

Подход к систематическому прогнозу землетрясений¹

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

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

Поступила в редколлегию 20.09.2021

Аннотация—Систематический прогноз землетрясений производится регулярно с шагом Δt в заранее выбранной зоне анализа. На каждом шаге дополняется выборка эпицентров целевых землетрясений и загружаются новые данные о сейсмическом процессе. Данные обрабатываются, преобразуются в сеточные поля и производится машинное обучение по всем данным, имеющимся от начала обучения до момента прогноза t^* . Результатом является карта зоны тревоги, в которой на интервале $(t^*, t^* + \Delta t]$ ожидается эпицентр целевого землетрясения. На следующем шаге интервал обучения увеличивается на величину Δt . Обучение базируется на методе минимальной области тревоги. Математическая модель метода позволяет выбирать в сеточных полях предшествующие целевым землетрясениям аномалии, которые могут являться предвестниками землетрясений, сравнивать эти аномалии между собой и затем использовать аналогичные аномалии при прогнозе. Модель вводит ограничения на класс решающих правил прогноза. Это помогает компенсировать малое количество сильных землетрясений, используемых для обучения.

В наших предыдущих работах качество прогноза оценивалось долей обнаруженных целевых событий при заданной средней величине зоны тревоги. Здесь мы рассматриваем обобщение метода минимальной области тревоги, предназначенное для улучшения другой характеристики качества прогноза: вероятности появления хотя бы одного целевого события в ожидаемой зоне тревоги. Отличие метода состоит в том, что в момент прогноза t^* принимаются два решения: прогноз по времени и по пространству. Первое решение определяет возможность появления эпицентра землетрясения с целевой магнитудой на интервале прогноза. Если принято решение о том, что целевое событие возможно, то вычисляется карта с зоной тревоги. Целевое землетрясение предсказано, если его эпицентр попадает в вычисленную зону тревоги.

Моделирование прогноза выполнено для регионов Камчатки и Калифорнии. Исходные данные Камчатки состоят из каталога землетрясений. Для региона Калифорнии использовались данные каталога землетрясений и временные ряды GPS. Результаты показали достаточно высокие оценки вероятности обнаружения целевых событий и весьма малые значения оценок вероятности прогноза появления эпицентра целевого события в зоне тревоги. При этом оценки вероятности систематического прогноза намного превышают аналогичные вероятности прогноза по случайным полям.

КЛЮЧЕВЫЕ СЛОВА: прогноз землетрясений, машинное обучение, метод минимальной области тревоги, качество прогноза, пространственно-временные процессы.

DOI: 10.53921/18195822_2021_21_3_182

1. ВВЕДЕНИЕ

Сильные землетрясения вызывают большой социальный и экономический ущерб. Своевременное предупреждение о разрушительном землетрясении может уменьшить человеческие жертвы и материальные потери. При этом ущерб от землетрясений определяется потерями как

¹ Работа частично поддержана грантом РФФИ 20-07-00445.

от разрушений пропущенного при прогнозе события, так и от проведения защитных мероприятий при ложной тревоге. Задача прогноза землетрясений обычно состоит в том, чтобы указать ограниченную по размеру зону тревоги, в которой в течение определенного времени ожидается эпицентр землетрясения с магнитудой больше некоторого порога. Представления о возможности прогноза землетрясений базируются на данных физического моделирования процесса разрушения горных пород и реальных наблюдениях, которые показывают, что землетрясению предшествуют процессы, формирующие аномальные изменения геологической среды в области очага ожидаемого землетрясения [1, 2, 3]. Так, например, в области подготовки сильного землетрясения зафиксированы аномальные отклонения в частоте и силе сейсмических событий, в деформациях земной поверхности, в химическом составе флюидов, в уровне грунтовых вод, во времени прохождения сейсмических волн, в значениях электрических и геомагнитных полей и др. [1, 3, 4, 5, 6, 7]. Это свидетельствует в пользу предположения о том, что землетрясение может быть предсказано по локальным изменениям геологической среды.

Работы по прогнозу землетрясений ведутся во многих направлениях. Они включают исследование механизмов разрушения горных пород, изучение различных предвестников землетрясений, разработку математических моделей и методов прогноза. В то же время известно мнение о том, предсказание землетрясений невозможно [8, 9, 10]. Трудность прогноза землетрясений обусловлена в основном двумя факторами. Методы, базирующиеся на физических особенностях сейсмологического и геодинамического режима, нуждаются в том, чтобы модели подготовки очага землетрясения определялись количественными показателями, доступными для инструментальных измерений. Статистические методы прогноза требуют совершенствования сети мониторинга сейсмических и геодинамических процессов, увеличения типов инструментальных наблюдений и повышения их точности.

В настоящее время для прогноза землетрясений широко используются методы машинного обучения [11, 12, 13, 14, 15, 16, 17, 18]. В ряде работ для прогноза используются искусственные нейронные сети [19, 20, 21], а также их гибридные и рекуррентные модификации [22, 23]. Эти методы требуют для обучения достаточно больших выборок целевых событий. В то же время известно, что для ряда сейсмически активных регионов количество сильных сейсмических событий на интервалах обучения невелико. Поэтому, некоторые более простые модели могут иметь аналогичные или лучшие предсказательные возможности. В данной статье мы развиваем подход к систематическому прогнозу землетрясений, представленный в [24, 25, 26]. Алгоритм систематического прогноза с постоянным шагом по времени регулярно обучается по произошедшим землетрясениям и вычисляет карту зоны тревоги, в которой ожидается целевое землетрясение. Зона тревоги определяется функцией сеточных пространственно-временных полей признаков, рассчитанных на основе исходных данных. Подход основан на методе машинного обучения, названного методом минимальной области тревоги. При обучении метод использует математическую модель, которая позволяет выбирать предшествующие целевым землетрясениям аномалии в полях признаков, сравнивать эти аномалии друг с другом и затем выбирать аналогичные аномалии в полях признаков при прогнозе. Эта модель вводит ограничения на правило прогноза, которые помогают до некоторой степени компенсировать малое количество сильных землетрясений, используемых для обучения.

В предыдущих наших работах основной характеристикой качества систематического прогноза являлась доля обнаруженных эпицентров землетрясений с целевыми магнитудами при заданной величине зоны тревоги (вероятность обнаружения). Здесь мы рассматриваем обобщение метода минимальной области тревоги, предназначенное для улучшения другой характеристики качества прогноза: вероятности появления целевого события в ожидаемой зоне тревоги. Базовые элементы подхода изложены в разделе 2. В разделе 3 мы рассматриваем обобщенный

метод минимальной области тревоги. В разделах 4 и 5 приводятся результаты применения подхода к прогнозу землетрясений в регионах Камчатки и Калифорнии.

2. БАЗОВЫЕ ЭЛЕМЕНТЫ ПОДХОДА

2.1. Постановка задачи

Основы технологии автоматической системы систематического прогноза землетрясений разработаны в 2018 г. [24]. Демонстрационная версия системы доступна на сайте <https://distcomp.ru/geo/prognosis/>. Прогноз землетрясений дается с шагом Δt в заранее выбранной зоне анализа. На каждом шаге с удаленных серверов дополняется выборка эпицентров целевых землетрясений и загружаются новые исходные данные о сейсмическом процессе. Данные обрабатываются, преобразуются в пространственно-временные сеточные поля признаков прогноза и производится машинное обучение по всем данным, имеющимся от начала обучения до момента прогноза t^* . В результате обучения вычисляется зона тревоги, в которой на интервале $(t^*, t^* + \Delta t]$ ожидается эпицентр целевого землетрясения. На следующем шаге интервал обучения увеличивается на величину Δt .

Основу подхода к систематическому прогнозу землетрясений составляет метод минимальной области тревоги. В работах [25, 26, 27] качество прогноза оценивалось по вероятности обнаружения эпицентров сильных землетрясений, которая определялась как доля предсказанных целевых событий. Эта характеристика позволила исследовать возможности использования метода минимальной области тревоги на малых учебных выборках целевых событий, показать возможность применения метода для прогноза магнитуд землетрясений и проанализировать эффективность прогноза землетрясений по ряду пространственно-временных полей, в том числе полей, вычисленных по временным рядам данных космической геодезии. В данной работе мы обобщаем метод минимальной области тревоги с тем, чтобы приблизить подход к практическим применениям.

Для практики важно знать оценку вероятности, с которой в вычисленную зону тревоги попадет хотя бы одно целевое событие. Если прогноз дается регулярно с интервалом Δt , число прогнозов N и в зоне анализа все целевые события содержатся в M интервалах, то оценка вероятности выбора интервала с целевым событием $P = M/N$. Это значение является оценкой сверху для вероятности того, что при прогнозе хотя бы одно целевое событие попадет в ограниченную по размеру зону тревоги. В ряде сейсмически активных регионов при прогнозе сильных землетрясений на относительно небольших временных интервалах Δt оценка вероятности прогноза интервала, содержащего целевое событие, слишком мала для принятия практических решений об объявлении тревоги. Этим объясняется необходимость развития подхода к систематическому прогнозу землетрясений и обобщения метода минимальной области тревоги.

Отличие обобщенного метода минимальной области тревоги состоит в том, что в момент прогноза t^* алгоритм принимает два решения. Первое решение определяет возможность появления эпицентра землетрясения с целевой магнитудой в зоне анализа на интервале прогноза $(t^*, t^* + \Delta t]$. Это решение не зависит от пространственного положения эпицентра ожидаемого землетрясения, а определяется временным рядом тревоги Ψ . Появление целевого события в зоне анализа ожидается в момент t^* , если значение временного ряда значений тревоги $\psi(t^*) \leq \psi_0$, где $\psi_0 = N^*/N$, N^* — число интервалов Δt , на которых от начала обучения до момента t^* объявлялась тревога, N — число всех интервалов Δt от начала обучения до момента t^* . Будем называть это решение *прогнозом по времени*. Если принято решение о том, что целевое событие в зоне анализа возможно, то принимается второе решение о прогнозе по пространству: на всем интервале обучения вычисляется пространственно-временное поле тревоги Φ и для момента t^* определяется зона тревоги площадью $s(t^*)$, в которой на интер-

вале $(t^*, t^* + \Delta t]$ ожидается эпицентр целевого землетрясения. Зону тревоги для интервала $(t^*, t^* + \Delta t]$ в узлах координатной сетки определяют значения поля тревоги $\varphi(t^*) \leq \varphi_0$, где, $\varphi_0 = L^*/L$, L^* — среднее от начала обучения до момента t^* число узлов сетки зоны тревоги, L — число узлов сетки зоны анализа.

Целевое землетрясение предсказано, если его эпицентр попадает в вычисленную зону тревоги. Прогноз тем успешнее, чем больше произведение $s(t^*)\Delta t$. Вместе с тем очевидно, что при прогнозе величина этой пространственно-временной области должна быть разумно ограничена.

2.2. Модель метода минимальной области тревоги

Модель метода минимальной области тревоги состоит из следующих положений.

1. *Поля признаки прогноза.* Процессы, связанные подготовкой сильных землетрясений, могут быть представлены пространственными и пространственно-временными сеточными полями.
2. *Данные машинного обучения.* Обучение использует доступные к моменту прогноза значения сеточных полей и эпицентры землетрясений с целевыми магнитудами. Значения сеточных полей в узлах сетки соответствуют векторам многомерного пространства признаков.
3. *Условие аномальности.* Сильным землетрясениям предшествуют аномальные значения сеточных полей, близкие к максимальным или минимальным. Этим аномалиям соответствуют векторы пространства признаков, которые называются *предвестниками землетрясений*.
4. *Цилиндр предвестника.* Узел сетки, соответствующий возможному предвестнику землетрясения, находится в предшествующем этому землетрясению цилиндре с центром основания в точке эпицентра.
5. *Условие монотонности.* Пусть известна классификация полей признаков на поля, у которых аномальные значения близки к минимальным, и поля, у которых аномальные значения близки к максимальным. Вектор пространства признаков, абсолютные значения компонент которого в соответствии с данной классификацией больше или равны компонентам вектора предвестника, также являются предвестниками аналогичных событий.

В более строгой формулировке условие монотонности выглядит следующим образом. Пусть вектор $f^{(q)}$ является предвестником землетрясения q . Пусть известны поля признаков, аномальные значения которых f_i^+ близки к максимальным, и поля, аномальные значения которых f_i^- близки к минимальным. Тогда любой вектор f , у которого значения компонент $f_i^+ \geq f_i^{(q)+}$ или $f_j^- \leq f_j^{(q)-}$, также может быть предвестником аналогичного события.

Условие монотонности соответствует идеальной ситуации, в которой пространственно-временные поля признаков содержат полную информацию о процессах подготовки землетрясений. В реальной ситуации эта информация не полна. Поэтому для некоторых предвестников землетрясений условие монотонности не выполняется. Однако условие монотонности позволяет ввести меру аномальности предвестника землетрясений. Эту меру алгоритм обучения использует для нахождения правила прогноза, которое позволяет обнаружить близкое к максимальному число целевых событий при заранее заданном среднем размере зоны тревоги. Далее в разделе 2 для упрощения объяснения метода минимальной области тревоги без ограничения общности будем предполагать, что *аномальные компоненты предвестников землетрясений принимают только значения полей признаков прогноза, близкие к максимальным*.

2.3. Обобщенный метод минимальной области тревоги

Основные положения. Метод минимальной области тревоги обучается по известным к моменту прогноза значениям сеточных пространственно-временных полей и эпицентрам землетрясений с целевыми магнитудам. Обучение начинается с того, что алгоритм метода находит

в сеточных полях аномальные значения, которые далее рассматриваются в качестве возможных предвестников землетрясений $q = 1, \dots, Q$. В результате учебная выборка состоит из векторов предвестников землетрясений $\mathbf{f}^{(q)} = (f_1^{(q)}, f_2^{(q)}, \dots, f_I^{(q)})$ и немаркированных векторов $\mathbf{f}^{(g)}$ значений полей во всех остальных узлах координатной сетки, среди которых могут находиться предвестники землетрясений, соответствующие эпицентрам землетрясений, не представленным в учебной выборке. По этим данным в процессе обучения требуется научиться обнаруживать близкое к наибольшему число эпицентров землетрясений при условии, что число выделенных немаркированных векторов ограничено. Алгоритмы машинного обучения в данной постановке относятся к методам одноклассовой классификации [28, 29, 30].

Пусть известно, что событию q предшествует предвестник $\mathbf{f}^{(q)}$. С предвестником связано множество $w^{(q)}$ векторов пространства признаков \mathbb{R}^I , которые покомпонентно больше или равны $\mathbf{f}^{(q)}$, т.е. $w^{(q)} = \{\mathbf{f}^{(g)} \in \mathbb{R}^I : f_i^{(g)} \geq f_i^{(q)}, i = 1, 2, \dots, I\}$. Будем называть это множество ортантом $w^{(q)}$ с вершиной в точке $\mathbf{f}^{(q)}$, а векторы $\mathbf{f}^{(g)} \in w^{(q)}$ — *базовыми векторами* целевого события q .

Базовые векторы события q согласно условию монотонности также являются предвестниками аналогичных событий. В географических координатах каждый базовый вектор формирует *цилиндр тревоги* радиуса R и образующей T . Цилиндр тревоги базового вектора $\mathbf{f}^{(g)}$ имеет центр основания в узле сетки g с координатами $(x^{(g)}, y^{(g)}, t^{(g)})$, радиус основания R и образующую $[(x^{(g)}, y^{(g)}, t^{(g)}), (x^{(g)}, y^{(g)}, t^{(g)} + T)]$, где $t^{(g)} \in [t_0, t^*]$, t_0 — время начала обучения, t^* — момент прогноза. Землетрясение может быть предсказано только в том случае, если его эпицентр попадет в один из цилиндров тревоги. Объединение цилиндров тревоги, сформированных базовыми векторами ортанта $w^{(q)}$, выделяет множество узлов сетки $W^{(q)}$, $|W^{(q)}| = L^{(q)}$. Алгоритм обучения формирует из цилиндров тревоги пространственно-временное поле тревоги.

Теперь заметим, что землетрясение q с координатами эпицентра $(x^{(g)}, y^{(g)}, t^{(g)})$ может попасть в цилиндр тревоги, а следовательно, может быть предсказано тогда и только тогда, если его предвестнику $\mathbf{f}^{(q)}$ соответствует какой-либо узел сетки из цилиндра с центром основания в точке $(x^{(g)}, y^{(g)}, t^{(g)})$, радиусом R и образующей $[x^{(g)}, y^{(g)}, t^{(g)} - T), (x^{(g)}, y^{(g)}, t^{(g)})]$. Этот цилиндр назовем *цилиндром предвестника* события q . Предвестником события q является вектор $\mathbf{f}^{(q)}$ цилиндра предвестника, который имеет минимальное значение $v^{(q)} = L^{(q)}/L$ среди всех векторов, соответствующих узлам сетки этого цилиндра, где $|W^{(q)}| = L^{(q)}$, L — число всех узлов сетки пространственно-временной области анализа на интервале $[t_0, t^*]$. Величина $v^{(q)}$ (*объем предвестника*) определяет меру аномальности предвестника события q .

Схема алгоритма обучения обобщенного метода минимальной области тревоги выглядит следующим образом. Прогноз целевого события в момент t^* на интервале $(t^*, t^* + \Delta t]$ осуществляется в два этапа. На 1-м этапе решается задача выборе для прогноза временного интервала, на котором целевое событие ожидается в пределах зоны анализа. Если принято решение, что целевое событие на данном интервале возможно, то выполняется 2-й этап прогноза: решается задача определения зоны тревоги, в которой ожидается целевое событие. Для обучения используются все данные, доступные к моменту прогноза t^* : пространственно-временные сеточные поля $\mathbf{F}_1, \mathbf{F}_2, \dots, \mathbf{F}_I$ и целевые события $q = 1, 2, \dots, Q$. На каждом шаге обучения Δt количество векторов в пространстве признаков увеличивается на число узлов пространственной сетки зоны анализа. Параметрами каждого этапа алгоритмов являются временные интервалы и радиусы цилиндров тревоги и предвестника, а также значения порогов функции тревоги, ψ_0 для принятия решения об объявлении интервала тревоги на 1-м этапе и φ_0 для определения размера зоны тревоги на 2-м этапе.

Этап 1: обучение прогнозу по времени. На 1-м этапе радиус R цилиндров тревоги и предвестника выбираются таким, чтобы круг с центром в любом узле сетки зоны анализа покрывал

всю зону анализа. Выбор указанного значения радиуса цилиндра приводит к тому, предвестник землетрясения, может относиться к любому узлу сетки области анализа. Таким образом, предвестник теряет пространственную привязку. При этом цилиндр тревоги вырождается во временной интервал, а пространственно-временное поле тревоги вырождается во временной ряд тревоги Ψ .

К моменту очередного прогноза t^* в пространстве признаков находятся векторы, соответствующие значениям полей признаков в узлах сетки зоны анализа на всех интервалах Δt за время от начала обучения до момента прогноза. По этим данным алгоритм обучения вычисляет значение временного ряда тревоги для момента прогноза t^* . Решение о тревоге на интервале $(t^*, t^* + \Delta t]$ объявляется в том случае, если значение $\psi(t^*)$ не превышает заранее заданного порога ψ_0 . Далее будет показано, что значения временного ряда тревоги Ψ принимают значения от 0 до 1 и равны отношению числа N^* интервалов, на которых объявлена тревога, к всем N интервалам от начала прогнозирования до момента t^* .

Алгоритм вычисления временного ряда тревоги для момента времени t^* состоит в следующем.

1. Найти предвестники землетрясений $f^{(q)}$ для каждого целевого события $q = 1, 2, \dots, Q$. Пусть событие q произошло в момент $t^{(q)}$. Предвестник $f^{(q)}$ события q ищется на интервале $[t^{(q)} - T, t^{(q)})$ среди векторов, соответствующим узлам сетки $g = 1, 2, \dots, G$ цилиндра предвестника. Каждый из этих векторов $f^{(q)}$ выделяет ортант $w^{(q)}$ базовых векторов. Каждый базовый вектор ортанта $w^{(q)}$ генерирует интервал тревоги T . Эти базовые векторы и генерируемые ими интервалы тревоги могут относиться к любому отсчету $\psi(t)$ временного ряда тревоги Ψ . Интервалы тревоги, которые соответствуют базовым векторам ортанта $w^{(q)}$, выделяют множество отсчетов временного ряда тревоги $W^{(q)}$, $|W^{(q)}| = N^{(q)}$. Предвестником события q является такой вектор $f^{(q)}$, который имеет минимальное значение объема тревоги предвестника $v^{(q)} = N^{(q)}/N$ среди векторов, соответствующих узлам сетки цилиндра предвестника, где $N^{(q)}$ — число интервалов тревоги при выборе ортанта $w^{(q)}$ с вершиной $f^{(q)}$, N — число всех интервалов от начала обучения до момента t^* .
2. Упорядочить предвестники $f^{(q)}$ в соответствии с увеличением объемов тревоги предвестников: $v^{(1)} \leq v^{(2)} \leq \dots \leq v^{(q)} \leq \dots \leq v^{(Q)}$.
3. Вычислить временной ряд тревоги Ψ .
 - (а) Присвоить отсчетам $\psi(t)$ временного ряда тревоги Ψ значение 1.
 - (б) Заменить значения 1 ряда значением $\psi(1) = v^{(1)}$ в множестве $W^{(1)}$ отсчетов ряда, соответствующих базовым точкам предвестника $f^{(1)}$; заменить значение 1 значением $\psi(2) = |W^{(1)} \cup W^{(2)}|/N$ в множестве $W^{(2)} \setminus W^{(1)}$; заменить значение 1 значением $\psi(3) = |W^{(1)} \cup W^{(2)} \cup W^{(3)}|/N$ в множестве $W^{(3)} \setminus (W^{(1)} \cup W^{(2)})$ и затем последовательно заменять значения 1 на соответствующие значения объемов тревоги. Полученный в результате временной ряд тревоги Ψ принимает значения $\psi(1) \leq \psi(2) \leq \dots \leq \psi(q) \leq \dots \leq \psi(Q) \leq 1$. Величина $\psi(1)$ относится к отсчетам ряда тревоги $W^{(1)}$, $\psi(2)$ относится к отсчетам ряда тревоги $W^{(2)} \setminus W^{(1)}$, $\psi(3)$ относится к отсчетам ряда тревоги $W^{(3)} \setminus (W^{(1)} \cup W^{(2)})$ и т.д.

Таким образом, в момент прогноза t^* обновляются значения временного ряда тревоги Ψ . Решение об объявлении тревоги на интервале $(t^*, t^* + \Delta t]$ принимается, если $\psi(t^*) \leq \psi_0$, где ψ_0 — величина заранее заданного порога. Интервалы Δt ряда тревоги Ψ со значениями отсчетов $\psi(t^*) \leq \psi_0$ будем называть интервалами тревоги. Число обнаруженных интервалов тревоги, в которых присутствуют целевые события M^* , зависит от значения $\psi = N^*/N$, где N^* — число интервалов тревоги, N — число всех интервалов, на которых принималось решение о прогнозе. Качество прогноза на 1-м этапе определяется следующими статистиками: $U^* = M^*/M$ — оценка вероятности успешного обнаружения интервала, в котором произошло хотя бы одно целевое событие, где M — число всех интервалов, которые содержат хотя бы одно целевое

событие, и $P_1 = M^*/N^*$ — оценка вероятности того, что в прогнозируемом интервале тревоги произойдет хотя бы одно целевое событие.

Этап 2: обучение прогнозу по пространству. Пространственный прогноз определяет зоны тревоги, в которых ожидаются целевые землетрясения. Зоны тревоги вычисляются только на интервалах тревоги со значениями $\psi(t^*) \leq \psi_0$. При этом, как и на предыдущем этапе, для обучения используются все имеющиеся к моменту прогноза t^* данные: сеточные поля F_1, F_2, \dots, F_I и целевые события $q = 1, 2, \dots, Q$.

Алгоритм обучения использует следующие параметры:

1. Область анализа и шаг координатной сетки по времени по сравнению с 1-м этапом не изменяются.
2. Пространственная координатная сетка определяется точностью данных.
3. Параметры цилиндров тревоги и предвестника одинаковы и зависят от точности данных, шага сетки и особенностей региональной сейсмичности.

Алгоритм вычисления поля тревоги состоит из трех шагов:

1. Вычислить векторы предвестников землетрясений $f^{(q)}, q = 1, \dots, Q$ (см раздел 2.3.1.)
2. Сгенерировать учебную выборку $f^{(q)}, v^{(q)}$. Упорядочить прецеденты $f^{(q)}, q = 1, \dots, Q$, в соответствии с увеличением объемов тревоги целевых событий $v^{(1)} \leq v^{(2)} \leq \dots \leq v^{(q)} \leq \dots \leq v^{(Q)}$.
3. Вычислить поле тревоги Φ .
 - (а) Присвоить узлам сетки поля тревоги значения, равные 1.
 - (б) Заменить значение 1 поля Φ на $\varphi(1) = v^{(1)}$ для множества $W^{(1)}$ узлов сетки, соответствующих базовым точкам прецедента $f^{(1)}$; заменить значение 1 на $\varphi(2) = |W^{(1)} \cup W^{(2)}|/L$ в узлах сетки, $W^{(2)} \setminus W^{(1)}$; заменить значение 1 на $\varphi(3) = |W^{(1)} \cup W^{(2)} \cup W^{(3)}|/L$ в узлах сетки $W^{(3)} \setminus (W^{(1)} \cup W^{(2)})$, и далее последовательно заменять значения 1 на соответствующие значения объемов тревоги. Полученное поле Φ принимает значения $\varphi(1) \leq \varphi(2) \leq \dots \leq \varphi(1q) \leq \dots \leq \varphi(Q) \leq 1$. Значение $\varphi(1)$ относится к узлам сетки из множества $W^{(1)}$, $\varphi(2)$ относится к узлам сетки из множества $(W^{(2)} \setminus W^{(1)})$, $\varphi(3)$ относится к узлам сетки из множества $(W^{(3)} \setminus (W^{(1)} \cup W^{(2)}))$, и т.д.

Таким образом, на каждом шаге прогноза Δt вычисляются новые значения пространственно-временного поля тревоги Φ . Временной срез поля тревоги в момент t^* используется для прогноза. Решение об успешности прогноза землетрясения с целевой магнитудой на интервале $(t^*, t^* + \Delta t]$ принимается, если его эпицентр попадает в зону тревоги $\varphi(t^*) \leq \varphi_0$, где φ_0 — величина заранее заданного порога. Число событий M^{**} в зонах тревоги, вычисленных в выбранных на 1-м этапе интервалах тревоги, зависит от значения объема тревоги $\varphi_0 = L^*/L$, где L^* — среднее число узлов сетки в зонах тревоги, L — число узлов сетки в зоне анализа всех интервалов, на которых принималось решение о прогнозе. Качество прогноза на 2-м этапе определяется следующими статистиками: $U^{**} = Q^*/Q$ — оценка вероятности успешного обнаружения целевого события на интервалах тревоги, где M — число интервалов с целевыми событиями в зоне анализа, M^{**} — число интервалов тревоги с целевыми событиями в зоне тревоги, и $P_2 = M^{**}/N^*$ — оценка вероятности того, что интервале тревоги хотя бы одно целевое событие попадет в прогнозируемую зону тревоги.

3. МОДЕЛИРОВАНИЕ

3.1. Исходные данные

Моделирование прогноза землетрясений выполнено на данных регионов Камчатки и Калифорнии. Исходные данные региона Камчатки состоят из каталога землетрясений Камчатского

филиала Геофизической службы РАН для 04.04.1986–05.05.2021 гг. с магнитудами $m \geq 3.5$ и глубинами гипоцентров $H \leq 160$ км (<http://sdis.emsd.ru/info/earthquakes/catalogue.php>) [31]. Прогноз вычислялся для эпицентров землетрясений с целевыми магнитудами $m \geq 5.8$ и с глубинами гипоцентров $H \leq 60$ км.

Исходные данные Калифорнии включают в себя два типа исходных данных: каталог землетрясений Национального центра данных (National Earthquake Information Center, NEIC) [32] для 01.01.1995–14.11.2020 гг. с магнитудами $m \geq 2.0$ и глубинами гипоцентров $H \leq 160$ км, а также временные ряды суточных горизонтальных смещений поверхности земли на интервале 01.01.2008–14.11.2020. Временные ряды получены из Nevada Geodetic Laboratory (NGL), <http://geodesy.unr.edu/about.php> [33]. Среднее расстояние между приемными станциями GPS составляет 9.38 км при стандартном отклонении 5.74 км. Прогноз вычислялся для эпицентров землетрясений с целевыми магнитудами $m \geq 5.5$ и с глубинами гипоцентров $H \leq 60$ км.

Зона анализа для Камчатки определяется по сейсмологическим данным. Зона ограничена пороговым значением пространственной плотности эпицентров землетрясений на интервале 1987–1999 гг., при котором для объема тревоги $\varphi_0 = 0.2$ вероятность обнаружения эпицентров землетрясений с магнитудами $m \geq 5.8$ по этому полю была бы не более 0.3–0.4. Для Калифорнии зона анализа определяется пересечением двух зон А и В. Зона А вычисляется по пространственной плотности эпицентров землетрясений по той же методике, что и для Камчатки. Зона В ограничивается территорией, для которой расстояние до наземных приемных станций GPS не превышает 50 км. Зоны анализа регионов Камчатки (слева) и Калифорнии (справа) показаны на Рис. 1. Кружками показаны эпицентры целевых землетрясений с магнитудами $m \geq 5.8$ для Камчатки и с магнитудами $m \geq 5.5$ для Калифорнии, используемые для обучения (желтый цвет) и прогноза (красный цвет). Черные точки — приемные станции GPS.

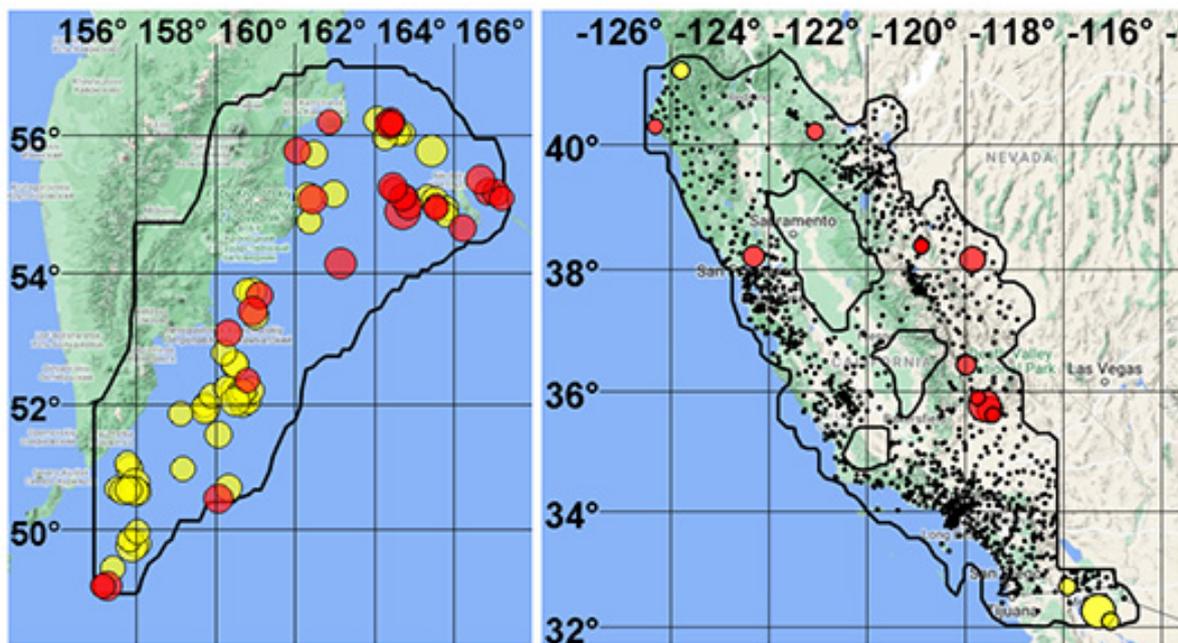


Рис. 1. Зоны анализа регионов Камчатки (слева) и Калифорнии (справа). Кружки — эпицентры целевых землетрясений с магнитудами $m \geq 5.8$ для Камчатки и с магнитудами $m \geq 5.5$ для Калифорнии, используемые для обучения (желтый цвет) и прогноза (красный цвет). Черные точки — приемные станции GPS.

3.2. Препроцессинг

3.2.1. Сейсмологические данные. По данным каталогов землетрясений были вычислены базовые пространственные и пространственно-временные сеточные поля, аналогичные тем, которые использовались в наших предыдущих работах [24, 25, 26]: поля плотности эпицентров землетрясений, средней магнитуды землетрясений, положительных и отрицательных вариаций значений полей во времени.

- S_1 — 3D поле плотности эпицентров землетрясений.
- S_2 — 3D поле средних магнитуд землетрясений.
- S_3 — 3D поле вариаций плотности эпицентров землетрясений во времени. Значения поля $s_{3n}(t)$ в узле сетки n равны отношению разности средних значений плотности эпицентров $((\overline{s_{1n}}) - (\overline{s_{2n}}))$ в двух последовательных интервалах T_1 и T_2 к выборочному стандартному отклонению этой разности $\sigma_n(s_1)$.
- S_4 — 3D поле вариаций средних магнитуд землетрясений во времени. Значения поля вычисляются аналогично значениям поля S_3 .
- S_5 — 3D поле квантилей фоновой плотности эпицентров землетрясений на интервале от начала анализа до начала обучения, значения которого соответствуют плотности эпицентров землетрясений на интервале анализа.
- S_6 — 2D поле плотности эпицентров землетрясений на интервале от начала анализа до начала обучения.

При выборе наиболее информативных полей признаков кроме полей, указанных в списке, использовались поля, равные их произведениям и отношениям.

3.2.2. Геодинамические данные. Геодинамические данные представлены временными рядами компонент $x(t)$ и $y(t)$ координат суточных горизонтальных смещений приемных станций GPS в направлениях Э-В и С-Ю. Эффективность использования данных космической геодезии для прогноза землетрясений исследована в [26]. Вычисление полей по временным рядам GPS выполняется в два шага. На первом шаге из временных рядов приемных станций GPS выделяется полезный сигнал, а далее вычисляются пространственно-временные поля признаков прогноза.

Суточная скорость смещений поверхности земли $g_x(t)$ и $g_y(t)$ вычисляем по смещениям на интервале $T_0 = 30$ суток: $g_x(t) = (x(t) - x(t - T_0))/T_0$ и $g_y(t) = (y(t) - y(t - T_0))/T_0$. Изменения координат станций для этой величины интервала сравнимо с величиной шума суточных измерений. По каждой координате временных рядов имеются большое число разрывов. При вычислении скоростей на интервалах T_0 количество пропущенных значений значительно возрастает. Чтобы ограничить число пропусков в рядах скоростей, мы линейно интерполируем значения координат станций при разрывах не более T_0 . При разрывах в рядах координат более T_0 , мы заканчиваем вычисление скорости на последнем значении до разрыва и вновь продолжаем оценивание скоростей, начиная от первого значения после разрыва.

Пространственно-временные поля вычисляются в сетке $\Delta x \times \Delta y \times \Delta t = 0.1^\circ \times 0.075^\circ \times 1$ сутки. Значения временных рядов интерполируются с помощью известного метода взвешенных обратных расстояний. В результате интерполяции разрывы во временных рядах скоростей, относящихся к узлам пространственной сетки, заполняются за счет данных от соседних приемных станций. После интерполяции поля компонент скоростей V_x и V_y сглаживались со скользящим окном 25×25 км².

Мы предполагаем, что предвестники сильных землетрясений могут быть найдены в пространственно-временных полях аномальных изменений режима деформаций поверхности земли. Аномалии во времени вычислялись в трех полях инвариантов скорости деформаций:

– F_1 — поле дивергенции скорости деформаций:

$$\operatorname{div}V_g = \frac{\partial V_{xg}}{\partial x} + \frac{\partial V_{yg}}{\partial y} \quad (1)$$

– F_2 — поле ротора скорости деформаций:

$$\operatorname{rot}V_g = \frac{\partial V_{xg}}{\partial y} - \frac{\partial V_{yg}}{\partial x} \quad (2)$$

– F_3 — поле сдвиговых деформаций:

$$\operatorname{sh}V_g = \frac{1}{2} \sqrt{\left(\frac{\partial V_{xg}}{\partial x} - \frac{\partial V_{yg}}{\partial y}\right)^2 + \left(\frac{\partial V_{xg}}{\partial y} + \frac{\partial V_{yg}}{\partial x}\right)^2} \quad (3)$$

Поля вариаций инвариантов скорости деформаций во времени вычислялись в координатной сетке $\Delta x \times \Delta y \times \Delta t = 0.1^\circ \times 0.075^\circ \times 30$ суток.

– F_4 — поле вариаций дивергенции скорости деформаций. Значения поля $f_{4g}(t)$ в узле сетки g равны отношению разности средних значений дивергенции ($\overline{(\operatorname{div}_{2g})} - \overline{(\operatorname{div}_{1g})}$) в двух последовательных интервалах T_1 и T_2 к стандартному отклонению этой разности $\sigma_g(\operatorname{div})$, $T_1 = T_2 = 361$ суток,

$$f_{4g}(t) = (\overline{\operatorname{div}_{2g}} - \overline{\operatorname{div}_{1g}}) / \sigma_g(\operatorname{div}), \quad (4)$$

где $\overline{\operatorname{div}_{2g}}$ вычисляется по значениям поля F_1 на интервале $(t - T_2, t)$, $\overline{\operatorname{div}_{1g}}$ вычисляется на интервале $(t - T_2 - T_1, t - T_2)$.

– F_5 — поле вариаций ротора скорости деформаций. Значения поля $f_5(t)$ вычисляются аналогично значениям поля F_4 ,

$$f_{5g}(t) = (\overline{\operatorname{rot}_{2g}} - \overline{\operatorname{rot}_{1g}}) / \sigma_g(\operatorname{rot}). \quad (5)$$

– F_6 — поле вариаций сдвиговых деформаций.

$$f_{6g}(t) = (\overline{\operatorname{sh}_{2g}} - \overline{\operatorname{sh}_{1g}}) / \sigma_g(\operatorname{sh}). \quad (6)$$

3.3. Прогноз

Обнаружение целевых событий на всех интервалах прогноза. Оценка вероятности обнаружения целевых событий определяется долей целевых событий $U = Q^*/Q$, попавших в прогнозную зону тревоги, средний размер которой ограничен величиной объема тревоги φ_0 , где Q — число прогнозируемых целевых событий, Q^* — число успешных прогнозов.

При обнаружении целевых событий используются следующие параметры алгоритма минимальной области тревоги. Интервалы обучения для регионов Камчатки и Калифорнии начинаются 01.01.1999 г. и 23.09.2009 г. и заканчиваются в момент очередного прогноза. Прогноз выполняется регулярно с шагом $\Delta t = 30$ суток на интервале тестирования 24.12.2011–05.05.2021 гг. для Камчатки и 19.01.2013–14.11.2020 гг. для Калифорнии. На интервале тестирования в зону анализа Камчатки попадает 49 эпицентров целевых землетрясений с магнитудами $m \geq 5.8$, а в зону анализа Калифорнии попадают 12 эпицентров землетрясений с магнитудами $m \geq 5.5$. Радиусы цилиндров тревоги и предвестника составляют для Камчатки $R = 10$ км и для Калифорнии $R = 18$ км. Интервалы образующих цилиндров тревоги и предвестника для обоих регионов одинаковы: $T = 61$ суток. Для вычисления поля тревоги в регионе Камчатки использовались 3 поля: (1) поле $S_9 = S_5 \times S_2$ произведений квантилей поля S_5 на значения средних

магнитуд землетрясений поля S_2 ; (2) поле S_{10} максимальных значений вариаций плотности эпицентров землетрясений во времени с параметрами окон $T_1 = T_2 = 1000$ суток; (3) поле S_{11} — максимальных значений вариаций плотности эпицентров землетрясений во времени с параметрами $T_1 = 3650$ и $T_2 = 365$ суток. В Калифорнии использовались те же 4 поля, что и в статье [26]: (1) поле F_6 — максимальных значений вариаций во времени сдвиговых деформаций поверхности земли с параметрами $T_1 = T_2 = 365$ суток; (2) поле F_4 — максимальных значений вариаций во времени для дивергенции поверхности земли с параметрами $T_1 = T_2 = 365$ суток; (3) поле S_{12} — минимальных значений вариаций поля S_2 средних магнитуд землетрясений во времени с параметрами $T_1 = 1095$ и $T_2 = 730$ суток; (4) поле $S_{13} = S_1/S_5$ — отношений значений плотности эпицентров землетрясений поля S_1 и квантилей поля S_5 . Для сравнения прогноз вычислялся по полям S_6 пространственной плотности эпицентров землетрясений. Основные результаты моделирования показаны на Рис 2 и сведены в Таблицы 1 и 2. Рис 2 показывает графики зависимостей вероятности обнаружения целевых событий $U = Q^*/Q$ от объема тревоги φ для прогноза по пространственно-временным полям признакам и прогноза по пространственному полю плотности эпицентров землетрясений.

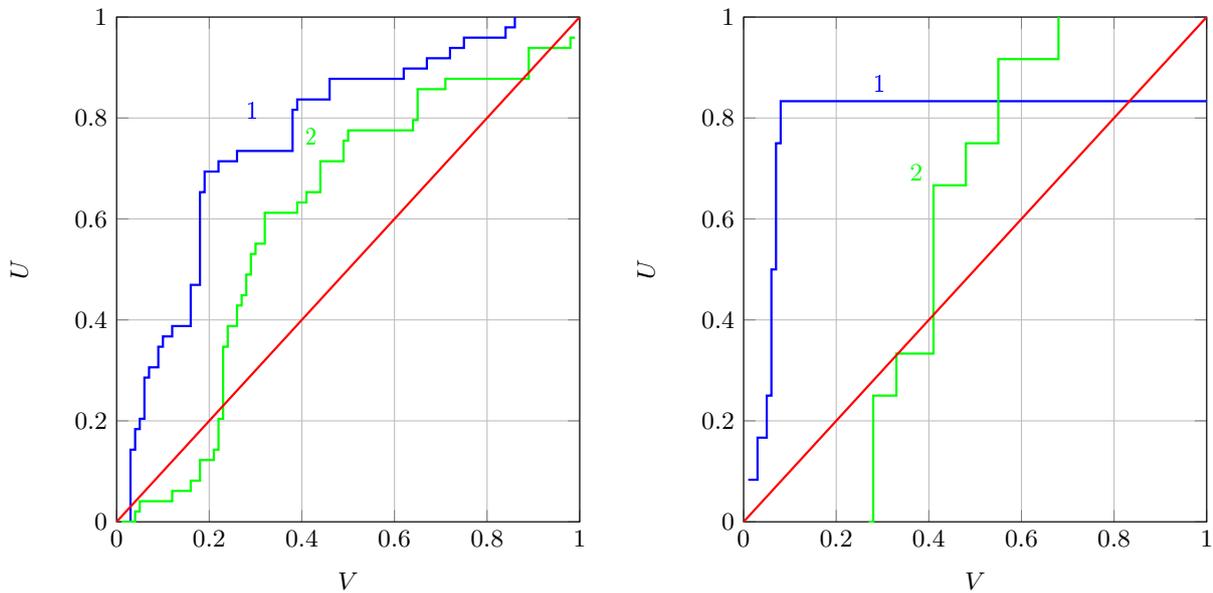


Рис. 2. Графики зависимостей $U(\varphi)$ вероятностей обнаружения целевых событий U от объема тревоги φ : (1) прогноз по выбранным пространственно-временным полям и (2) прогноз по пространственному полю плотности эпицентров землетрясений. Слева — Камчатка, 49 событий, $m \geq 5.8$; Справа — Калифорния, 12 событий, $m \geq 5.5$.

Таблица 1. Камчатка: вероятность обнаружения целевых событий.

Объем тревоги φ	Поля признаков S_9, S_{10} и S_{11}		Поле 2D плотности эпицентров S_6	
	Число событий Q^*	Вероятность $U = Q^*/Q$	Число событий Q^*	Вероятность $U = Q^*/Q$
0.05	10	0.20	2	0.04
0.10	18	0.37	2	0.04
0.15	19	0.39	3	0.06
0.20	34	0.69	6	0.12
0.25	35	0.71	19	0.39
0.30	37	0.76	27	0.55

Таблица 2. Калифорния: вероятность обнаружения целевых событий.

Объем тревоги φ	Поля признаков F_4, F_6, S_{12} и S_{13}		Поле 2D плотности эпицентров S_6	
	Число событий Q^*	Вероятность $U = Q^*/Q$	Число событий Q^*	Вероятность $U = Q^*/Q$
0.05	3	0.39	0	0
0.10	10	0.83	0	0
0.15	10	0.83	0	0
0.20	10	0.83	0	0
0.25	10	0.83	0	0
0.30	10	0.83	3	0.25

Из рисунка и таблиц можно видеть, что значения вероятности обнаружения целевых событий при использовании указанных пространственно-временных полей значительно выше, чем при использовании поля пространственной плотности эпицентров землетрясений. Это свидетельствует о пространственной однородности целевых событий в выбранных зонах анализа.

Рассмотрим теперь оценки вероятностей прогноза. Интервал тестирования Камчатки 24.12.2011–05.05.2021 содержит $N = 115$ интервалов прогноза $\Delta t = 30$ суток. Все $Q = 49$ целевых событий с магнитудами $m \geq 5.8$ попадают в $M = 23$ интервала. Число интервалов, для которых целевые события попадают в зоны с объемом тревоги $\varphi = 0.2$ составляет $M^{**} = 16$. Оценка вероятности того, что на интервале прогноза Δt хотя бы одно целевое событие попадает в зону анализа, равно доле интервалов прогноза с целевыми событиями, $P_1 = M/N = 23/115 = 0.2$. Оценка вероятности того, что на интервале прогноза Δt хотя бы одно целевое событие попадает в зону тревоги с объемом тревоги $\varphi = 0.2$, равно доле интервалов прогноза, в которых зоны тревоги содержат целевые события, $P_2 = M^{**}/N = 16/115 = 0.14$. Аналогичными величинами для Калифорнии являются $Q = 12$, $m \geq 5.5$, $M = 7$ из $N = 95$, $P_1 = M/N = 7/95 = 0.07$.

Обнаружение целевых событий на интервалах тревоги

Этап 1. Вероятность того, что на интервале прогноза Δt в зону тревоги попадает хотя бы одно целевое событие, оценивается как отношение числа интервалов тревоги, на которых хотя бы одно целевое событие попало в зону тревоги к общему числу интервалов тревоги. При этом пространственная точность прогноза определяется объемом тревоги φ , равным средней доле зоны тревоги от площади зоны анализа. Зона тревоги вычисляется в 2 этапа. На 1-м этапе выбирается интервал тревоги, в течение которого целевое событие ожидается в зоне анализа. На 2-м этапе в выбранном интервале тревоги вычисляется зона тревоги, в которой ожидается целевое событие.

Моделирование для региона Камчатки выполнялось по следующей схеме. Интервал 01.01.1999–12.01.2014 гг. использован для подбора параметров алгоритма: выбора полей признаков, параметров цилиндра тревоги, цилиндра предвестника, и порога ψ_0 .

Тестирование выполняется на интервале 12.01.2014–05.05.2021. Значения пространственно-временных полей, вычисленных по данным GPS для региона Калифорнии, известны на очень коротком интервале 23.09.2009–14.11.2020 гг. Весь этот интервал использован для тестирования. Поэтому поля и параметры алгоритма выбраны по результатам обнаружения целевых событий такими же, как в [26]. Значение порога ψ_0 выбрано по результату тестирования на 1-м этапе.

Камчатка. Для вычисления временного ряда тревоги для региона Камчатки поля признаков были сглажены по пространству с окном радиуса 130 км и преобразованы в сетку $\Delta x \times \Delta y \times \Delta t = 2^\circ \times 1.5^\circ \times 30$ суток. Результаты обучения показали, что для выбора

временных интервалов тревоги наиболее информативными являются следующие сглаженные пространственно-временные поля: (1) поле $S_{9smoothed} = S_5 \times S_2$ произведений квантилей поля S_5 на значения средних магнитуд землетрясений поля S_2 ; (2) поле $S_{10smoothed}$ произведений квантилей поля S_5 на значения плотности эпицентров землетрясений поля S_1 ; (3) поле $S_{11smoothed}$ произведений значений плотности эпицентров землетрясений поля S_1 на значения средних магнитуд землетрясений поля S_2 . Параметры алгоритма на 1-м этапе: $T = 61$ суток, $R = 3000$ км. Решение о назначении интервала прогноза интервалом тревоги принимается, если значения временного ряда не превышают порога $\psi_0 = 0.30$.

Калифорния. Для выполнения 1-го этапа поля были сглажены по пространству с окном радиуса 200 км и преобразованы в сетку $\Delta x \times \Delta y \times \Delta t = 1.85^\circ \times 1.85^\circ \times 30$ суток. Для прогноза во времени использовались сглаженные поля: (1) поле $F_{6smoothed}$ максимальных значений поля вариаций сдвиговых деформаций поверхности земли во времени с параметрами $T_1 = T_2 = 365$ суток; (2) поле $F_{4smoothed}$ минимальных значений поля вариаций дивергенции поверхности земли во времени с параметрами $T_1 = T_2 = 365$ суток. На 1-м этапе параметры алгоритма минимальной области тревоги для региона Калифорнии составляют $T = 61$ суток, $R = 3000$ км. Решение о назначении интервала прогноза интервалом тревоги принимается, если значения временного ряда Ψ не превышают порога $\psi_0 = 0.35$.

Данные тестирования относятся к $N = 89$ интервалам прогноза Δt для Камчатки и $N = 95$ интервалам для Калифорнии. Тестовые землетрясения Камчатки, $Q = 26$, содержатся в $M = 19$ интервалах. Тестовые землетрясения Калифорнии, $Q = 12$, содержатся в $M = 7$ интервалах. Вероятности того, что при тестировании в случайно выбираемый интервал Δt попадает хотя бы одно целевое землетрясение определяются, как $P = M/N$: для Камчатки $P = 19/89 = 0.21$, а для Калифорнии, $P = 7/95 = 0.07$.

Результаты моделирования 1-го этапа прогноза для Камчатки и Калифорнии показаны на Рис 3 и сведены в Таблицы 3 и 4.

Рис 3 показывает график зависимости оценки вероятности $U^* = M^*/M$ обнаружения тестовых землетрясений в интервалах тревоги Δt от доли интервалов тревоги $\psi = N^*/N$, где N^* — число интервалов тревоги, N — число всех интервалов прогноза, M — число всех интервалов с целевыми событиями, M^* — число интервалов тревоги, в которых целевые события присутствуют, P_1 - вероятность того, что на интервале тревоги в зону анализа попадает хотя бы одно целевое событие.

На Рис 4 показан график временного ряда тревоги Ψ и последовательность целевых землетрясений на интервале обучения. На Рис 5 и 6 показаны временные ряды интервалов тревоги и тестовая временная последовательность целевых землетрясений с $m \geq 5.8$ для Камчатки и $m \geq 5.5$ для Калифорнии. В момент прогноза t^* решение о тревоге на интервале $(t^*, t^* + T]$ принимается, если значение $\psi(t)$ не превышает порога $\psi_0 = 0.30$ для Камчатки и $\psi_0 = 0.35$ для Калифорнии.

Таблица 3. Камчатка. Результаты тестирования 1-го этапа прогноза.

Объем тревоги ψ	Камчатка $m \geq 6.0$			
	N^*	M^*	$U^* = M^*/M$	$P_1 = M^*/N^*$
0.30	46	13	0.68	0.28
0.45	70	16	0.84	0.23

Из таблицы можно видеть, что для Камчатки выбранному порогу $\psi_0 = 0.30$ соответствует вероятность успешного прогноза на интервале тревоги $P_1 = 0.28$, а для Калифорнии поро-

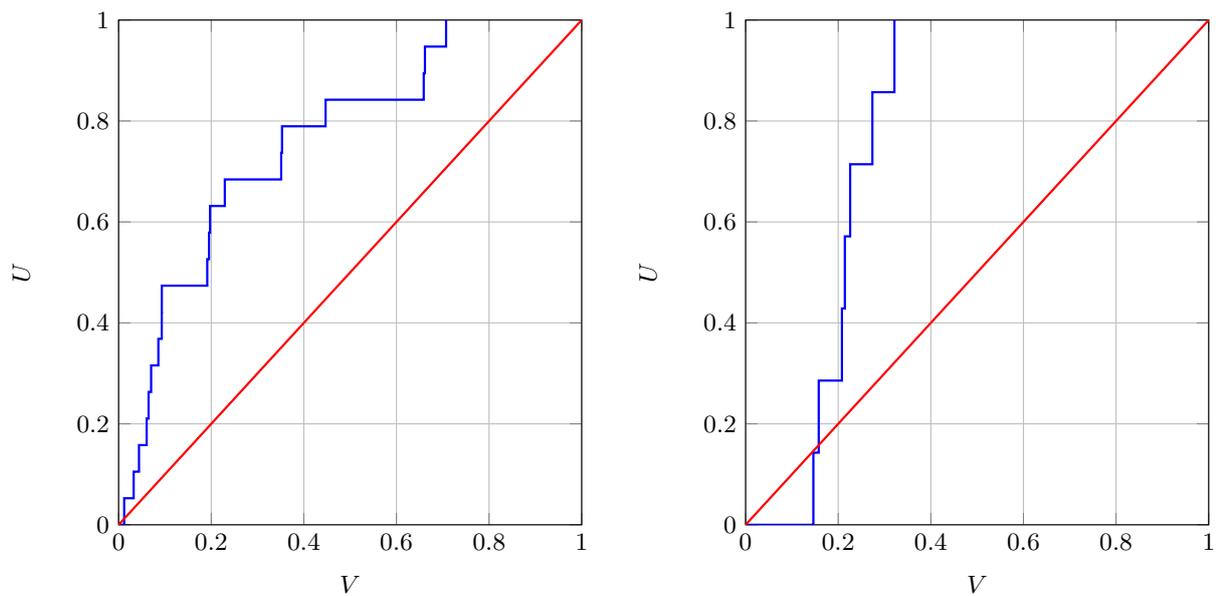


Рис. 3. Графики зависимостей $U^*(\psi)$ вероятностей обнаружения интервалов с целевыми событиями $U^* = M^*/M$ от отношения числа интервалов тревоги к общему числу интервалов прогноза на интервале тестирования $\psi = N^*/N$. Слева: Камчатка, прогноз интервалов тревоги с событиями $m \geq 5.8$. Справа: Калифорния, прогноз интервалов тревоги с событиями $m \geq 5.5$.

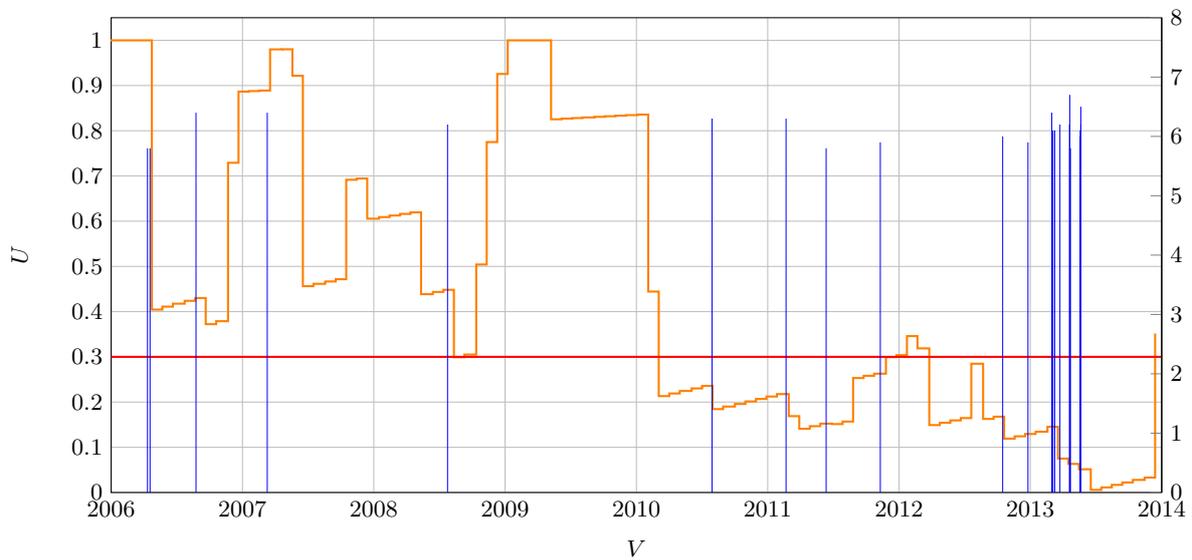


Рис. 4. Камчатка. Временной ряд интервалов тревоги Ψ и последовательность целевых землетрясений с $m \geq 5.8$ на интервале обучения. Решение о тревоге принимается, если значения временного ряда не выше прямой со значением $\psi_0 = 0.30$.

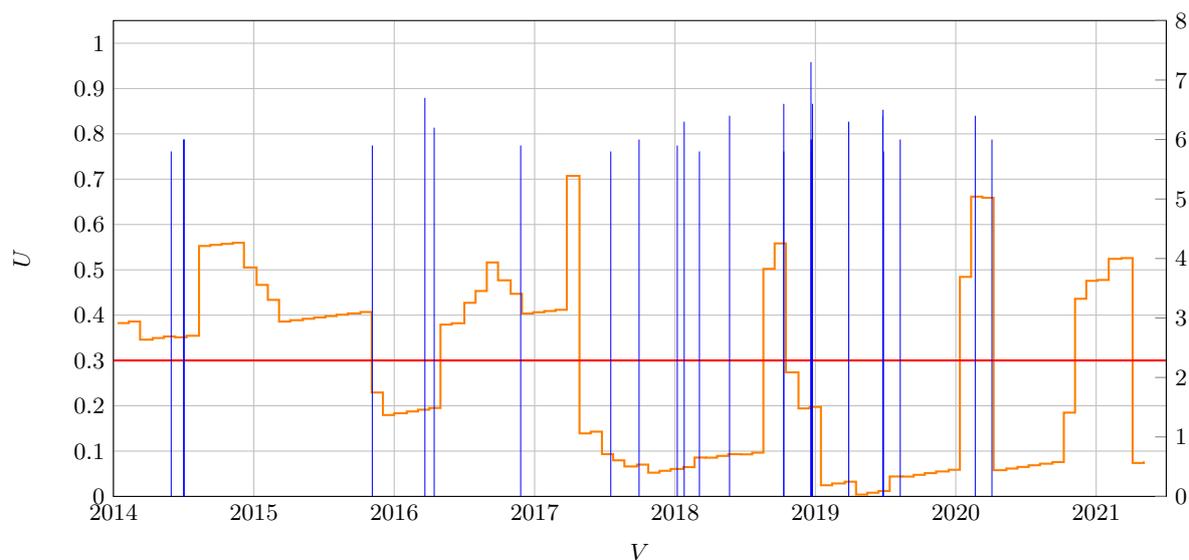


Рис. 5. Камчатка. Временной ряд интервалов тревоги Ψ и тестовая последовательность целевых землетрясений с $m \geq 5.8$. Решение о тревоге принимается, если значения временного ряда не выше прямой со значением $\psi_0 = 0.30$.

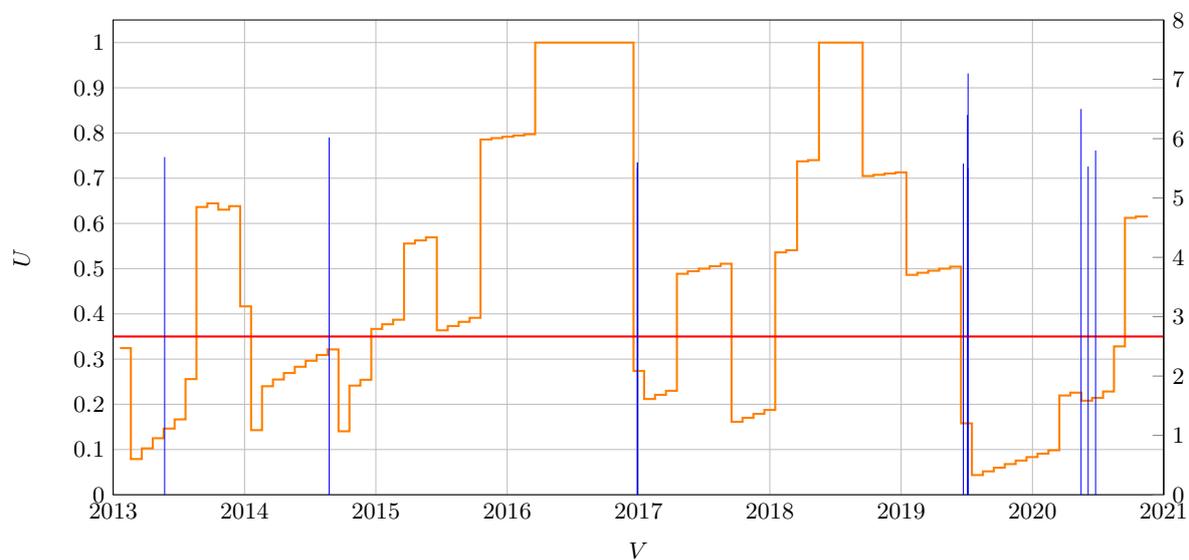


Рис. 6. Калифорния. Временной ряд интервалов тревоги Ψ и тестовая последовательность целевых землетрясений с $m \geq 5.5$. Решение о тревоге принимается, если значения временного ряда не выше значения $\psi_0 = 0.35$.

Таблица 4. Калифорния. Результаты тестирования 1-го этапа прогноза.

Объем тревоги ψ	Калифорния $m \geq 5.5$			
	N^*	M^*	$U^* = M^*/M$	$P_1 = M^*/N^*$
0.25	30	5	0.71	0.17
0.35	41	7	1.00	0.17

гу $\psi_0 = 0.35$ — вероятность $P_1 = 0.17$. Эти величины превосходят соответствующие оценки вероятностей успешного прогноза землетрясений в зоне анализа при случайном выборе интервалов: $P = 0.21$ и $P = 0.07$. Они являются верхними границами для вероятностей того, что на 2-м этапе хотя бы одно целевое событие попадает в зону тревоги, вычисленную при выбранных на 1-м этапе значениях ψ_0 .

Этап 2. Для пространственно-временного прогноза использовались те же пространственно-временные поля, что и в разделе 3.2.2. Результаты моделирования 2-го этапа прогноза сведены в Таблицу 5. В таблице используются следующие обозначения: обозначения φ — доля узлов сетки поля тревоги V при обучении; N^* — число интервалов тревоги при тестировании; $M^{**}(\psi_0, \varphi_0)$ — число интервалов тревоги, в которые при прогнозе хотя бы одно целевое событие попало в зону тревоги; $U^{**} = M^{**}(\psi_0, \varphi_0)/M$ — вероятность обнаружения интервалов тревоги, для которых при тестировании в зону тревоги попало хотя бы одно событие; $P_2 = M^{**}(\psi_0, \varphi_0)/N^*(\psi_0)$ — вероятность того, что при прогнозе на интервале тревоги, выбранному с порогом ψ_0 , в зону с объемом тревоги φ_0 попадает хотя бы одно целевое событие.

Таблица 5. Калифорния. Результаты тестирования 1-го этапа прогноза.

Объем тревоги φ	Камчатка $m \geq 6$, $\psi_0 = 0.30, M = 19, N^*(\psi_0) = 46$			Калифорния $m \geq 5.5$, $\psi_0 = 0.35, M = 7, N^*(\psi_0) = 41$		
	M^{**}	U^{**}	P_2	M^{**}	U^{**}	P_2
0.10	5	0.26	0.10	6	0.86	0.15
0.20	7	0.37	0.15	6	0.86	0.15
0.30	7	0.37	0.15	6	0.86	0.15

Из таблицы 5 можно видеть, что для Камчатки при пороге $\psi_0 = 0.30$ вероятность того, что в интервале тревоги хотя бы одно целевое событие попадет в зону тревоги при средней величине зоны тревоги, равной 20% и 30% от зоны анализа, равна $P_2 = 0.15$. Аналогичные результаты для Калифорнии при пороге $\psi_0 = 0.35$ составляют $P_2 = 0.15$ при средней величине зоны тревоги от 10% до 30% от зоны анализа.

4. ДИСКУССИЯ

Основные результаты моделирования можно представить с помощью следующих четырех показателей:

1. Оценка вероятности обнаружения целевых событий U при заданном объеме тревоги φ , равная доле обнаруженных событий Q^* в прогнозных зонах тревоги, площадь которых в среднем от начала обучения до момента прогноза составляет часть φ_0 от площади зоны анализа: $U = Q^*(\varphi_0)/Q$.
2. Оценка вероятности обнаружения интервалов тревоги с целевыми событиями M^* при заданной доле интервалов тревоги ψ_0 от всех интервалов M : $U^* = M^*(\psi_0)/M$.

3. Оценка вероятности того, что хотя бы одно целевое событие появится на прогнозируемом интервале тревоги Δt в зоне анализа (вероятность прогноза по времени): $P_1 = M^*(\psi_0)/N^*(\psi_0)$.
4. Оценка вероятности того, что хотя бы одно целевое событие появится на интервале тревоги Δt в прогнозируемой зоне тревоги (вероятность пространственно-временного прогноза): $P_2 = M^{**}(\psi_0, \varphi_0)/N^*(\psi_0)$.

Для сравнения будем использовать значения этих же показателей при прогнозе целевых событий по случайным полям.

1. Вероятность обнаружения целевых событий при выборе зоны тревоги случайным полем с объемом тревоги φ : $U_{rand}(\varphi) = \varphi = L^*/L$.
2. Вероятность обнаружения интервалов тревоги Δt с целевыми событиями при выборе N^* интервалов из N по значениям случайного поля: $U_{rand}^*(\psi) = \psi = N^*/N$.
3. Вероятность того, что в предположении о равномерном распределении интервалов Δt с целевыми событиями при случайном выборе интервала тревоги в нем окажется хотя бы одно целевое событие: $P_{1rand} = M/N$.
4. Вероятность P_{2rand} того, что при случайном выборе зоны тревоги хотя бы одно из целевых событий, в предположении о равномерном их распределений в зоне анализа, появится в случайно сгенерированной зоне тревоги. Пусть X — случайная величина, равная числу событий в зоне тревоги. Среднее значение $\bar{X} = \varphi Q/N$, где объем тревоги φ равен отношению средней величины площади зоны тревоги к площади зоны анализа, число интервалов N равно отношению времени тестирования к величине интервала прогноза Δt . Полагая, что эмпирическое среднее X равно математическому ожиданию EX можно оценить P_{2rand} с помощью неравенства Маркова [34]: $P_{2rand}(X \geq 1) \leq \varphi Q/N$.

В Таблицах 6 и 7 сопоставляются результаты систематического прогноза землетрясений и результаты прогноза по случайным полям.

Таблица 6. Камчатка. Результаты систематического и случайного прогнозов.

№	Показатель	Систематический прогноз	Случайный прогноз
1.	Обнаружение событий $U(\varphi) = Q^*(\varphi)/Q$	$U(0.10) = 18/49 = 0.37$ $U(0.20) = 34/49 = 0.69$ $U(0.30) = 37/49 = 0.76$	$U_{rand}(\varphi) = \varphi = 0.10$ $U_{rand}(\varphi) = \varphi = 0.20$ $U_{rand}(\varphi) = \varphi = 0.30$
2.	Обнаружение интервалов $U^*(\psi) = M^*(\psi)/M$	$U^*(0.30) = 13/19 = 0.68$ $U^*(0.45) = 16/19 = 0.84$	$U_{rand}^*(\psi) = \psi = 0.30$ $U_{rand}^*(\psi) = \psi = 0.45$
3.	Прогноз интервалов $P_1(\psi) = M^*(\psi)/N^*(\psi)$	$P_1(0.30) = 13/46 = 0.28$ $P_1(0.45) = 16/70 = 0.23$	$P_{1rand} = M/N = 19/89 = 0.21$
4.	Прогноз событий $P_2(\varphi \psi) = M^{**}(\varphi)/N^*(\psi)$	$P_2(20 30) = 7/46 = 0.15$ $P_2(30 30) = 7/46 = 0.15$	$P_{2rand} \leq \varphi Q/N = 0.20 \times 26/89 = 0.06$ $P_{2rand} \leq \varphi Q/N = 0.30 \times 26/89 = 0.09$

Таблица 7. Калифорния. Результаты систематического и случайного прогнозов.

№	Показатель	Систематический прогноз	Случайный прогноз
1.	Обнаружение событий $U(\varphi) = Q^*(\varphi)/Q$	$U(0.10) = 10/12 = 0.83$ $U(0.20) = 10/12 = 0.83$ $U(0.30) = 10/12 = 0.83$	$U_{rand}(\varphi) = \varphi = 0.10$ $U_{rand}(\varphi) = \varphi = 0.20$ $U_{rand}(\varphi) = \varphi = 0.30$
2.	Обнаружение интервалов $U^*(\psi) = M^*(\psi)/M$	$U^*(0.25) = 5/7 = 0.71$ $U^*(0.35) = 7/7 = 1.00$	$U_{rand}^*(\psi) = \psi = 0.25$ $U_{rand}^*(\psi) = \psi = 0.35$
3.	Прогноз интервалов $P_1(\psi) = M^*(\psi)/N^*(\psi)$	$P_1(0.25) = 5/30 = 0.17$ $P_1(0.35) = 7/41 = 0.17$	$P_{1rand} = M/N = 7/95 = 0.07$
4.	Прогноз событий $P_2(\varphi \psi) = M^{**}(\varphi)/N^*(\psi)$	$P_2(0.2 0.35) = 6/41 = 0.15$ $P_2(0.3 0.35) = 6/41 = 0.15$	$P_{2rand} \leq \varphi Q/N = 0.20 \times 12/95 = 0.03$ $P_{2rand} \leq \varphi Q/N = 0.30 \times 12/95 = 0.04$

Результаты показывают, что при систематическом прогнозе оценки вероятности обнаружения целевых событий $U(\varphi)$ и интервалов тревоги с событиями $U^*(\psi)$ принимают достаточно высокие значения. Однако для практического применения методов прогноза землетрясений более важными являются оценки вероятностей прогноза $P_1(\psi)$ и $P_2(\psi, \varphi)$, вероятности появления хотя бы одного целевого события в прогнозируемых интервалах тревоги и зоне тревоги. Можно видеть, что экспериментальные оценки этих величин для Камчатки и Калифорнии принимают малые значения. В то же время эти оценки вероятностей прогноза выше аналогичных значений, полученных при получаемых при использовании для прогноза случайных полей признаков, а также выше тех, которые получены в разделе 3.2.2 при использовании предыдущей версии метода минимальной области тревоги: для Камчатки $P_1 = 0.2$, $P_2 = 0.14$ при $\varphi = 0.2$, для Калифорнии $P_1 = 0.07$, $P_2 = 0.06$ при $\varphi = 0.2$.

5. ЗАКЛЮЧЕНИЕ

Мы рассмотрели базовые элементы подхода к систематическому прогнозу землетрясений. Решение о прогнозе принимается регулярно для временного интервала Δt . Качество прогноза мы определяем с помощью двух типов оценок. Мы оцениваем: (1) вероятность обнаружения целевых событий, равную доле обнаруженных событий при прогнозе, и (2) вероятность прогноза целевых событий, которая определяет меру возможности появления целевого события на очередном интервале тревоги Δt в зоне анализа и в прогнозируемой зоне тревоги. Первая оценка определяет эффективность исходных данных, методов их обработки и метода прогноза. Вторая оценка нужна для принятия практических решений, связанных с прогнозом сейсмической опасности.

Решение второй задачи потребовало развития метода минимальной области тревоги. Обобщенный метод минимальной области тревоги состоит из двух этапов. На первом этапе прогнозируется возможность появления эпицентра землетрясения с целевой магнитудой в зоне анализа, и принимается решение об объявлении тревоги на интервале Δt . На втором этапе в интервале тревоги вычисляется пространственное положение зоны тревоги, в которой ожидается появление эпицентра этого землетрясения.

Метод экспериментально исследован для регионов Камчатки и Калифорнии. Прогноз землетрясений Камчатки выполнен по данным каталога землетрясений. Для Калифорнии использовались два типа данных: каталоги землетрясений и временные ряды смещений поверхности земли, полученные с помощью GPS. Результаты моделирования показывают, что при зонах тревоги, площадь которых в среднем составляет от 20% до 30% от площади зоны анализа, оценки вероятности обнаружения целевых событий составляют 0.7–0.8. При этом вероятности того, что в прогнозируемой зоне тревоги ожидается появление эпицентра целевого землетрясения принимают весьма малые значения порядка 0.15. В то же время можно видеть, что оценки вероятности систематического прогноза значительно превышают аналогичные вероятности прогноза по случайным полям.

Мы полагаем, что для развития подхода к систематическому прогнозу землетрясений в направлении решения прикладных задач целесообразно проведение следующих исследований: (1) совершенствование методов систематического прогноза землетрясений во времени, (2) поиск новых сеточных полей, представляющих характеристики подготовки очага сильного землетрясения, (3) развитие методов анализа дополнительных типов исходных данных.

Авторы выражают благодарность Е.Н. Петровой и С.А. Пирогову за полезные замечания.

СПИСОК ЛИТЕРАТУРЫ

1. Соболев Г. Основы прогноза землетрясений. – 1993.

2. Соболев Г. А., Пономарев А. В. Физика землетрясений и предвестники. — Общество с ограниченной ответственностью Международная академическая издательская компания "Наука/Интерпериодика 2003.
3. Завьялов А. Среднесрочный прогноз землетрясений: основы, методика, реализация. т. 254.—М.: Наука, 2006.
4. Lighthill J. A critical review of VAN: earthquake prediction from seismoelectrical signals.—1996.
5. King C.-Y. Gas geochemistry applied to earthquake prediction: An overview // *Journal of Geophysical Research: Solid Earth*.—1986.—т. 91, В12. —с. 12269–12281.
6. Matsumoto N., Koizumi N. Recent hydrological and geochemical research for earthquake prediction in Japan // *Natural hazards*.—2013.—т. 69, No 2.—с. 1247–1260.
7. Murai S. Can we predict earthquakes with GPS data? // *International Journal of Digital Earth*.—2010.—т. 3, No 1.—с. 83–90.
8. Earthquakes cannot be predicted / R. J. Geller [и др.] // *Science*.—1997.—т. 275, No 5306.—с. 1616–1616.
9. Короновский Н., Наймарк А. Непредсказуемость землетрясений как фундаментальное следствие нелинейности геодинамических систем // *Вестник Московского университета. Серия 4. Геология*.—2012.—No 6.
10. Гуфельд И. Л., Матвеева М. И., Новоселов О. Н. Почему мы не можем осуществить прогноз сильных коровых землетрясений // *Геодинамика и тектонофизика*.—2011.—т. 2, No 4.1
11. Keilis-Borok V., Soloviev A. A. Nonlinear dynamics of the lithosphere and earthquake prediction.—Springer Science & Business Media, 2013.
12. Kossobokov V., Shebalin P. Earthquake prediction // *Nonlinear dynamics of the lithosphere and earthquake prediction*.—Springer, 2003.—с. 141–207.
13. Machine learning can predict the timing and size of analog earthquakes / F. Corbi [и др.] // *Geophysical Research Letters*.—2019.—т. 46, No 3. —с. 1303–1311.
14. Combining earthquake forecasts using differential probability gains / P. N. Shebalin [и др.] // *Earth, Planets and Space*.—2014.—т. 66, No 1. —с. 1–14.15.
15. Amei A., Fu W., Ho C.-H. Time series analysis for predicting the occurrences of large scale earthquakes // *International Journal of Applied Science and Technology*.—2012.—т. 2, No 7.
16. Marzocchi W., Zechar J. D. Earthquake forecasting and earthquake prediction: different approaches for obtaining the best model // *Seismological Research Letters*.—2011.—т. 82, No 3.—с. 442–448.
17. Moustira M., Avraamides M., Christodoulou C. Artificial neural networks for earthquake prediction using time series magnitude data or seismoelectric signals // *Expert systems with applications*.—2011.—т. 38, No 12.—с. 15032–15039.
18. Rhoades D. A. Mixture models for improved earthquake forecasting with short-to-medium time horizons // *Bulletin of the Seismological Society of America*.—2013.—т. 103, No 4.—с. 2203–2215.
19. Alves E. I. Earthquake forecasting using neural networks: results and future work // *Nonlinear Dynamics*.—2006.—т. 44, No 1.—с. 341–349.
20. Priambodo B., Mahmudy W. F., Rahman M. A. Earthquake Magnitude and Grid-Based Location Prediction using Backpropagation Neural Network // *Knowledge Engineering and Data Science*.—2020.—т. 3, No 1.—с. 28–39.
21. Mignan A., Broccardo M. Neural network applications in earthquake prediction (1994–2019): Meta-analytic and statistical insights on their limitations // *Seismological Research Letters*.—2020.—т. 91, No 4.—с. 2330–2342.
22. Earthquake prediction model using support vector regressor and hybrid neural networks / K. M. Asim [и др.] // *PloS one*.—2018.—т. 13, No 7. —e0199004.

23. Panakkat A., Adeli H. Neural network models for earthquake magnitude prediction using multiple seismicity indicators // International journal of neural systems. – 2007. – т. 17, No 01. – с. 13–33.
24. Gitis V. G., Derendyaev A. B. Web-Based GIS platform for automatic prediction of earthquakes // International Conference on Computational Science and Its Applications. – Springer. 2018. – с. 268–283.
25. Gitis V., Derendyaev A. The Method of the Minimum Area of Alarm for Earthquake Magnitude Prediction // Frontiers in Earth Science. – 2020. – т. 8. – с. 482.
26. Gitis V., Derendyaev A., Petrov K. Analyzing the Performance of GPS Data for Earthquake Prediction // Remote Sensing. – 2021. – т. 13, No 9. – с. 1842.
27. Gitis V. G., Derendyaev A. B. Machine learning methods for seismic hazards forecast // Geosciences. – 2019. – т. 9, No 7. – с. 308.2
28. Bishop C. M. Pattern recognition // Machine learning. – 2006. – т. 128, No 9.
29. Supervised machine learning: A review of classification techniques / S. B. Kotsiantis, I. Zaharakis, P. Pintelas [и др.] // Emerging artificial intelligence applications in computer engineering. – 2007. – т. 160, No 1. – с. 3–24.
30. Khan S. S., Madden M. G. A survey of recent trends in one class classification // Irish conference on artificial intelligence and cognitive science. – Springer. 2009. – с. 188–197.
31. Чебров В., Салтыков В., Серафимова Ю. Опыт выявления предвестников сильных (м 6.0) землетрясений на Камчатке в 1998–2011 гг. По материалам КФ РЭС // Вулканология и сейсмология. – 2013. – No 1. – с. 85–85.
32. Barnhart W. D., Hayes G. P., Wald D. J. Global earthquake response with imaging geodesy: Recent examples from the USGS NEIC // Remote Sensing. – 2019. – т. 11, No 11. – с. 1357.
33. Blewitt G., Hammond W. C., Kreemer C. Harnessing the GPS data explosion for interdisciplinary science // Eos. – 2018. – т. 99, No 10. – с. 1029. – с. 485.
34. Кремер Н. Теория вероятностей и математическая статистика // М.: ЮНИТИ: ЮНИТИ-Дана. – 2004. – с. 573.

An approach to systematic earthquake prediction

Valeri Gitis , Alexander Derendyaev and Konstantin Petrov

The systematic forecast of earthquakes is regularly performed with a step of Δt in a predetermined analysis zone. At each step, the training set of target earthquakes is augmented, new data about the seismic process is loaded, the data is processed, transformed into grid-based fields, and the machine learning program calculates a solution using all available data from the beginning of the training to the moment t^* of the earthquake forecast. The result is a map of the alarm zone in which the epicenter of the target earthquake is expected in the interval $(t^*, t^* + \Delta t]$. At the next step, the learning interval is increased by Δt . The training is based on the method of minimum area of alarm. The mathematical model of the method makes it possible to distinguish in the grid fields the outliers that precede target earthquakes, which may be precursors of earthquakes, to compare these outliers with each other and then use similar outliers in the forecast. The model introduces restrictions on the class of prediction decision rules. This helps to compensate to some extent for the small number of strong earthquakes used for training.

In our previous works, the forecast quality was assessed by the proportion of detecting target events at a given average value of the alarm zone. Here we consider a generalization of the method of minimum area of alarm, designed to improve another characteristic of the forecast quality: the probability of occurrence of at least one target event in the expected alarm zone. The difference between the methods is that at the moment of forecasting t^* two decisions are made: forecast in time and forecast in space. The first solution determines the possibility of an earthquake epicenter with a target magnitude in the forecast interval. If it is decided that the target event is possible, then a map with an alarm zone is calculated. A target earthquake is predicted if its epicenter falls into the calculated alarm zone.

Predictive modeling is carried out for the regions of Kamchatka and California. The initial data for Kamchatka is a catalog of earthquakes. For the California region, earthquake catalog data and GPS time series are used. The results showed rather high estimates of the probability of detecting target events and very small values of estimates of the probability of the appearance of the predicted epicenter of a target event in the alarm zone. At the same time, the estimates of this probability for a systematic forecast are much higher than the similar ones for random fields.

KEYWORDS: earthquake forecast, machine learning, method of minimum area of alarm, forecast quality, spatio-temporal processes.