGill University, Montreal, Quebec, H3A 2K6.

single magnitude scale (m) and the sequence of events is reasonably Poisson.

The main innovations of the proposed method are best illustrated through comparison with standard procedures, which differs little from the analysis first prepared by Cornell (1968). In traditional analysis, the regions of interest is divided into homogeneous Poisson sources, also referred to as seismogenic provinces. Such provinces are identified judgementally, based in part on historical data, in part on seismotectonic principles. The rate at which earthquakes are generated in each source is often assumed to depend exponentially on magnitude, so that for the i^{th} source.

$$\nu_i(m)=10^{a_i-b_i m}, m_o \le m \le m_u \tag{1}$$

where ν_i: the recurrence rate density, (rate per unit time, unit magnitude, and unit area) and a_i and b_i are source-specific parameters. Frequently, the magnitude is limited by an upper-bound m_{u} and the recurrence model is considered valid above a specified earthquake size m_0. The parameters a_i and b_i are estimated using only the complete portion of the historical record. Recognizing that the time interval of complete reporting depends on magnitude, Stepp (1992) has proposed to estimate the period of completeness T(m) based on the stability of the observed rate of events with magnitude close to m, as a function of the period of observation. Weichert (1980) has developed a maximum likelihood method for the estimation of a_i and b_i that accounts for unequal periods of completeness.

The procedure outlined above suffers from three main deficiencies: 1. estimation of a_i and b_i are based only on the complete portion of the catalog; 2. the method does not recognize that the quantification of incompleteness, the estimation of the recurrence laws, and the correction for estimation errors are mathematically coupled problems; and 3. the model with homogeneous seismogenic provinces may not be compatible with the historical data. Each of these aspects is discussed next in greater detail. Incompleteness of the earthquake catalog is a major concern in the estimation of recurrence rates. Unless recognized and properly modeled, incompleteness introduces bias in the seismicity parameters a and b and spurious spatial variations of seismicity, if the rate of reporting itself varies on the geographical plan. Current procedures avoid these problems by limiting the data used for rate estimation to the period when the catalog is judged to be complete. Spatial variation of incompleteness is sometimes accounted for by specifying a different function T(m) for each earthquake source or group of sources. One limitation of this procedure is that incompleteness is difficult to estimate, for large magnitudes because the number of events is small and for low magnitudes because even below the catalog may be incomplete. These problems are equally severe in regions of low seismicity.

The second problem with current methods is the sequential estimation of incompleteness and rates. It is clear that the two are interdependent: for example, if the rates were known, it would be simple to estimate incompleteness. A third and in practice more important limitation of current statistical models is their inability to represent the spatial variation of severity other than by partitioning the region of study into homogeneous sources. Extended regions inside which the historical data would pass a statistical test of homogeneity are seldom found. Partitioning the region of study into very many sources would seem to be a solution but doing so severely limits the accuracy of the estimates, especially for the parameters b_i in Equation 1 (Bender, 1984).

The models and estimation methods discussed in this paper attempt to reduce the three aforementioned problems as follows: the entire historical data set can be used by explicitly modelling incompleteness. The method can simultaneously estimate incompleteness of the data and recurrence rates. The influence of reporting errors is also accounted for. The seismicity model allows the recurrence parameters to vary smoothly in space, either within predefined "seismosimilar" provinces or over the entire region of study. The user can specify prior distributions on the parameters and can fit the truncated exponential recurrence law over a selected range of magnitudes (more in general, different weights can be assigned to different values of m). The latter capability is useful when the exponential model is not exact but is sufficiently accurate if fitted locally, over the magnitude range of interest.

PROBABILISTIC MODEL OF SEISMICITY AND INCOMPLETENESS

After conversion to a single magnitude scale, the earthquakes in the catalog can be thought of as points in a multidimensional space $(\underline{x},t,m)$; for earthquake i, $\underline{x}_i$ is the epicentral location, t_i is the time of occurrence, and m_i is the magnitude. The probability model is as follows: Earthquakes are assumed generated by a Poisson process which is stationary in time and has rate density $\nu(\underline{x},m)$ in space and magnitude. This function is defined such that $\nu(\underline{x},m)\, d\underline{x}\, dm\, dt$ is the expected number of earthquakes in the neighbourhood ($d\underline{x}$ dm, dt) of $(\underline{x},m,t)$. Nonstationarity of the observed earthquake sequence is attributed to incomplete reporting and incompleteness is modelled through a function $P_D(x,t,m)$, which gives the rate density of events reported. If one assumes independence in the reporting of different events, then the sequence of historical earthquakes observed is a realization of a Poisson process in $(\underline{x},t,m)$ space with rate density,

$$\nu(\underline{x},m)=A(\underline{x})\cdot e^{a(\underline{x})-b(\underline{x})m},\quad m=m_o,\ldots,m_{1\underline{x}} \tag{2}$$

Here, it is assumed that ν varies nonparametrically in space but exponentially in magnitude. The variables are discretized with $\underline{x}$ and m that denote respectively cells in space and constant-amplitude intervals of magnitude where ν is now the expected number of events per unit time inside the cell associated with $\underline{x}$ and the magnitude interval centred at m, $a(\underline{x})$ and $b(\underline{x})$ are functions to be estimated, and $A(\underline{x})$ is the area of the geographical cell around $\underline{x}$. If the geographical cells are large (e.g., if they are associated with extended regions or very active sources), then estimation of $a(\underline{x})$ and $b(\underline{x})$ may rely on local seismicity. However, as the discretization of the geographical plane becomes more detailed, the number of parameters increases and some constraint on the functions $a(\underline{x})$ and $b(\underline{x})$ becomes necessary to obtain reliable results. This can be done by limiting the roughness of the functions. For example, a convenient measure of the roughness of $a(\underline{x})$ is

$$P_a=\sum_{\underline{x}}[a(\underline{x})-\hat{a}(\underline{x})]^2 \tag{3}$$

where $\hat{a}(\underline{x})$ is an estimate of $a(\underline{x})$ based on the value of the function at neighbouring locations. The neighbouring locations are either fixed to the eight neighbouring cells, or are selected within the eight surrounding cells on the basis of a test of Poisson homogeneity (Chouinard 1989, Veneziano and Chouinard 1987). A nonparametric representation of the probability of detection is to divide time, space, and magnitude into intervals over which P_D may be considered constant. If P_D does not vary rapidly in space and time, large geographical regions and long time intervals can be used in the discretization.

PENALIZED LIKELIHOOD FORMULATION

The criterion of estimation is maximization of a likelihood function, penalized and modified to include prior information on the parameters and to produce "smooth" solutions. Under the previous assumptions, the number of events n_c for a combination $c=\{\underline{x},t,m\}$ of the discretized variables has Poisson distribution with probability mass function

$$f(n_c)=\frac{E[n_c]^{n_c}}{n_c!}e^{-E[n_c]} \tag{4}$$

The mean value $E[n_c]$ depends on the unknown parameters $a(\underline{x})$ and $b(\underline{x})$ in Equation 4 and on the probability of detection P_D

$$E[n_c]=T_c\cdot P_D\cdot e^{a(\underline{x})-b(\underline{x})m} \tag{5}$$

where T_c is the volume in $(\underline{x},t,m)$ space associated with category c and P_D is the corresponding detection probability. Depending on the discretization of $\underline{x}$, m, and t used to model $a(\underline{x})$, $b(\underline{x})$, the likelihood may be a function of very many parameters, which therefore can be estimated with less accuracy. The solution can be improved by introducing constraints, in the form of multiplicative penalty terms on the original likelihood function. Such terms have a dual interpretation: in a Bayesian series, they represent prior knowledge on the parameters. Alternatively and more pragmatically, the same terms can be viewed as penalties or preference factors over the various possible solutions. The penalties are used here to impose smoothness to the solution.

The MPL conditions form a system of coupled nonlinear equations, with a large number of unknown seismicity and incompleteness parameters. A numerical technique must therefore be used for their solution. An important feature of the present model is that following discretization of seismicity and incompleteness, minimal assumptions are made so that the estimates are almost exclusively data based. Given the small amount of earthquake data typically available and the large number of parameters to be estimated, the uncertainty on the estimates may however be large. For this reason, prior information if available should be used. Contrary to traditional methods where such information is judgementally embedded into the model (through the notions of seismogenic provinces and periods of completeness), the present method provides a more general framework where prior information is explicitly stated and combines with historical data to obtain best estimates. Through examination of the goodness-of-fit of the resulting model, the compatibility of a priori statements with data can be verified and, if judged necessary, such statements can be modified.

The parameter b in Equation 2 is often considered to be less variable in space than the parameter a. This opinion may be influenced by physical interpretations of b, by worldwide observations (e.g., Utsu 1969, 1979, 1971), or by the statistical consideration that a large number of earthquakes is necessary to obtain reliable estimates of b (Bender, 1983).

In some cases, information exists on the value of a(x) and b(x), other than in the historical data. For example, one might find the value of b in other regions to be relevant to the estimation of that parameter in the region of interest. In other cases, local information on seismicity comes form tectonic models, geologic characteristics, on paleoseismicity. The inclusion of prior knowledge in the form of independent normal distributions N(b(x), σ_b^2(x)) is also easy: the maximum likelihood equations should in this case include the additive term

$$-\frac{1}{\sigma_b^2(x)}[b(x)-\hat{b}(x)] \qquad (6)$$

GOODNESS-OF-FIT AND UNCERTAINTY OF THE ESTIMATORS

Examination of the goodness-of-fit is of importance to validate the assumption underlying a model to detect possible deficiencies, and to compare the relative performance of alternative models. Uncertainty on the parameters is of concern in the prediction of future earthquake activity, hence in the evaluation of seismic hazard. Both problems are complex and mainly the result of two characteristics of the data and the models:

1. The data is sparse and prohibits the use of asymptotic properties of usual goodness-of-fit statistics or asymptotic expressions for maximum likelihood estimators.
2. The estimated parameters can be strongly dependent due to the use of smoothing, constraints and other a-priori conditions. Consequently, the number of degrees of freedom, which are necessary to judge the usual goodness-of-fit statistics, are not well defined and the likelihood function, which play a key role in calculating uncertainty of the estimators, have a complicated form.

A simple Poisson test is useful for this purpose. Given that the expected count in a certain category i is n_i, the probability that the Poisson count N_i does not exceed the observed value n_i is

$$P[N_i<n_i]=\sum_{k=0}^{n_i}\frac{E[n_i]^{n_i}e^{-E[n_i]}}{n_i!}=\alpha_i \qquad (7)$$

Very low and very high values of α_i indicate that the expected count is too high or too low respectively. It should be emphasized that no strict interpretation must be given to α_i, because the expected count used in the test is data dependent. The true α_i is more extreme than the calculated

α_i. These approximate significant levels can be used to compare model predictions with observations in an intelligible way, by flagging categories i associated with very low or very high values of α_i. Both the fraction of cells that are flagged and the pattern of flagging should then be examined.

Finally, two data-based procedures have been developed for the selection of the optimal penalties on a($\underline{x}$) and b($\underline{x}$). The first is based on the distribution of some target statistics and the second is based on the optimization of cross-validated measures of goodness-of-fit (Chouinard 1989, Veneziano and Chouinard 1987).

SEISMICITY OF EASTERN CANADA

Estimation of the parameters a_k and b_k in each province is relatively straightforward if only earthquake data within the periods of completeness are used. Seismogenic provinces are typically identified by experts based on an analysis of the historical seismicity and the geological and tectonic setting of the region. However, there is a lot of uncertainty on the exact configuration of these zones which gives rise to many competing alternatives.

The model can optionally be used with or without the external specification of sources. The effect of the latter are not as severe on the estimates of the seismic hazards as in the case of traditional seismic source estimates because of the smooth variation of the parameters inside each source. An alternative to the specification of sources by experts is to identify statistically homogeneous regions through hypothesis testing. The hypothesis testing is not used to estimate the location and size of seismogenic provinces but to define the size of the cells over which an average is computed for the penalized likelihood. When these are used, significant discontinuities in seismicity are preserved.

The earthquake catalog used is the one compiled in the context of EPRI (1985) and the region analyzed is between longitudes 62°W-80°W and latitudes 41°N-52°N. Earthquakes identified as aftershocks in the catalog are removed prior to the analysis. The spatial dsicretization for this application is half degree cells and the probabilities of detection are fixed to those obtained in the context of EPRI (1985). Magnitude is discretized into 0.6 unit intervals from 3.3 to 7.5. Maximum magnitude has been set equal to 7.5 everywhere for the estimation of seismicity. For all regions and for magnitudes greater than 3.3, the catalog is assumed complete since 1975, which is the time when the instrument network was considerably improved.

Estimates of a($\underline{x}$) and b($\underline{x}$) are shown in Figure 1 for given penalties on a($\underline{x}$) and b($\underline{x}$). Note that significant discontinuities in the rate function are preserved even for large penalties. In particular, Charlevoix, the Ottawa River Valley, and Newburyport are identified as three regions with significantly higher seismic activity.

CONCLUSIONS

The proposed model is one which allows deviations of the model of seismicity from the traditional assumptions of a homogeneous Poisson process, and exponential distribution for the intensity of earthquakes within a source. The model can be used with or without the external specification of sources. When used without sources, the model can be used to identify statistically significant features in the historical data set in terms of size and magnitude. This information can then be used directly in seismic hazard estimates or in conjunction with other pertinent information in the formulation of a seismicity model.

ACKNOWLEDGEMENTS

The authors would like to acknowledge the financial support of the National Science and Engineering Research Council of Canada. The original project was conducted at M.I.T under the supervision of Professor D. Veneziano and the collaboration of Dr. J. VanDyck and the lead author under a grant from EPRI.

REFERENCES

Bender, B. (1983) Maximum Likelihood estimation of b Values for Magnitude Grouped Data", Bulletin of the Seismological Society of America, 73(3), 831-851.

Chouinard, L.E. (1989) ' Estimation of Seismicity Models in Regions of Intraplate Seismicity", D.Sc. Thesis, Department of Civil Engineering, Massachusetts Institute of Technology, Cambridge, Massachusetts.

Cornell, C.A. (1968) "Engineering Seismic Risk Analysis", Bulletin of the Seismological Society of America, 54(5), 1583-1606.

Stepp, J.C. (1972) "Analysis of the Completeness of the Earthquake Sample in the Puget Sound Area and its Effect on Statistical Estimates of Earthquake Hazard", Proceedings , International Conference on Microzonation, 2, 897-910.

Utsu, T. (1969, 1970, 1971) "Aftershocks and Earthquake Statistics, Parts I-III",Journal of the Faculty of Science, Hokkaido University, Ser. VII, Vol, III, No. 3 (1969(, NO. 4 (1970), and No. 5 (1971).

VanDyck, J. (1985) " ", Ph.D. thesis, Department of Civil Engineering, Massachusetts Institute of Technology, Cambridge, Massachusetts

Veneziano, D. and Chouinard, L.E. (1987) "Local Models of Seismicity and their Estimation", Proceedings Symposium on Seismic Hazards, Ground Motion, Soil Liquefaction in Engineering Practice in the Eastern United States", Sterling Forest, New York.

Veneziano, D. and VanDyck, J. (1987) "Statistical Analysis of Earthquake Catalogs for Seismic Hazard", in Y.K. Lin and R. Minai (Eds.), Stochastic Approaches in Earthquake Engineering, Lecture Notes in Engineering, New York, Springer-Verlag.

Weichert, D.H. (1980) "Estimation of the Earthquake Recurrence Parameters for Unequal Observation Periods for Different Magnitudes", Bulletin of the Seismological Society of America, 70, 1337-1346.

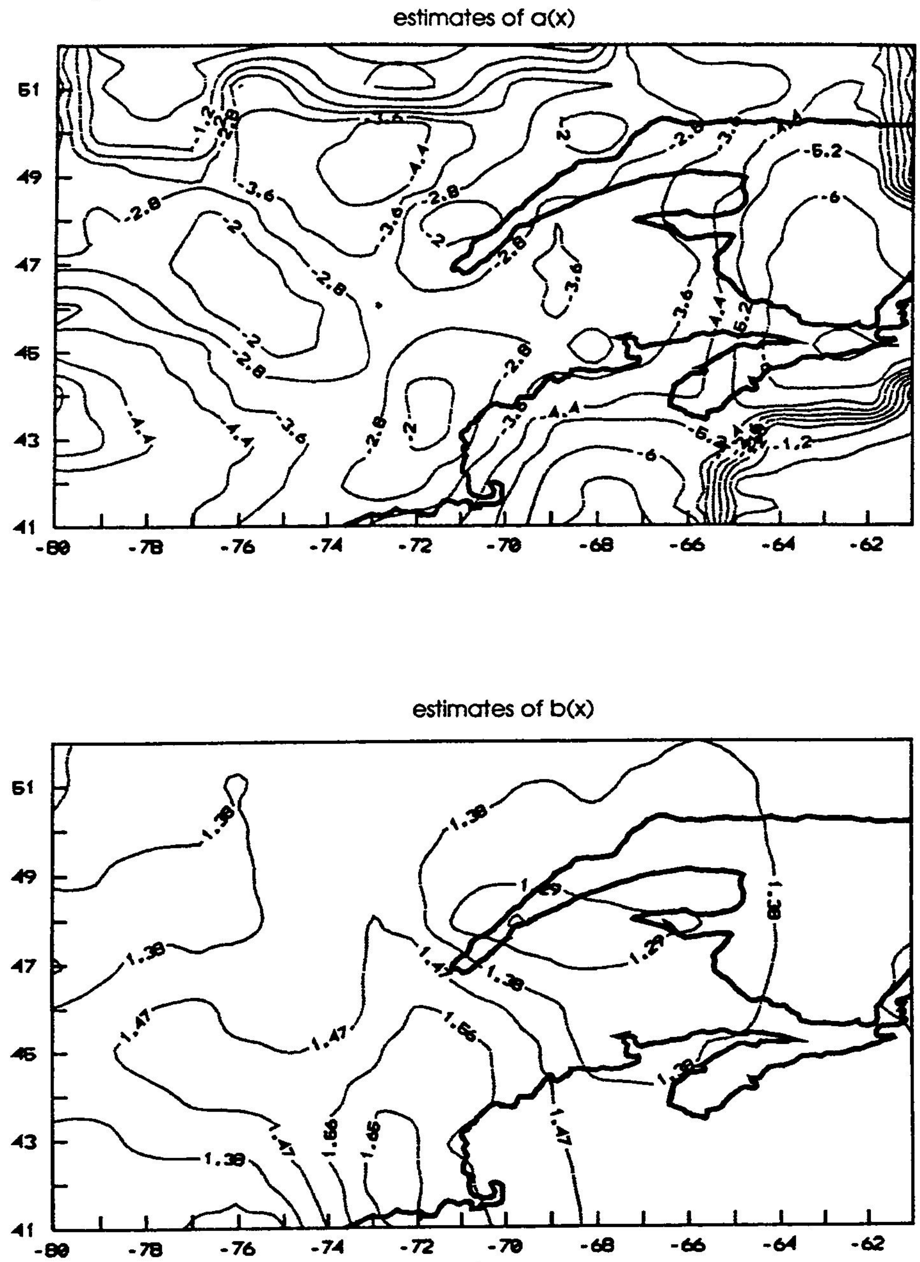

Figure 1: Estimates of a(x) and b(x) for Pa=1.0 and Pb=50.