

Анализ влияния очистки каталогов от афтершоков на эффективность систематического прогноза землетрясений¹

В.Г. Гитис, А.Б. Дерендяев, К.Н. Петров

Институт проблем передачи информации, Российская академия наук, Москва, Россия

Поступила в редколлегию 07.06.2019

Аннотация—В статье проведена оценка эффективности прогноза землетрясений методом минимальной области тревоги при очистке каталогов землетрясений от афтершоков. Тестирование проведено с каталогами землетрясений Японии и Средиземноморья за период 1988–2019 гг. Сравниваются результаты прогноза целевых землетрясений с афтершоковыми последовательностями и прогноз только основных толчков, полученные по сеточным полям, которые вычислялись по очищенным и неочищенным от афтершоков каталогам землетрясений. В результате тестирование показало, что очистка каталога целевых землетрясений от афтершоков ухудшает качество прогноза. Возможная причина состоит в уменьшении числа примеров для обучения. Очистка каталогов землетрясений при вычислении полей признаков прогноза целесообразна при наличии в регионе сильной афтершоковой активности.

КЛЮЧЕВЫЕ СЛОВА: анализ данных, машинное обучение, метод минимальной области тревоги, прогноз землетрясений, афтершоки.

1. ВВЕДЕНИЕ

Сильные землетрясения вызывают большой социальный и экономический ущерб. При систематическом прогнозе землетрясений требуется регулярно определять ограниченные по объему пространственно-временные области сейсмической тревоги, в которых ожидается эпицентр землетрясения с целевой магнитудой больше некоторого заранее заданного порога. Большинство землетрясений происходят в результате разрывов геологической среды под действием накопленных тектонических деформаций. Такие землетрясения называются тектоническими. Часто после достаточно сильного землетрясения вокруг плоскости разрыва возникают повторные толчки, называемые афтершоками. Обычно магнитуда афтершоков на 1–2 единицы меньше магнитуды основного события, но, тем не менее, они представляют серьезную сейсмическую опасность [1].

Серии афтершоков следуют почти за всеми тектоническими землетрясениями. Они обычно тесно связаны с областью очага основного землетрясения и по сейсмическим параметрам не отличаются ни от них, ни от фоновой сейсмичности [2, 3]. В то же время предполагается, что временная последовательность повторных толчков для всех последовательностей афтершоков статистически одинакова [4, 5]. Это позволяет разделить каталог землетрясений на афтершоки и основные толчки.

В статье мы рассматриваем результаты анализа свойств метода минимальной области тревоги, связанных с наличием в каталоге землетрясений сильных афтершоков с целевыми магнитудами. Обычно при систематическом прогнозе землетрясений целевыми событиями являются

¹ Авторы выражают свои благодарности Смирнову В.Б. за помощь в написании программы и Пирогову С.А. за консультации и полезные замечания. Исследование выполнено в ИППИ РАН частично за счет гранта РФФИ 17-07-00494.

все землетрясения с магнитудой больше заданного порога. Они включают как главные толчки, так и их сильные афтершоки. При этом на оценку качества прогноза не влияет, какое землетрясение было обнаружено. Может оказаться так, что главное землетрясение не было спрогнозировано, а были спрогнозированы в основном зависимые от главного толчка афтершоки. В работе мы сравниваем три варианта прогноза: (1) прогноз всех целевых землетрясений, (2) прогноз главных толчков и (3) прогноз афтершоков. В каждом случае прогноз выполнялся дважды: по полям признакам, вычисленным по полному каталогу землетрясений, и по каталогу, очищенному от афтершоков.

Для исследований применялась ГИС ГеоТайм 3 [6], для которой был разработан плагин, выполняющий разбиение каталога землетрясений на основные события и афтершоки. Для его разработки использовались результаты, опубликованные в работах [7, 8]. Идея метода идентификации афтершоковых последовательностей описана в разделе 2. Тестирование прогноза землетрясений выполнялось с помощью метода минимальной области тревоги [9]. Метод кратко описан в разделе 3. Результаты моделирования приведены в разделе 4.

2. МЕТОД ИДЕНТИФИКАЦИИ АФТЕРШОКОВ

Основные предположения при идентификации афтершоков состоят в том, что афтершоки концентрируются в пространственно-временной области вокруг основного события, смешиваются с фоновой сейсмичностью и их временная последовательность статистически однородна. Эта модель позволяет выделять афтершоки алгоритмически.

Метод идентификации афтершоков предложен в [7]. Идея алгоритма состоит в следующем. Вводится функция потерь от ошибок I рода (ложная классификация фонового события вместо афтершока) и II рода (классификация пропуска афтершока). Задача идентификации состоит в нахождении правила, которое минимизирует функцию потерь. В качестве такого правила выбрана дискриминантная функция. Предполагается, что землетрясения в очищенном от афтершоков каталоге, распределены по закону Пуассона, а афтершоки, по отношению к основному событию, имеют степенное распределение во времени (закон Омори) и двумерное распределение Гаусса в пространстве. Для выделения афтершоков, данный метод локально сравнивает отношение пространственно-временного распределения афтершоковой и фоновой активности с порогом, который строится на основе правила минимакса для некоторой функции потерь.

3. МЕТОД МИНИМАЛЬНОЙ ОБЛАСТИ ТРЕВОГИ

Метод минимальной области тревоги подробно рассмотрен в [10, 11]. Метод решает задачу одноклассовой классификации [12–14]. Для обучения используются ретроспективные данные: эпицентры землетрясений с магнитудами $m \geq M$ и набор пространственных и пространственно-временных сеточных полей, значениям которых соответствуют векторы пространства признаков прогноза.

Модель данных содержит два предположения.

1. Условие аномальности: целевым землетрясениям предшествуют векторы пространства признаков (предвестники землетрясений), у которых значения некоторых компонент (значения некоторых полей признаков) маловероятны и близки к максимальным (или минимальным).
2. Условие монотонности: векторы пространства признаков, которые покомпонентно больше (или меньше) вектора предвестника землетрясений, также могут быть предвестниками аналогичных целевых событий. Для упрощения изложения будем считать, что аномалии близки только к максимальным значениям полей.

Модель позволяет вести меру аномальности векторов пространстве признаков. Эту меру алгоритм использует, чтобы выделить из всех векторов пространства признаков те, которые можно считать предвестниками произошедших целевых землетрясений. После этого алгоритм классификации обучается обнаруживать аномальные объекты по выборке, в которой присутствуют помеченные целевые аномальные объекты и непомеченная смесь аномальных и нормальных объектов. В результате обучения строится поле тревоги, для которого, как и для полей признаков, выполняются предположения (1) и (2). Землетрясение обнаружено, если его эпицентр попадает в область с допустимыми значениями объема тревоги. Качество прогноза определяют два показателя: (1) оценка вероятности обнаружения U , равная доле правильно обнаруженных целевых событий Q^* от всех целевых Q событий, $U = Q^*/Q$, и (2) объем тревоги V , равный доле числа узлов сетки в области тревоги L^* от числа всех узлов сетки L области, в которой производится анализ, $V = L^*/L$. Обычно качество прогноза определяется по зависимости $U(V)$, которая практически совпадает с кривой ошибок, представляемых ROC-кривой [15].

Алгоритм метода минимальной области тревоги является непараметрическим. Имеются три наиболее важные версии алгоритма. Первая версия алгоритма состоит в построении поля тревоги, с помощью которого целевые события обнаруживаются в той последовательности, в которой увеличиваются объемы тревоги, относящиеся к прогнозу по каждому предвестнику землетрясений в отдельности, $v(1) \leq v(2) \leq \dots \leq v(Q)$ (эта версия выбрана для тестирования). Вторая версия алгоритма позволяет оптимизировать поле тревоги так, чтобы при добавлении каждого предвестника объем тревоги увеличивался на минимальную величину. Для этого на каждом шаге от события $q + 1$ к q , производится перебор, с помощью которого выбирается такой предвестник, при котором объем тревоги увеличивается минимально. Третья версия алгоритма позволяет оптимизировать поле тревоги так, что оно обнаруживает максимальное число событий при объеме тревоги не более заданного. В этом случае для выбора предвестников требуется выполнить полный перебор.

4. ТЕСТИРОВАНИЕ

4.1. Методика

Тестирование имитирует рабочий режим прогноза. Как и прогноз, оно выполняется с постоянным шагом Δt . Обозначим: t_0 — начало вычислений полей признаков, t_1 — начало обучения, t_2 — начало тестирования (прогноза), t_3 — конец тестирования и момент проведения теста t (момент прогноза), $t_3 \geq t > t_2$. Обучение по алгоритму минимальной области тревоги сначала выполняется на интервале (t_1, t_2) , а затем продолжается с шагом Δt на каждом интервале (t_1, t) . При этом на каждом шаге область тревоги обновляется. В момент t временной срез поля тревоги показывает карту тревоги на интервале $(t, t + \Delta t)$. Для целевого события на интервале $(t, t + \Delta t)$ по карте тревоги определяется значение поля тревоги V в точке эпицентра события. Если значение V менее допустимого, например, менее 0.2, то считается, что землетрясение предсказано. В результате тестирования мы получаем две оценки показателей прогноза: объем тревоги $V = L^*/L$, вычисленный на полном интервале обучения (t_1, t_3) , и вероятность (оценка вероятности) обнаружения целевых событий $U = Q^*/Q$, вычисленная на интервале тестирования (t_2, t_3) по данным, не участвовавшим в обучении. Заметим здесь, что алгоритм метода минимальной области тревоги работает так, что зона тревоги для интервала $(t, t + \Delta t)$ увеличивается во время сейсмической активизации и уменьшается при сейсмическом затишье. При этом в целом на всем интервале обучения размер пространственно-временной области тревоги $L^* = VL$ не изменяется.

Активность сейсмического процесса в области анализа как правило неоднородна. Если зона, где происходят целевые землетрясения, мала по сравнению с размером области анализа, то возможно тривиальное решение. Оно состоит в том, что объявление постоянной тревоги в

относительно малой подобласти анализа может дать весьма высокую вероятность прогноза при малом объеме тревоги. Поэтому при прогнозе землетрясений весьма важен выбор области анализа. Для тестирования область анализа выбрана по пространственному полю плотности эпицентров землетрясений с представительными магнитудами, вычисленному на интервале (t_0, t_2) . Область анализа определяют границы поля пространственной плотности эпицентров землетрясений со значениями выше такого порога, при котором поле обнаруживает целевые события с вероятностью порядка 0.4–0.5.

Тестирование выполнено для регионов Средиземноморья и Японии.

Для прогноза анализировались следующие 6 полей:

1. F_1 — 3D поле плотности эпицентров землетрясений.
2. F_2 — 3D поле средних магнитуд землетрясений.
3. F_3 — 3D поле отрицательных временных аномалий плотности эпицентров.
4. F_4 — 3D поле положительных временных аномалий плотности эпицентров.
5. F_5 — 3D поле положительных временных аномалий средних магнитуд землетрясений.
6. F_6 — 2D поле плотности эпицентров землетрясений: ядерное сглаживание на интервале 1988–2008 гг. с параметром $R = 50$ км.
7. F_7 — 3D поле минимальных значений квантилей распределения сейсмической активности на интервале (t_0, t_1) которые соответствуют значениям сейсмической активности, вычисленными на текущем шаге $t \in (t_1, t_3)$ с помощью ядерной функции.

Оценивание полей F_1, F_7, F_8 и F_2 выполнялось с помощью метода локальной ядерной регрессии. Ядерная функция для q -го землетрясения имеет вид $K_q = [ch^2(r_q/R)^2 ch^2(t_q/T)]^{-1}$, где $r_q < R\epsilon$, $t_q < T\epsilon$ — расстояние и временной интервал между q -м эпицентром землетрясения и узлом сетки 3D поля, $\epsilon = 2$, $R = 50$ км, $T = 100$ дня для F_1, F_7 и F_8 , а для F_2 , $R = 100$ км, $T = 730$ дня. Поле F_6 , оценивалось на интервале 1988–2008 гг. Для оценивания полей использовалось ядро $K_q = [ch^2(r_q/R)^2]$, $r_q < R\epsilon$, с параметрами $\epsilon = 2$, $R = 50$ км. Поля F_3, F_4, F_5 представляют значение статистик Стьюдента от полей F_1 и F_2 , которые вычислялись для каждого узла сетки как отношение разности средних значений поля для текущего момента времени на 196 днях и для фонового прошлого на 3650 днях к оценке среднеквадратичного отклонения этой разности. Положительное значение аномалий соответствует увеличению значений полей F_1 и F_2 на текущем интервале.

Для разделения каталогов землетрясений на основные события и афтершоки использовалась ГИС GeoТайм 3. При очистке каталогов от афтершоков, в качестве основных событий рассматривались все землетрясения с магнитудами $M \geq 4$. Для пересчёта магнитуды в энергию землетрясения использовалась формула $E = 10^{(1.5+4.8m)}$.

Тестирование выполнялось по следующим данным. Поля признаков прогноза вычислялись по полному каталогу и по каталогу без афтершоков. Целевые землетрясения также выбирались из полного и очищенного от афтершоков каталогов. Наиболее информативные поля признаков выбирались по полным каталогам землетрясений. Для выбора использовался шаговый метод: выбор наилучшего признака прогноза, выбор второго признака, дающего вместе с первым наилучший прогноз и т.д. Выбор признаков заканчивается, если наилучший добавляемый признак не вносит значимого улучшения в качество прогноза.

Регион Японии характеризуется высокой сейсмической активностью. Для экспериментов был выбран полигон 128° – 148° в.д., 28° – 45° с.ш. Сейсмический процесс представлен каталогом землетрясений на интервале 18.05.1988–03.09.2019 гг., взятом с сайта Японского Метеорологического Агенства [16, 17]. Каталог был обрезан по величине минимальной представительной магнитуды, от $1.5(M \geq 1.5)$ и по глубине гипоцентров землетрясений $H \leq 160$ км. Итоговый размер каталога в области анализа составил 1035372 события, из которых 392291 (37.9 %) являются афтершоками. Сеточные поля вычислены с шагом $\Delta x \times \Delta y \times \Delta t = 0.2^\circ \times 0.17^\circ \times 100$ суток.

Интервал тестирования $(t_2, t_3) = (14.01.2013, 03.09.2019)$. Дата начала тестирования выбрана так, чтобы ограничить влияние на прогноз афтершоков Великого восточного японского землетрясения в префектуре Фукусима с магнитудой 9.0, которое произошло 11 марта 2011 г. Тем не менее, афтершоки этого землетрясения оказывают существенное влияние на результаты обучения на интервале (08.08.2000, 14.01.2013). Прогнозное поле обучено при параметрах цилиндра тревоги: радиусе $R = 25$ км и размере цилиндра по времени $T = 102$ дня. Тестировались целевые землетрясения с магнитудами $m \geq 6.0$ и $m \geq 5.5$. Наиболее информативными для региона Японии оказались поля признаков F_1, F_2 и F_7 . Результаты тестирования сведены в таблицы 1 и 2.

Таблица 1. Результаты обнаружения целевых событий с $m \geq 6.0$.

	Объём тревоги	Целевые события — все землетрясения, $Q = 49$		Целевые события — основные толчки, $Q = 42$	
		Поле F_6	Поля F_1, F_2, F_7	Поле F_6	Поля F_1, F_2, F_7
Поля признаки из полного каталога	5%	0.16	0.28	0.14	0.29
	10%	0.26	0.45	0.26	0.38
	20%	0.41	0.57	0.40	0.55
Поля признаки из очищенного каталога	5%	0.16	0.35	0.14	0.26
	10%	0.29	0.41	0.28	0.43
	20%	0.43	0.67	0.45	0.69

Более развернутые результаты для прогноза землетрясений с магнитудой $m \geq 6.0$ представлены на рисунке 1, а с магнитудой $m \geq 5.5$ на рисунке 2, как графики зависимости вероятностей обнаружения от объема тревоги $U(V)$.

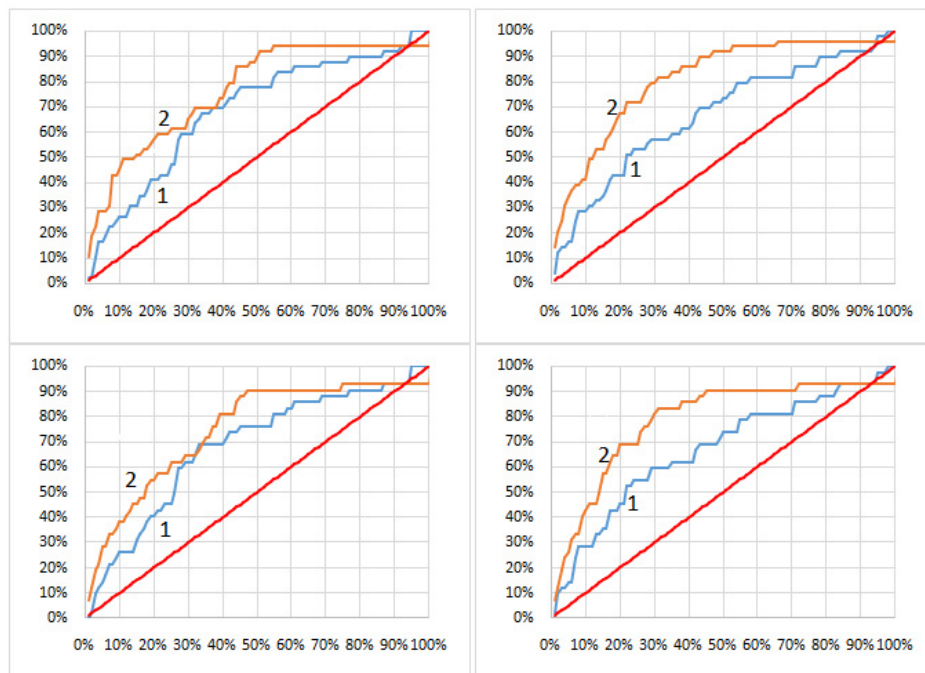


Рис. 1. Графики зависимостей $U(V)$ вероятностей обнаружения от объема тревоги для региона Япония, для магнитуды $m \geq 6.0$. Левые графики посчитаны по каталогу неочищенному от афтершоков, правые по каталогу очищенному от афтершоков. На первой паре графиков целевые события из неочищенного каталога, у второй пары из очищенного. Цифра 2 — графики регулярного прогноза по полям F_1, F_2 и F_7 , цифра 1 — графики прогноза по полю F_6 .

Из таблицы 1 и рисунка 1 видно, что при прогнозе целевых событий с магнитудами $m \geq 6.0$ при объеме тревоги $V = 0.2$ наилучшие результаты, равные вероятностям обнаружения 0.68 И 0.69, получены при прогнозе при использовании полей-признаков, вычисленных по каталогу землетрясений очищенных от афтершоков. Можно предположить, что это такой результат объясняется очисткой каталога землетрясений от афтершоков землетрясения в префектуре Фукусима. Этому предположению не противоречит результат прогноза целевых событий с магнитудами $m \geq 5.5$ (см таблицу 2 и рисунок 2).

Таблица 2. Результаты обнаружения целевых событий с $m \geq 5.5$.

	Объём тревоги	Целевые события — все землетрясения, $Q = 49$		Целевые события — основные толчки, $Q = 42$	
		Поле F_6	Поля F_1, F_2, F_7	Поле F_6	Поля F_1, F_2, F_7
Поля признаки из полного каталога	5%	0.20	0.36	0.21	0.38
	10%	0.30	0.54	0.34	0.52
	20%	0.44	0.70	0.52	0.68
Поля признаки из очищенного каталога	5%	0.24	0.48	0.25	0.43
	10%	0.32	0.60	0.35	0.61
	20%	0.45	0.76	0.55	0.72

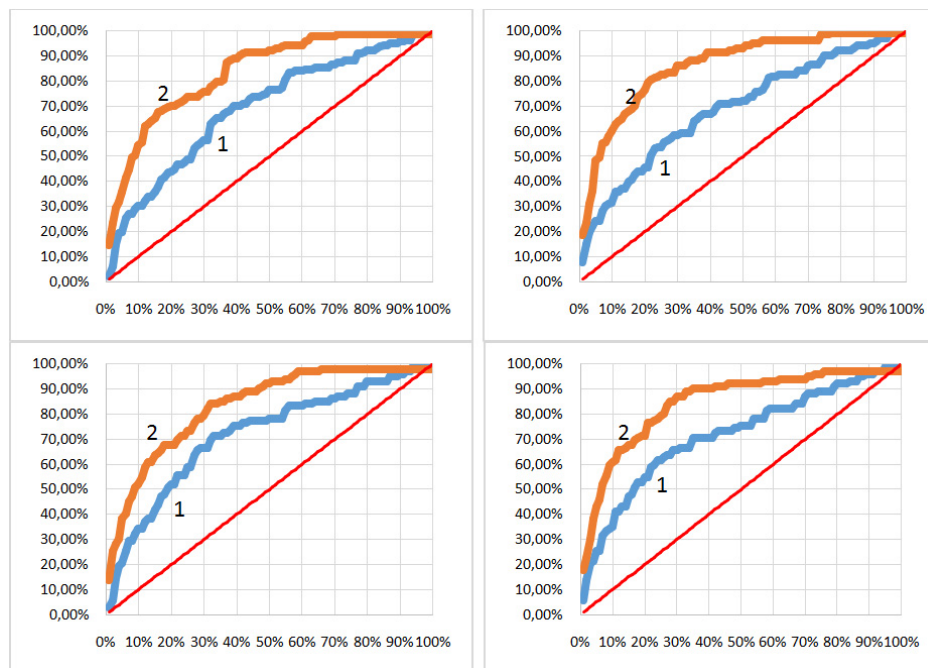


Рис. 2. Графики зависимостей $U(V)$ вероятностей обнаружения от объема тревоги для региона Японии, для магнитуды $m \geq 5.5$. Левые графики посчитаны по каталогу неочищенному от афтершоков, правые по каталогу очищенному от афтершоков. На первой паре графиков целевые события из неочищенного каталога, у второй пары из очищенного. Цифра 2 — графики регулярного прогноза по полям F_1, F_2 и F_7 , цифра 1 — графики прогноза по полю F_6 .

Регион Средиземноморье: 10° – 30° в.д., 34° – 47° с.ш. Сейсмический процесс представлен каталогом землетрясений, взятом с сайта Международного Сейсмологического центра (ISC — <http://www.isc.ac.uk/>). Каталог представляет события на интервале 18.05.1988–03.09.2019 гг. Каталог обрезан по величине минимальной представительной магнитуды $m \geq 2.7$ и по глубине гипоцентров землетрясений до 160 км. Итоговый размер каталога в области анализа составил 184372 события, из которых 50913 (27.6%) — афтершоки. Сеточные поля вычислены с шагом

$\Delta x \times \Delta y \times \Delta t = 0.2^\circ \times 0.17^\circ \times 100$ дня. Интервал тестирования $(t_2, t_3) = (18.12.2012, 03.09.2019)$. Прогнозное поле обучено при параметрах цилиндра тревоги: радиусе $R = 25$ км и размере цилиндра по времени $T = 100$ дня.

В качестве целевых событий выбраны землетрясения с магнитудой больше или равной 5.7 ($m \geq 5.7$) и глубиной гипоцентров меньшей или равной 60 км.

Наиболее информативными для региона Средиземноморья оказались поля признаков F_1 и F_3 . Результаты тестирования сведены в таблицу 3.

Таблица 3. Результаты обнаружения целевых событий с $m \geq 5.7$.

	Объем тревоги	Целевые события — все землетрясения, $Q = 17$		Целевые события — основные толчки, $Q = 14$	
		Поле F_6	Поля F_1, F_3	Поле F_6	Поля F_1, F_3
Поля признаки из полного каталога	5%	0.06	0.53	0.0	0.43
	10%	0.06	0.59	0.07	0.5
	20%	0.47	0.71	0.50	0.64
Поля признаки из очищенного каталога	5%	0.0	0.0	0.0	0.0
	10%	0.0	0.29	0.08	0.29
	20%	0.18	0.47	0.21	0.50

Более развернутые результаты представлены на рисунке 3 графиками зависимостей $U(V)$ вероятностей обнаружения от объема тревоги.

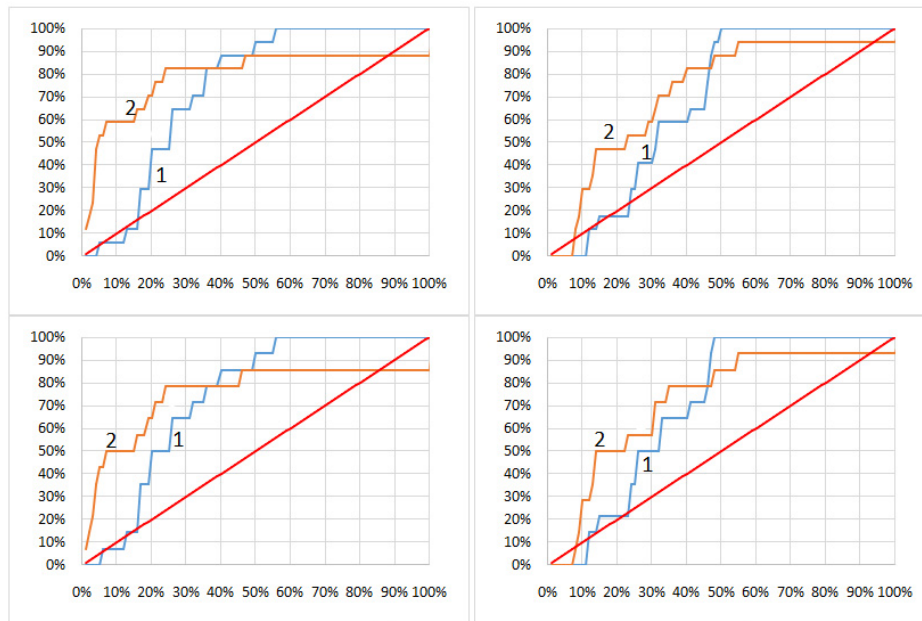


Рис. 3. Графики зависимостей $U(V)$ вероятностей обнаружения от объема тревоги для региона Средиземноморья, для магнитуды $m \geq 5.7$. Левые графики посчитаны по каталогу неочищенному от афтершоков, правые по каталогу очищенному от афтершоков. На первой паре графиков целевые события из неочищенного каталога, у второй пары графиков целевые события из очищенного. Цифра 2 — графики регулярного прогноза по полям F_1 и F_3 , цифра 1 — графики прогноза по полю F_6

Из таблицы 3 и рисунка 3 видно, что при объеме тревоги $V = 0.2$ наилучший результат, равный вероятности обнаружения $U = 0.71$, получен при прогнозе всех целевых событий по полям признаков, которые вычислялись по полному каталогу землетрясений.

5. ВЫВОД

Исследование показало, что при удалении афтершоков из списка целевых событий, успешность прогноза падает для обоих регионов. Возможная причина состоит в уменьшении количества целевых событий, на которых на каждом шаге прогноза происходит обучение по всем ретроспективным данным.

Тестирование показало, что очистка каталога от афтершоков для вычисления полей признаков может неоднозначно влиять на результаты. Можно видеть, что для региона Японии при вычислении полей признаков по каталогу, очищенному от афтершоков, результат прогноза улучшился, в то время как для Средиземноморья результат ухудшился. Это можно объяснить сильным влиянием на сейсмический процесс большого числа афтершоков, индуцированных землетрясением в префектуре Фукусима с магнитудой 9.0. В регионе Средиземноморья на интервале тестирования также наблюдалось существенное увеличение сейсмической активности. Но в отличие от предыдущего примера этот процесс не был локализован одной относительно небольшой зоной региона, а относился весьма большой пространственной области. Очистка каталога от афтершоков здесь могла привести к уменьшению точности оценивания полей признаков.

По результатам эксперимента можно сделать три предварительных вывода: (1) Прогноз землетрясений более эффективен, если при обучении поля тревоги используются все целевые землетрясения; (2) Оценка качества прогноза является более информативной, если подсчет оценок вероятности обнаружения событий ведется отдельно для основных толчков и для афтершоков; (3) При вычислении полей признаков очистка каталогов землетрясений от афтершоков целесообразна, если в них имеются большие афтершоковые последовательности.

СПИСОК ЛИТЕРАТУРЫ

1. Татевосян Р. Э., Аптекман Ж. Я. Этапы развития афтершоковых последовательностей сильнейших землетрясений мира. *Физика Земли*, 2008, № 12, стр. 3–23.
2. Гульельми А. В., Зотов О. Д., Завьялов А. Д. Динамика афтершоков Суматра-Андаманского землетрясения. *Физика Земли*, 2014, № 1, стр. 66–66
3. Лукк А. А. Последовательность афтершоков Джурмского глубокофокусного землетрясения 14 марта 1965 г. *Изв. АН СССР. Физика Земли*, 1968, № 5, стр. 83.
4. Omori F. On aftershocks. *Seismological journal of Japan*, 1894, vol. 19, pp. 71–80.
5. Utsu T. A statistical study on the occurrence of aftershocks. *Geophysics Magazine*, 1961, vol. 30, pp. 521–605.
6. Gitis V. G., Derendyaev A. B., Metrikov P. A., Shogin, A. Network geoinformation technology for seismic hazard research. *Natural hazards*, 2012, vol.62 no. 3, pp. 1021–1036.
7. Молчан Г. М., Дмитриева О. Е. Идентификация афтершоков: обзор и новые подходы. *Вычислительная сейсмология*, 2003, стр. 19–50.
8. Смирнов В. Б. Прогностические аномалии сейсмического режима. I. Методические основы подготовки исходных данных. *Геофизические исследования*, 2009, том 10, № 2, стр. 7–22.
9. Gitis V. G., Derendyaev A. B. Geomatics for Environmental Monitoring, Analysis and Forecast. *International Conference on Computational Science and Its Applications*, 2019. pp. 200–215, 2019.
10. Gitis V.G., Derendyaev A.B. Web-Based GIS Platform for Automatic Prediction of Earthquakes. *International Conference on Computational Science and Its Applications*, 2018, pp. 268–283.
11. Gitis V. G., Derendyaev A. B. Machine Learning Methods for Seismic Hazards Forecast. *Geosciences*, 2019, vol. 9, no. 7, с. 308.
12. Bishop C. *Artificial hells: Participatory art and the politics of spectatorship*. Verso Books, 2012.

13. Kotsiantis S. B., Zaharakis I., Pintelas P. Supervised machine learning: A review of classification techniques. *Emerging artificial intelligence applications in computer engineering*, 2007, vol. 160, pp. 3–24.
14. Khan S. S., Madden M. G. A survey of recent trends in one class classification. *Irish conference on artificial intelligence and cognitive science*, Berlin: Heidelberg, 2009, pp. 188–197.
15. Fawcett T. An introduction to ROC analysis. *Pattern recognition letters*, 2006, vol. 27, no. 8, pp. 861–874.
16. Okada Y., Kasahara K., Hori S., Obara K., Sekiguchi S., Fujiwara H., Yamamoto A. Recent progress of seismic observation networks in Japan — Hi-net, F-net, K-NET and KiK-net. *Earth, Planets and Space*, vol. 56, no. 8, pp. xv–xxviii.
17. Obara K., Kasahara K., Hori S., Okada Y., A densely distributed high-sensitivity seismograph network in Japan: Hi-net by National Research Institute for Earth Science and Disaster Prevention. *Review of Scientific Instruments*, 2005, vol. 76, no. 2, 021301.

Analysis of the impact of cleaning catalogs from aftershocks on the systematic earthquake prediction effectiveness

V.G. Gitis, A.B. Derendyaev, K.N. Petrov

The article assesses the effectiveness of earthquake prediction using the method of minimum area of alarm with removed aftershocks. We conducted testing with catalogs of earthquakes in Japan and the Mediterranean for the period 1988-2019. We compare the forecast of the target earthquakes obtained from the full catalog and the forecast of the main shocks. For the forecast we use the grid fields obtained from the full catalog and from the catalog with removed aftershocks. The result of testing is that removing aftershocks from the catalog spoils the quality of the forecast. A possible reason is reducing the number of training examples. Removing aftershocks from the catalog to calculate grid fields of prediction features is advisable if there is a strong aftershock activity in the region.

KEYWORDS: data analysis, machine learning, method of minimum area of alarm, earthquake prognosis, aftershocks.