

СРАВНИТЕЛЬНЫЙ АНАЛИЗ МЕТОДОВ ОЦЕНКИ МАГНИТУДЫ ПРЕДСТАВИТЕЛЬНОЙ РЕГИСТРАЦИИ ЗЕМЛЕТРЯСЕНИЙ

© 2022 г. В. А. Павленко^{1,*}, А. Д. Завьялов^{1,**}

¹Институт физики Земли им. О.Ю. Шмидта РАН, г. Москва, Россия

*E-mail: pavlenko.vasily@gmail.com

**E-mail: zavyalov@ifz.ru

Поступила в редакцию 09.10.2020 г.

После доработки 18.05.2021 г.

Принята к публикации 11.06.2021 г.

Магнитуда представительной регистрации землетрясений M_c — важный параметр, характеризующий полноту сейсмических данных. Известно, что пространственно-временные вариации M_c необходимо учитывать для получения корректных оценок параметров сейсмического режима. В работе сравниваются шесть современных методов оценки M_c . Для сравнения методов использованы выборки из реальных инструментальных каталогов землетрясений, а также синтетические каталоги, сгенерированные на основе трех моделей распределения магнитуды. Проанализированы зависимости двух первых моментов распределений оценок M_c от формы распределения магнитуды и объема выборки. Используются три модели, отвечающие выборочным распределениям, встречающимся при анализе инструментальных каталогов землетрясений. На основе полученных результатов сформулированы рекомендации по выбору подходящего метода оценки M_c .

Ключевые слова: магнитуда представительной регистрации, полнота сейсмических данных, инструментальные каталоги, синтетические каталоги.

DOI: 10.31857/S000233722010069

ВВЕДЕНИЕ

Одной из важнейших характеристик сейсмического процесса является частота возникновения землетрясений разной силы. Зависимость частоты землетрясений от магнитуды называют законом повторяемости землетрясений. В сейсмологии наиболее широко известен закон повторяемости Гутенберга–Рихтера [Gutenberg, Richter, 1944]:

$$\lg_{10}(N) = a - bM, \quad (1)$$

где: N — число землетрясений с магнитудой $m \geq M$; a — параметр сейсмической активности; b — наклон графика повторяемости землетрясений.

Закон Гутенберга–Рихтера применим для описания сейсмичности различных тектонических режимов. Параметры a и b имеют принципиальное значение для анализа сейсмической опасности. Кроме того, временное снижение величины b неоднократно наблюдалось перед возникновением сильных землетрясений [Gibowicz, 1973; Завьялов, 1984; 2006; Nuannin et al., 2005]. Временные снижения и нарастания сейсмической активности, известные как сейсмические затишья и форшоковые активизации, также рассматриваются в качестве возможных предвестников сильных землетрясений [Wyss et al., 1999; Zöller et al., 2002].

Определение значений a и b существенно осложняется тем, что обычно график повторяемости землетрясений соответствует теоретической зависимости (1) лишь на ограниченном интервале значений магнитуды (см. рис. 1). Выраженный излом кумулятивного графика повторяемости в области малых магнитуд обычно связывают с неполной регистрацией слабых землетрясений [Писаренко, 1989; Smirnov, 1998; Смирнов, Габсатарова, 2000; Wiemer, Wyss, 2000].

Магнитуда представительной регистрации M_c — это минимальная магнитуда, для которой землетрясения в заданной области регистрируются в полном объеме. Значение M_c соответствует началу линейной части графика повторяемости, поэтому знание M_c становится критически важным для получения адекватных оценок параметров Гутенберга–Рихтера.

Из соотношения (1) следует, что магнитуда имеет экспоненциальное распределение с функцией плотности:

$$f_M(m) = \beta e^{-\beta(m-M_c)}, \quad m \geq M_c, \quad (2)$$

где $\beta = b \ln(10)$.

Существующие методы оценки параметра b включают метод наименьших квадратов [Guttorp,

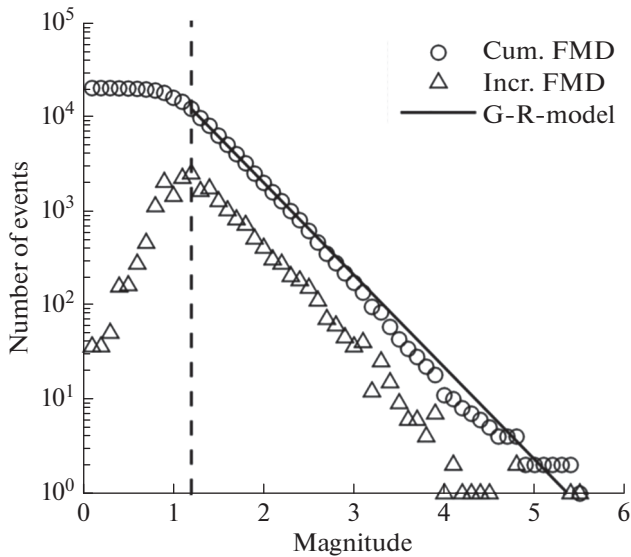


Рис. 1. Метод МАХС. Дискретный и кумулятивный графики повторяемости для выборки из каталога NCSN (Northern California Seismic Network). Сплошной линией показана аппроксимация законом Гутенберга–Рихтера, вертикальная штриховая линия показывает оценку M_c , взятую в точке максимальной кривизны кумулятивного графика повторяемости. Полученные оценки параметров $M_c = 1.2$, $a = 5.25$, $b = 0.98$.

1987; Nuannin et al., 2005], метод максимального правдоподобия [Aki, 1965; Гусев, 1974; Bender, 1983] и метод, основанный на минимизации статистики критерия согласия Колмогорова [Goldstein et al., 2004]. Прекрасный обзор наиболее распространенных на практике методов оценки b представлен в работе [Marzocchi, Sandri, 2003].

В статье [Bengoubou-Valérius, Gibert, 2013] был выполнен сравнительный анализ этих методов. Авторы исследования пришли к выводу о том, что наиболее надежным является метод максимального правдоподобия. С учетом поправки за дискретизацию магнитуды [Utsu, 1966], оценка максимального правдоподобия b будет иметь вид:

$$\hat{b} = \frac{\lg_{10}(e)}{\bar{M} - (M_c - \Delta M/2)}, \quad (3)$$

где \bar{M} – среднее выборочное значение магнитуды при $m \geq M_c$, ΔM – интервал группировки, т.е. точность определения магнитуды (в современных каталогах обычно $\Delta M = 0.1$).

Недооценка, т.е. занижение величины M_c приводит к получению некорректных оценок параметров сейсмичности и ошибочной интерпретации данных. Так, включение событий с непредставительной регистрацией в оценку наклона графика повторяемости b может привести к суще-

ственному занижению этой оценки (см. рис. 2б), что в свою очередь влечет недооценку частоты возникновения слабых землетрясений и переоценку частоты возникновения сильных землетрясений. Переоценка, т.е. завышение M_c может быть не столь критична [Chen et al., 2011], но приводит к снижению объема анализируемой выборки, а значит повышает неопределенность всех дальнейших результатов анализа. Значение M_c изменяется в пространстве и во времени и эти вариации необходимо учитывать при работе с каталогами землетрясений для оценки различных параметров сейсмического режима.

Таким образом, анализ пространственно-временных вариаций M_c становится неотъемлемым элементом практически любого исследования, опирающегося на каталоги землетрясений.

Одна из первых методик оценки M_c была предложена в работе [Stepp, 1972] и была основана на предположении о том, что последовательность основных землетрясений соответствует стационарному Пуассоновскому процессу. В таком случае средняя частота землетрясений λ должна быть стабильна в представительной части каталога, а ее стандартное отклонение обратно пропорционально взятому для оценки λ интервалу времени: $\sigma_\lambda = \sqrt{\lambda/T}$. По графику зависимости $\sigma_\lambda(T)$ для каждого значения магнитуды можно определить момент, когда график начинает отклоняться от степенной зависимости $1/\sqrt{T}$, что соответствует окончанию периода представительной регистрации.

Позднее был предложен метод оценки M_c , основанный на сравнении числа землетрясений, зарегистрированных в дневное и в ночное время [Rydelek, Sacks, 1989]. Этот метод предназначен для анализа локальных каталогов. Предполагается, что вероятность обнаружения землетрясений возрастает в ночное время вследствие снижения уровня промышленных шумов. Таким образом, если преобладающее число землетрясений в заданном интервале значений магнитуды регистрируется ночью, каталог считается непредставительным в этом интервале.

В дальнейшем проблеме оценки M_c уделялось существенное внимание и со временем число методов оценки M_c значительно возросло. Подробный обзор современных методов оценки M_c приводится в работе [Mignan, Woessner, 2012]. Большинство этих методов можно отнести к одной из двух категорий:

1. Методы, основанные на анализе каталогов землетрясений.
2. Методы, анализирующие возможности регистрации землетрясений станциями сейсмиче-

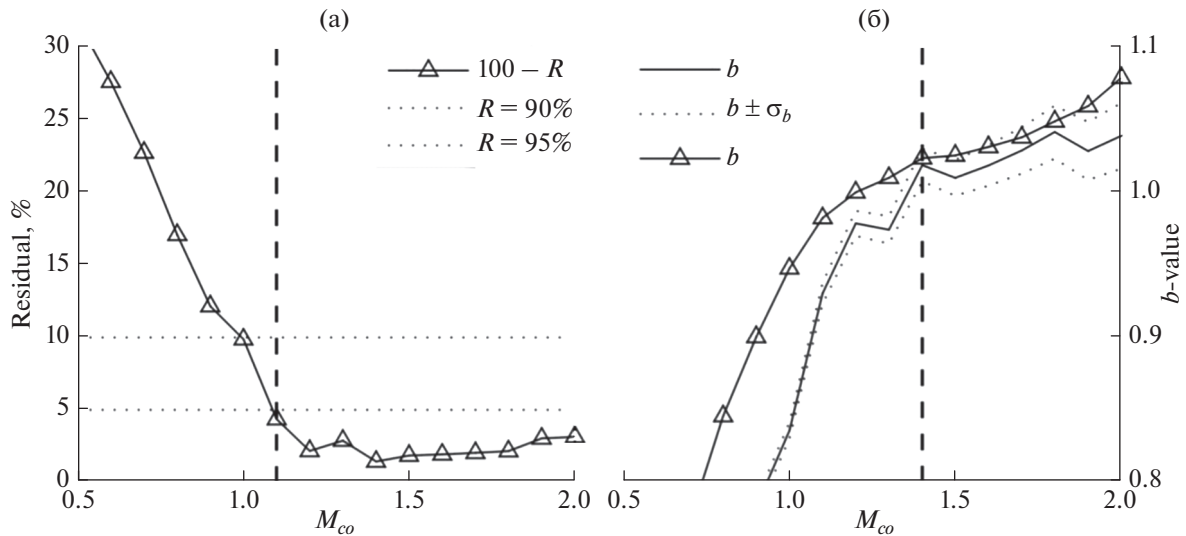


Рис. 2. (а) – Метод GFT. График величины $100 - R$, построенный по выборке из каталога NCSN. Горизонтальные пунктирные линии показывают уровни величины критерия R , при достижении которых берется оценка M_c ; вертикальная штриховая линия показывает оценку M_c ; (б) – метод MBS. Графики текущих и усредненных значений b , построенные по выборке из каталога NCSN. Сплошной линией показан график значений b , пунктиром показаны границы одного стандартного отклонения b , линия с треугольными маркерами показывает график усредненных значений b , вертикальная штриховая линия показывает оценку M_c .

ской сети заданной конфигурации [Ringdal, 1975; Gomberg, 1991].

Сопоставление результатов этих двух подходов выполнялось, в частности, в работе [Смирнов, Габсатарова, 2000], где по данным каталога землетрясений Северного Кавказа было показано, что оценки минимального представительного класса, полученные с помощью двух подходов, хорошо согласуются между собой и адекватно отражают изменения сейсмической сети.

Методы из первой категории менее трудоемкие, требуют меньшего объема информации и чаще применяются на практике. Кроме того, информация, необходимая для использования методов из второй категории, не всегда доступна для исследователей, вследствие чего эти методы не всегда применимы. Поэтому в данной работе рассматриваются только методы из первой категории. Существующее разнообразие приводит к тому, что сам по себе выбор предпочтительного метода оценки M_c перестает быть тривиальным.

В настоящей работе рассмотрены шесть современных методов оценки M_c , проанализировано поведение оценок этих методов на синтетических каталогах землетрясений в зависимости от объема выборки и формы исходного распределения числа событий по magnitude, использованного при создании синтетического каталога. Для анализа вариаций оценок M_c использован метод бутстрап анализа.

МЕТОДЫ ОЦЕНКИ МАГНИТУДЫ ПРЕДСТАВИТЕЛЬНОЙ РЕГИСТРАЦИИ

Оценка M_c в точке максимальной кривизны кумулятивного графика повторяемости землетрясений [Wiemer, Wyss, 2000]. Это самая простая и быстрая процедура оценки M_c . На практике точка максимальной кривизны кумулятивного графика повторяемости соответствует magnitude, на которую приходится максимальное число событий в выборке. Несмотря на простоту и надежность, данная процедура имеет тенденцию занижать значение M_c , в особенности в тех случаях, когда дискретный график повторяемости не имеет выраженного максимума. Предположительно, такая форма дискретного графика повторяемости, а именно кривизна непредставительной его части, может указывать на неоднородность анализируемой выборки [Mignan, 2012]. Тем не менее, для однородных выборок такая оценка будет адекватной [Mignan et al., 2011]. Процедура будет обозначаться аббревиатурой MAXC (от *Maximum curvature*), а соответствующая оценка – M_c^{MAXC} . Пример такой оценки показан на рис. 1.

Оценка M_c посредством расчета критерия согласия [Wiemer, Wyss, 2000]. Данный метод основан на сравнении анализируемой выборки значений magnitude и созданной на ее основе синтетической выборки. На первом этапе выбирается значение magnitude нижней отсечки M_{co} (это

значение не должно превышать M_c^{MAXC}), по части выборки с магнитудами $m \geq M_{co}$ методом максимального правдоподобия оцениваются параметры a, b из (1). На основе полученных оценок a, b создается синтетическая выборка из распределения Гутенберга–Рихтера. Далее сравниваются кумулятивные графики повторяемости для анализируемой и синтетической выборки и рассчитывается критерий согласия:

$$R(a, b, M_{co}) = 100 - \left(\frac{\sum_{i=1}^{M_{\max}} |B_i - S_i|}{\sum_i B_i} \times 100 \right), \quad (4)$$

где B_i, S_i – кумулятивное число событий в i -ом магнитудном интервале для анализируемой и синтетической выборки соответственно.

Величина R показывает, какой процент от объема анализируемой выборки можно описать соотношением (1) при данном значении M_{co} . Значение M_{co} повышается на величину ΔM , соответствующую точности определения магнитуды, и вычисления повторяются. В качестве оценки M_c выбирается наименьшее значение M_{co} , для которого величина критерия согласия R достигает заданного уровня (обычно это 90 или 95%). Процедура будет обозначаться аббревиатурой GFT (от *Goodness-of-fit test*), а соответствующая оценка – M_c^{GFT} . Пример применения данной процедуры показан на рис. 2а.

Оценка M_c по стабилизации значений b -value. Данная процедура была предложена в работах [Арефьев и др, 1989; Сао, Гао, 2002], в ее основе лежит наблюдение, согласно которому величина b возрастает при $M_{co} < M_c$ и при $M_{co} \gg M_c$, но остается практически неизменной при $M_{co} \geq M_c$. Соответственно, в качестве оценки M_c можно взять магнитуду, для которой разность двух последовательных значений b достаточно мала (например, меньше 0.03). В работе [Woessner, Wiemer, 2005] процедура была доработана и был предложен формализованный критерий стабилизации b . Модифицированная процедуры выполняется следующим образом: для некоторого диапазона значений M_{co} по формуле (3) оцениваются значения b и рассчитываются среднеквадратичные ошибки этих оценок σ_b по формуле [Shi, Bolt, 1982]:

$$\sigma_b = \frac{b^2}{\lg_{10}(e)} \sqrt{\frac{\sum_{i=1}^N (M_i - \bar{M})^2}{N(N-1)}}, \quad (5)$$

где \bar{M} – среднее выборочное значение магнитуды, N – число событий.

Затем в окне шириной $dM = 5\Delta M$ рассчитываются усредненные значения b :

$$\bar{b} = \sum_{M_{co}}^{M_{co}+5\Delta M} b(M_{co})/5. \quad (6)$$

В качестве оценки M_c выбирается наименьшее значение M_{co} , для которого разность между текущим и усредненным значениями b не превосходит среднеквадратичную ошибку оценки этого параметра, т.е. выполняется условие $\Delta b = |\bar{b} - b| \leq \sigma_b$. Процедура обозначается аббревиатурой MBS (M_c by b -value stability), а соответствующая оценка – M_c^{MBS} . Пример применения метода показан на рис. 2б.

Оценка M_c методом медианного анализа последовательности значений сдвигов дискретного графика повторяемости [Amorèse, 2007]. Метод основан на итерационной процедуре поиска точек, в которых происходит излом дискретного графика повторяемости. Для этого в каждой точке M_i , ($i > 1$) рассчитывается величина сдвига:

$$s(M_i) = \frac{\lg_{10}[N(M_{i-1})] - \lg_{10}[N(M_i)]}{M_{i-1} - M_i}, \quad (7)$$

где $N(M_i)$ – число событий в интервале $M_i \pm \Delta M/2$.

Сдвиги $s(M_i)$ характеризуют направленность и выраженность изменений дискретного графика повторяемости. Каждому значению $s(M_i)$ присваивается ранг R_i так, что наименьшее значение получает минимальный ранг – 1. Затем ранги суммируются в каждой точке $SR_i = \sum_{k=1}^i R_k$ и рассчитывается модифицированная сумма рангов SA_i :

$$SA_i = |2SR_i - i(n+1)|, \quad (8)$$

где n – длина последовательности.

Последовательность SA_i достигает максимума в точке n_1 , это наиболее вероятная точка излома дискретного графика повторяемости. В этой точке последовательность $s(M_i)$ делится на две части: $i_1 = 2, \dots, n_1$ и $i_2 = n_1 + 1, \dots, n + 1$. Проверяется гипотеза о том, что в точке n_1 нет излома дискретного графика повторяемости, против альтернативной гипотезы о том, что излом есть. Решение принимается на основе непараметрического критерия суммы рангов Уилкоксона [Wilcoxon, 1945], также известного как U-критерий Манна–Уитни [Mann, Whitney, 1947]. Если основная гипотеза отвергается при заданном уровне значимости,

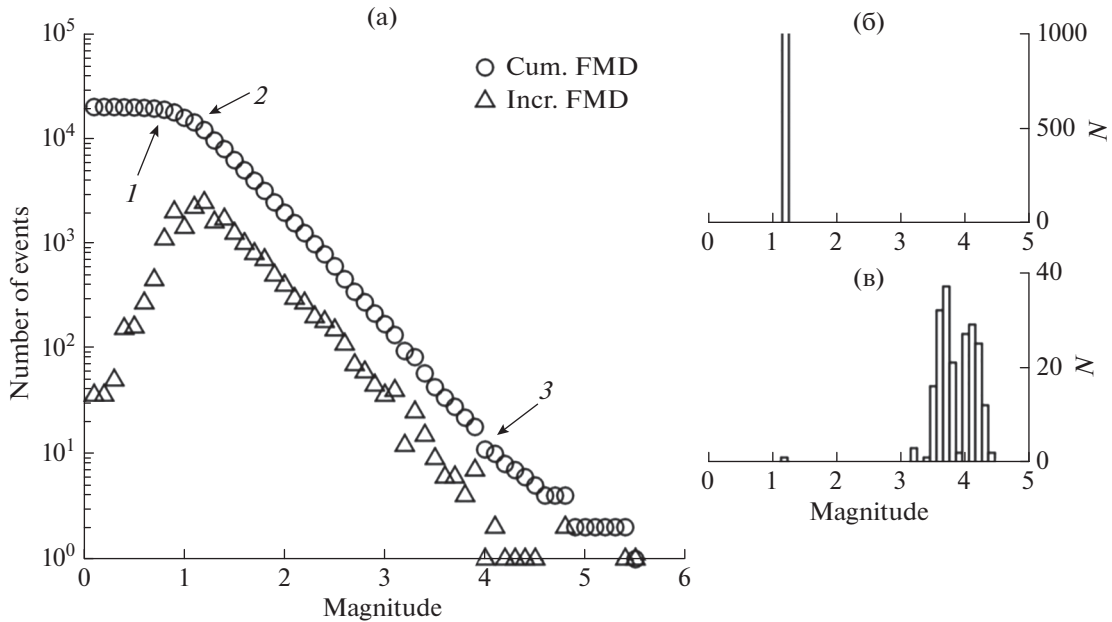


Рис. 3. Метод MBASS: (а) – дискретный и кумулятивный графики повторяемости для выборки из каталога NCSN. Стрелками показаны типичные точки излома графика повторяемости. Точка 2 соответствует магнитуде M_c . Бутстрэп-распределения (б) основной точки излома (точка 2) и (в) второстепенной точки излома (точки 1 и 3).

точка n_1 заносится в список и итерации продолжают. Процедура может обнаружить несколько значимых точек излома (рис. 3), магнитуде M_c соответствует та точка, для которой вероятность ошибки первого рода при проверке гипотезы минимальна. Процедура обозначается аббревиатурой MBASS (Median-based analysis of the segment slope), а соответствующая оценка – M_c^{MBASS} .

Оценка M_c методом моделирования полного диапазона значений магнитуды [Ogata, Katsura, 1993]. Согласно этому методу, для любого значения m наблюдаемую интенсивность потока событий $\lambda(m)$ можно представить в виде:

$$\lambda(m) = \lambda_0(m)q(m), \quad (9)$$

где $\lambda_0(m) = e^{-\beta m}$ – теоретическая интенсивность, $q(m)$ – вероятность обнаружения событий, которая задается нормальным распределением:

$$q(m) = \frac{1}{\sqrt{2\pi\sigma}} \int_{-\infty}^m \exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right) dx, \quad (10)$$

где μ и σ – среднее и стандартное отклонение.

С учетом сделанных предположений в работе [Ogata, Katsura, 1993] была получена функция плотности распределения магнитуды на всем диапазоне ее значений:

$$f_M(m) = \beta \exp\left(-\beta(m-\mu) - \beta^2 \frac{\sigma^2}{2}\right) q(m). \quad (11)$$

Параметры этой модели оцениваются методом максимального правдоподобия. Магнитуда M_c не входит в модель в явном виде и задается через параметры μ и σ : $M_c(n) = \mu + n\sigma$, n – показатель уровня доверия. При $n = 0$ регистрируются лишь 50% событий с магнитудой $m \geq M_c$ при $n = (1, 2, 3)$ регистрируются 84, 98 и 99% событий, соответственно. Вопрос о том, как соотносятся уровни представительности $M_c(n)$ модели (11) с оценками M_c остальных методов, рассматривался в статье [Mignan, Woessner, 2012], где на синтетических и реальных каталогах землетрясений было показано, что оценки M_c преимущественно попадают в интервал $(\mu, \mu + \sigma)$, в редких случаях могут достигать уровня $\mu + 2\sigma$, то есть 95%-ой вероятности обнаружения этой модели. Модель (11) обозначается аббревиатурой ОК (от Ogata and Katsura), пример применения этой модели показан на рис. 4.

В дальнейшем метод был модифицирован для того, чтобы явно включить в модель M_c [Woessner, Wiemer, 2005]. Была предложена составная модель, в которой нормальное распределение описывает вероятность обнаружения непредставительных событий:

$$q(m) = \begin{cases} \frac{1}{\sqrt{2\pi\sigma}} \int_{-\infty}^m \exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right) dx & m < M_c \\ 1 & m \geq M_c \end{cases}. \quad (12)$$

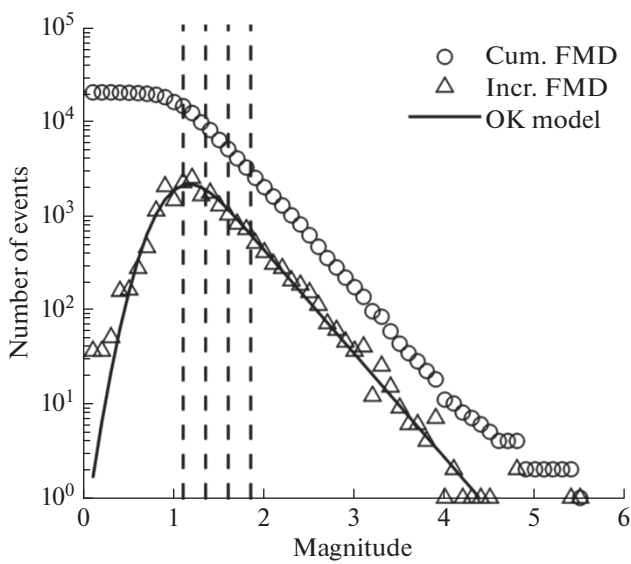


Рис. 4. Модель ОК: дискретный и кумулятивный графики повторяемости для выборки из каталога NCSN. Сплошной линией показана аппроксимация моделью ОК, вертикальными штриховыми линиями показаны оценки $M_c(n)$ для $n = (0, 1, 2, 3)$.

$$\lambda(m) = \begin{cases} \frac{1}{\sqrt{2\pi}\sigma} \int_{-\infty}^m \exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right) dx & m < M_c \\ \exp(-\beta(m - M_c)) & m \geq M_c \end{cases} \quad (13)$$

Эта модель будет обозначаться аббревиатурой WW (от *Woessner and Wiemer*), метод будет обозначаться аббревиатурой EMR (от *Entire Magnitude Range*), а соответствующая оценка – M_c^{EMR} . Параметры μ и σ оцениваются методом нелинейной регрессии по данным выборки в диапазоне $m < M_c$, параметры a и b оцениваются методом максимального правдоподобия по данным в диапазоне $m \geq M_c$. На основе этих оценок и уравнения (13) рассчитывается теоретический дискретный график повторяемости на всем диапазоне значений магнитуды и оценивается правдоподобие получения такой выборки. Вычисления повторяются для некоторого диапазона значений M_c , итоговая оценка берется в максимуме функции правдоподобия. Такой подход дает наиболее всестороннее описание сейсмичности, но является наиболее затратным в плане времени вычислений. Применение метода показано на рис. 5.

Важно отметить, что для этой модели уравнение (9) не выполняется, интенсивность потока событий задается следующей функцией:

Оценка M_c в точке начала линейной части графика повторяемости [Писаренко, 1989; Смирнов, 2009]. Процедура основана на проверке гипотезы о прямолинейности графика повторяемости. На первом этапе выбирается начальное значение

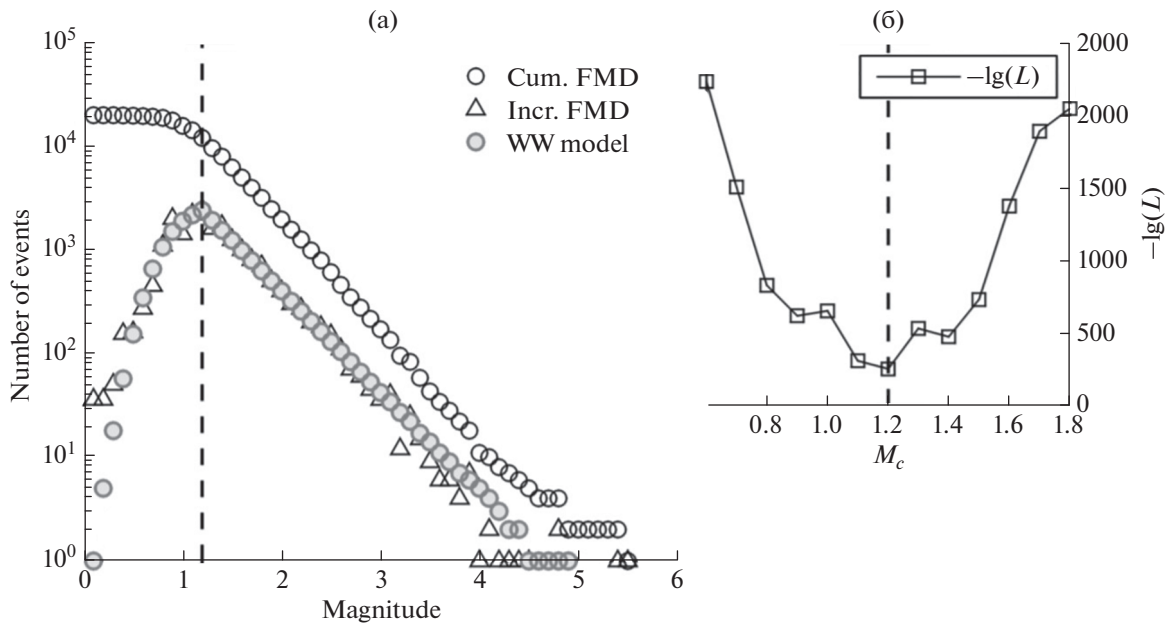


Рис. 5. Метод EMR: (а) – дискретный и кумулятивный графики повторяемости для выборки из каталога NCSN. Серыми кружками показана аппроксимация моделью WW, вертикальная штриховая линия показывает оценку M_c ; (б) – график зависимости логарифма функции правдоподобия, взятого с минусом, от величины M_c , вертикальная штриховая линия показывает оценку M_c , взятую в максимуме функции правдоподобия.

магнитуды нижней отсечки M_{co} (это значение не должно превышать M_c^{MAXC}), по части выборки с магнитудами $m \geq M_{co}$ методом максимального правдоподобия оцениваются параметры закона повторяемости. Проверяется гипотеза H_1 о прямолинейности графика повторяемости в области $m \geq M_{co}$ против гипотезы H_2 о том, что график повторяемости в этой области нелинейный. Для этого рассчитываются наблюдаемые частоты попадания значений магнитуды в интервалы $M_i \pm \Delta M/2$, где $M_i = M_{co} + i\Delta M$, $i = 0, \dots, n$:

$$p_i = \frac{N_i}{Q_0}, \quad (14)$$

где N_i – число событий в интервале $M_i \pm \Delta M/2$, $Q_0 = \sum_{i=0}^n N_i$.

И теоретические частоты, соответствующие закону Гутенберга–Рихтера с параметром \hat{b} :

$$\pi_i = \frac{10^{-\hat{b}(M_i - M_{co})}}{\sum_{i=0}^n e^{-\hat{b}(M_i - M_{co})}}. \quad (15)$$

Рассчитывается выборочная статистика, характеризующая отклонение гипотезы H_2 от гипотезы H_1 :

$$\hat{I} = \sum_{i=0}^n p_i \ln \left(\frac{p_i}{\pi_i} \right). \quad (16)$$

Если верна гипотеза H_1 , то величина $2Q_0\hat{I}$ имеет асимптотически при $Q_0 \rightarrow \infty$ распределение χ^2 с $(n-1)$ степенью свободы. Для проверки гипотезы H_1 определяется уровень значимости μ величины $2Q_0\hat{I}$. Если μ меньше заданного уровня (например, 0.01 или 0.05), то гипотеза H_1 отвергается. В этом случае на следующем этапе решается вопрос о причинах нелинейности графика повторяемости. Если же гипотеза H_1 принимается, то M_{co} – искомая оценка M_c и процедура на этом завершается.

Вероятность регистрации землетрясений с $m = M_{co}$ обозначается как p , если $p = 1$ регистрация представительна, если $p < 1$ имеются пропуски землетрясений с $m = M_{co}$. Для решения вопроса о представительности регистрации землетрясений с $m = M_{co}$ строится оценка максимального правдоподобия \hat{p} для p и проверяется гипотеза об отличии этой оценки от 1. Оценка \hat{p} определяется отношением наблюдаемой величины N_0 к оценке числа событий с $m = M_{co}$, построенной по старшим магнитудам с $m \geq M_{co} + \Delta M$. Пусть \tilde{x}^i –

оценка максимального правдоподобия, полученная по части выборки с $m \geq M_{co} + \Delta M$, $\tilde{x}^i = 10^{-\hat{b}(M_i - M_{co})}$. Тогда оценка \hat{p} имеет вид:

$$\hat{p} = \frac{N_0\Psi_0}{Q_0 - N_0}, \quad (17)$$

а ее асимптотическая дисперсия оценивается как:

$$Var(\hat{p}) = \frac{N_0\Psi_0^2}{(Q_0 - N_0)^2} + \frac{N_0^2\Psi_0^3\Psi_2}{(Q_0 - N_0)^3(\Psi_0\Psi_2 - \Psi_1^2)^2}, \quad (18)$$

где $\Psi_k = \sum_{i=1}^n i^k \tilde{x}^i$, $k = 0, 1, 2$.

Для проверки гипотезы H_0 о представительной регистрации землетрясений с $m = M_{co}$ рассчитывается величина $t = \frac{1 - \hat{p}}{(Var(\hat{p}))^{1/2}}$, имеющая распределение Стьюдента, и определяется ее уровень значимости μ_0 . Гипотеза H_0 отвергается если $\mu_0 \geq 1 - \varepsilon$, где ε – некоторое малое число (например 0.05 или 0.1). В этом случае значение M_{co} повышается на ΔM и вся процедура повторяется. Если гипотеза H_0 принимается, то M_{co} – искомая оценка M_c . Метод будет обозначаться аббревиатурой LLS (от Lower end of linear segment), а соответствующая оценка – M_c^{LLS} . Пример применения метода показан на рис. 6.

Бутстреп [Efron, Tibshirani, 1993] – непараметрический метод исследования распределений искомого параметра, основанный на многократном извлечении методом Монте-Карло повторных выборок из имеющейся выборки. Метод позволяет быстро и просто оценивать разнообразные статистики (например, дисперсию, доверительные интервалы) для сложных моделей. Суть метода состоит в том, что из имеющейся выборки случайным выбором с повторением формируется некоторое множество n_b повторных выборок заданного размера. На множестве повторных выборок оцениваются искомые параметры и по полученным эмпирическим распределениям определяются все необходимые статистики.

В частности, бутстреп применялся для анализа дисперсии оценок M_c в работах [Woessner, Wiemer, 2005; Amorèse, 2007; Mignan et al., 2011; Mignan, Woessner, 2012]. В книге [Chernick, 1999] в качестве рекомендованного числа повторных выборок, достаточного для надежной оценки дисперсии, приводится цифра $n_b = 100$, однако при наличии достаточных вычислительных мощностей автор рекомендует использовать $n_b = 1000$. В работе [Woessner, Wiemer, 2005] авторы отмечают, что оценки дисперсии M_c стабилизируются при $n_b \geq 200$. Эту рекомендацию использовали в рабо-

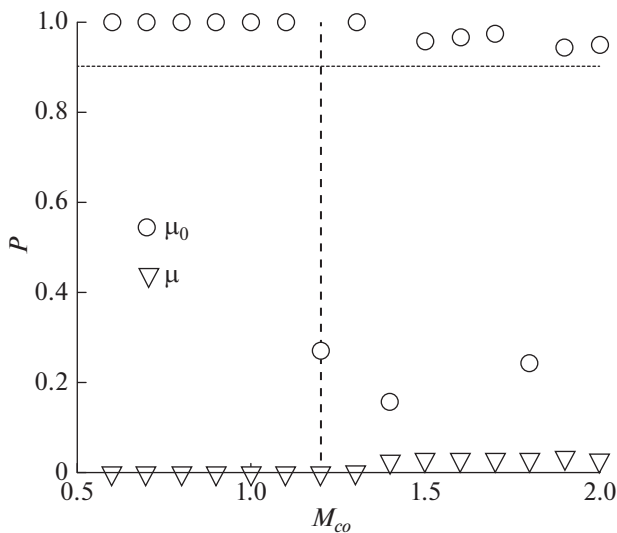


Рис. 6. Метод LLS. Треугольниками показаны уровни значимости статистики $2Q_0\hat{J}$, кружками показаны уровни значимости величины t , вертикальная штриховая линия показывает оценку M_c . При $M_{co} < 1.4$ гипотеза H_1 отвергается с высоким уровнем значимости. При $M_{co} < 1.2$ гипотеза H_0 также отвергается, но при $M_{co} = 1.2$ уже нет достаточных оснований считать, что регистрация событий непредставительная, это значение принимается в качестве оценки M_c .

тах [Mignan et al., 2011; Mignan, Woessner, 2012]. В работе [Amorèse, 2007] автор использовал $n_b = 1000$ повторных выборок. В настоящей работе бутстрап применяется с $n_b = 500$ повторных выборок.

АНАЛИЗ ПРЕДСТАВИТЕЛЬНОСТИ ВЫБОРОК ИЗ ИНСТРУМЕНТАЛЬНЫХ КАТАЛОГОВ ЗЕМЛЕТРЯСЕНИЙ

Все описанные процедуры были программно реализованы В.А. Павленко в MATLAB. В качестве проверки корректности работы процедур была предпринята попытка воспроизвести результаты, полученные в работах [Woessner, Wiemer, 2005; Amorèse, 2007]. Для этого были взяты те же выборки из каталогов землетрясений, которые использовали авторы этих работ в качестве демонстрационного материала:

1. Выборка из регионального каталога Швейцарской сейсмологической службы, the Earthquake Catalog of Switzerland (ECOS; <http://ecos09.seismo.ethz.ch/query>) за 1992–2002 гг.
2. Выборка из глобального каталога the Global Centroid Moment Tensor Catalog (CMT; <https://www.globalcmt.org/CMTsearch>) за 1983–2002 гг.
3. Выборка из регионального каталога землетрясений Северной Калифорнии, the Northern California Seis-

mic Network (NCSN; <https://www.ncedc.org/ncedc/catalog-search>) за 1998–2000 гг.

4. Выборка из каталога землетрясений вулканической зоны сейсмичности Канто, the National Research Institute for Earth Science and Disaster Prevention Earthquake Catalog (NIED; <http://evrrss.eri.u-tokyo.ac.jp/db/nied>) за 1992–2002 гг.

Параметры получившихся выборок сведены в табл. 1, построенные по выборкам графики повторяемости показаны на рис. 7. Сравнение графиков повторяемости и объемов выборок с представленными в работах [Woessner, Wiemer, 2005; Amorèse, 2007], показывает, что ни одна из полученных выборок не совпадает в полной мере с теми выборками, которые использовали эти авторы. Вероятнее всего, причиной различий стали изменения, внесенные в каталоги со времени публикации этих работ.

По каждой выборке с помощью описанных процедур и метода бутстрап оценивались средние значения и дисперсии оценок M_c . Полученные результаты представлены в табл. 1. Результаты для выборок из каталогов ECOS, NCSN и NIED хорошо согласуются с результатами, представленными в работах [Woessner, Wiemer, 2005; Amorèse, 2007]. Оценки M_c , полученные по выборке из каталога CMT, существенно отличаются от тех оценок, которые представлены в работе [Woessner, Wiemer, 2005], но это расхождение, по всей видимости, обусловлено различиями выборочных данных.

В этих результатах проявляются некоторые характерные особенности рассматриваемых методов: методы МАХС и GFT обычно дают самые низкие оценки M_c , оценки метода MBS обычно оказываются наиболее консервативными. Оценки остальных методов занимают промежуточные позиции.

Из табл. 1 видно, что для всех методов, кроме метода MBS, дисперсии оценок M_c , полученных по выборкам из каталогов NCSN и CMT, оказывались минимальными или близкими к минимальным. Форма распределения выборки из каталога NCSN близка к характерной угловой форме с выраженным максимумом, которая присуща выборкам с равномерным уровнем регистрации. Распределение выборки из каталога CMT имеет сглаженный максимум, для этой выборки разброс оценок M_c заметно выше, чем для выборки из каталога NCSN.

Распределения выборок из каталогов ECOS и NIED имеют сглаженную форму с заметной кризической непредставительной частью дискретного графика повторяемости, которая может свидетельствовать о неоднородности выборочных данных. В целом это приводит и к возрастанию дисперсий

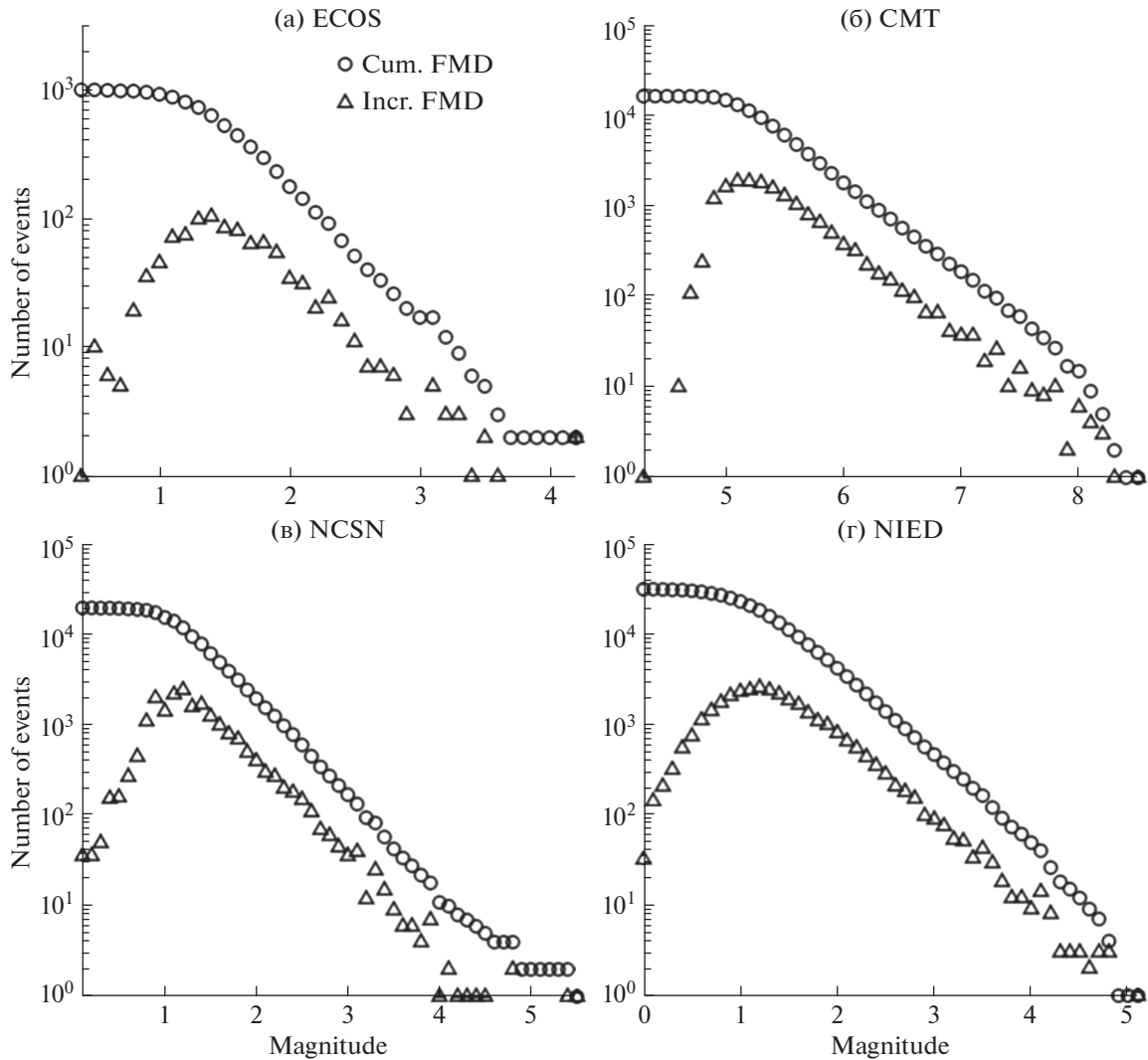


Рис. 7. Графики повторяемости, построенные по выборкам из каталогов: (а) – The Earthquake Catalog of Switzerland за 1992–2002 гг.; (б) – The Global Centroid Moment Tensor Catalog за 1983–2002 гг.; (в) – The Northern California Seismic Network за 1998–2000 гг.; (г) – The National Research Institute for Earth Science and Disaster Prevention Earthquake Catalog за 1992–2002 гг.

оценок M_c и к более значительным различиям в оценках M_c , полученных разными методами.

Полученные результаты демонстрируют чувствительность методов оценки M_c к форме распределения анализируемой выборки. Этот эффект более детально рассмотрен на синтетических каталогах землетрясений в следующих секциях.

МОДЕЛИ РАСПРЕДЕЛЕНИЯ МАГНИТУДЫ

Создание синтетического каталога землетрясений требует применения модели распределения магнитуды, позволяющей описать как представительную, так и не представительную часть каталога. Помимо уже упоминавшихся моделей

ОК и WW, существуют еще две модели распределения, подходящие для этих целей.

Модель, предложенная в работе [Mignan, 2012], описывает элементарный график повторяемости, построенный по выборке значений магнитуды, в которой пространственно-временные неоднородности M_c сведены к минимуму. Вероятность обнаружения событий при $m < M_c$ изменяется по экспоненте:

$$q(m) = \begin{cases} \exp(\kappa(m - M_c)) & m < M_c \\ 1 & m \geq M_c \end{cases}, \quad (19)$$

где: $\kappa = k \ln(10) > \beta$ – параметр обнаружения; β – показатель экспоненциального распределения (2).

Таблица 1. Параметры выборок из инструментальных каталогов землетрясений, оценки магнитуды M_c , оценки параметров и модели WW

Каталог	ECOS	CMT	NCSN	NIED
Объем выборки	988	16472	19833	31 372
Диапазон значений магнитуды	0.4–4.2	4.3–8.5	0.1–5.5	0.0–5.1
Границы по долготе	6.8°–8.4° в.д.		120.5°–123° з.д.	138.95°–139.35° в.д.
Границы по широте	45.9°–46.65° с.ш.		36.0°–39.0° с.ш.	34.8°–35.05° с.ш.
Начало выборки	01.01.1992	01.01.1983	01.01.1998	01.01.1992
Окончание выборки	31.12.2002	31.12.2002	31.12.2000	31.12.2002
M_c^{MAXC}	1.37 ± 0.06	5.15 ± 0.06	1.20 ± 0.00	1.20 ± 0.01
M_c^{GFT}	1.54 ± 0.16	5.20 ± 0.01	1.11 ± 0.03	1.35 ± 0.06
M_c^{MBS}	1.66 ± 0.13	5.42 ± 0.06	1.46 ± 0.16	1.98 ± 0.10
M_c^{MBASS}	1.39 ± 0.27	5.39 ± 0.07	1.20 ± 0.00	1.39 ± 0.10
M_c^{EMR}	1.55 ± 0.14	5.31 ± 0.07	1.20 ± 0.03	1.27 ± 0.05
M_c^{LLS}	1.55 ± 0.15	5.44 ± 0.05	1.20 ± 0.01	1.84 ± 0.15
μ	0.93	4.87	0.84	0.62
σ	0.16	0.08	0.22	0.26

Модель будет обозначаться аббревиатурой AN (от английского *Angular*).

В статье [García-Hernández et al., 2019] авторы предложили модель, в которой вероятность обнаружения событий задается полиномом второй степени (модель POL):

$$q(m) = \begin{cases} 0 & m \leq M_i \\ 1 - \left(\frac{m - M_c}{M_c - M_i}\right)^2 & M_i < m < M_c, \\ 1 & m \geq M_c \end{cases} \quad (20)$$

где M_i – минимальная магнитуда обнаружения.

Такая форма распределения характерна для ранних периодов инструментальных наблюдений, когда ввиду сравнительно невысокой чувствительности приборов слабые события не регистрировались. Обе модели AN и POL подчиняются уравнению (9).

Авторы работы [García-Hernández et al., 2019] проанализировали формы выборочных распределений магнитуды в зависимости от объема выборки и радиуса области, из которой выбирались землетрясения. Для анализа были использованы данные Международного Сейсмологического Центра [ISC; www.isc.ac.uk/iscbulletin/search/catalogue/; Storchak et al., 2013], модели оценивались по критерию согласия Колмогорова и информационному критерию Акаике [Akaike, 1974]. Результаты этого исследования показали, что для всех моделей степень согласия с данными снижалась с возрастанием объема выборки, но практи-

чески не зависела от радиуса области выборки событий. Выраженное снижение согласия с данными наблюдалось для выборок, содержащих более 1000 событий. В целом модель ОК чаще других оказывалась наилучшей в плане аппроксимации данных; модель AN показывала хорошие результаты для малых выборок объемом 100–300 событий; модель WW показывала хорошие результаты для более объемных выборок, включающих 1000–3000 событий; модель POL оказалась существенно хуже, показав относительно неплохой результат только для выборок объемом 100 событий.

Таким образом, для создания синтетических каталогов были выбраны модели WW, AN и POL. Модель WW описывает наиболее распространенную ситуацию, при которой выборка значений магнитуды неоднородна по M_c . Модель AN описывает выборку, в которой неоднородности M_c сведены к минимуму. Модель POL описывает выборку, из которой удалены слабые землетрясения с магнитудами $m \leq M_i$. Различие этих трех моделей продемонстрировано на рис. 8.

ТЕСТИРОВАНИЕ МЕТОДОВ ОЦЕНКИ M_c НА СИНТЕТИЧЕСКИХ КАТАЛОГАХ ЗЕМЛЕТРЯСЕНИЙ

При выборе значений параметров модели WW учитывались их оценки, полученные по выборкам из каталогов ECOS, CMT, NCSN и NIED (табл. 1). Из этих результатов видно, что магнитуда μ располагается примерно посередине между

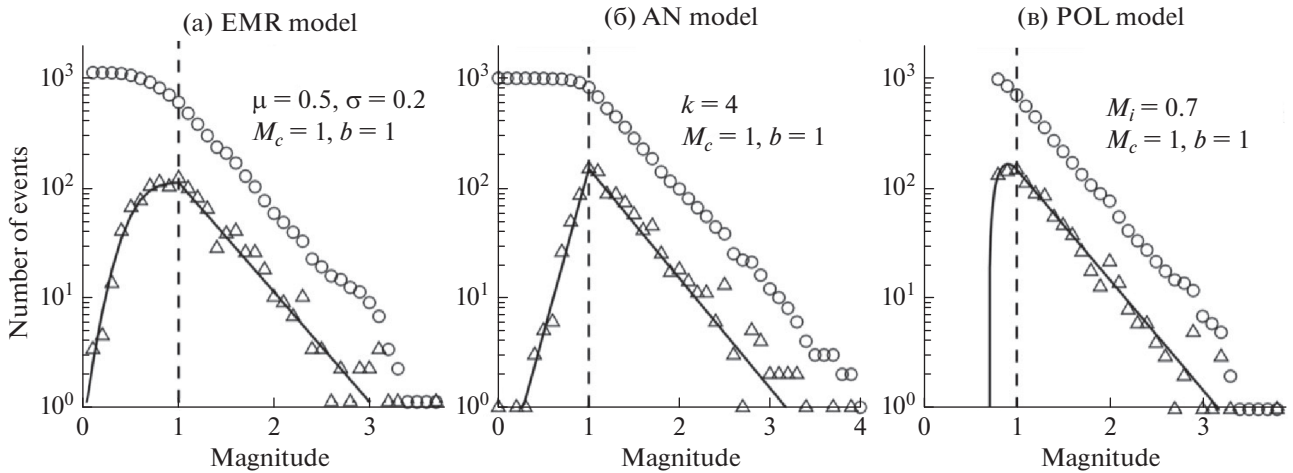


Рис. 8. Примеры случайных выборок объемом по 1000 событий из синтетических каталогов, демонстрирующие разницу формы графиков повторяемости, получаемых при использовании трех описанных в тексте моделей распределения магнитуды: (а) модель WW с параметрами $M_c = 1$, $b = 1$, $\mu = 0.5$, $\sigma = 0.2$; (б) модель AN с параметрами $M_c = 1$, $b = 1$, $k = 4$; (в) модель POL с параметрами $M_c = 1$, $b = 1$, $M_i = 0.7$.

минимальной магнитудой в выборке и оценкой M_c^{EMR} , а σ имеет величину порядка 0.1–0.3.

Выборки из каталогов ECOS и NIED имеют типичную для модели WW форму распределения (рис. 7а, 7г). При фиксированном μ , снижение σ приводит к тому, что левый хвост распределения приобретает более крутой наклон, максимум распределения сглаживается, форма распределения приближается к форме выборки из каталога СМТ (рис. 7б). При смещении μ ближе к M_c при фиксированном σ форма распределения становится близка к форме выборки из каталога NCSN (рис. 7в) с острым максимумом и крутым наклоном левого хвоста распределения.

На основе этих трех вариантов формы распределения WW были выбраны следующие комбинации значений параметров для создания синтетических каталогов: $\mu_1 = 0.5$, $\sigma_1 = 0.25$ (такие значения параметров использовали для создания синтетического каталога авторы [Woessner, Wiemer, 2005]), $\mu_2 = 0.5$, $\sigma_2 = 0.1$ и $\mu_3 = 0.75$, $\sigma_3 = 0.25$.

В работе [Mignan, 2012] приводятся оценки параметра k модели AN, полученные по данным каталогов землетрясений Невады и Южной Калифорнии. Для Невады оценки k лежат в пределах от 1.4 до 5.4, максимум плотности распределения приходится на $\hat{k} = 3$, для Южной Калифорнии оценки лежат в пределах от 2 до 5 с максимумом плотности при $\hat{k} = 2.9$. По выборке из каталога NCSN была получена оценка максимального правдоподобия $\hat{k} = 2.95$. С повышением k наклон левого хвоста распределения становится все более крутым, а число событий в непредставитель-

ной части распределения снижается. Для создания синтетических каталогов с распределением AN были использованы значения $k_1 = 2$, $k_2 = 3$, $k_3 = 4$.

Значения параметра M_i модели POL были заданы произвольно: $M_{i1} = 0.5$, $M_{i2} = 0.6$, $M_{i3} = 0.7$.

Для представительной части каталогов для всех трех моделей распределения магнитуды были заданы значения параметров $M_c = 1$, $b = 1$. Для каждой модели и для каждой комбинации значений параметров был сгенерирован синтетический каталог объемом 10000 событий.

Далее по синтетическим каталогам с помощью метода бутстрап оценивались зависимости двух первых моментов распределений M_c от объема выборки N . Объем выборки варьировался в диапазоне от $N_{\min} = 20$ до $N_{\max} = 2000$. Нижняя граница диапазона соответствует минимальному объему выборки, позволяющему применять описанные процедуры оценки M_c . Верхняя граница выбиралась как достаточно большое значение объема выборки, позволяющее оценить асимптотические свойства оценок M_c при $N \rightarrow \infty$.

Как известно, состоятельной называется оценка, которая сходится по вероятности к оцениваемому параметру. Признаком состоятельной оценки служит асимптотическая несмещенность и убывание дисперсии с ростом объема выборки. Таким образом, от оценок M_c ожидалось, что с ростом объема выборки средние значения будут приближаться к истинному значению M_c , а их дисперсии будут снижаться.

РЕЗУЛЬТАТЫ И ОБСУЖДЕНИЕ

Результаты тестирования методов оценки M_c на синтетических каталогах показаны на рис. 9–рис. 11. На этих рисунках слева направо по столбцам показаны результаты для каталогов с номерами 1, 2, 3. Сверху вниз по строкам показаны результаты для методов MAXC, GFT, MBS, MBASS, EMR и LLS. На графиках горизонтальной сплошной чертой показано истинное значение M_c , точками показаны средние значения оценок M_c , пунктирными линиями обозначены границы одного стандартного отклонения от среднего.

Результаты для модели WW показаны на рис. 9. Оценка M_c^{MAXC} для каталога WW₁ медленно сходится к истинному значению M_c и, следовательно, является состоятельной. Оценка M_c^{MAXC} для каталога WW₃ также является состоятельной. Оценка M_c^{MAXC} для каталога WW₂ недооценивает истинное значение M_c на величину порядка 0.12 и имеет неубывающую дисперсию в силу особенностей формы распределения.

Оценка M_c^{GFT} для всех трех каталогов WW имеет при $N \rightarrow \infty$ асимптотическое смещение размером порядка -0.1 , -0.11 и -0.05 соответственно. Методы MBS и LLS дают схожие результаты, оба метода несколько недооценивают M_c на малых выборках, но с ростом N оценки постепенно сходятся к истинному значению M_c .

Результаты метода MBASS очень похожи на результаты метода MAXC, однако оценка M_c^{MBASS} для каталога WW₂ все же оказывается состоятельной. Метод EMR дает состоятельные оценки M_c для каталогов WW₁ и WW₂. Оценка M_c^{EMR} для каталога WW₃ переоценивает M_c на малых выборках, но с ростом N постепенно сходится к истинному значению M_c .

Результаты для модели AN показаны на рис. 10. Метод MAXC дает состоятельные оценки M_c для всех трех каталогов AN. Метод GFT занижает значение M_c для каталогов AN₁ и AN₂ на величину порядка 0.08 и 0.02 соответственно. Для каталога AN₃ метод GFT дает состоятельную оценку M_c . Методы MBS и LLS вновь демонстрируют схожие результаты, однако оценка M_c^{MBS} для всех трех каталогов AN имеет при $N \rightarrow \infty$ малое асимптотическое смещение размером порядка 0.02.

Оценки M_c^{MBASS} в среднем соответствуют оценкам M_c^{MAXC} , но имеют более высокую дисперсию. Для всех трех каталогов AN метод EMR переоценивает M_c на малых выборках, но с ро-

стом N оценка M_c^{EMR} постепенно сходится к истинному значению M_c . Результаты слабо зависят от параметра k .

Результаты для модели POL показаны на рис. 11. Описанные методы оказываются не в состоянии корректно определить значение M_c для модели POL. Особенно ярко это демонстрируют результаты метода MBASS, который катастрофически завышает значение M_c для всех трех каталогов POL. Ни одна из оценок не является состоятельной. Результаты слабо зависят от параметра M_i .

Полученные результаты показывают, что все описанные методы хорошо справляются с задачей определения M_c для модели AN, имеющей резкий переход между непредставительной и представительной частями выборки. Несколько хуже это получается в случае модели WW, со сглаженным максимумом распределения. Наконец, в случае модели POL, у которой этот переход практически никак не выражен, описанные методы оказываются не в состоянии корректно определить M_c .

С одной стороны, такой результат подсказывает, что следует с особым вниманием подходить к анализу представительности данных, относящихся к началу инструментальных каталогов. С другой стороны, наблюдения [García-Hernández et al., 2019], согласно которым модель POL в реальных каталогах встречается достаточно редко, дают надежду на то, что, возможно, на практике с таким распределением сталкиваться не придется.

Метод MAXC оказывается незаменимым при анализе представительности выборок из малых пространственно-временных объемов, для которых характерной является форма распределения AN. Это единственный метод, применимый для анализа представительности выборок экстремально малого объема $N \geq 4$ [Mignan et al., 2011]. Этот метод применялся для создания карт пространственных вариаций M_c высокого разрешения на Тайване [Mignan et al., 2011]. Для более объемных выборок, включающих сотни и более событий, в которых максимум распределения магнитуды обычно сглажен, метод MAXC занижает M_c и не рекомендуется к использованию.

Методы GFT и MBS – несложные методы, дающие реалистичные оценки M_c . Подходят для анализа представительности выборок среднего размера, содержащих от нескольких сотен до нескольких тысяч событий. Метод GFT применялся для картирования M_c на Аляске, в Западных США и Японии [Wiemer, Wyss, 2000]. Учитывая, что метод GFT, как правило, дает более мягкую оценку M_c , а метод MBS – более консервативную, следует выбирать один из них, основываясь на тех требованиях к полноте каталога, которые

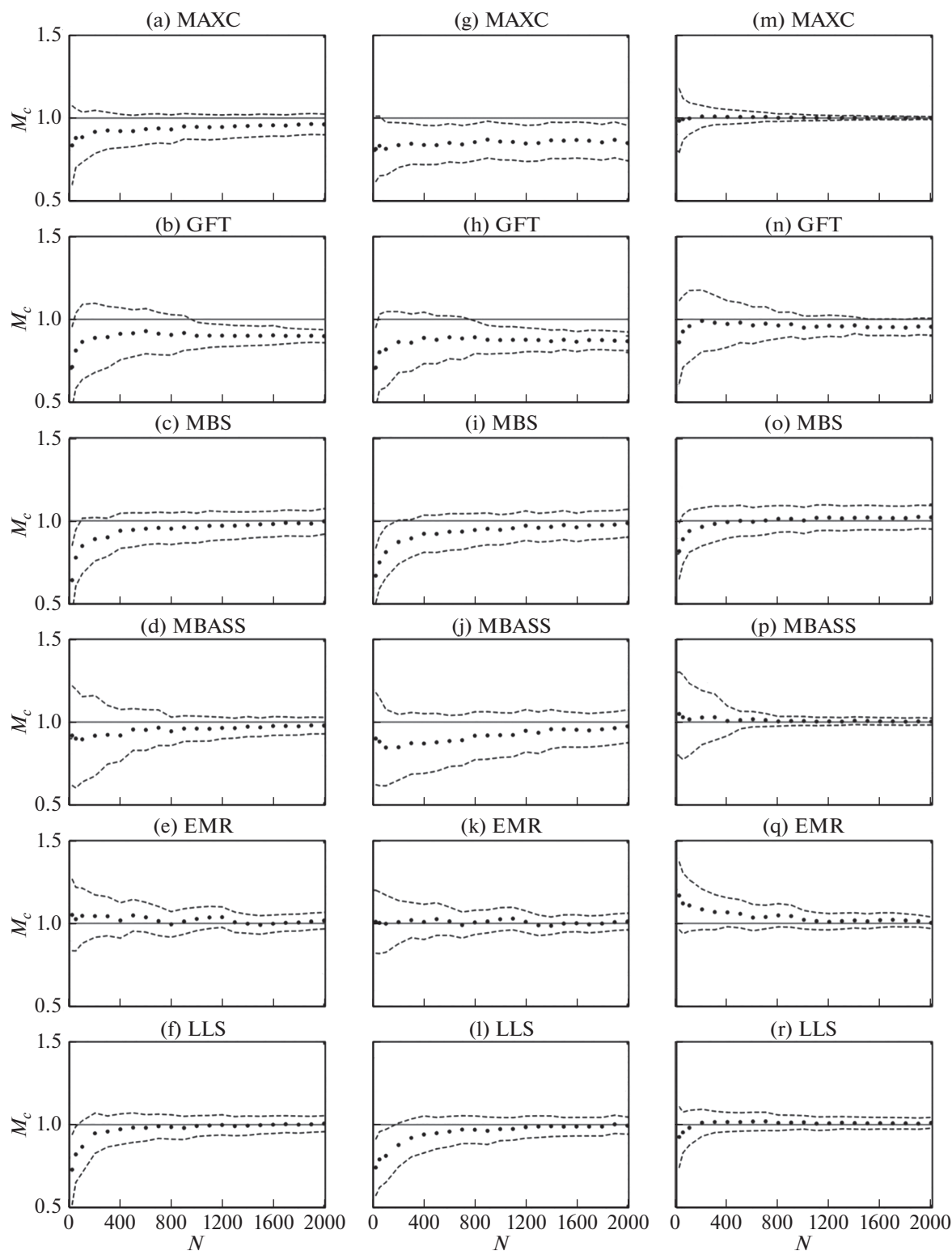


Рис. 9. Результаты тестирования методов оценки M_c на синтетических каталогах землетрясений, сгенерированных на основе модели WW. Слева направо по столбцам показаны результаты для каталогов с номерами 1, 2, 3. Сверху вниз по строкам показаны результаты для методов MAXC, GFT, MBS, MBASS, EMR и LLS. На графиках горизонтальной чертой показано истинное значение M_c , точками показаны средние значения оценок M_c , пунктирными линиями обозначены границы одного стандартного отклонения от среднего.

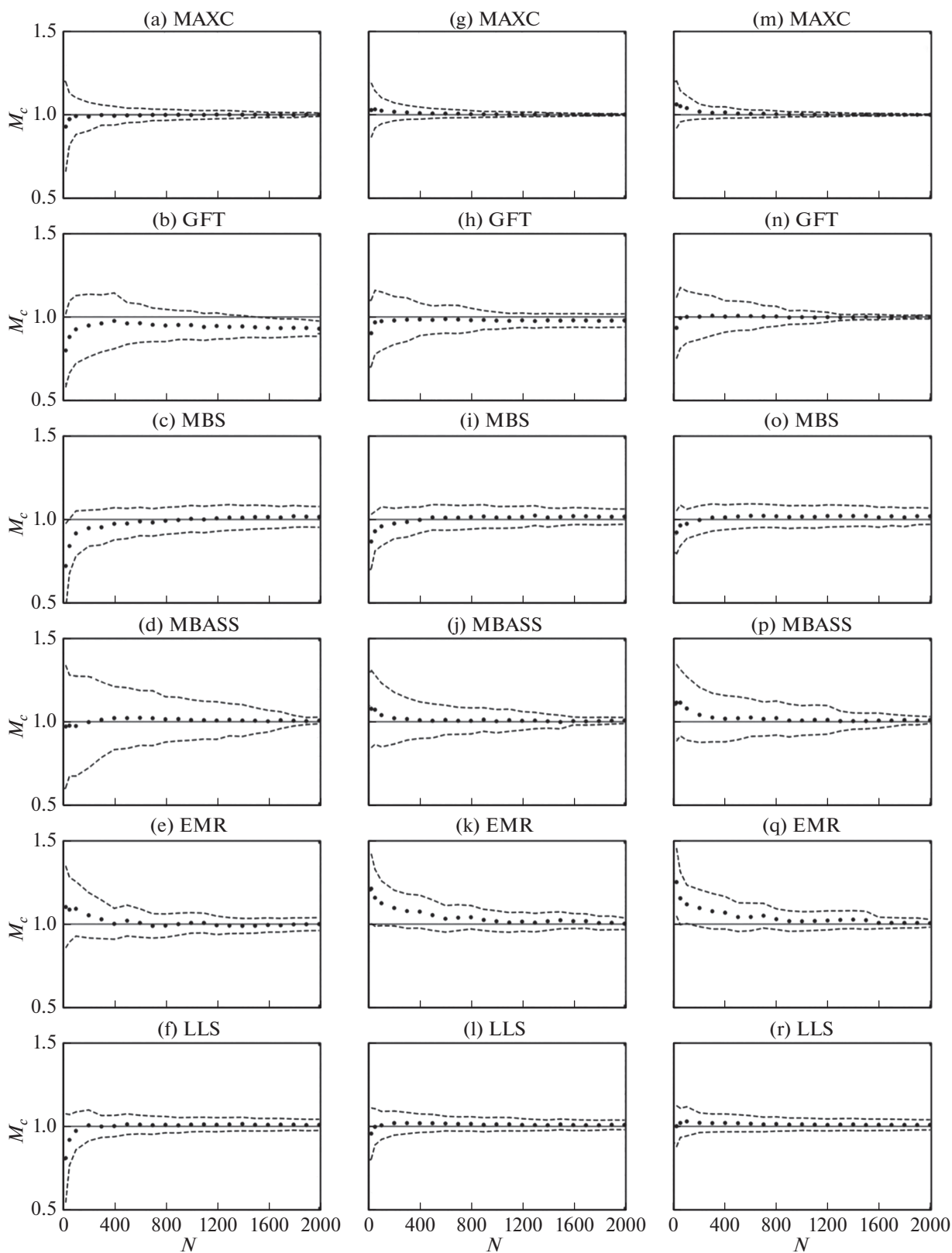


Рис. 10. Результаты тестирования методов оценки M_c на синтетических каталогах землетрясений, сгенерированных на основе модели AN.

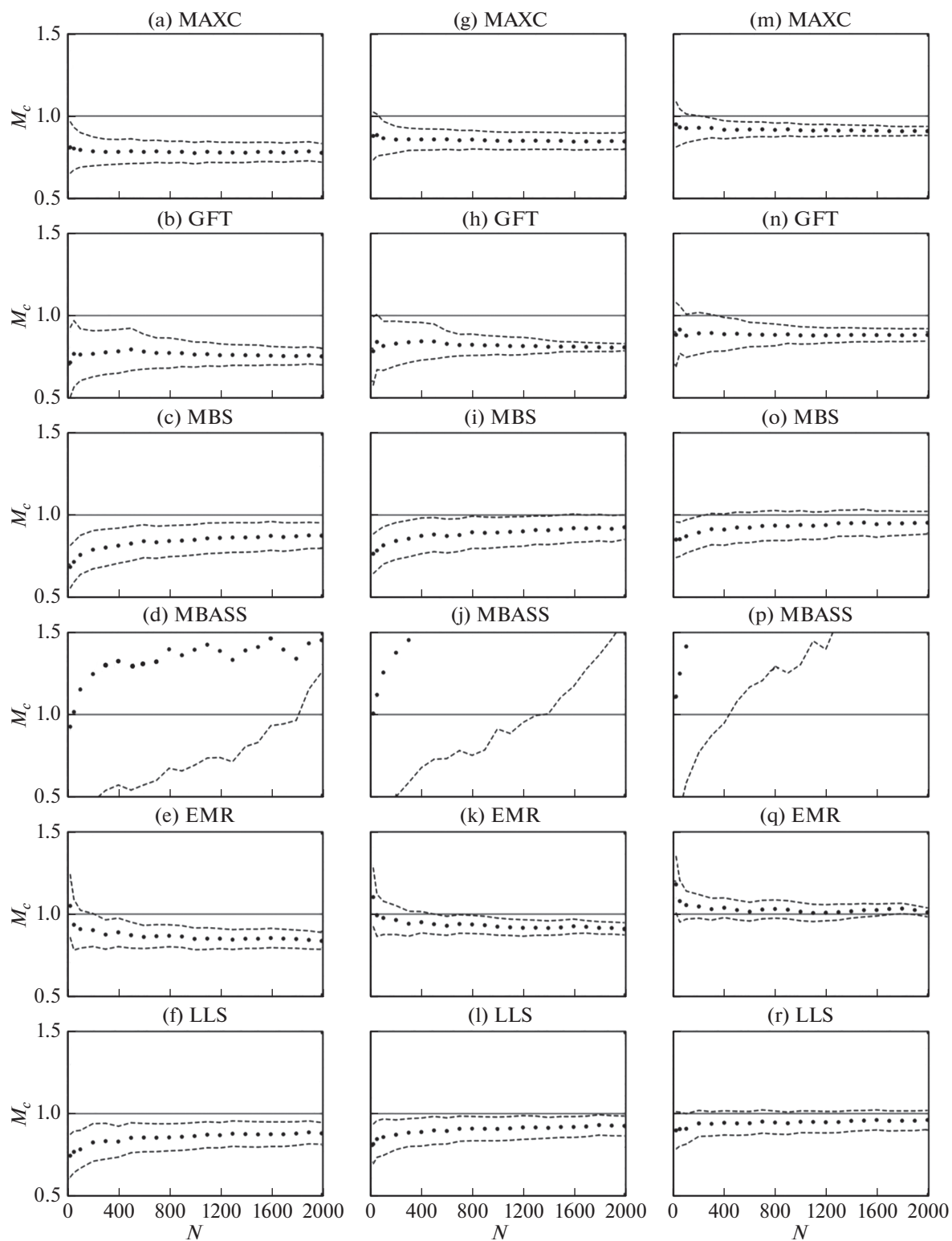


Рис. 11. Результаты тестирования методов оценки M_c на синтетических каталогах землетрясений, сгенерированных на основе модели POL.

предъявляет поставленная задача. Применять эти методы к более объемным выборкам следует с большой осторожностью.

Метод MBASS не обладает преимуществами перед более простым непараметрическим методом МАХС. В тех случаях, когда метод MBASS корректно определяет M_c , оценки оказываются близки к тем, что дает метод МАХС. При этом дисперсия оценок M_c^{MBASS} как правило оказывается выше. По этой причине метод MBASS не рекомендуется к использованию.

Метод EMR долгое время позиционировался как лучший метод для анализа представительности каталогов землетрясений. В частности, он применялся для картирования значений M_c по мировым каталогам ISC и CMT [Woessner, Wiemer, 2005], а также по каталогам землетрясений Южной Калифорнии [Hutton et al., 2010] и Японии [Nanjo et al., 2010]. Однако в итоге сами авторы метода EMR признали его неудачным [Mignan, Woessner, 2012] и призвали отказаться от его дальнейшего использования.

Метод LLS демонстрирует замечательные результаты: оценки M_c^{LLS} достаточно быстро сходятся к истинному значению M_c и имеют сравнительно невысокую дисперсию. Этот метод был внедрен в практику в организациях ЕГС РАН и применялся для картирования пространственно-временных неоднородностей M_c по каталогам землетрясений Новой Зеландии, Греции, Камчатки, Кавказа, Киргизии и Северного Китая [Smirnov, 1998]. При этом, судя по публикациям, этот метод остается практически неизвестным за рубежом. Метод LLS можно использовать для анализа представительности выборок среднего и большого размера.

В заключение стоит отметить, что картирование M_c в областях с высокими вариациями уровня регистрации, таких как Камчатка или Япония, должно опираться на анализ локальных распределений магнитуды, как было сделано, например, в работе [Mignan, 2012].

ВЫВОДЫ

В работе выполнен сравнительный анализ шести методов оценки магнитуды представительной регистрации землетрясений M_c . Для сравнения методов использовались как выборки из реальных инструментальных каталогов, так и синтетические каталоги землетрясений, сгенерированные на основе трех моделей распределения магнитуды. Для создания синтетических каталогов были использованы модели, соответствующие выборочным распределениям, встречающимся при анализе реальных данных инструментальных

каталогов: модель неоднородной выборки, модель однородной выборки, модель выборки, из которой удалены слабые землетрясения. Показано, что результаты применения описанных методов оценки M_c в значительной мере зависят от формы распределения и объема анализируемой выборки. По результатам анализа сформулированы рекомендации по выбору подходящих методов оценки M_c .

БЛАГОДАРНОСТИ

Авторы благодарят профессора В.Б. Смирнова за помощь в реализации метода LLS.

Авторы благодарят рецензентов Д.А. Сторчака и И.А. Воробьеву, а также анонимного куратора статьи за их ценные замечания.

СПИСОК ЛИТЕРАТУРЫ

- Гусев А.А. Прогноз землетрясений по статистике сейсмичности. Сейсмичность, сейсмический прогноз, свойства верхней мантии и их связь с вулканизмом на Камчатке. М.: Наука. 1974. С. 109–119.
- Завьялов А.Д. Наклон графика повторяемости как предвестник сильных землетрясений на Камчатке. Прогноз землетрясений. 1984. С. 173–184.
- Завьялов А.Д. Среднесрочный прогноз землетрясений: основы, методика, реализация. М.: Наука. 2006. 254 с.
- Писаренко В.Ф. Дискретные свойства геофизической среды. О законе повторяемости землетрясений. М.: Наука. 1989. С. 47–60.
- Смирнов В.Б., Габсатарова И.П. Представительность каталога землетрясений Северного Кавказа: расчетные данные и статистические оценки // Вестник ОГГГН РАН. 2000. № 4. С. 83–99.
- Смирнов В.Б. Прогностические аномалии сейсмического режима. Методические основы подготовки исходных данных // Геофизические исследования. 2006. № 2. С. 7–22.
- Akaike H. A new look at the statistical model identification // IEEE Trans Autom Control. 1974. V. 19. № 6. P. 716–723. <https://doi.org/10.1109/TAC.1974.1100705>
- Aki K. Maximum likelihood estimate of b in the formula $\log N = a - bM$ and its confidence limits // Bull. Earthq. Res. Inst. Univ. Tokyo. 1965. V. 43. P. 237–239.
- Amorèse D. Applying a change-point detection method on frequency-magnitude distributions // Bull. Seismol. Soc. Am. 2007. V. 97. № 5. P. 1742–1749. <https://doi.org/10.1785/0120060181>
- Bender B. Maximum likelihood estimation of b -values for magnitude grouped data // Bull. Seismol. Soc. Am. 1983. V. 73. № 3. P. 831–851.
- Bengoubou-Valérius M., Gibert D. Bootstrap determination of the reliability of b -values: an assessment of statistical estimators with synthetic magnitude series // Nat. Hazards. 2013. V. 65. № 1. P. 443–459. <https://doi.org/10.1007/s11069-012-0376-1>

- Cao A.M., Gao S.S.* Temporal variation of seismic b -values beneath northeastern Japan island arc // *Geophys. Res. Lett.* 2002. V. 29. №9. P. 48–1–48–3.
<https://doi.org/10.1029/2001GL013775>
- Chen K.-P., Tsai Y.-B., Amorèse D., Chang W.-Y.* Incorporating change-point detection updates of frequency-magnitude distributions within the Taiwan earthquake catalog // *Terr. Atmos. Ocean. Sci.* 2011. V. 22. № 3. P. 261–269.
[https://doi.org/10.3319/TAO.2010.09.17.01\(T\)](https://doi.org/10.3319/TAO.2010.09.17.01(T))
- Chernick M.R.* Bootstrap methods: a practitioner's guide. Wiley and Sons. 1999. 288 p.
- Efron B., Tibshirani R.* An introduction to the bootstrap. Chapman and Hall. 1993. 436 p.
- García-Hernández R., D'Auria L., Barrancos J., Padilla G.D.* On the functional expression of frequency-magnitude distributions: a comprehensive statistical examination // *Bull. Seismol. Soc. Am.* 2019. V. 109. № 1. P. 482–486.
<https://doi.org/10.1785/0120180197>
- Gibowicz S.J.* Variation of the frequency-magnitude relation during earthquake sequences in New Zealand // *Bull. Seismol. Soc. Am.* 1973. V. 63. № 2. P. 517–528.
- Goldstein M.L., Morris S.A., Yena G.G.* Problems with fitting to the power-law distribution // *Eur. Phys. J. B.* 2004. V. 41. № 2. P. 255–258.
<https://doi.org/10.1140/epjb/e2004-00316-5>
- Gomberg J.* Seismicity and detection/location threshold in the Southern Great Basin seismic network // *J. Geophys. Res.* 1991. V. 96. № B10. P. 16401–16414.
<https://doi.org/10.1029/91JB01593>
- Gutenberg B., Richter C.F.* Frequency of earthquakes in California // *Bull. Seismol. Soc. Am.* 1944. V. 34. № 4. P. 185–188.
- Guttorp P.* On least-squares estimation of b values // *Bull. Seismol. Soc. Am.* 1987. V. 77. № 6. P. 2115–2124.
- Hutton K., Woessner J., Hauksson E.* Earthquake monitoring in Southern California for seventy seven years (1932–2008) // *Bull. Seismol. Soc. Am.* 2010. V. 100. № 2. P. 423–446.
<https://doi.org/10.1785/0120090130>
- Mann H.B., Whitney D.R.* On a test of whether one of two random variables is stochastically larger than the other // *Ann. Math. Statist.* 1947. № 18. P. 50–60.
- Marzocchi W., Sandri L.* A review and new insights on the estimation of the b -value and its uncertainty // *Ann. Geophys.* 2003. V. 46. № 6. P. 1271–1282.
<https://doi.org/10.4401/ag-3472>
- Mignan A., Woessner J.* Estimating the magnitude of completeness for earthquake catalogs // *Community Online Resource for Statistical Seismicity Analysis.* 2012.
<https://doi.org/10.5078/corssa-00180805>
- Mignan A.* Functional shape of the earthquake frequency-magnitude distribution and completeness magnitude // *J. Geophys. Res.* 2012. V. 117. № B08302.
<https://doi.org/10.1029/2012JB009347>
- Mignan A., Werner M. J., Wiemer S., Chen C.-C., Wu Y.-M.* Bayesian estimation of the spatially varying completeness magnitude of earthquake catalogs // *Bull. Seismol. Soc. Am.* 2011. V. 101. № 3. P. 1371–1385.
<https://doi.org/10.1785/0120100223>
- Nanjo K. Z., Ishibe T., Tsuruoka H., Schorlemmer D., Ishigaki Y., Hirata N.* Analysis of the completeness magnitude and seismic network coverage of Japan // *Bull. Seismol. Soc. Am.* 2010. V. 100. № 6.
<https://doi.org/10.1785/0120100077>
- Nuannin P., Kulhanek O., Persson L.* Spatial and temporal b value anomalies preceding the devastating off coast of NW Sumatra earthquake of December 26, 2004 // *Geophys. Res. Lett.* 2005. V. 32. № L11307.
<https://doi.org/10.1029/2005GL022679>
- Ogata Y., Katsura K.* Analysis of temporal and spatial heterogeneity of magnitude frequency distribution inferred from earthquake catalogs. // *Geophys. J. Int.* 1993. V. 113. P. 727–738.
- Ringdal F.* On the estimation of seismic detection thresholds // *Bull. Seismol. Soc. Am.* 1975. V. 65. № 6. P. 1631–1642.
- Rydelek P.A., Sacks I.S.* Testing the completeness of earthquake catalogues and the hypothesis of self-similarity // *Nature.* 1989. 337. P. 251–253.
<https://doi.org/10.1038/337251a0>
- Shi Y., Bolt B.A.* The standard error of the magnitude-frequency b -value // *Bull. Seismol. Soc. Am.* 1982. V. 72. № 5. P. 1677–1687.
- Smirnov V.B.* Earthquake catalogs: evaluation of data completeness // *Volc. Seism.* 1998. V. 19. P. 497–510.
- Stepp J. C.* Analysis of completeness of the earthquake sample in the Puget Sound area and its effects on statistical estimates of earthquake hazard. Proceedings of the International conference on microzonation for safer construction research and application. Seattle, USA. 1972. P. 897–909.
- Storchak D.A., Giacomo D.D., Bondár I., Engdahl E.R., Harris J., Lee W.H.K., Villaseñor A., Bormann P.* Public release of the ISC-GEM global instrumental earthquake catalog (1900–2009) // *Seismol. Res. Lett.* 2013. V. 84. № 5. P. 810–815.
<https://doi.org/10.1785/0220130034>
- Utsu T.* A statistical significance test of the difference in b -value between two earthquake groups // *J. Phys. Earth.* 1966. V. 14. № 2. P. 37–40.
<https://doi.org/10.4294/jpe1952.14.37>
- Wiemer S., Wyss M.* Minimum magnitude of completeness in earthquake catalogs: examples from Alaska, the Western United States, and Japan // *Bull. Seismol. Soc. Am.* 2000. V. 90. № 4. P. 859–869.
<https://doi.org/10.1785/0119990114>
- Wilcoxon F.* Individual comparisons by ranking methods // *Biometrics Bulletin.* 1945. V. 1. № 6. P. 80–83.
- Woessner J., Wiemer S.* Assessing the quality of earthquake catalogues: estimating the magnitude of completeness and its uncertainty // *Bull. Seismol. Soc. Am.* 2005. V. 95. № 2. P. 684–698.
<https://doi.org/10.1785/0120040007>
- Wyss M., Hasegawa A., Wiemer S., Umino N.* Quantitative mapping of precursory seismic quiescence before the 1989, M 7.1 off-Sanriku earthquake, Japan // *Ann. Geofisc.* 1999. V. 42. № 5. P. 851–869.
<https://doi.org/10.4401/ag-3765>
- Zoller G., Hainzl S., Kurths J., Zschau J.* A systematic test on precursory seismic quiescence in Armenia // *Nat. Hazards.* 2002. V. 26. № 3. P. 245–263.
<https://doi.org/10.1023/a:1015685006180>

Comparative Analysis of the Methods for Estimating the Magnitude of Completeness of Earthquake Detection

V. A. Pavlenko^{a,*} and A. D. Zavyalov^{a,**}

^a*Schmidt Institute of Physics of the Earth, Russian Academy of Sciences, Moscow, Russia*

^{*}*e-mail: pavlenko.vasily@gmail.com*

^{**}*e-mail: zavyalov@ifz.ru*

The magnitude of completeness M_c of the earthquakes above which 100% of events are thought to be reliably detected is a vital parameter characterizing the completeness of seismic data. A known fact is that to obtain correct estimates of the parameters of seismicity, it is compulsory to take into account variations in space and time. In this work, we compare six modern methods of M_c estimation. To compare the methods, we use event samples from real instrumental earthquake catalogs and synthetic catalogs generated based on three models of magnitude distribution. We analyze the dependences of the two first moments of the distributions of M_c estimates on the shape of magnitude distributions and the sample size. We use three models corresponding to sample distributions that occur in the analysis of instrumental earthquake catalogs. Based on the obtained results, we formulate recommendations on selecting the suitable method for estimating the magnitude of completeness M_c .

Keywords: magnitude of completeness of earthquake detection, seismic data completeness, instrumental catalogs, synthetic catalogs