doi:10.26443/seismica.v2i2.384

Assessing Earthquake Rates and *b*-value given Spatiotemporal Variation in Catalog Completeness: Application to Atlantic Canada

Alexandre P. Plourde * [1]

[1]Geological Survey of Canada (Atlantic), Natural Resources Canada, Dartmouth, Nova Scotia, Canada

Abstract Spatiotemporal variations in the magnitude of completeness M_c make it challenging to confidently assess seismic hazard or even to simply compare earthquake rates between regions. In this study, we introduce new techniques to correct for heterogeneous M_c in a treatment of the eastern and Atlantic Canada earthquake catalog (1985–2023). We first introduce new methodology to predict $M_c(x,t)$ based on the distribution of seismometers. Second, we introduce a modified maximum-likelihood estimator (MLE) for b (the *b*-value) that accounts for spatiotemporal M_c variation, allowing the inclusion of more earthquakes. Third, we compute the ratio of detected/predicted $M{>}1$ earthquakes as a function of M_c and apply it to create a calibrated $M{>}1$ event-rate map. The resulting map has advantages over a moment-rate map, which is effectively sensitive only to the very largest earthquakes in the dataset. The new MLE results in a modestly more precise b when applied to the Charlevoix Seismic Zone, and a substantial increase in precision when applied to the full Atlantic Canada region. It may prove useful in future hazard assessments, particularly of regions with highly heterogeneous M_c and relatively sparse catalogs.

Non-technical summary Earthquake hazard assessments, and earthquake science in general, can be complicated by the uneven distribution of the seismometers used to detect earthquakes. This study examines the earthquake catalog from eastern and Atlantic Canada (from 1985 to 2023) and introduces new methods to deal with the uneven seismometer distribution. We first analyze what magnitude of earthquake we are able to detect as a function of location and time. Second, we introduce a new way to estimate the "*b*-value", which describes the ratio of the number of large earthquakes to small earthquakes. We apply the new method to the full map region and, separately, to the earthquake-dense Charlevoix Seismic Zone in Quebec. Finally, we produce an earthquake map that is calibrated for the historical distribution of seismometers. These methods may be useful in future earthquake hazard assessments, particularly for regions with highly-uneven seismometer coverage and low to moderate earthquake rates.

Production Editor:
Gareth Funning
Handling Editor:
Andrea Llenos
Copy & Layout Editor:
Kirsty Bayliss

Signed reviewer(s):
Maria Mesimeri

Received:
February 02, 2023
Accepted:
July 07, 2023
Published:
September 13, 2023

1 Introduction

The rate of earthquake occurrence in a given region is generally reported as either an event rate, a moment rate, or, most formally, with a Gutenberg-Richter (GR) model (Ishimoto and Iida, 1939; Gutenberg and Richter, 1944). An event rate is the simplest way to communicate earthquake density, it is the number of earthquakes per unit time, generally considering only those above some threshold magnitude. A moment rate sums the seismic moment of all earthquakes in the region, and is effectively only sensitive to the largest earthquakes in the region. GR models express the number of earthquakes as a function of magnitude $N(M)$ through a log-linear relation $\log_{10} N(M) = a - bM$, where the constant a describes the overall abundance of earthquakes and b (the b-value) describes the relative abundance of small earthquakes to large ones, and typically $b \approx 1$. This is a central equation to probabilistic seismic hazard assessments, and reliable estimates of a and b are therefore critical to seismic hazard analysis.

The magnitude-frequency distribution (MFD) in actual earthquake catalogs always deviates from the strict log-linear model. They are often characterized by a double-truncated GR model, which has an upper magnitude limit $M_{\max}$ based on the maximum fault size in the region, as well as a magnitude of completeness M_c, below which there will be fewer earthquakes detected than predicted because of our limited ability to detect them. M_c can be affected by noise conditions as well as geological factors that affect seismic attenuation, but it primarily depends on the distribution of seismometers in the region. High M_c increases uncertainty in hazard assessments and complicates even event-rate estimates. Knowledge of background rates of naturally-occurring earthquakes is critical to the responsible management of any activity that can pose risk of induced seismicity, such as hydraulic fracturing and wastewater injection in the oil and gas industry, as well as geological carbon storage (Schultz et al., 2020; Cheng et al., 2023).

This paper introduces new methodologies to calculate b and compare event rates across regions that account for spatiotemporal M_c variations. It will focus on

*Corresponding author: ap.plourde@dal.ca

eastern and Atlantic Canada, which generally has low to moderate levels of seismicity and seismometer coverage, but does contain several prominent onshore and offshore seismic zones.

2 Data and regional seismicity

This study examines the Canadian National Earthquake Database (CNED, see section "Data and code availability") from 1985–2023. We selected this timespan for two reasons: i) 1985 is the first year for which the catalog is publicly available, and ii) it is challenging to determine what seismometers were operational and used in routine earthquake detection at any given time before approximately the mid-1990s. Seismometer locations and operating dates were downloaded from the IRIS database, and supplemented with an annual publication on the network that ended in the late 1980s (Munro et al., 1988). Note that these sources indicate when stations became active, but not periods of time they were nonoperational. A map of epicenters and a magnitude-time plot are shown in Figure 1, and Figure 2a shows the location of all stations that may have contributed to the catalog.

The catalog contains 13612 earthquakes, with a variety of magnitude types. We follow the procedure of Halchuk et al. (2015) to convert all magnitudes to M_W; the method is partly based on the m_N–M_W relation computed by Bent (2011), but does not follow it directly. We note that the linear scaling suggested by Bent (2011) would directly affect b estimates; more details on the treatment of magnitudes is found in Appendix 1. The magnitude-time plot (Figure 1b) shows a drastic change around 1995, when the seismograph network and detection routines underwent several major changes (Bent, 2011). It also has notably few events in the range of approximately $0 < M_W < 1$ (the range is higher pre-1995 than post-1995), whereas there are lobes of relatively abundant events both above and below. The lower-magnitude lobe consists mainly of onshore events reported with only a local magnitude M_L, for which the conversion to M_W may be more dubious. Proper scaling of onshore M_L to M_W may warrant future investigation, but for this study we opt to simply ignore $M_W < 0$ for subsequent analyses, removing the vast majority of these onshore M_L events.

Following the approach taken in Canadian Seismic Hazard models (Kolaj et al., 2020), we do not attempt to decluster the earthquake catalog. Declustering is often performed as part of probabilistic seismic hazard analysis to remove foreshocks and aftershocks, such that the remaining earthquakes can be considered a Poisson process in time (Gerstenberger et al., 2020). However, declustering techniques require arbitrary thresholds to define what constitutes a foreshock or aftershock, and can cause unintended bias of b and hazard level estimates (Gerstenberger et al., 2020; Mizrahi et al., 2021). We also assume, given the vast geographic scale and generally low event rates and high M_c of this dataset, that temporal variation of M_c due to short-term aftershock incompleteness (Stallone and Falcone, 2021; van der Elst, 2021) is not a significant issue.

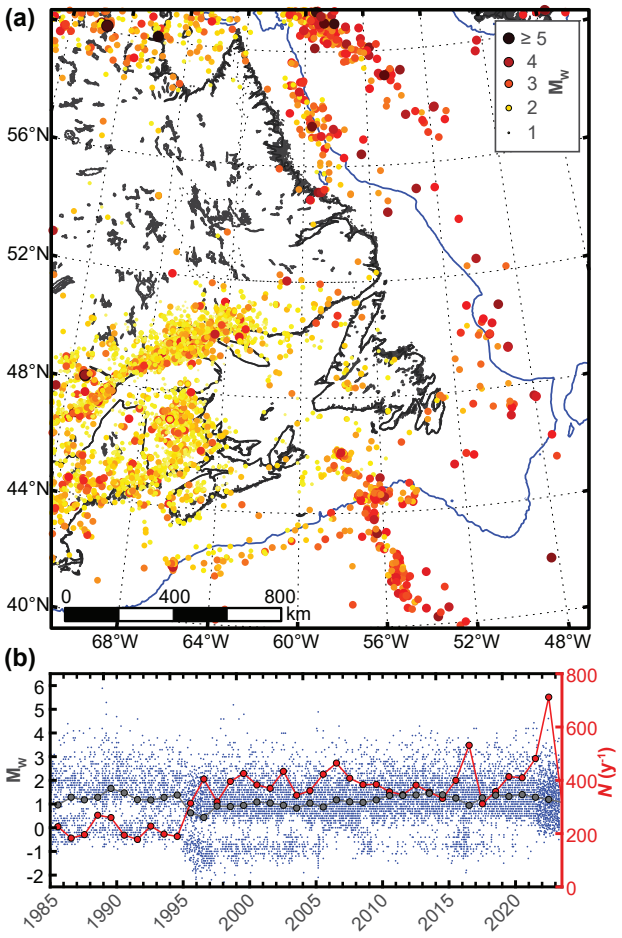

Figure 1 (a) Map of $M_W \geq 1$ earthquakes, with dots coloured and sized by magnitude. A 600 m bathymetric contour, corresponding roughly to the continental shelf edge, is plotted in blue. (b) Scatter plot of earthquake magnitudes in time (blue dots). The grey markers and line mark the mean magnitude by year. The red markers and line indicate the total number of earthquakes per year, corresponding to the second y-axis on the right side.

Map boundaries were chosen in order to include the Charlevoix Seismic Zone (CSZ, Lamontagne et al., 2003a; Yu et al., 2016) and Lower St. Lawrence Seismic Zone (LSZ, Lamontagne et al., 2003b; Plourde and Nedimović, 2021) in the west, as well as the less-studied seismic zones at the Laurentian slope and fan in the south (Adams and Basham, 1989; Bent, 1995), and in the the Labrador Sea to the north (Bent and Hasegawa, 1992; Bent and Voss, 2022). These areas are labeled in the earthquake density map in Figure 2b. The map includes all M_W>1 earthquakes and is significantly affected by spatial M_c variations. The CSZ and LSZ stand out as the most earthquake-dense regions on the map, with another prominent area in northern New Brunswick around the epicenter of the 1982 Miramichi M 5.7 event (Wetmiller et al., 1984).

Different trends emerge if we map moment-density of the same earthquake catalog (Figure 2c). The CSZ and LSZ are still prominent, but they have much more similar moment-rates to the offshore seismic zones than in the event-rate map. Spatial M_c variation biases moment-rates much less than event-rates, but the trade-

off is that moment-rate is controlled almost entirely by the largest, infrequent events, so it is more highly affected by the limited catalog duration. The highest mapped moment-rate results from the largest event in the catalog, the M_W 6.3 Ungava Bay earthquake of December 1989 (Bent, 1994), which falls in the northwestern corner of the map. The anomaly just east of it is a M_W 5.7 that occurred in March 1989. Another prominent anomaly lies just north of the CSZ, and results from the second-largest earthquake in the dataset, the 1988 Saguenay M_W 5.9 (Haddon, 1995).

3 Earthquake density mapping in Atlantic Canada

We assume that the MFD in Atlantic Canada follows a GR model allowing us to predict the distribution of undetected small earthquakes based on the distribution of larger earthquakes. This requires i) knowledge of M_c as a function of both space and time, i.e. $M_c(x,t)$, ii) a method for estimating b given spatiotemporal variations in M_c, and iii) a function to predict the ratio of undetected to detected earthquakes for a given $M_c(x,t)$. New methodologies to address these three issues are presented in the following subsections, along with results from their application to Atlantic Canada.

3.1 Estimating $M_c(x,t)$

In practice, it is challenging to estimate $M_c(x,t)$ precisely. M_c is typically estimated from GR plots either by visual inspection or using any number of algorithms (several popular ones are reviewed by Woessner and Wiemer, 2005), but that requires many earthquakes (typically hundreds or more). We resolve this by using a predictive $M_c(x,t)$ model based on the distribution of seismometers (Mignan et al., 2011), which uses the distance to the n^{th} closest seismometer $d_n = d_n(x,t)$. Mignan et al. (2011) produces empirical power-law models of the form $M_c^{\text{pred}} = c_1 d_n^{c_2} + c_3$, where c_i are constants, based on a dense earthquake catalog from Taiwan. They produce models for distance to the 3$^{\text{rd}}$-, 4$^{\text{th}}$-, and 5$^{\text{th}}$-nearest seismometer, each with similarly close fits. Here we form a similar model for our dataset, but to partially reduce the effect of temporarily nonoperational stations we define a (admittedly arbitrary) weighted station distance metric as a weighted sum of the distances to the 4$^{\text{th}}$–6$^{\text{th}}$ nearest stations:

$$d(x,t) = 0.70d_4 + 0.25d_5 + 0.05d_6. \quad (1)$$

We evaluate $d(x,t)$ for each earthquake in the dataset using the distribution of seismometers that were active when it occurred.

Earthquake density in our Atlantic Canada catalog is orders of magnitude lower than the Taiwan catalog used by Mignan et al. (2011)—it has about one tenth the number of earthquakes in a region 20 times larger. As such, there are few areas with enough earthquakes in a small radius (e.g. 50 km) to reliably estimate a, b, or M_c. We therefore need an alternative way to fit the power-law model, and take the following approach. We first compute $d(x,t)$, as defined above, for each earthquake given its origin time and epicentre. We then sort the earthquakes by their $d(x,t)$ and bin into groups of 300 with 50% overlap, resulting in 78 groups with maximum $d(x,t)$, or $d_{\max}$, of 21 to 1310 km. For each group of 300 earthquakes, we apply the method of Ogata and Katsura (1993) to estimate M_c. The method assumes the number detected earthquakes of a given magnitude $N(M)$ depends on the actual number $N_0(M)$ and a "thinning" function q:

$$N(M) = q(M|\sigma,\mu)N_0(M). \quad (2)$$

The thinning function is assumed to be a cumulative normal distribution function:

$$q(M|\sigma,\mu) = \frac{1}{\sigma\sqrt{2\mu}} \int_{-\infty}^{M} \exp\left(-\frac{(M-\mu)^2}{2\sigma^2}\right) dM, \quad (3)$$

where μ is the magnitude at which 50% are detected and σ describes the width of the thinning function; a low σ indicates a steep falloff in detection below M_c, whereas a high σ indicates a more gradual falloff. We use the nonlinear optimization toolbox of MATLAB® to find the set of b, σ, and μ that maximizes the log-likelihood function defined by Ogata and Katsura (1993) (see their Equation 8 or Si and Jiang, 2019 for details). $M_c(d_{\max})$ is then taken to be $\mu + 2.4\sigma$, i.e. the magnitude where we expect 99% of earthquakes are detected. We provide the optimizer limits on b, selecting $0.85 \leq b \leq 1.05$, as we found the output b to vary dramatically otherwise. Also, given that we threshold our catalog at $M_W \geq 0$, we set 0 as the lower integral bound in Eq. 3, rather than $-\infty$.

Results for three example distance bins with $d_{\max} =$ 23, 192, and 684 km are shown in Figure 3a–c, and the overall results shown in Figure 3d. Note that we attempted to estimate uncertainties of each M_c by repeating the Ogata and Katsura (1993) method in 200 bootstrap iterations (and these are shown in Figure 3). However, we find that systematic trends in the $M_c(d)$ data are more relevant to model fits rather than "noise" that is characterized by the bootstrap confidence intervals, so we opt not to use these uncertainties in the model-fitting process. The thinning width $\sigma(d)$ generally covaries with $M_c(d)$, although it has a prominent peak at d<100 km, where there is little range between M_c and the cutoff magnitude of M_W 0, so σ is less well constrained; this results in an overall σ–M_c correlation coefficient of 0.60.

The resulting $M_c(d)$ is poorly fit by a power law due to a bend to unexpectedly low M_c in the d range of ~100–500 km, and we therefore omit that distance range to compute the power-law model shown in Figure 3d (red curve). We compute an alternative, non-analytical $M_c(d)$ function as a best-fit smooth, continually increasing model (hereafter smooth-increasing, shown by the black dashed line in Figure 3d). The curve minimized an L1 data misfit $||M_c^{\text{pred}}(d) - M_c^{\text{meas}}(d)||_1$, plus a second-derivative smoothing term which was weighted subjectively by trial-and-error. Note that M_c^{pred} is the $M_c(d)$ predicted by the smooth-increasing model and M_c^{meas} is the input $M_c(d)$ as estimated with the Ogata and Katsura (1993) method. This model has

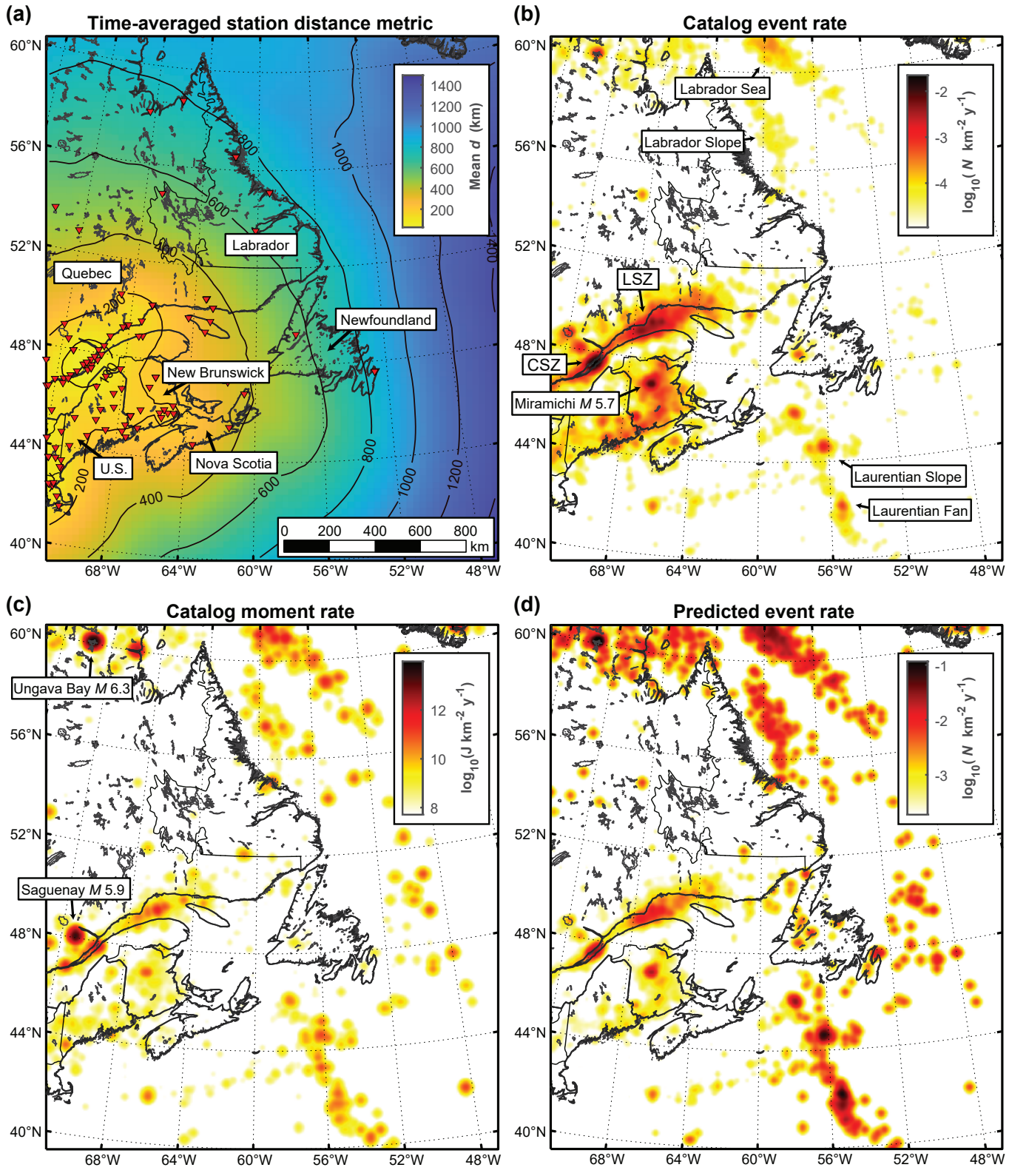

Figure 2 (a) Time-averaged weighted station-distance metric d over the study period of 1985–2023. Contours indicate distance in kilometres. Red triangles indicate locations of seismometers active for $\geq$3 years of the study period. Canadian provinces/regions and the United States (U.S.) are labeled. (b) Uncorrected yearly $M>1$ earthquake density (N km^{-2} y^{-1}) from the CNED catalog. (c) Moment density (J km^{-2} y^{-1}) for the same catalog. (d) The main result of this study (Section 3): Predicted yearly earthquake density based on the CNED catalog and the magnitude-of-completeness analysis of this study. All maps were first computed on a coarse grid ($\sim$15 km spacing), then converted to a finer grid and smoothed with a 2D Gaussian filter.

far greater freedom than the power-law, and as a result fits the data much more closely. For the remaining sections we use the smooth-increasing $M_c(d)$ model. However, we repeat the analyses using the power-law model and plot the results in Supplemental Figures S1–S4, which demonstrate that the choice has little impact on our overall conclusions.

3.2 Estimating b given spatiotemporal M_c variations

The standard and most-accepted way to estimate b is the Aki-Utsu maximum-likelihood estimator (MLE) (Utsu, 1965; Aki, 1965):

$$b = \frac{\log_{10}(e)}{\langle M \rangle - M_c}, \tag{4}$$

where $\langle M \rangle$ is the mean magnitude of the catalog, including only earthquakes with $M \geq M_c$. There have been several techniques introduced to allow time-varying M_c in the MLE (e.g. Weichert, 1980; Kijko and Smit, 2012; van der Elst, 2021). Taroni (2021) presented a convenient modification of the MLE that does not require the evaluation of b for subcatalogs. Ignoring two minor corrections, their MLE can be written:

$$b = \frac{\log_{10}(e)}{\langle M - M_{cE} \rangle}, \tag{5}$$

where M_{cE} is $M_c(t)$ evaluated for each earthquake. In this study, we further generalize their method by allowing spatial variation of M_c in addition to temporal variation, i.e. we consider $M_{cE} = M_c(x,t)$, evaluated using $d(x,t)$ for each earthquake. Derivation of Equation 5 is shown in Appendix 2, beginning from the magnitude probability density function of Aki (1965) and including an extension to incorporate a maximum magnitude (Page, 1968). The notion of considering the completeness level for each earthquake is not only useful in applying the MLE, but more generally we can examine the MFD using the relative magnitude $M^* = M - M_{cE}$.

Figure 4 displays regular M_W and M^* GR plots for both the CSZ and the full Atlantic Canada region, as defined by the map area in Figures 1 and 2. In both cases, M^*-derived b are lower than the M_W-derived estimate, but not significantly so according to the 95% confidence limits from bootstrapping. Confidence limits are highly dependent on the number of events included, and thus the M_c chosen. We therefore plot b vs. M_c (or b^* vs. M_c^*) for each GR plot (Figure 4c,d,g,h). These plots demonstrate that b^* is more stable over M_c^* than their equivalent M_W-derived estimates. Note that here we are ignoring potential variations of b to get average results over broad areas and times, despite observing varying b in the previous section.

3.3 Correcting density maps for undetected earthquakes

Here we apply our new MLE (Eq. 5) in order to estimate $r = r(M_c)$, defined as the ratio (total $M_W \geq 1$ earthquakes)/(recorded $M_W \geq 1$ earthquakes) expected for a given $M_c(x,t)$. If the thinning parameter σ was consistent in space and time, we could estimate the function $r(M_c)$ using only the MFD in Figure 4d. However, because we noted in Section 3.1 that σ is not constant, we expect more accurate results if we estimate $r(M_c)$ from multiple MFDs, formed using narrower ranges of M_c (or, equivalently, narrower ranges of d). We therefore consider the following procedure for a series of M_c spanning 1.0 to 5.2, using increments of 0.1:

1. Select all earthquakes with $M_{cE} \leq \max(M_c, 2)$ to form the M^* MFD.
2. Estimate b using the MLE described in Eq. 5 and Appendix 2.
3. Fit a thinning function $q(M^*|\sigma,\mu)$ on the MFD, with b constrained to the MLE value.
4. Compute the ratio $r(M_c)$ as:

$$r(M_c) = \frac{\int_{1-M_c}^{M^*_{\max}} 10^{-bM^*} dM^*}{\int_{1-M_c}^{M^*_{\max}} q(M^*) 10^{-bM^*} dM^*}, \tag{6}$$

where the integral limits define the range between $M_W = 1$ and the maximum magnitude in the MFD.

The resulting $r(M_c)$ function is shown in Figure 5; note that it approaches a log-linear relationship with slope of ~1, which is expected as $b \approx 1.0$ in Figure 4f,h. As an alternative model that does not depend on fitting a thinning function, we can simply extrapolate the GR model fit in step 2 to predict the total number events $N_{\text{pred}} = N(M^* \geq 0) 10^{-b(M_c-1)}$ and divide by $N_{\text{obs}} = N(M^* \geq 1 - M_c)$, which is equivalent to $N(M_W \geq 1)$. This ratio is plotted as the green curve in Figure 5 and produces similar results.

We opt not to fit a best-fit curve and instead directly interpolate the $r(M_c)$ results to compute $r_E = r(M_{cE})$ for each earthquake. The ratios r_E represent the predicted number of $M_W \geq 1$ earthquakes each (recorded) earthquake represents, and can be used to make the calibrated event-rate map shown in Figure 2d. This is in practice very similar to how the scalar moment of each earthquake is used to make the moment-rate map. Note that we cannot simply multiply the grid-cell values from Figure 2b by a ratio like $r(M_c)$ because M_c varies in time, as well as space. Finally, the resulting map shows that offshore seismic zones have comparable earthquake densities to the CSZ and LSZ, and it is much smoother than the moment-rate map because it is not dominated by the infrequent, largest earthquakes.

4 Discussion and Conclusions

Although we opted to use the smooth-increasing $M_c(d)$ model rather than the power-law fit (which was highly sensitive to the particular data range included), we are not suggesting that a power-law is inappropriate for the region. The Mignan et al. (2011) $M_c(d)$ power-law fitting results show significant scatter for individual M_c estimates, but they converge to a best-fit model because they have sufficient data to average M_c from many MFDs

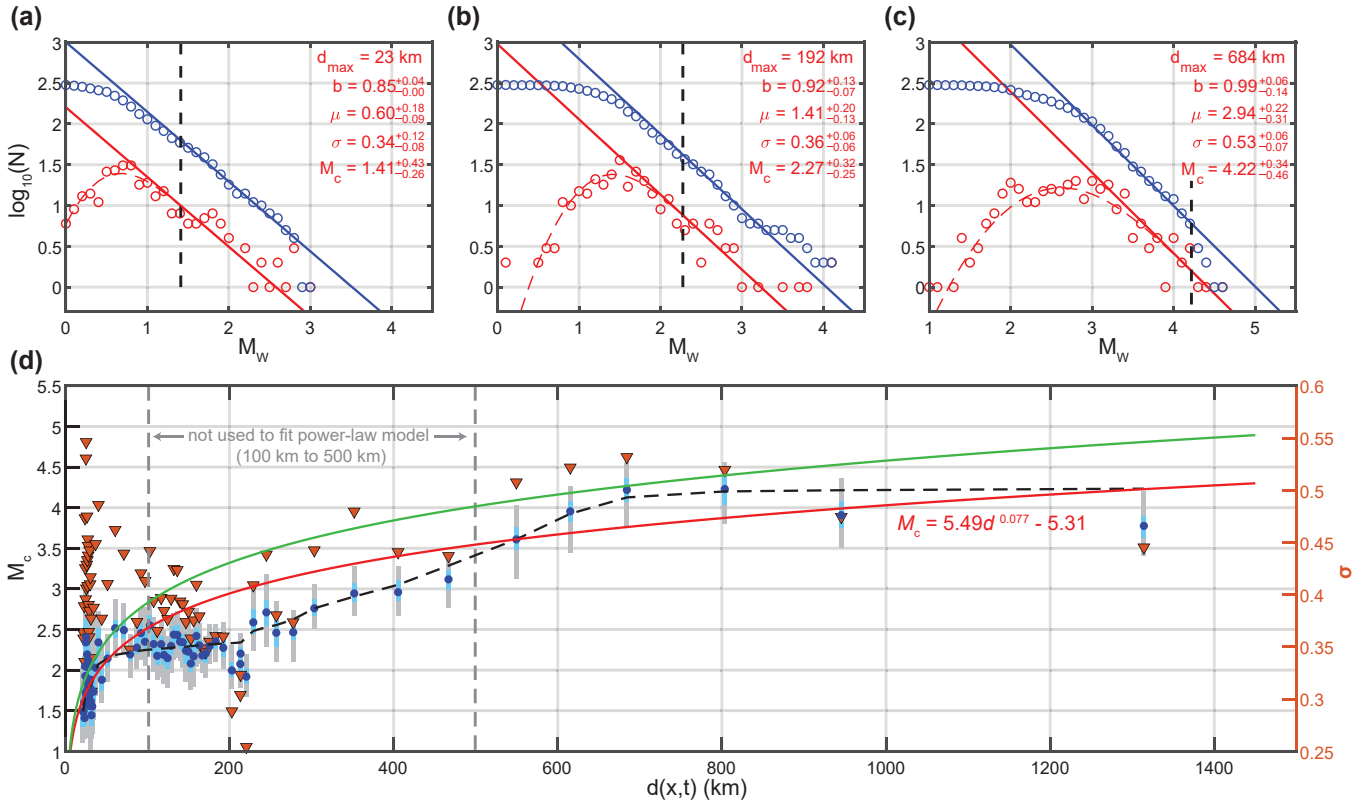

Figure 3 (a-c) The MFD and resulting Ogata and Katsura (1993) model fit for three distance bins of 300 earthquakes. 95% confidence intervals from bootstrapping are reported for each model parameter. (d) Overall $M_c(d)$ results (blue dots) with a power-law fit (red curve) and a best-fit smooth, increasing model (black dashed curve). Cyan and grey error bars indicate the 50% and 95% confidence intervals from bootstrapping, respectively. The green curve indicates the (4th-nearest station) model of Mignan et al. (2011), $M_c = 5.96d_4^{0.0803} - 5.80$. Orange triangles indicate the corresponding σ for each distance bin, as indicated on the right-hand side y-axis.

per distance bin. It is therefore unsurprising our $M_c(d)$ data show substantial scatter, but this does not fully explain why our $M_c(d)$ seems to have systematic deviations for a power-law, and remains constant over a range of $\sim 50 < d < 200$ km. It could be that the true $M_c(d)$ follows a power-law more closely and that our $M_c(d)$ estimates for this distance range are underestimated due to non–log-linear effects in their MFDs, although we have no hypothesis to suggest why this should be the case over an extensive range of $d(x,t)$. We demonstrate with Supplemental Figures S1–S4 that this issue does not substantially affect our following results or conclusions, but nevertheless this topic may warrant further investigation in future studies. The assessment of $M_c(x,t)$ more generally is discussed further near the end of this section.

Upon visual inspection, the GR plot produced by Equation 5 for the CSZ is quite similar to the raw M_W plot (Figure 4a,b), but the b vs. M_c plots show a modest improvement in the stability of b when using M^* (Figure 4c,d). M^* provides a much more dramatic improvement for the full Atlantic Canada catalog, as it produces much smaller confidence intervals on b, more stable b vs. M_c, and a lower thinning width σ (Figure 4e–h). The lower σ (0.39 for M^* vs. 0.65 for M_W) suggests a more angular MFD, rather than the gradual curvature that is caused by spatiotemporal heterogeneity in M_c (Mignan, 2012). The full Atlantic Canada catalog is an extreme case, where M_c varies greatly in both space and time, and there may be interesting b variation within the sample. Nevertheless, these observations suggest that M^*-derived b are more reliable than traditional estimates and that future studies, even those in areas where earthquakes and seismometers are more abundant, should consider spatiotemporal M_c variation when estimating b.

The predicted event-rate map (Figure 2d), as well as the associated methodology introduced here, are a robust way to compare earthquake rates between regions with different levels of seismometer coverage. It may be preferable to moment rate maps because it is sensitive to all earthquakes, instead of (effectively) only very large ones. With regard to the regulation of geological fluid-injection activities and induced-earthquake risk, the method is no replacement for local seismic monitoring efforts, but it may provide the best-available baseline earthquake rate estimates. Due to variable b and non–log-linear effects in MFDs, event rates will not always correlate perfectly with hazard. We also note that, for the purposes of hazard assessment, any map of recorded earthquakes is only useful to the extent that previous earthquakes locations can help predict the sites of future large earthquakes; our analysis does nothing to account for the possibility that areas of elevated intraplate seismicity today are extended aftershock sequences (Basham and Adams, 1983; Toda and

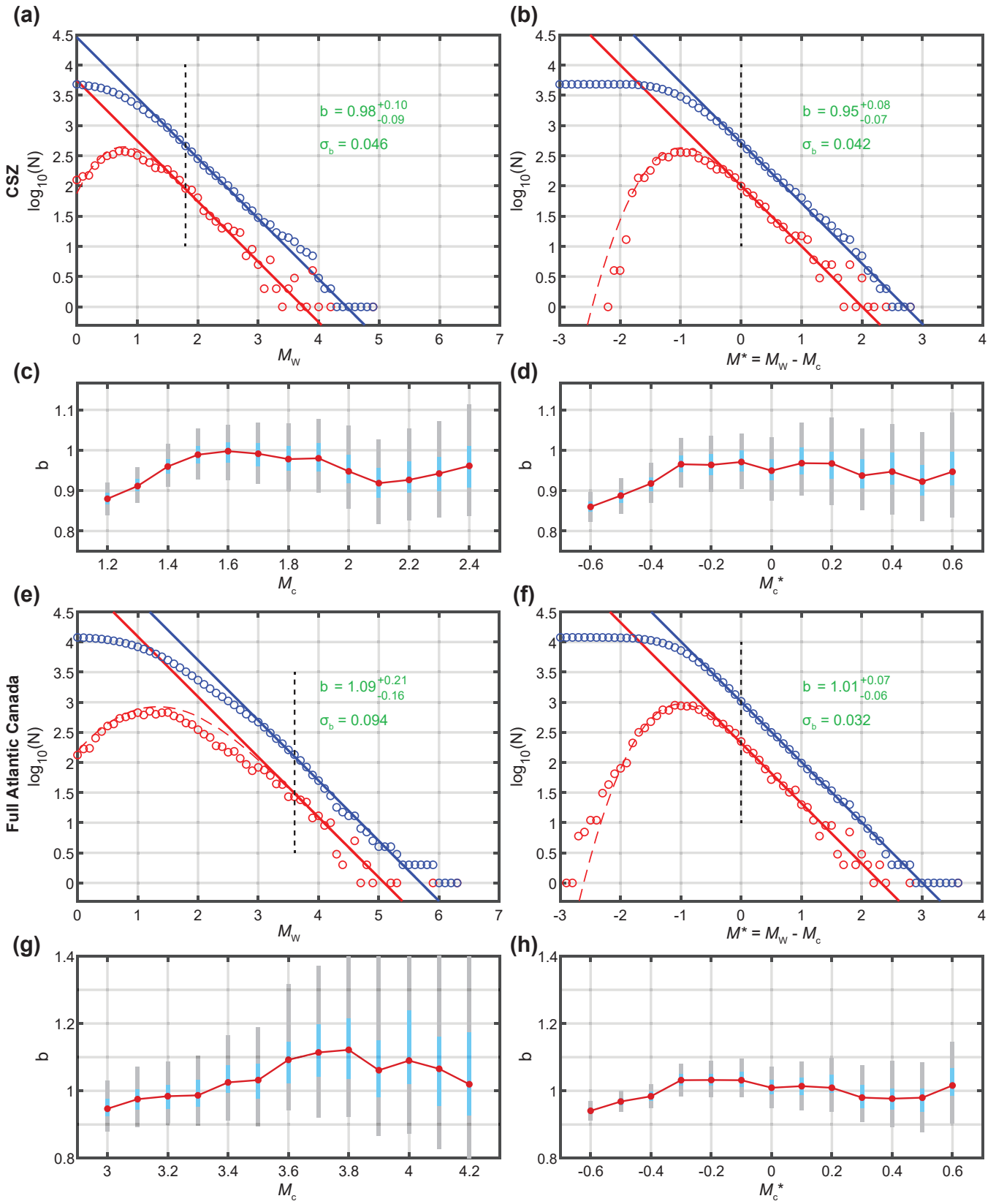

Figure 4 (a) Standard GR plot for the CSZ. Red and blue markers show the non-cumulative and cumulative MFDs, respectively. MLE-derived b is printed with 95% confidence intervals from 500 bootstrap iterations, as well as the standard error σ_b, which depends directly on the number of events with $M \geq M_c$ (N): $\sigma_b = b/\sqrt{N}$. The M_c used is shown with the dashed black line. (b) CSZ GR plot using M^* to account for variable $M_c(x,t)$. (c) b vs. M_c for the CSZ, using raw M_W, cyan and grey error bars show the 50% and 95% intervals from 500 bootstrap iterations. (d) b vs. M_c^* for the CSZ, using M^*, (e–h) like a–d except for the full map region of Figure 1. The Ogata and Katsura (1993) model fits in panels a, b, e, and f, were computed with b and M_c fixed to be consistent with the MLE fit and resulted in thinning widths σ of 0.39, 0.39, 0.65, and 0.39, respectively.

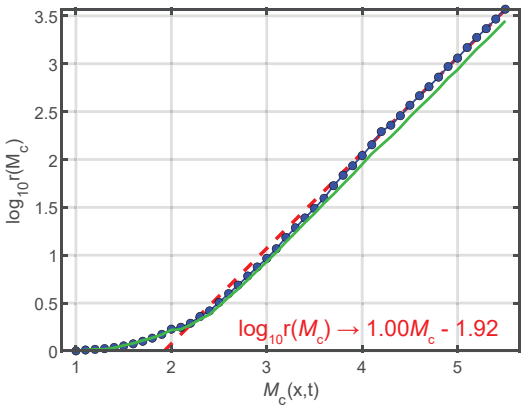

Figure 5 The predicted ratio of total/detected $M_W \geq 1$ earthquakes r, as a function of M_c, computed using Equation 6 for $1.0 \leq M_c \geq 5.2$ using increments of 0.1. The ratio approaches the log-linear relationship $\log_{10} r = 1.00 M_c - 1.92$ (computed using least-squares over the range $3.9 \geq M_c \leq 5.2$). The green curve is an alternative $r(M_c)$ model computed by extrapolating the GR model (without the thinning function), as described in the text.

Stein, 2018).

Finally, we must acknowledge some of the limitations of our $M_c(x,t)$ analysis. In addition to uncertainty in station metadata and the "noise" caused by periods where a seismometer was nonoperational that we have not accounted for, treating all seismometers equally is also a severe limitation. Noise levels, and signal-to-noise ratio, vary between seismometers for many reasons; geology at the site, instrument type, and proximity to anthropogenic or ocean-wave noise sources all being important factors. Instrumentation quality and noise levels also generally improve throughout the study period. Although we do not expect it to be a major factor in this dataset, short-term aftershock incompleteness (Stallone and Falcone, 2021) can also cause M_c variation in time. Schorlemmer and Woessner (2008) use a full phase pick catalog to empirically determine the likelihood of an earthquake being picked at a particular seismometer as a function of magnitude and distance; then, taking these functions at all seismometers, they equate $M_c(x,t)$ to a threshold probability of the earthquake being detected at four or more seismometers. Mahani et al. (2016) measure ambient noise levels at each seismometer and compare them with theoretical earthquake amplitudes in order to spatially map M_c. We expect that incorporating either system in place of, or in combination with, our simplified d–M_c relation would further improve the predicted event rate and b estimates.

Acknowledgements

This work was funded by NRCan (Contribution Number 20220415). We would like to thank Nicholas Ackerley for his thorough internal review, as well as John Adams and Allison Bent for their guidance early on, which were critical to creating the methodology and forming the datasets necessary for this study. We further thank Maria Mesimeri and an anonymous reviewer for their constructive comments.

Data and code availability

The earthquake catalog analyzed in this study comes from the Canadian National Earthquake Database (CNED), and was accessed at https://earthquakescanada.nrcan.gc.ca/stndon/NEDB-BNDS/bulletin-en.php in April 2022. Digital seismometer operating dates were taken from the Incorporated Research Institutions for Seismology (IRIS) database, and were accessed using the Python ObsPy package in June 2022.

Competing interests

The authors declare that they have no competing interests.

References

Adams, J. and Basham, P. The seismicity and seismotectonics of Canada east of the cordillera. *Geoscience Canada*, 16(1):3–16, 1989.

Aki, K. Maximum likelihood estimate of b in the formula log N = a-bM and its confidence limits. *Bull. Earthq. Res. Inst., Tokyo Univ.*, 43:237–239, 1965.

Basham, P. and Adams, J. Earthquakes on the continental margin of eastern Canada: Need future large events be confined to the locations of large historical events. *U.S. Geological Survey Open File Report*, 83-843:456–467, 1983.

Bent, A. L. The 1989 (MS 6.3) Ungava, Quebec, earthquake: a complex intraplate event. *Bulletin of the Seismological Society of America*, 84(4):1075–1088, 1994. doi: 10.1785/BSSA0840041075.

Bent, A. L. A complex double-couple source mechanism for the Ms 7.2 1929 Grand Banks earthquake. *Bulletin of the Seismological Society of America*, 85(4):1003–1020, 1995. doi: 10.1785/BSSA0850041003.

Bent, A. L. Moment magnitude (Mw) conversion relations for use in hazard assessment in eastern Canada. *Seismological Research Letters*, 82(6):984–990, 2011. doi: 10.1785/gssrl.82.6.984.

Bent, A. L. and Hasegawa, H. S. Earthquakes along the Northwestern Boundary of the Labrador Sea. *Seismological Research Letters*, 63(4):587–602, 1992. doi: 10.1785/gssrl.63.4.587.

Bent, A. L. and Voss, P. Seismicity in the Labrador-Baffin Seaway and surrounding onshore regions. In Dafoe, L. T. e. and Bingham-Koslowski, N. e., editors, *Geological synthesis of Baffin Island (Nunavut) and the Labrador-Baffin Seaway*, volume Bulletin 608, pages 379–387. Geological Survey of Canada, 2022. doi: 10.4095/321857.

Cheng, Y., Liu, W., Xu, T., Zhang, Y., Zhang, X., Xing, Y., Feng, B., and Xia, Y. Seismicity induced by geological CO_2 storage: A review. *Earth-Science Reviews*, 239:104369, 2023. doi: 10.1016/j.earscirev.2023.104369.

Gerstenberger, M. C., Marzocchi, W., Allen, T., Pagani, M., Adams, J., Danciu, L., Field, E. H., Fujiwara, H., Luco, N., Ma, K.-F., Meletti, C., and Petersen, M. D. Probabilistic seismic hazard analysis at regional and national scales: State of the art and future challenges. *Reviews of Geophysics*, 58(2):e2019RG000653, 2020. doi: 10.1029/2019RG000653.

Gutenberg, B. and Richter, C. F. Frequency of earthquakes in California. *Bulletin of the Seismological Society of America*, 34(4): 185–188, 1944. doi: 10.1785/BSSA0340040185.

Haddon, R. A. W. Modeling of source rupture characteristics for the Saguenay earthquake of November 1988. *Bulletin of the Seismological Society of America*, 85(2):525–551, 1995. doi: 10.1785/BSSA0850020525.

Halchuk, S., Allen, T. I., Rogers, G. C., and Adams, J. Seismic Hazard Earthquake Epicentre File (SHEEF2010) used in the Fifth Generation Seismic Hazard Maps of Canada. *Geological Survey of Canada, Open File*, 7724, 2015. doi: 10.4095/296908.

Ishimoto, M. and Iida, K. Observations of earthquakes registered with the microseismograph constructed recently. *Bulletin of the Earthquake Research Institute*, 17:443–478, 1939.

Kijko, A. and Smit, A. Extension of the Aki-Utsu b-Value Estimator for Incomplete Catalogs. *Bulletin of the Seismological Society of America*, 102(3):1283–1287, 2012. doi: 10.1785/0120110226.

Kolaj, M., Adams, J., and Halchuk, S. Seismic hazard in southeastern Canada: uncertainty and controls on seismic hazard in a region of low-to-moderate seismicity. In *17th World Conference on Earthquake Engineering*, 2020.

Lamontagne, M., Beauchemin, M., and Toutin, T. Earthquakes in the Charlevoix seismic zone, Quebec. *International Journal on Hydropower & Dams*, 10(2):98–99, 2003a.

Lamontagne, M., Keating, P., and Perreault, S. Seismotectonic characteristics of the Lower St. Lawrence Seismic Zone, Quebec: insights from geology, magnetics, gravity, and seismics. *Canadian Journal of Earth Sciences*, 40(2):317–336, 2003b. doi: 10.1139/e02-104.

Mahani, A. B., Kao, H., Walker, D., Johnson, J., and Salas, C. Performance evaluation of the regional seismograph network in northeast British Columbia, Canada, for monitoring of induced seismicity. *Seismological Research Letters*, 87(3):648–660, 2016. doi: 10.1785/0220150241.

Mignan, A. Functional shape of the earthquake frequency-magnitude distribution and completeness magnitude. *Journal of Geophysical Research: Solid Earth*, 117(B8):B009347, 2012. doi: 10.1029/2012JB009347.

Mignan, A., Werner, M. J., Wiemer, S., Chen, C.-C., and Wu, Y.-M. Bayesian estimation of the spatially varying completeness magnitude of earthquake catalogs. *Bulletin of the Seismological Society of America*, 101(3):1371–1385, 2011. doi: 10.1785/0120100223.

Mizrahi, L., Nandan, S., and Wiemer, S. The effect of declustering on the size distribution of mainshocks. *Seismological Research Letters*, 92(4):2333–2342, 2021. doi: 10.1785/0220200231.

Munro, P. S., Halliday, R. J., Shannon, W. E., and Shieman, D. R. J. Canadian Seismograph Operations — 1986. *Geological Survey of Canada*, Paper 88-16, 1988. doi: 10.4095/130933.

Ogata, Y. and Katsura, K. Analysis of temporal and spatial heterogeneity of magnitude frequency distribution inferred from earthquake catalogues. *Geophysical Journal International*, 113 (3):727–738, 1993. doi: 10.1111/j.1365-246X.1993.tb04663.x.

Ogata, Y. and Yamashina, K. Unbiased estimate for b-value of magnitude frequency. *Journal of Physics of the Earth*, 34(2):187–194, 1986. doi: 10.4294/jpe1952.34.187.

Page, R. Aftershocks and microaftershocks of the great Alaska earthquake of 1964. *Bulletin of the Seismological Society of America*, 58(3):1131–1168, 1968. doi: 10.1785/BSSA0580031131.

Plourde, A. P. and Nedimović, M. R. Earthquake depths, focal mechanisms, and stress in the Lower St. Lawrence Seismic Zone. *Seismological Research Letters*, 92(4):2562–2572, 2021. doi: 10.1785/0220200429.

Schorlemmer, D. and Woessner, J. Probability of detecting an earthquake. *Bulletin of the Seismological Society of America*, 98 (5):2103–2117, 2008. doi: 10.1785/0120070105.

Schultz, R., Skoumal, R. J., Brudzinski, M. R., Eaton, D., Baptie, B., and Ellsworth, W. Hydraulic fracturing-induced seismicity. *Reviews of Geophysics*, 58(3):e2019RG000695, 2020. doi: https://doi.org/10.1029/2019RG000695.

Si, Z. and Jiang, C. Research on parameter calculation for the Ogata-Katsura 1993 model in terms of the frequency-magnitude distribution based on a data-driven approach. *Seismological Research Letters*, 90(3):1318–1329, 2019. doi: 10.1785/0220180372.

Stallone, A. and Falcone, G. Missing earthquake data reconstruction in the space-time-magnitude domain. *Earth and Space Science*, 8(8):e2020EA001481, 2021. doi: 10.1029/2020EA001481.

Taroni, M. Back to the future: old methods for new estimation and test of the Gutenberg-Richter b-value for catalogues with variable completeness. *Geophysical Journal International*, 224(1): 337–339, 2021. doi: 10.1093/gji/ggaa464.

Toda, S. and Stein, R. S. Why aftershock duration matters for probabilistic seismic hazard assessment. *Bulletin of the Seismological Society of America*, 108(3A):1414–1426, 2018. doi: 10.1785/0120170270.

Utsu, T. A method for determining the value of 'b' in a formula log n = a-bM showing the magnitude-frequency relation for earthquakes. *Geophys. Bull. Hokkaido Univ.*, 13:99–103, 1965.

Utsu, T. A statistical significance test of the difference in b-value between two earthquake groups. *Journal of Physics of the Earth*, 14(2):37–40, 1966. doi: 10.4294/jpe1952.14.37.

van der Elst, N. J. B-Positive: A robust estimator of aftershock magnitude distribution in transiently incomplete catalogs. *Journal of Geophysical Research: Solid Earth*, 126(2):e2020JB021027, 2021. doi: 10.1029/2020JB021027.

Weichert, D. H. Estimation of the earthquake recurrence parameters for unequal observation periods for different magnitudes. *Bulletin of the Seismological Society of America*, 70(4): 1337–1346, 1980. doi: 10.1785/BSSA0700041337.

Wetmiller, R. J., Adams, J., Anglin, F. M., Hasegawa, H. S., and Stevens, A. E. Aftershock sequences of the 1982 Miramichi, New Brunswick, earthquakes. *Bulletin of the Seismological Society of America*, 74(2):621–653, 1984. doi: 10.1785/BSSA0740020621.

Woessner, J. and Wiemer, S. Assessing the quality of earthquake catalogues: Estimating the magnitude of completeness and its uncertainty. *Bulletin of the Seismological Society of America*, 95 (2):684–698, 2005. doi: 10.1785/0120040007.

Yu, H., Liu, Y., Harrington, R. M., and Lamontagne, M. Seismicity along St. Lawrence paleorift faults overprinted by a meteorite impact structure in Charlevoix, Québec, eastern Canada. *Bulletin of the Seismological Society of America*, 106(6):2663–2673, 2016. doi: 10.1785/0120160036.

Appendices

Appendix 1: Converting to $\mathrm{M_W}$

The most commonly used type of magnitude used in eastern Canada is the Nuttli magnitude m_N; it is reported for 82% of the earthquakes in our catalog. Bent (2011) analyzed the relation between m_N and M_W in eastern Canada and estimated conversion formulas of $M_W = 0.99m_N - 0.36 \pm 0.16$ for pre-1995 earthquakes and $M_W = 0.93m_N - 0.22 \pm 0.19$ for 1995–present.

Halchuk et al. (2015) reference this study and opt to use simplified relations of $M_W = m_N - 0.4$ for pre-1995 and $M_W = m_N - 0.5$ for 1995–present. They also treat local magnitude M_L as equivalent to m_N, and all other magnitudes (which form <0.02% of our catalog) as equivalent to M_W. In this work, we follow the procedure of Halchuk et al. (2015). However, as we have not seen this explicitly discussed elsewhere, we will point out here that linear conversions directly affect the estimated b. If, for example, we consider $M_W \propto 0.93 m_N$, we should expect b_W and b_N (estimates of b derived from the M_W and m_N catalogs, respectively), to differ according to the proportionality $b_N \propto 0.93 b_W$ (although we cannot verify that the proportionality is statistically significant). The Bent (2011) formula therefore suggests that our analysis (for which data is mostly from 1995–present) underestimates b_W. Note that this does not suggest a bias in seismic hazard analyses since m_N and other local magnitudes are more closely related to local amplitudes than M_W.

Appendix 2: Modification of maximum-likelihood estimator

In this section we justify Equation 5 beginning from the (unnormalized) probability density function for earthquake magnitude in a Gutenberg-Richter distribution (Aki, 1965):

$$f(M|M_c, \beta) = \beta e^{-\beta(M - M_c)} \tag{A1}$$

where $\beta = \ln(10) b$. We begin with similar reasoning as Kijko and Smit (2012), who consider the total likelihood to be a product of likelihoods from N subcatalogs with distinct M_c, but in our case we effectively consider each earthquake its own subcatalog, such that N is the total number of earthquakes. The total likelihood (ignoring a normalization constant) can be expressed as:

$$\begin{aligned} L(\beta) &= \prod_i^N f(M_i|M_{c,i}, \beta) \\ &= \prod_i^N \beta \exp\Big(-\beta(M_i - M_{c,i}) \Big) \\ &= \beta^N \exp\left(-\beta \sum_i^N (M_i - M_{c,i}) \right). \end{aligned} \tag{A2}$$

We can then closely follow the original MLE derivation and differentiate to find the maximum:

$$\begin{aligned} \frac{\partial L(\beta)}{\partial \beta} &= 0 \\ &= \left(N\beta^{N-1} - \beta^N \sum_i^N M_i \right. \\ &\quad \left. - M_{c,i} \right) \exp\left(-\beta \sum_i^N (M_i - M_{c,i}) \right). \end{aligned} \tag{A3}$$

After eliminating the remaining exponential term (as it cannot be zero) we can rearrange to find:

$$\begin{aligned} \beta &= \frac{1}{\frac{1}{N}\sum_i^N M_i - M_{c,i}} \\ &= \frac{1}{\langle M - M_c \rangle}, \end{aligned} \tag{A4}$$

which is equivalent to Equation 5. As a minor correction, this result should also be multiplied by $(N-1)/N$ to achieve an unbiased results (Ogata and Yamashina, 1986), resulting in the estimator:

$$b = \frac{\frac{N-1}{N} \log_{10}(e)}{\langle M - M_{cE} \rangle}. \tag{A5}$$

This is equivalent to Eq. 6 of Taroni (2021), except we omit the correction for binned magnitudes (adding $\Delta M/2$ to the denominator; Utsu, 1966) which is unnecessary because $M^* = M - M_{cE}$ is effectively unbinned. In practice, we also incorporate an upper magnitude limit in the MLE, which is important when the maximum magnitude is less than $M_c + 2$ (Page, 1968). To do this, we effectively replace M with M^* and set $M_c^* = 0$, then follow the MLE of Page (1968), which becomes:

$$\frac{1}{\beta} = \langle M^* \rangle - \frac{M^*_{\max} \exp(-\beta M^*_{\max})}{1 - \exp(-\beta M^*_{\max})}. \tag{A6}$$

We solve this formula using a line search, and then apply the $(N-1)/N$ correction to b for our final estimate.