

УДК 550.343.2:681.3(794)

© Коллектив авторов

А.Д. Геишиани, С.М. Агаян, М.Н. Добровольский, Б.А. Дзедобеев

ОБЪЕКТИВНАЯ КЛАССИФИКАЦИЯ ЭПИЦЕНТРОВ И РАСПОЗНАВАНИЕ МЕСТ ВОЗМОЖНОГО ВОЗНИКНОВЕНИЯ СИЛЬНЫХ ЗЕМЛЕТРЯСЕНИЙ В КАЛИФОРНИИ

Исторический экскурс и постановка задачи

В 1972 г. И.М. Гельфанд, В.И. Кейлис-Борок, Е.Я. Ранцман и др. разработали алгоритмические основы и впервые применили распознавание образов для определения мест возможного возникновения сильных землетрясений. Этот метод (в дальнейшем ERA – Earthquake-prone areas recognition) [25] был успешно использован ими в регионе Памира и Тянь-Шаня для распознавания мест возможного возникновения эпицентров землетрясений с магнитудой $M \geq 6,5$.

В последующие годы подход ERA к определению потенциально высокосейсмичных зон в горных странах получил дальнейшее развитие. К исследованиям ERA активно подключились американские, итальянские и французские ученые. Распознавание мест возможного возникновения сильных землетрясений было выполнено в Калифорнии [21] (рис. 1), Малой Азии [20], Южной Америке [12] и на новых территориях в Средней (в сегодняшних названиях Центральной) Азии [22].

В методе ERA [6, 7, 13, 21, 25] рассматриваются, как правило, коровые землетрясения. Как задача распознавания образов ERA представляет собой классическую дихотомию с обучением [3, 5]. Распознавание мест возможного возникновения сильных землетрясений выполняется на базе векторного представления объектов распознавания в пространстве геолого-геофизических и геоморфологических параметров. При этом параметры должны быть априорно связаны с возможностью характеризовать места высокой сейсмичности. Для каждого региона подбираются и измеряются соответствующие параметры, что представляет собой нетривиальную и весьма трудоемкую процедуру.

Объекты распознавания ERA интерпретируются как окрестности пересечений линеаментов, построенных в результате морфоструктурного районирования (MCP) [24]. Другим вариантом объектов служат морфоструктурные узлы, содержащие

эти пересечения. Границы узлов устанавливаются в результате полевых наблюдений [24]. На площади объектов (узлов или окрестностей линеаментов) измеряются используемые геолого-геофизические параметры. Несмотря на логическую формализацию [24], процесс MCP остается не полностью однозначным.

Следует отметить, что набор используемых в ERA алгоритмов весьма узок. Это созданный еще в 1966 г. алгоритм «Кора-3» [4, 5], его модификация «Подклассы» и относящийся к 1981 г. алгоритм Хемминга.

В последующем метод ERA был обобщен и адаптирован для распознавания мест возможного возникновения землетрясений других пороговых магнитуд: как сильнейших ($M \geq 7,75$ и $M \geq 8,2$), так и существенно более слабых ($M \geq 6,0$; $M \geq 5,5$; $M \geq 5,0$). Был выполнен масштабный советско-французский проект распознавания мест возможного возникновения сильных землетрясений в трех регионах Альпийского пояса – Западные Альпы, Пиренеи, Большой Кавказ [15]. Возможность использования ERA для различных пороговых магнитуд сделала его полезным инструментом для уточнения сейсмического районирования.

Достоверность результатов ERA в различных регионах земного шара хорошо подтверждается расположением эпицентров землетрясений, произошедших после публикации соответствующих работ, и многочисленными контрольными экспериментами, включая статистическую проверку необходимых условий достоверности [14]. В то же время, аналитическая и вычислительная проверки необходимых условий достоверности [9, 10, 11, 14] также дают весомые аргументы в пользу достоверности результатов.

Анализ обширной базы знаний ERA естественно приводит к следующим вопросам:

Вопрос 1. Является ли обучение необходимой частью распознавания мест возможного возникно-

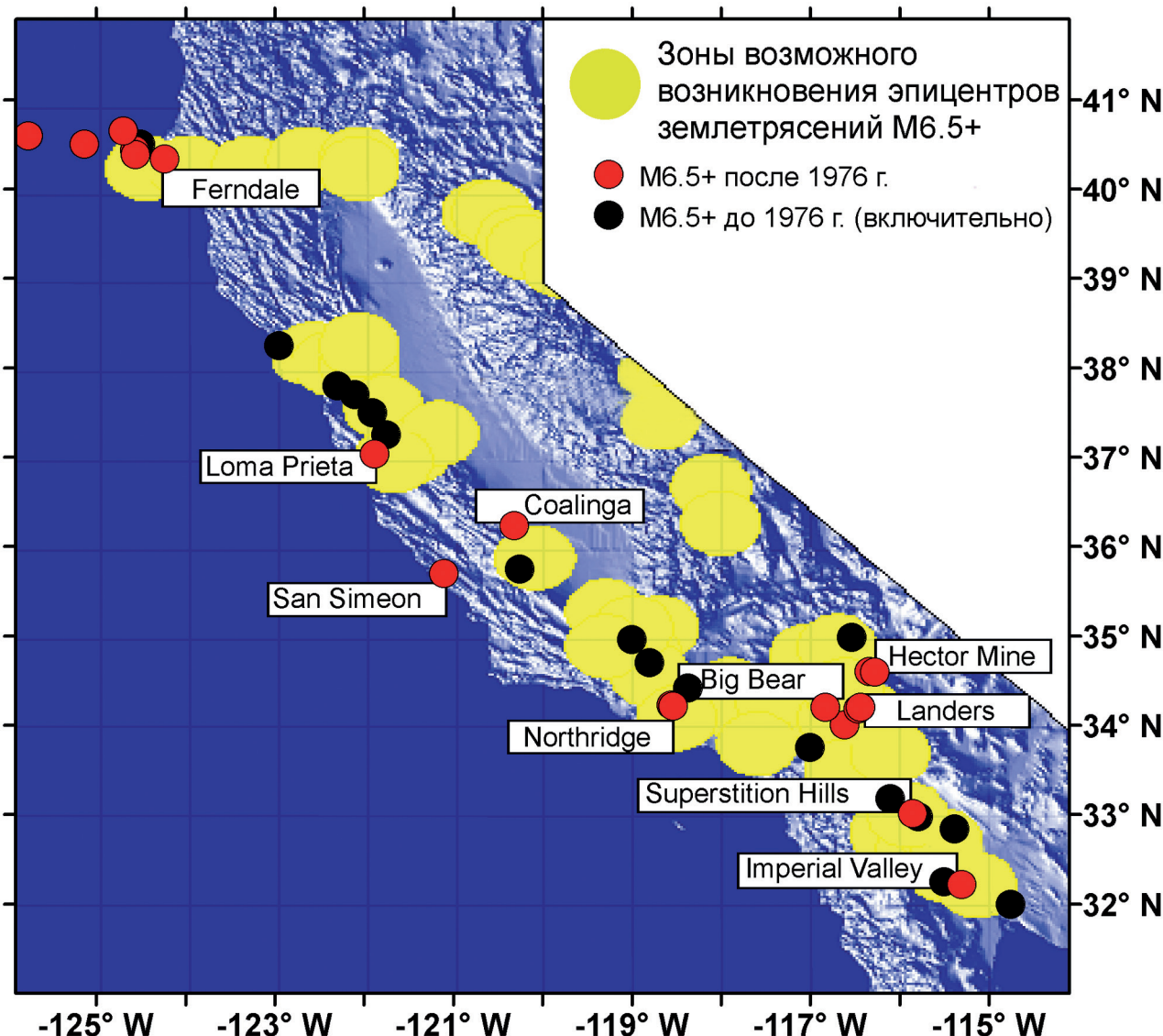


Рис. 1. Зоны возможного возникновения сильных землетрясений ($M \geq 6,5$) в Калифорнии, распознанные методом EPA в 1976 г. [21]

вения сильных землетрясений? Иными словами, может ли результат EPA быть получен с помощью алгоритма кластеризации (объективной классификации), не содержащего блока обучения, но сохраняющего оригинальный уровень достоверности? (Общие принципы объективной классификации изложены, например, в [3].)

В методе EPA в пространстве параметров строятся векторы, представляющие объекты распознавания, — морфоструктурные узлы или пересечения морфоструктурных линеаментов.

Вопрос 2. Можно ли избежать трудоемкого и неоднозначного процесса морфоструктурного районирования? Можно ли использовать другое множество объектов распознавания, приводящее к результату, не уступающему по качеству EPA? Можно ли при этом существенно упростить и уменьшить число геолого-геофизических параметров?

Зоны, распознанные EPA как высокосейсмичные, как правило, занимают существенную часть рассматриваемой территории. Так, например, в случае Калифорнии (рис. 1) потенциально высокосейсмичные зоны занимают примерно 30% территории. Это уменьшает информативность результата и затрудняет его использование для выбора мест расположения объектов повышенной ответственности.

Вопрос 3. Можно ли модифицировать метод EPA таким образом, чтобы общая площадь, объявленная в результате распознавания высокосейсмичной, составляла бы меньшую часть от общей площади региона? При этом этот более информативный результат должен иметь не меньшую степень достоверности.

Попытке продвинуться в ответах на эти вопросы и посвящена настоящая работа.

Исходные данные: каталоги землетрясений Калифорнии

В этой статье мы проводим альтернативное распознавание зон возможного возникновения сильных землетрясений в Калифорнии без использования ЕРА [21] и предшествующих ему морфо-структурного районирования и обучения. Более того, мы абстрагируемся от геолого-геофизических параметров, использованных в [21], и имеем дело только с информацией об эпицентрах землетрясений в Калифорнии, отраженную в подробных каталогах за немногим более чем 40 лет.

Мы использовали два каталога. Во-первых, это каталог землетрясений Центра сейсмических данных по Северной Калифорнии (Northern California Earthquake Data Center – NCEDC, <http://www.ncedc.org>), доступный по адресу <http://www.ncedc.org/ncedc/catalog-search.html>. Каталог был создан в результате совместного проекта NCEDC, Сейсмологической лаборатории Калифорнийского университета в Беркли (Berkeley Seismological Laboratory, University of California, Berkeley) и Геологической службы США в Менло Парке (US Geological Survey, Menlo Park).

В каталоге представлены данные по землетрясениям в Северной и Южной Калифорнии с 1967 года по настоящее время. При этом мы работали с событиями до 2011 года включительно. Для каждого землетрясения в каталоге указана предпочтительная магнитуда, которая использовалась в настоящей работе. Общее количество событий с 1967 по 2011 год составляет 759 582.

Следует отметить, что для Южной Калифорнии каталог NCEDC неполон. Поэтому для Южной Калифорнии мы также использовали каталог Центра сейсмических данных по Южной Калифорнии (Southern California Earthquake Data Center – SCEDC, <http://www.data.scec.org>), доступный по адресу http://www.data.scec.org/ftp/catalogs/SCEC_DC/. В нем представлены гипоцентры событий, начиная с 1932 года по настоящее время, определенные на основе регистраций Сейсмической сети Южной Калифорнии (Southern California Seismic Network – SCSN). С 1932 по 2011 год каталог насчитывает 591 306 событий. При этом из-за меньшей плотности сети наблюдений событий с 1932 по 1966 год в каталоге всего 13 186. Для каждого события в каталоге указана одна магнитуда.

В качестве объектов распознавания (кластеризации) мы используем эпицентры землетрясений с магнитудой $M \geq 3,0$ из этих двух каталогов. Вообще говоря, каталоги землетрясений по Северной и Южной Калифорнии покрывают пересекающиеся территории, а их объединение содержит дублирующую

информацию о некоторых событиях. Тем не менее, мы применяем алгоритм к объединению каталогов. В силу конструкции алгоритма DPS, приведенной ниже, это обстоятельство не оказывает существенного влияния на результат кластеризации.

Алгоритм объективной классификации Discrete Perfect Sets (DPS)

Для кластеризации эпицентров землетрясений применялся созданный авторами алгоритм Discrete Perfect Sets (DPS), впервые представленный на конференции «Математические методы распознавания образов» (ММРО-15) в 2011 году [2].

Алгоритм DPS выделяет в конечном множестве X евклидова пространства $E \supset X$ плотные области при уровне плотности $a \in [0, 1]$. При этом свободными параметрами алгоритма являются показатель $q < 0$ радиуса локализации r и параметр $\beta \in (-1, 1)$, определяющий вместе с q уровень плотности $a = a(\beta, q)$. На выходе получается множество точек $X(a(\beta, q))$, a -плотное в каждой своей точке. При этом $X(a(\beta, q))$ не является r -связным в том смысле, что не любые две его точки можно соединить между собой ломаной с длиной звена не больше r и вершинами в точках X . Однако $X(a(\beta, q))$ может быть разбито на конечное число попарно непересекающихся r -связных компонент:

$$X(a(\beta, q)) = X_1(a) \cup \dots \cup X_N(a).$$

Связные компоненты $X_i(a)$, $i = 1, \dots, N$ и понимаются как искомые кластеры. Описание алгоритма DPS будем вести согласно его блок-схеме, представленной на рис. 2.

Радиус локализации. Пусть d стандартное расстояние в евклидовом пространстве E . Обозначим через $D(X)$ множество всех нетривиальных попарных расстояний в X :

$$D(X) = \{d(x_1, x_2) : x_1, x_2 \in X, d(x_1, x_2) \neq 0\}.$$

Радиус локализации r для X определяется как степенное среднее всех расстояний из $D(X)$ при отрицательных значениях степени q :

$$r_q = r_q(X) = \left(\frac{\sum_{d \in D(X)} d^q}{|D(X)|} \right)^{1/q}, \quad q < 0. \quad (1)$$

Здесь, как обычно, через $|D(X)|$ обозначено количество элементов во множестве $D(X)$.

При обработке каталога землетрясений Калифорнии q полагалось равным -2 .

Плотность. Выбор радиуса локализации $r = r_q(X)$ дает возможность определить плотность $P = P(r)$ на множестве X . Плотность $P_A(X)$ произвольного подмножества $A \subseteq X$ в точке $x \in X$ зададим следующим образом. Рассмотрим r -окрестность точки x в A :

$$B_A(x, r) = \{y \in A : d(x, y) \leq r\}.$$



Рис. 2. Блок-схема этапов работы алгоритма кластеризации DPS (Discrete Perfect Sets). В параллелограмме – исходные данные, в прямоугольниках – этапы выполнения алгоритмом процедуры распознавания

В пространстве E это будут все точки множества A , лежащие внутри шара с центром в x радиуса r . Вычислим сумму расстояний до x от точек из r -окрестности $B_A(x, r)$ с весами, убывающими при удалении от ее центра x :

$$\Sigma_A(x, r) = \sum_{y \in B_A(x, r)} \left(1 - \frac{d(x, y)}{r} \right). \quad (2)$$

Чтобы нормировать (2) к отрезку $[0, 1]$, разделим найденную сумму на наибольшую из сумм (2) по окрестностям всего универсума X :

$$P_A(x) = \frac{\Sigma_A(x, r)}{\Sigma(X, r)}, \quad \Sigma(X, r) = \max_{x \in X} \Sigma_X(x, r). \quad (3)$$

Определение 1. Функцию $P_A(x)$, определяемую формулой (3), будем называть плотностью произвольного подмножества $A \subseteq X$ в точке x .

Единственным параметром при определении плотности P является радиус локализации r , однозначно определяемый свободным параметром алгоритма $q \in R$.

Как легко видеть, построенная плотность обладает свойством монотонности. Она тем больше, чем больше множество A :

$$A_1 \subseteq A_2 \Rightarrow \forall x \in X : P_{A_1}(x) \leq P_{A_2}(x).$$

Это свойство является необходимым условием для работы алгоритма DPS.

Абстрагируясь от алгоритма DPS, заметим, что предложенный способ построения плотности на дискретных множествах не единственно возможный. Общее определение плотности для дискретного случая и некоторые другие примеры плотностей приведены в [2].

Уровень плотности. Выбор уровня плотности α сильно влияет на результат работы алгоритма DPS. Удобным инструментом для выбора уровня α являются нечеткие сравнения [1, 23]. Нечеткое сравнение $n(a, b)$ двух чисел a и b есть мера превосходства числа b над числом a со значениями на отрезке $[-1, 1]$:
 $n(a, b) = mes(a < b) \in [-1, 1]$.

Нечеткое сравнение числа a и конечного множества чисел B можно определить как среднее нечетких сравнений a со всеми числами из множества B :

$$n(a, B) = \frac{\sum_{b \in B} n(a, b)}{|B|}, \quad n(B, a) = \frac{\sum_{b \in B} n(b, a)}{|B|}$$

и понимать как меру минимальности в случае $n(a, B)$ и меру максимальности в случае $n(B, a)$ числа a на фоне B .

При фиксированном B функции $n(a, B)$ и $n(B, a)$ монотонны и непрерывны по a . Таким образом, по результату сравнения a с B однозначно определяется само a .

В этой работе мы использовали нечеткое сравнение:

$$n(a, b) = \frac{b - a}{\max(|a|, |b|)}.$$

Мера максимальности $n(B, a)$ дает возможность сформулировать необходимое требование к будущему результату $A \subset X$ алгоритма DPS: плотность A в каждой своей точке x должна быть значительной (достаточно максимальной) на фоне X . Для этого сначала рассчитываем значения плотности всего универсума X во всех его точках:

$$P_X(X) = \{P_X(x) : x \in X\}.$$

Это «фон X ». Если $\beta \in [-1, 1]$ – необходимый уровень максимальности плотности P на фоне X , то непосредственный уровень $\alpha = \alpha(\beta)$ для P однозначно определяется по β из уравнения:

$$n(P_X(X), \alpha) = \beta. \quad (4)$$

Искать такое α можно как решение уравнения (4) методом деления отрезка пополам.

Таким образом, алгоритм DPS должен найти в X подмножество A β -экстремально P -плотное в каждой своей точке x :

$$n(P_X(X), P_A(x)) \geq \beta \Leftrightarrow P_A(x) \geq \alpha(\beta) \quad \forall x \in X.$$

Алгоритмическое ядро DPS. Пусть на пространстве X определены плотность P и задан ее уровень $\alpha = \alpha(\beta)$. Задача алгоритма DPS состоит в построении такого множества $X(\alpha)$, что его плотность будет не меньше уровня α во всех своих точках, даже если рассматривать только точки самого множества $X(\alpha)$:

$$X(\alpha) = \{x \in X : P_{X(\alpha)}(x) \geq \alpha\}. \quad (5)$$

Если понимать плотность $P_A(x)$ как меру предельности точки x для множества A , а точки x с плотностью, большей уровня α

$$\{x \in X : P_A(x) \geq \alpha\},$$

считать предельными для A , то множество $X(\alpha)$ будет совпадать со множеством своих предельных точек и являться в этом смысле неким аналогом совершенного множества из непрерывной математики. Отсюда название алгоритма Discrete Perfect Sets.

Множество $X(\alpha)$ будем строить следующим образом. Вначале выберем все точки, где плотность всего универсума X не меньше уровня α :

$$X^1(\alpha) = \{x \in X : P_X(x) \geq \alpha\}.$$

Заметим, что при выборе точек с плотностью большей заданной наше множество может только уменьшаться:

$$X^1(\alpha) \subseteq X.$$

Но мы не можем утверждать, что плотность выбранных точек по-прежнему будет не меньше уровня α , если мы ограничимся только ими. Поэтому продолжим выбор точек с плотностью не меньшей α до тех пор, пока множество с предыдущего шага $X^n(\alpha)$ не совпадет со множеством точек, где его плотность не меньше α :

$$X^2(\alpha) = \{x \in X : P_{X^1(\alpha)}(x) \geq \alpha\} \subseteq X^1(\alpha),$$

...

$$X^n(\alpha) = \{x \in X : P_{X^{n-1}(\alpha)}(x) \geq \alpha\} \subseteq X^{n-1}(\alpha),$$

$$X^{n+1}(\alpha) = \{x \in X : P_{X^n(\alpha)}(x) \geq \alpha\} \subseteq X^n(\alpha).$$

Равенство обязательно будет достигнуто на некотором шаге n в силу конечности универсума X . Но, вообще говоря, полученное множество $X^n(\alpha)$ может оказаться пустым при достаточно больших значениях уровня плотности α .

Построенное таким образом множество $X^n(\alpha)$ и будет нужным нам множеством $X(\alpha)$, обладающим свойством (5):

$$X(\alpha) = X^n(\alpha).$$

Смысл построения множества $X(\alpha)$ состоит в следующем. Мы можем просто выбрать в универсуме X все те точки, где его плотность P_X не меньше заданного порога α . Но полученное таким образом множество $X^1(\alpha)$:

$$X^1(\alpha) = \{x \in X : P_X(x) \geq \alpha\}$$

не обязано быть плотным. Мы не можем утверждать, что если будем рассматривать только точки множества $X^1(\alpha)$, то их плотность будет не меньше порога α . Построенное же алгоритмом DPS множество $X(\alpha)$ является плотным. Его плотность не меньше порога α , даже если рассматривать только точки самого множества $X(\alpha)$.

Отметим, что с ростом уровня плотности α результат работы алгоритма DPS $X(\alpha)$ будет убывать:

$$\alpha_1 < \alpha_2 \Rightarrow X(\alpha_1) \supseteq X(\alpha_2).$$

При α равном 0 мы получим все множество X , а при α равном 1 – пустое множество или ту часть множества X , где его плотность тождественно равна 1:

$$X(0) = X,$$

$$X(1) = \emptyset \text{ или } X(1) = \{x \in X : P_{X(1)}(x) = 1\}.$$

Числа α и β мы рассматриваем как свободные параметры (параметры подгонки) алгоритма DPS.

Кластеризация землетрясений Калифорнии алгоритмом DPS. Сравнение кластеров с зонами, распознанными методом EPA

Как уже было сказано, распознавание велось по каталогам землетрясений Северной и Южной Калифорнии, общее количество событий в которых составляло 1 350 888. Поэтому было принято решение обрабатывать только события, начиная с некоторой магнитуды. После проведения предварительных численных экспериментов мы ограничились землетрясениями, начиная с магнитуды $M \geq 3,0$. С 1980 по 2011 год каталоги содержали 25 118 таких событий. С целью упрощения расчетов каталоги по Северной и Южной Калифорнии обрабатывались отдельно. Затем бралось объединение результатов работы алгоритма DPS для получения итогового результата по всей Калифорнии.

Результаты применения алгоритма DPS к обработке эпицентров землетрясений с $M \geq 3,0$ приведены на рис. 3. Были выполнены три итерации алгоритма. Вначале был применен алгоритм DPS с уровнем плотности β_1 . Полученные плотные области $X(\alpha_1(\beta_1))$ были убраны из рассмотрения и алгоритм DPS был выполнен второй раз с уровнем плотности β_2 , что дало новые кластеры $X_1(\alpha_2(\beta_2))$, где $X_1 = X / X(\alpha_1(\beta_1))$. И, наконец, после удаления из каталога

$X(\alpha_1(\beta_1)) \cup X_1(\alpha_2(\beta_2))$ алгоритм был применен в третий раз с уровнем плотности β_3 , что дало кластеры $X_2(\alpha_3(\beta_3))$, где $X_2=X_1/X_1(\alpha_2(\beta_2))$. Уровни плотности равнялись $\beta_1 = -0,2, \beta_2 = 0,3, \beta_3 = 0,2$ для каталога землетрясений Северной Калифорнии и $\beta_1 = -0,4, \beta_2 = 0,2, \beta_3 = 0,1$ для каталога землетрясений Южной Калифорнии. Окончательно искомыми DPS кластерами были объявлены все r -связные компоненты, входящие в $X(\alpha_1(\beta_1)) \cup X_1(\alpha_2(\beta_2)) \cup X_2(\alpha_3(\beta_3))$. Несколько последовательных итераций применялись для уточнения результатов работы алгоритма.

Как видно из рис. 3, все сильные события с магнитудой $M \geq 6,5$ из каталога землетрясений (табл. 1), попали в пределы зон, выделенных алгоритмом DPS и изображенных на рис. 3 зеленым цветом.

Также на рис. 3 нанесены эпицентры сильных землетрясений с $M \geq 6,5$, которые приведены в ката-

логе статьи [21], но отсутствуют в использованных нами каталогах NCEDC и SCEDC, т.к. относятся к другим периодам времени. Они наряду с землетрясениями из табл. 1, произошедшими до 1976 г., изображены черными кругами. Всего было 11 таких событий. Первое из них относится к 1836 году, а последнее зарегистрировано 31 декабря 1934 года. Из рис. 3 видно, что эпицентры этих землетрясений тоже достаточно хорошо приурочены к зонам, выделенным алгоритмом DPS как опасные для $M \geq 6,5$. Исключение составляют лишь два землетрясения: землетрясение у Форт-Техона 9 января 1857 г. (№ 4 в табл. 2) с $M > 7$ и землетрясение в Сан-Франциско 18 апреля 1906 г. с $M = 8,25$ (№ 5 в табл. 2). Эти землетрясения произошли задолго до начала инструментальных наблюдений за сейсмичностью региона и достоверность данных об их эпицентрах не ясна.

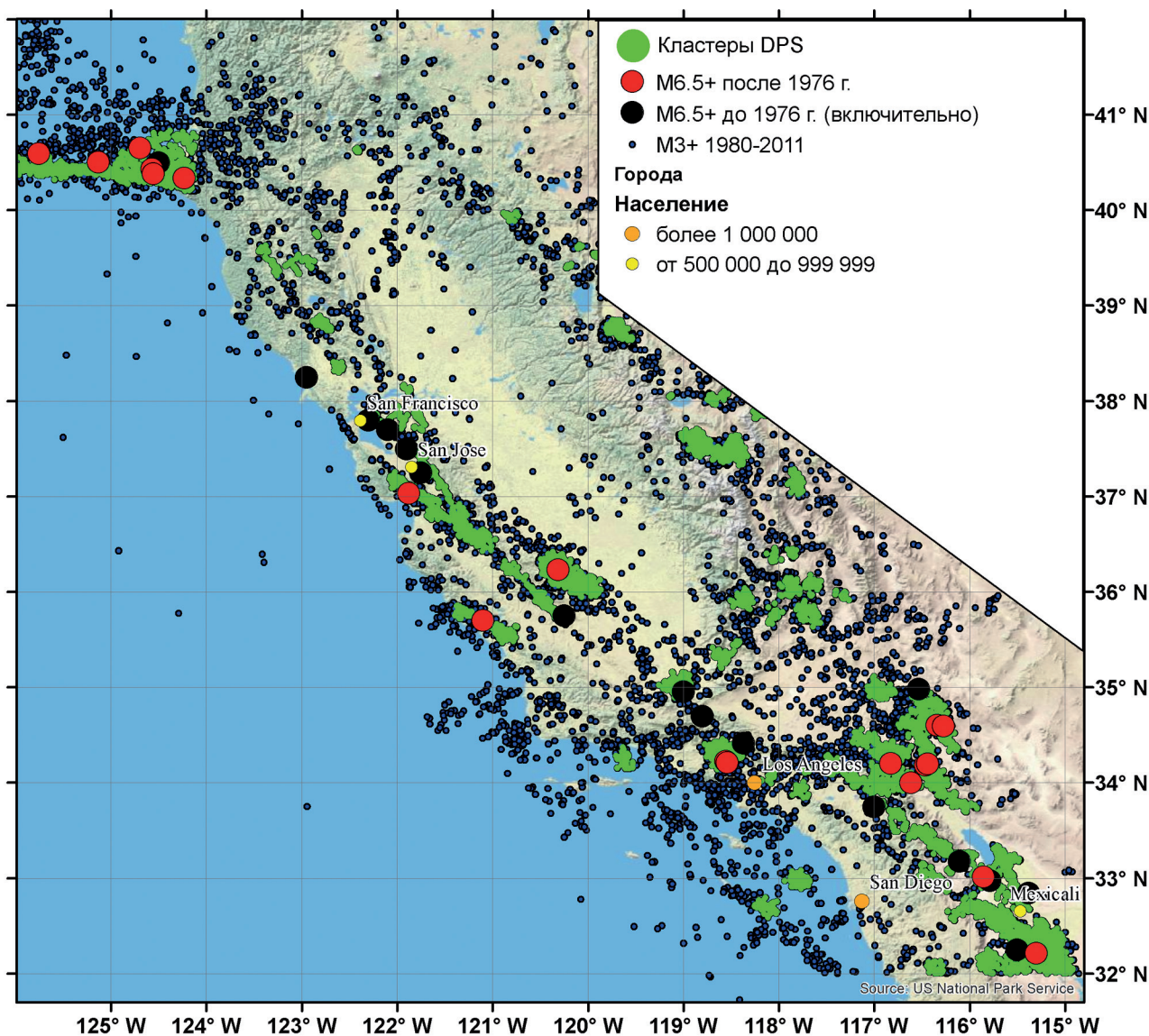


Рис. 3. Результат применения алгоритма DPS к кластеризации эпицентров землетрясений с магнитудой $M \geq 3,0$ в Калифорнии за 1980-2011 годы

Таблица 1

Сильные землетрясения с магнитудой $M \geq 6,5$ в Калифорнии с 1932 по 2011 гг. из каталогов NCEDC (Northern California Earthquake Data Center) и SCEDC (Southern California Earthquake Data Center). Горизонтальной линией отделены события до и после 1976 г., когда была опубликована работа [21]

№	Дата	$\varphi, ^\circ$	$\lambda, ^\circ$	h, км	M
1	19.05.1940	32,84	-115,38	6	6,9
2	21.10.1942	32,98	-115,79	6	6,6
3	10.04.1947	34,98	-116,53	6	6,5
4	21.07.1952	34,96	-119,00	6	7,5
5	09.04.1968	33,18	-116,10	10	6,6
6	09.02.1971	34,42	-118,37	8,9	6,6
7	02.05.1983	36,23	-120,31	10,18	6,7
8	10.09.1984	40,50	-125,13	4,9	6,6
9	08.07.1986	34,00	-116,61	10,11	6,6
10	24.11.1987	33,02	-115,85	11,2	6,6
11	18.10.1989	37,04	-121,88	17,43	7
12	25.04.1992	40,33	-124,23	10,55	6,95
13	26.04.1992	40,43	-124,57	19,57	6,5
14	26.04.1992	40,38	-124,56	22,62	6,6
15	28.06.1992	34,19	-116,46	20,12	7,4
16	28.06.1992	34,20	-116,83	5,21	6,6
17	17.01.1994	34,23	-118,55	14,26	6,9
18	01.09.1994	40,41	-126,30	5,35	7
19	19.02.1995	40,59	-125,76	5	7,6
20	16.10.1999	34,60	-116,34	24	7
21	22.12.2003	35,70	-121,10	8,7	6,5
22	10.01.2010	40,65	-124,69	29,33	6,5
23	04.04.2010	32,21	-115,30	10	7,2

Таблица 2

Сильные землетрясения с магнитудой $M \geq 6,5$ в Калифорнии из каталога статьи [21], отсутствующие в использованных каталогах NCEDC и SCEDC

№	Дата	$\varphi, ^\circ$	$\lambda, ^\circ$	M
	1836	37,5	121,9	>7
	1836	37,7	122,1	>7
	1836	37,8	122,3	>7
	1857	34,7	118,8	>7
	18.04.1906	38,25	122,95	8,25
	01.07.1911	37,25	121,75	6,6
	21.04.1918	33,75	117	6,8
	10.03.1922	35,75	120,25	6,5
	22.01.1923	40,5	124,5	7,2
	30.12.1934	32,25	115,5	6,5
	31.12.1934	32	114,75	7,0

Поэтому мы не имеем достаточного числа объектов (эпицентров землетрясений) для их распознавания.

Сравнение зон, выделенных алгоритмом DPS, с зонами, полученными методом EPA, приведено на рис. 4. Видно, что зеленые зоны, полученные кластеризацией эпицентров землетрясений, занимают меньшую площадь, чем желтые зоны, выделенные методом EPA. Площадь зеленых зон оказывается примерно в 7 раз меньше площади желтых. При этом по большей части зеленые зоны находятся внутри желтых. Также внутри зеленой зоны находится и эпицентр землетрясения вблизи г. Сан-Симеон 22 декабря 2003 года с магнитудой $M = 6,5$ (№ 21 в табл. 1), не попавший в зоны, полученные методом EPA.

В северо-западном углу карты внутри кластеров DPS находятся эпицентры двух морских землетрясений, тоже не вошедшие в зоны, распознанные

как высокосейсмичные алгоритмом в [21]. Таким образом, если говорить о сильных событиях материала обучения [21], то зоны, выделенные алгоритмом DPS, занимают меньшую площадь и лучше согласуются с эпицентрами землетрясений с магнитудой $M \geq 6,5$ на территории Калифорнии, чем зоны, полученные методом EPA в [21].

В качестве контрольного эксперимента в дополнение к кластеризации эпицентров за все доступное время каталога была проведена кластеризация эпицентров за 20 лет перед сильными событиями с магнитудой $M \geq 6,5$. Всего рассматриваемые каталоги содержат 23 таких события. Их эпицентры и магнитуды приведены в табл. 1. Чтобы перед землетрясением была доступна история наблюдений за предыдущие 20 лет, рассматривались только землетрясения, начиная с 1987 года.

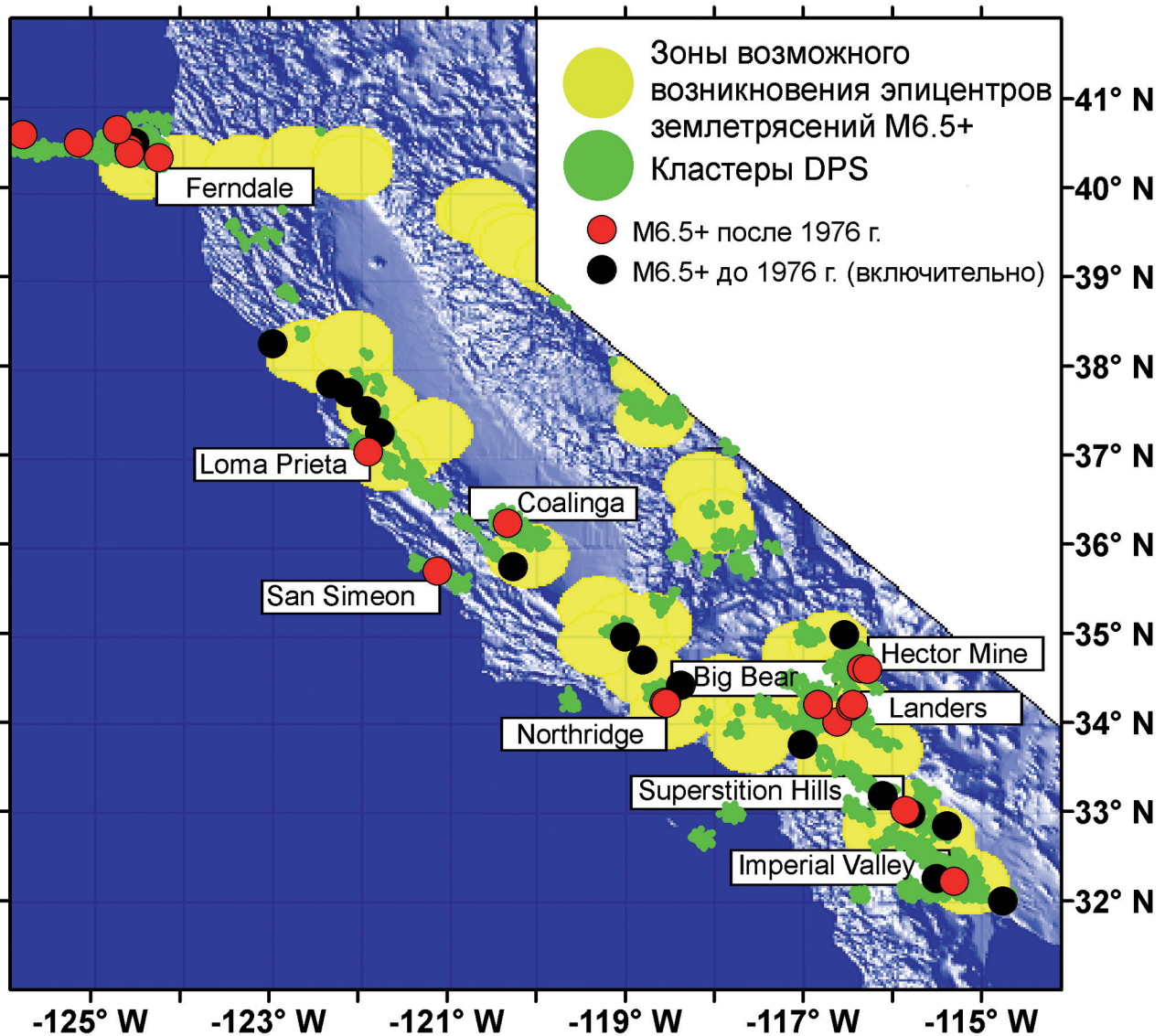


Рис. 4. Сравнение зон, полученных алгоритмом DPS, с зонами, полученными методом EPA [21]

Такие сильные землетрясения происходили в Калифорнии в 1987, 1989, 1992, 1994, 1995, 1999, 2003 и 2010 годах. Для кластеризации были выделены эпицентры всех событий с магнитудой $M \geq 3,0$ за 20 лет перед каждым годом, когда происходили указанные землетрясения с магнитудой $M \geq 6,5$. К полученным таким образом эпицентрам был применен алгоритм DPS с едиными для всех восьми случаев значениями параметров. На первой итерации $\beta_1 = -0,3$, на второй итерации $\beta_2 = 0,1$. Значение параметра q при этом бралось равным -2 , также как и при кластеризации всего каталога.

В качестве примера приведем результаты кластеризации для землетрясений 1987 и 2010 годов. На рис. 5 приведены результаты обработки алгоритмом DPS эпицентров землетрясений с магнитудой $M \geq 3,0$ перед землетрясением 24 ноября

1987 года в Superstition Hills с магнитудой $M = 6,6$. А на рис. 6 – результаты обработки эпицентров землетрясений с магнитудой $M \geq 3,0$ перед двумя землетрясениями, произошедшими в 2010 году, морским 10 января 2010 года с магнитудой $M = 6,5$ и землетрясением в Baja California 4 апреля 2010 года с магнитудой $M = 7,2$.

Из этих рисунков видно, что пространственное распределение эпицентров землетрясений с магнитудой $M \geq 3,0$ за промежутки с 1967 по 1986 год и с 1990 по 2009 год, хотя и различно, но достаточно подобно. При этом рассматриваемые события с магнитудой $M \geq 6,5$ оказались внутри выделенных алгоритмом DPS кластеров эпицентров $M \geq 3,0$, либо в непосредственной близости от них (рис. 5, 6). То же верно и для остальных шести лет, когда происходили землетрясения с магнитудой $M \geq 6,5$.

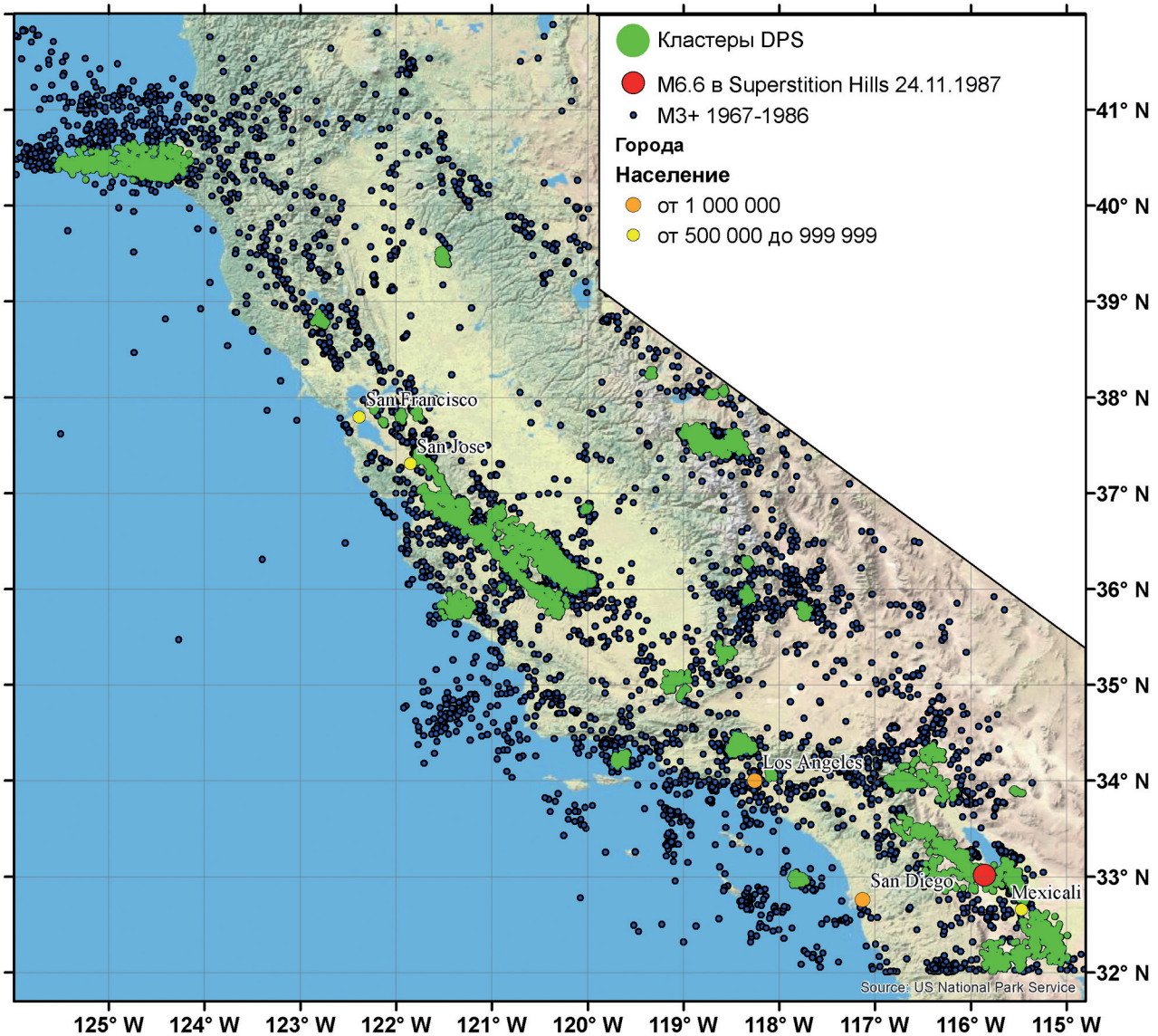


Рис. 5. Результат кластеризации алгоритмом DPS эпицентров землетрясений с магнитудой $M \geq 3,0$ за 20 лет (1967-1986 годы) перед сильным землетрясением 24 ноября 1987 года в Superstitions Hills с магнитудой $M = 6,6$

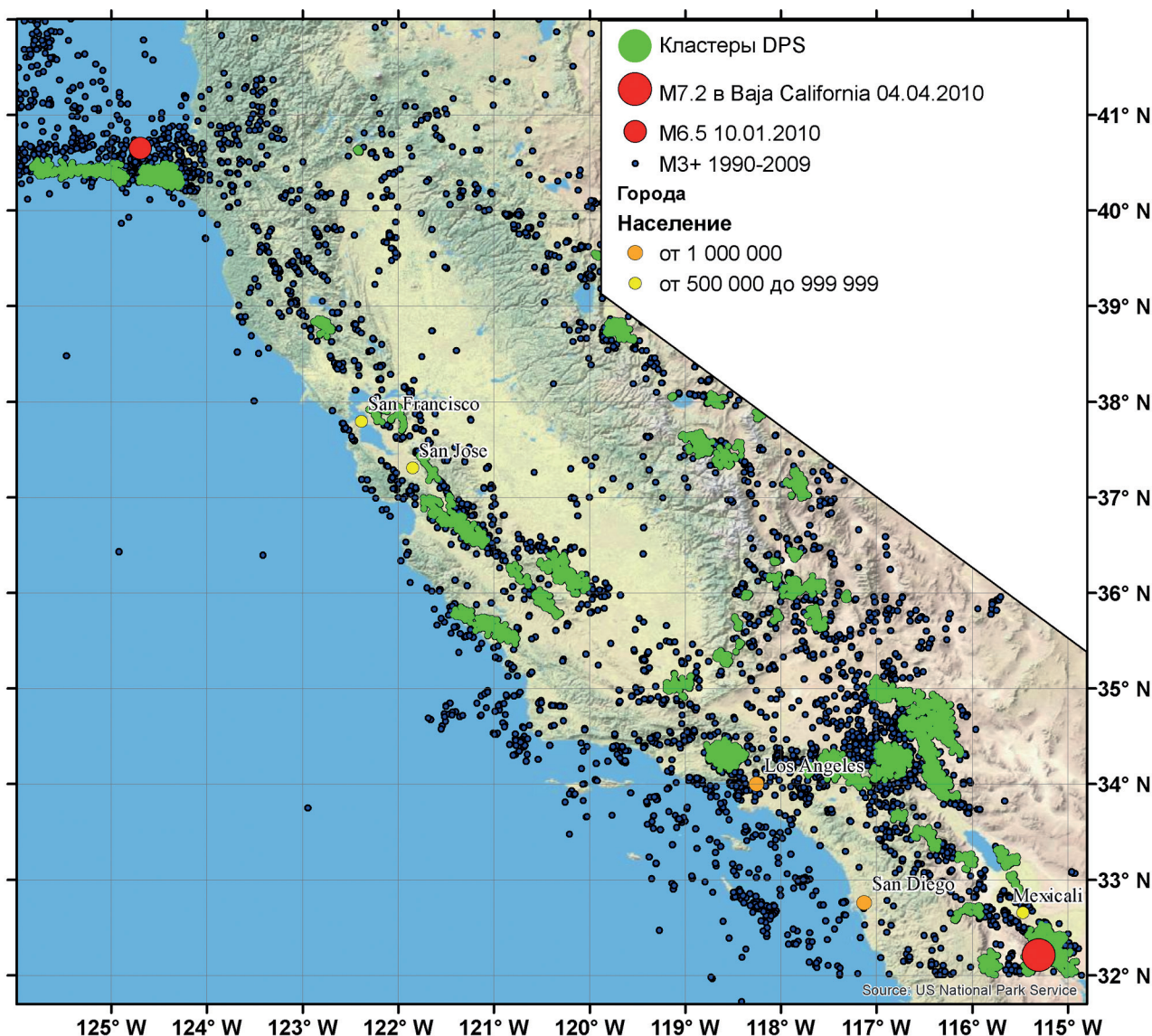


Рис. 6. Результат кластеризации алгоритмом DPS эпицентров землетрясений с магнитудой $M \geq 3,0$ за 20 лет (1990-2009 годы) перед сильными землетрясениями 10 января 2010 года с магнитудой $M = 6,5$ близ северного побережья Калифорнии и 4 апреля 2010 года в Baja California, Мексика с магнитудой $M = 7,2$

Единственным исключением является землетрясение, произошедшее на разломе Мендосино (Mendocino Fault earthquake) 1 сентября 1994 года (№ 18 в табл. 1), оказавшееся на чуть большем расстоянии от зеленых зон, чем все остальные землетрясения с магнитудой $M \geq 6,5$.

Сравнение основного результата кластеризации (рис. 3) и этих восьми контрольных экспериментов (рис. 5, 6) показывает их достаточную близость. Это обстоятельство, с одной стороны, дает существенный аргумент в пользу достоверности кластеризации на рис. 3 как варианта распознавания мест возможного возникновения эпицентров

землетрясений с $M \geq 6,5$ в Калифорнии (аналог эксперимента «сейсмическая история» в [21]). С другой стороны, мы видим, что достоверность результата носит итерационный характер.

Обсуждение результатов

Использованный в этой работе новый алгоритм объективной классификации DPS (Discrete Perfect Sets) является частью созданного в Геофизическом центре РАН оригинального подхода к анализу дискретных данных под общим названием дискретный математический анализ (ДМА) [17, 19]. DPS входит в его блок ДМА-кластеризация. В рамках

этого блока ранее А.Д. Гвишиани, С.М. Агаяном, Ш.Р. Богоутдиновым, А.А. Соловьевым (младшим) и др. были созданы алгоритмы «Роден», «Кристалл» и «Монолит», хорошо зарекомендовавшие себя при решении важных геофизических задач [1, 16, 17, 18, 19, 26].

Проведенная объективная классификация (DPS кластеризация) эпицентров землетрясений с магнитудой $M \geq 3,0$ в Калифорнии показывает, что для этого региона и сильных землетрясений, определенных как имеющие магнитуду $M \geq 6,5$, результат И.М. Гельфанда и др. [21] может быть получен без обучения. При этом повышается его информативность. Действительно, в кластеры DPS попадает суммарная площадь, составляющая менее 13% от площади зон, объявленных в результате EPA опасными в смысле возможности возникновения в их пределах эпицентров сильных землетрясений.

На рис. 4 зоны, распознанные как высокосейсмичные модификацией алгоритмов «Кора-3» и «Подклассы» [4] для $M \geq 6,5$, показаны желтым цветом [21]. На этом же рисунке зеленым цветом выделены кластеры эпицентров, полученные алгоритмом DPS. При этом объектами распознавания DPS являются собственно эпицентры землетрясений, в то время как в [21] – это пересечения морфоструктурных линейментов. Таким образом, в предложенном в настоящей работе кластеризационном подходе отпадает необходимость в трудоемком процессе морфоструктурного районирования для определения пересечений линейментов или узлов как объектов распознавания [21]. Тем самым, не ухудшая результата, мы избегаем наиболее трудоемкого и неавтоматизированного шага в процедуре EPA.

Сравнивая кластеры DPS с высокосейсмичными желтыми зонами EPA, мы видим, что почти все эпицентры землетрясений $M \geq 6,5$ располагаются внутри кластеров или на их границах. Это касается как эпицентров, использованных при обучении алгоритма «Подклассы» (черные круги на рис. 4), так и эпицентров землетрясений, происшедших после опубликования результатов [21] в 1976 г. (красные круги на рис. 4). Внутри кластеров DPS оказываются и три эпицентра землетрясений с магнитудой $M \geq 6,5$ после 1976 г. (№№ 8, 19, 21 в табл. 1), пропущенных алгоритмом «Подклассы» в [21].

Землетрясение № 21 из табл. 1, эпицентр которого находится на суше в районе города Сан-Симеон, не использовалось при формировании материала обучения и тем самым составляет элемент независимого экзамена. Ближайшие пересечения линейментов не были распознаны в [21] как высокосейсмичные. Таким образом, этот эпицентр создает ошибку распознавания типа пропуска цели метода

в [21] (см. рис. 1). В то же время этот эпицентр попадает в кластер DPS. Два других эпицентра сильных землетрясений, произошедших после публикации [21] (№ 8 и № 19 в табл. 1), расположены в северной части региона в океане западнее распознанной высокосейсмичной зоны EPA (рис. 1). Эта высокосейсмичная зона захватывает океанскую акваторию, однако не включает в себя эпицентры № 8 и № 19.

Тем самым эти два эпицентра также могут трактоваться как пропуски цели распознавания EPA [21]. Так же как и эпицентр землетрясения № 21, эпицентры № 8 и № 19 располагаются в пределах кластеров DPS. Таким образом, интерпретировав кластеры DPS как зоны возможного возникновения эпицентров сильных землетрясений, полученные алгоритмом DPS, мы получаем отсутствие пропусков цели. Это служит весомым аргументом в пользу такой интерпретации.

Приведенные аргументы говорят за то, что кластеры, полученные алгоритмом DPS (рис. 3, 4), определяют зоны возможного возникновения эпицентров землетрясений магнитуды $M \geq 6,5$ в Калифорнии. Достоверность кластеров DPS, как высокосейсмичных зон в Калифорнии, подтверждается и расположением эпицентров землетрясений магнитуды $M \geq 6,5$. При этом мы получаем лучшую степень подтверждения по сравнению с [21], так как все эпицентры землетрясений из табл. 1 с магнитудой $M \geq 6,5$ оказываются внутри кластеров DPS.

Результаты эксперимента «сейсмическая история» для DPS показывают, что кластеры эпицентров землетрясений с магнитудой $M \geq 3,0$ на территории Калифорнии слабо меняются со временем. За единственным исключением землетрясения 1994 года, эпицентры событий с магнитудой $M \geq 6,5$, из рассматриваемых каталогов землетрясений, попадают в пределы кластеров, выделенных алгоритмом DPS среди событий с магнитудой $M \geq 3,0$ за 20 лет перед сильными землетрясениями.

При этом для всех восьми лет, начиная с 1987 года, когда происходили сильные события, алгоритм DPS применялся с одними и теми же параметрами – радиусом локализации r , полученными при степени $q = -2$, и уровнями плотности $\beta_1 = -0,3$ и $\beta_2 = 0,1$ на первой и второй итерациях алгоритма соответственно. Это позволяет ожидать, что если применять алгоритм DPS с указанными параметрами к эпицентрам землетрясений с магнитудой $M \geq 3,0$, произошедшим за предыдущие 20 лет, то мы получим зоны, где могут произойти сильные события с магнитудой $M \geq 6,5$ в текущем году.

Таким образом, качество результатов контроля достоверности с помощью эксперимента «сейсмическая история» для результатов алгоритмов

«Кора-3»/«Подклассы» [21] и DPS оказывается равносильным. Таким образом, мы получаем серьезные аргументы в пользу положительных ответов на вопросы 1 и 2, поставленные в начале статьи.

Обратимся к сравнению информативности результатов. Как видно из рис. 4, зоны G, распознанные алгоритмом DPS, как правило, входят в зоны Y, распознанные методом EPA в [21]. Иногда зоны G продолжают зоны Y в том или ином направлении.

На рис. 4 результаты распознавания EPA и DPS сопоставляются с помощью геоинформационной системы ArcGIS 10.1. Оценивая возможное отношение площадей G/Y как отношение соответствующих чисел пикселей в слоях ГИС, получаем, что площадь G составляет менее 13% от площади Y.

Таким образом, если интерпретировать зоны G как места возможного возникновения землетрясений $M \geq 6,5$ в Калифорнии, то результат может оказаться существенно информативнее, чем [21]. Таким образом, мы получаем аргумент в пользу положительного ответа на вопрос 3, поставленный в первом разделе статьи.

Резюмируя, отметим следующее. Геология и геофизика Калифорнии хорошо изучены. Здесь развернута одна из лучших в мире систем сейсмологических наблюдений. В результате для этого региона созданы каталоги землетрясений высшего стандарта качества. Они прямо или косвенно отражают в себе параметры, влияющие на возможность возникновения сильных землетрясений. В этой ситуации, полученная DPS кластеризация показывает, что в Калифорнии каталога землетрясений оказывается достаточно для распознавания мест возможного возникновения эпицентров сильных ($M \geq 6,5$) событий.

Анализируя расположение эпицентров землетрясений, происшедших до и после проведения распознавания EPA [21], можно заключить, что сильных землетрясений следует ожидать внутри кластеров менее сильных и слабых ($M \geq 3,0$) событий. Этот результат имеет, по крайней мере, сравнимую степень достоверности с результатом EPA [21]. При этом для его получения не нужно трудоемкого морфоструктурного районирования и обучения [24]. Достаточно для распознавания оказывается высококачественный каталог землетрясений. Тем самым мы получаем, в случае Калифорнии, положительный ответ на поставленный выше вопрос 2.

Является ли полученное соответствие кластеров DPS и расположения эпицентров сильных землетрясений закономерностью, свойственной именно для Калифорнии, или эта закономерность носит более общий характер? Иными словами, может ли менее трудоемкий подход к распознаванию мест возможного возникновения сильных землетрясений

с помощью объективной кластеризации DPS использоваться наряду (или вместо) метода EPA в других сейсмичных регионах?

Ответ на этот вопрос требует серьезного дальнейшего исследования. Качество и полнота каталога землетрясений в других регионах должны иметь существенное значение. Действительно, проводя кластеризацию, мы не используем никаких других параметров кроме собственно параметров эпицентров землетрясений. При этом они являются как объектами распознавания, так и его предметом. Таким образом, каталог должен быть достаточно качественным, чтобы отражать в себе тренды изменений геолого-геофизических параметров региона, связанных прямо или косвенно с местами возможного возникновения сильных землетрясений.

Даже при наличии высококачественных каталогов найденная закономерность соответствия кластеров DPS и расположения эпицентров сильных землетрясений может носить региональный характер и не иметь места в других сейсмичных регионах. Поэтому метод EPA а priori является более универсальным. Действительно, для его применения не требуется столь подробного каталога землетрясений, как это имеет место в Калифорнии.

Как уже было отмечено, для DPS объектами распознавания являются эпицентры землетрясений, возможное расположение части которых изучается. Тем самым снимается необходимость обоснования адекватности тех или иных объектов распознавания для данной задачи.

В то же время, в методе EPA объекты распознавания – это пересечения морфоструктурных линейментов. Возможность их использования для распознавания мест возможного возникновения землетрясений необходимо обосновывать для каждого региона [8, 24]. При этом для некоторых регионов (например, горный пояс Анд [8]) это обоснование представляет собой нетривиальную задачу.

Наконец, алгоритм DPS, напрямую работая с эпицентрами, имеет существенно меньше свободных параметров, чем алгоритмы «Кора-3» и «Подклассы», использованные в [21]. Кроме определенных выше параметров q и β , результат DPS зависит только от выбора исходного множества эпицентров землетрясений (то есть от выбора нижнего порога магнитуды в каталоге в нашем случае) и числа итераций. Результат DPS очевидно не зависит и от материала обучения в силу его отсутствия в этом алгоритме.

Для построения карт, приведенных в статье, было использовано программное обеспечение ArcGIS 10.1.

Результаты этой статьи получены в рамках проектов РФФИ № 12-05-00583, № 12-05-90716

и Программы Отделения наук о Земле РАН № 7 «Геофизические данные: анализ и интерпретация». Они докладывались на Первой международной конференции по рискам (First International Conference on Moldavian Risks – From Global to Local Scale, <http://mgconference.ub.ro>), 16–19 мая 2012 года в Бакау (Вацау), Румыния, и на 33-й Генеральной ассамблее Европейской сейсмологической комиссии (European Seismological Commission 33-rd General Assembly, <http://www.esc2012-moscow.org>), 19–24 августа 2012 года в Москве.

Авторы благодарны Н.А. Сергеевой, директору Мирового центра данных по физике твердой Земли ГЦ РАН, за предоставленные каталоги землетрясений, А.И. Горшкову и Ш.Р. Богоутдинову за полезные обсуждения и Р.В. Сидорову за помощь в оформлении рисунков.

Ключевые слова: распознавание мест возможного возникновения сильных землетрясений, кластеризация, earthquake-prone areas (EPA), дискретный математический анализ (ДМА), алгоритм Discrete Perfect Sets (DPS), нечеткие сравнения, выделение плотных областей.

ЛИТЕРАТУРА

1. Агаян С.М., Соловьев А.А. Выделение плотных областей в метрических пространствах на основе кристаллизации // System Research and Information Technologies. – 2004. – № 2. С. 7-23.
2. Агаян С.М., Богоутдинов Ш.Р., Добровольский М.Н. Об одном алгоритме поиска плотных областей и его геофизических приложениях // Математические методы распознавания образов : 15-я Всероссийская конференция, г. Петрозаводск, 11-17 сентября 2011 г. : сб. докл. – М. – С. 543-546.
3. Айвазян С.А., Бухштабер В.М., Енюков И.С., Мешалкин Л.Д. Классификация и снижение размерности : справочное издание. – М. : Финансы и статистика. 1989. – 607 с. – (Прикладная статистика).
4. Бонгард М.М., Вайнцвайг М.Н., Губерман Ш.А., Извекова М.Л., Смирнов М.С. Использование обучающейся программы для выявления нефтеносных пластов // Геология и Геофизика. – 1966. – № 6 (II). – С. 15-29.
5. Бонгард М.М. Проблема узнавания. – М. : Наука, 1967. – 320 с.
6. Вебер К., Гвишиани А.Д., Годфруа П., Горшков А.И., Кособоков В.Г., Ламбер С., Ранцман Е.Я., Саллантен Ж., Сальдано А., Соловьев А.А. Распознавание мест возможного возникновения сильных землетрясений. XII. Два подхода к прогнозу мест возможного возникновения сильных землетрясений в Западных Альпах // Теория и анализ сейсмической информации : сб. науч. ст. / под ред. В.И. Кейлис-Борока. – М. : Наука, 1985. – С. 139-154. – (Вычислительная сейсмология. Вып. 18).
7. Гвишиани А.Д., Зелевинский А.В., Кейлис-Борок В.И., Кособоков В.Г. Распознавание мест возникновения сильнейших землетрясений Тихоокеанского пояса // Методы и алгоритмы интерпретации сейсмологических данных : сб. науч. ст. / под ред. В.И. Кейлис-Борока. – М. : Наука, 1980. – С. 30-45. – (Вычислительная сейсмология. Вып. 13).
8. Гвишиани А.Д., Соловьев А.А. О приуроченности эпицентров сильных землетрясений к пересечениям морфоструктурных линейментов на территории Южной Америки // Методы и алгоритмы интерпретации сейсмологических данных : сб. науч. ст. / под ред. В.И. Кейлис-Борока. – М. : Наука, 1980. – С. 45-50. – (Вычислительная сейсмология. Вып. 13).
9. Гвишиани А.Д., Кособоков В.Г. К обоснованию результатов прогноза мест сильных землетрясений, полученных методами распознавания // Изв. АН СССР. Физика Земли. – 1981. – № 2. – С. 21-36.
10. Гвишиани А.Д. Устойчивость по времени прогноза мест сильных землетрясений. I. Юго-Восточная Европа и Малая Азия // Изв. АН СССР. Физика Земли. – 1982. – № 8. – С. 13-19.
11. Гвишиани А.Д., Гурвич В.А. Устойчивость по времени прогноза мест сильных землетрясений. II. Восток Средней Азии // Изв. АН СССР. Физика Земли. – 1982. – № 9. – С. 30-38.
12. Гвишиани А.Д., Жидков М.П., Соловьев А.А. Распознавание мест возможного возникновения сильных землетрясений. X. Места землетрясений магнитуды $M \geq 7,75$ на Тихоокеанском побережье Южной Америки // Математические модели строения Земли и прогноза землетрясений : сб. науч. ст. / под ред. В.И. Кейлис-Борока. – М. : Наука, 1982. – С. 56-66. – (Вычислительная сейсмология. Вып. 14).
13. Гвишиани А.Д., Жидков М.П., Соловьев А.А. К переносу критериев высокой сейсмичности горного пояса Анд на Камчатку // Изв. АН СССР. Физика Земли. – 1984. – № 1. – С. 20-33.
14. Гвишиани А.Д., Гурвич В.А. Динамические задачи распознавания образов I. Условия стабильности для прогноза мест сильных землетрясений // Математическое моделирование и интерпретация геофизических данных : сб. науч. ст. / под ред. В.И. Кейлис-Борока. – М. : Наука, 1984. – С. 70-88. – (Вычислительная сейсмология. Вып. 17).
15. Гвишиани А.Д., Горшков А.И., Ранцман Е.Я., Систернас А., Соловьев А.А. Прогнозирование мест землетрясений в регионах умеренной сейсмичности. М. : Наука, 1988. – 176 с.
16. Гвишиани А.Д., Диаман М., Михайлов В.О., Гальдеано А., Агаян С.М., Богоутдинов Ш.Р., Граева Е.М. Алгоритмы искусственного интеллекта для

- кластеризации магнитных аномалий // Физика Земли. – 2002. – № 7. – С. 13-28.
17. Гвишиани А.Д., Агаян С.М., Богоутдинов Ш.Р. О новом подходе к кластеризации // Кибернетика и системный анализ. – 2002. – № 2. – С. 104-122.
18. Гвишиани А.Д., Белов С.В., Агаян С.М., Родкин М.В., Морозов В.Н., Татаринцов В.Н., Богоутдинов Ш.Р. Геоинформационные технологии: методы искусственного интеллекта при оценке тектонической стабильности Нижнеканского массива // Инженерная экология. – 2008. – № 2. – С. 3-14.
19. Гвишиани А.Д., Агаян С.М., Богоутдинов Ш.Р. Дискретный математический анализ и мониторинг вулканов // Инженерная экология. – 2008. – № 5. – С. 26-31.
20. Гельфанд И.М., Губерман Ш.А., Жидков М.П., Калецкая М.С., Кейлис-Борок В.И., Ранцман Е.Я., Ротвайн И.М. Распознавание мест возможного возникновения сильных землетрясений. II. Четыре региона Малой Азии и Юго-Восточной Европы // Машинный анализ цифровых сейсмических данных : сб. науч. ст. / под ред. В.И. Кейлис-Борока. – М. : Наука, 1974. – С. 3-40. – (Вычислительная сейсмология. Вып. 7).
21. Гельфанд И.М., Губерман Ш.А., Кейлис-Борок В.И., Кнопов Л., Пресс Ф.С., Ранцман Е.Я., Ротвайн И.М., Садовский А.М. Условия возникновения сильных землетрясений (Калифорния и некоторые другие регионы) // Исследование сейсмичности и моделей Земли : сб. науч. ст. / под ред. В.И. Кейлис-Борока. – М. : Наука, 1976. – С. 3-91. – (Вычислительная сейсмология. Вып. 9).
22. Жидков М.П., Кособоков В.Г. Распознавание мест возможного возникновения сильных землетрясений. VIII. Пересечения линеаментов Востока Средней Азии // Вопросы прогноза землетрясений и строения Земли : сб. науч. ст. / под ред. В.И. Кейлис-Борока. – М. : Наука, 1978. – С. 48-71. (Вычислительная сейсмология. Вып. 11).
23. Соловьев А.А., Агаян С.М., Гвишиани А.Д., Богоутдинов Ш.Р., Шульга А. Распознавание возмущений с заданной морфологией на временных рядах. II. Выбросы на секундных магнитограммах // Физика Земли. – 2012. – № 5. – С. 37-52.
24. Alekseevskaya M., Gabrielov A., Gelfand I., Gvishiani A., Rantsman E. Formal morphostructural zoning of mountain territories // Geophysics. – 1977. – 42(2). – P. 227-233.
25. Gelfand I.M., Guberman Sh.I., Izvekova M.L., Keilis-Borok V.I., Ranzman E.Ja. Criteria of high seismicity, determined by pattern recognition // Tectonophysics. – 1972. – V. 13. – № 1-4. – P. 415-422.
26. Mikhailov V.O., Galdeano A., Diament M., Gvishiani A.D., Agayan S.M., Bogoutdinov Sh.R., Graeva E.M., Sailhac P. Application of artificial intelligence for Euler solutions clustering // Geophysics. – 2003. – V. 68. – № 1. – P. 168-180.