

МЕТОД ФЛИККЕР-ШУМОВОЙ СПЕКТРОСКОПИИ В ЗАДАЧЕ ОЦЕНКИ СЕЙСМИЧЕСКОЙ ОПАСНОСТИ ПО ДАННЫМ ГИДРОГЕОХИМИЧЕСКИХ НАБЛЮДЕНИЙ. АНАЛИЗ ПРОГНОСТИЧЕСКОЙ ЭФФЕКТИВНОСТИ МЕТОДА

Рябинин Г.В.

*Камчатский филиал Учреждения Российской академии наук Геофизической службы РАН,
г. Петропавловск-Камчатский, gena@emsd.iks.ru*

Введение

При анализе данных наблюдений нередко возникает необходимость оценить насколько тесно связаны между собой отдельные явления изучаемых процессов и существует ли возможность прогнозировать возникновение одного явления по факту наблюдения другого. Такая задача может иметь место, в частности, на начальной стадии исследований, когда о наблюдаемых процессах мало что известно, но имеются основания предполагать, что отдельные экстремальные значения рассматриваемых процессов могут быть связаны во времени. Одной из задач подобного рода может являться задача анализа тесноты связи между отдельными сильными землетрясениями и предшествующими им аномальными эффектами в вариациях самых различных показателей. В представленной работе такая задача рассматривается применительно к данным гидрогеохимических наблюдений, проводимых на Камчатке, а точнее, к результатам их обработки методом фликкер-шумовой спектроскопии.

Сущность метода фликкер-шумовой спектроскопии

Метод фликкер-шумовой спектроскопии (ФШС) был предложен С.Ф. Тимашевым (Научно-исследовательский физико-химический институт им. Л.Я. Карпова, г. Москва) как общий феноменологический подход к анализу хаотических сигналов разной сущности [6]. Появление в сложной динамической системе фликкер-шума связывается с эффектом перемежаемости (интермиттанса), который представляет собой смену относительно длительных участков ламинарного или регулярного во времени поведения характеристик неравновесной динамической системы, – участками хаотических всплесков и скачков. Сущность ФШС подхода заключается в придании информационной значимости корреляционным взаимосвязям, которые реализуются в последовательностях нерегулярностей сигнала – всплесках, скачках, изломах производных различных порядков – как носителях информации об изменениях, происходящих на каждом пространственно-временном уровне иерархической организации исследуемой динамической системы [6]. В качестве базового образа для извлечения информации из сложных сигналов в ФШС методе используется корреляционная функция. Для классификации информации, заключенной в корреляционной функции, предлагается анализировать не саму эту функцию, а некоторые ее преобразования («проекции»), такие как спектр мощности $S(f)$ (f – частота) и разностный момент («переходная структурная функция») $\Phi^{(2)}(\tau)$ второго порядка. Информация, извлекаемая из анализа зависимостей $S(f)$ и $\Phi^{(2)}(\tau)$, построенных на основе временных рядов $V(t)$, имеет смысл времен корреляций или параметров, характеризующих потерю «памяти» (корреляционных связей) для рассматриваемых нерегулярностей типа всплесков и скачков. При этом в формирование зависимостей $\Phi^{(2)}(\tau)$ вносят вклад только нерегулярности типа скачков динамической переменной $V(t)$, а в формирование $S(f)$ – и скачки, и всплески (выбросы) хаотических серий $V(t)$ [6]. Одна из задач ФШС метода заключается в анализе зависимостей $S(f)$ и $\Phi^{(2)}(\tau)$, а точнее, в анализе динамики их изменения в пределах скользящего интервала длиной T , последовательно смещаемого в каждый дискретный момент времени $k\Delta T$ ($k = 0, 1, 2, 3, \dots$) на величину ΔT . В качестве меры изменения зависимостей $S(f)$ и $\Phi^{(2)}(\tau)$ во времени ФШС - процедура предполагает использование безразмерного критерия – параметра нестационарности, выражение для которого задается соотношением вида:

$$C(t_{k+1}) = 2 \cdot \frac{Q_{k+1} - Q_k}{Q_{k+1} + Q_k} \cdot \frac{\Delta T}{T}, \quad (1)$$

$$Q_k = \int_0^{\tau_{\max}} [\Phi^{(2)}(\tau)]_k d\tau \quad \text{или} \quad Q_k = \int_0^{f_{\max}} [S(f)]_k df \cdot (2)$$

Как следует из самого названия, вводимое соотношение характеризуют «меру нестационарности» анализируемого временного ряда при перемещении интервала усреднения T по оси времени на величину ΔT . Очевидно, что, теоретически, для стационарных процессов при неизменных значениях τ_{max} или f_{max} величина параметра нестационарности будет равняться нулю на любом интервале осреднения. Фактически же его величина должна варьировать от низких значений на интервалах, где наблюдается регулярные вариации, до весьма значительных там, где обнаруживаются нерегулярности типа всплесков, скачков и изломов производных. При этом возможность идентификации нерегулярностей различного масштаба связана с самой возможностью вариации величиной интервала усреднения T . Моменты времени значимых всплесков параметра нестационарности наряду с моментами времени сильных сейсмических событий являлись объектом исследований данной работы. В качестве главной задачи работы рассматривалась оценка возможности прогноза сильных землетрясений (ретроспективная оценка прогностической эффективности) по вариациям параметра нестационарности, рассчитанного по данным гидрогеохимических наблюдений.

Фактическая основа работы

В работе использовались данные гидрогеохимических наблюдений по 6 самоизливающимся скважинам (ГК-1, №1, ГК-15, №44, ГК-5, Г-1) и данные каталога землетрясений Камчатского филиала Геофизической службы РАН за период с 1985 (1993) по 2008 г. Данные гидрогеохимических наблюдений были представлены временными рядами вариаций различных показателей химического и газового состава подземных вод: pH, HCO_3^- , H_3BO_3 , H_4SiO_4 , Cl, Ca^{2+} , SO_4^{2-} , Na^+ , He, H_2 , O_2 , N_2 , CO_2 , CH_4 , C_2H_4 , C_3H_6 , $n\text{C}_4\text{H}_{10}$, $i\text{C}_4\text{H}_{10}$. Фактическая дискретность временных рядов гидрогеохимических показателей неравномерна и составляет, для разных пунктов наблюдения, от 3 до 6 суток. Процедура предварительной обработки данных гидрогеохимических наблюдений включала в себя удаление одиночных выбросов и приведение временных рядов к равномерной дискретности равной одним суткам. Далее по предварительно обработанным временным рядам гидрогеохимических показателей рассчитывались значения параметра нестационарности. Для каждого временного ряда параметр нестационарности рассчитывался в скользящих окнах размерами в 100, 200, 400 и 800 суток. Таким образом, получалось, что для каждого временного ряда того или иного гидрогеохимического показателя рассчитывалось по четыре временных ряда вариаций параметра нестационарности.

Временной ряд последовательности землетрясений формировался на основе каталога КФ ГС РАН путем выбора из него сейсмических событий с энергетическим классом $K_s \geq 10$ и эпицентральным расстоянием $\Delta \leq 500$ км от г. Петропавловска-Камчатского. После этого для каждого из выбранных землетрясений рассчитывалась величина параметра Добровольского (ϵ) [2], представляющая собой интегральную характеристику энергии землетрясения и его эпицентрального расстояния. В тех случаях, когда промежуток времени между моментами сейсмических событий оказывался меньше приведенной дискретности временных рядов гидрогеохимических показателей (меньше суток), из последовательности землетрясений выбиралось событие, имеющее наибольшую величину параметра ϵ .

Алгоритм оценки прогностической эффективности

В основу алгоритма оценки прогностической эффективности положен анализ таблиц сопряженности двух дихотомических признаков. Подобный подход использовался, к примеру, в работе [1] для оценки эффективности прогноза по статистике сейсмичности и в работе [4] для оценки эффективности прогноза землетрясений по данным наблюдений за вариациями уровня воды в скважине Е1. В представленной работе в качестве двух дихотомических признаков рассматривались: 1 – наличие или отсутствие в определенном временном интервале значимых вариаций параметра нестационарности, 2 – наличие или отсутствие в том же интервале сильных сейсмических событий. Поскольку интерес представляет оценка возможности прогноза сильных землетрясений по вариациям параметра нестационарности, в алгоритм вводилось дополнительное условие, определяющее, что значимые всплески параметра нестационарности внутри временного интервала должны находиться перед сейсмическими событиями.

Анализ таблиц сопряженности относится к классу разведочного анализа данных. В связи с этим, задачей анализа было сделать грубую оценку того, на сколько вариации параметра нестационарности связаны во времени с наиболее значительными сейсмическими событиями. При этом в качестве начальных условий принималась ситуация полной априорной неопределенности. Имеется в виду, что априори не известны ни пороговые значения для вариаций параметра нестационарности, ни пороговые значения для пространственно – энергетических характеристик

сейсмических событий, ни временной интервал, в пределах которого возможно наличие статистической связи между рассматриваемыми признаками. При таких начальных условиях задача становится многопараметрической, а анализ приобретает вид достаточно рутинной, оптимизационной процедуры. Для некоторого упрощения задачи и алгоритма анализа имело смысл использовать только несколько значений пороговых уровней для вариаций параметра нестационарности и величин параметра Добровольского, а связь между ними искать на ограниченном множестве величин временных интервалов. Для иллюстрации работы алгоритма введем следующие обозначения. $C_k(T)$ – временные ряды параметра нестационарности, рассчитанные для интервалов усреднения $T = 100, 200, 400$ и 800 суток, $k = 1, 2, 3, 4$. TL_i^C – пороговые уровни для параметра нестационарности. $TL^C = 1, 3, 5$, $i = 1, 2, 3$. TL_j^ε – пороговые уровни для величин параметра Добровольского. $TL^\varepsilon = 1 \cdot 10^{-8}, 5 \cdot 10^{-8}, 1 \cdot 10^{-7}$, $j = 1, 2, 3$. W_l – последовательность величин неперекрывающихся временных интервалов (времена тревоги), которые тестировались на наличие или отсутствие в них значимых вариаций параметра нестационарности $C_k(T) \geq TL_i^C$ и величин параметра Добровольского $\varepsilon \geq TL_j^\varepsilon$. $W_l = 10, 20, 30, \dots, 500$ суток, $l = 1, 2, 3, \dots, 50$. Результаты теста неперекрывающихся интервалов W_l сводились в таблицы сопряженности (СТ) двух дихотомических признаков, которые в самом общем виде представляют собой двухмерные таблицы вида

Признак 1

		есть	нет	
<i>Признак 2</i>	есть	a	b	$a+b$
	нет	c	d	$c+d$
		$a+c$	$b+d$	n

где n – общее количество неперекрывающихся интервалов.

В данном случае, для совокупности таблиц сопряженности можно ввести следующее обозначение $СТ_{l,k,i,j}(W_l, C_k(T) \geq TL_i^C, \varepsilon \geq TL_j^\varepsilon)$. К примеру, обозначение $СТ_{5,1,1,1}(W_5, C_1(100) \geq TL_1^C, \varepsilon \geq TL_1^\varepsilon)$ следует рассматривать как таблицу сопряженности, сконструированную по последовательности неперекрывающихся временных интервалов W размером в 50 суток, которые тестировались на наличие или отсутствие в их пределах вариаций параметра нестационарности C (рассчитанного в окне осреднения 100 суток), превышающих пороговый уровень $TL^C = 1$, и величин параметра Добровольского ε , превышающих пороговый уровень $TL^\varepsilon = 1 \cdot 10^{-8}$. Нетрудно подсчитать, что для каждого временного ряда из анализируемых гидрогеохимических показателей (с учетом того, что для каждого из них рассчитывалось по 4 временных ряда параметра нестационарности) конструировалось в общей сложности 1800 таблиц сопряженности.

Следующий этап алгоритма заключался в оценке статистик, рассчитываемых по таблицам сопряженности и характеризующих тесноту связи между дихотомическими признаками. Поскольку наибольший интерес представляла оценка возможности прогноза сильных землетрясений по вариациям параметра нестационарности, в качестве основной статистики использовался коэффициент эффективности прогноза, представленный в работе [1], выражение для которого, в обозначениях вышеприведенной таблицы, задается соотношением вида

$$J = n \frac{a}{(a+b)(a+c)} \quad (3).$$

Значение статистики $J = 1$, соответствует ситуации "случайного угадывания". Кроме коэффициента эффективности прогноза по таблицам сопряженности рассчитывался коэффициент Φ , представляющий собой, для таблиц сопряженности, аналог классического коэффициента корреляции. Выражение для коэффициента Φ задается соотношением вида [5]

$$\Phi = \frac{ad - bc}{\sqrt{(a+c)(b+d)(a+b)(c+d)}} \quad (4).$$

Для каждого значения статистик $J_{l,k,i,j}$ и $\Phi_{l,k,i,j}$, рассчитанных по таблицам сопряженности $СТ_{l,k,i,j}$, вычислялись значения их дисперсий $D_{l,k,i,j}^J$ и $D_{l,k,i,j}^\Phi$, выражения для которых даны в работах [1] и [3] соответственно. Дальнейший анализ сводился к поиску, для каждого из временных рядов гидрогеохимических показателей, значений статистик эффективности прогноза $J_{l,k,i,j} > 2$. Величина статистики $J > 2$ означает, что при определенных фиксированных значениях параметров (W, T, TL^C и TL^ε) по вариациям параметра нестационарности возможен прогноз в два раза большего количества

сильных землетрясений, чем это можно было бы сделать методом "случайного угадывания" при $J = 1$. Логично предположить, что при различном сочетании величин параметров (W , TL^C и TL^ε) существует вероятность, что условию $J > 2$ могут соответствовать сразу несколько оценок эффективности прогноза, полученных для вариаций параметра нестационарности $C_k(T)$. В таких случаях за окончательный принимался результат (сочетание параметров W , TL^C и TL^ε) с величиной статистики $J > 2$, имеющий минимальную дисперсию D^J и максимальное значение коэффициента Φ . Поскольку для каждого временного ряда показателей химического и газового состава рассчитывалось по четыре временных ряда вариаций параметра нестационарности $C_k(T)$ ($T = 100, 200, 400$ и 800 суток), количество оценок эффективности прогноза $J > 2$ могло варьировать от 0 до 4.

На последнем этапе алгоритма осуществлялась проверка гипотезы о том, что полученные оценки эффективности прогноза являются следствием случайного совпадения, т.е. случайного взаимного расположения на временной оси экстремальных вариаций параметра нестационарности ($C_k(T) \geq TL_i^C$) и моментов сильных землетрясений ($\varepsilon \geq TL_j^\varepsilon$). Для этого использовалась итерационная процедура, суть которой заключалась в следующем. На каждом итерационном шаге $h = 1, 2, 3, \dots, 1000$ последовательность значений анализируемого ряда параметра нестационарности перетасовывалась на временной оси в случайном порядке. После этого для каждого из полученных модельных рядов параметра нестационарности и ряда последовательности величин параметра Добровольского при заданных значениях параметров (W , TL^C и TL^ε) рассчитывалась статистика эффективности прогноза J_m . Проверка гипотезы заключалась в оценке значимости отклонения величины статистики J , полученной для исходных данных, от среднего значения $\mu(J_m)$ совокупности величин статистик J_m , полученных для модельных рядов параметра нестационарности при условии, что эта совокупность имеет нормальное распределение. Другими словами, оценивалась вероятность $P(|J_m - \mu(J_m)| \geq J)$. Результаты анализа показали, что значения вероятностей $P(|J_m - \mu(J_m)| \geq J)$ для всей совокупности оценок эффективности прогноза ($J > 2$) оказались в среднем не выше 1%. Столь низкие значения вероятностей указывают на то, что рассматриваемая гипотеза должна быть отвергнута и полученные оценки эффективности прогноза не являются "игрой случая".

Обсуждение результатов

На рисунке 1 схематически представлены результаты ретроспективной оценки эффективности прогноза землетрясений по временным рядам параметра нестационарности $C(T)$ ($T = 100, 200, 400$ и 800 суток), которые рассчитывались для различных компонентов химического и газового состава воды самоизливающихся скважин.

Компонент	Скважина					
	№1	Г-1	ГК-1	ГК-15	ГК-44	ГК-5
pH						
HCO3						
H3BO3						
H4SiO4						
Cl						
Ca						
SO4						
Na						
He						
H2						
O2						
N2						
CO2						
CH4						
C2H4						
C3H6						
C4H10n						
C4H10i						

Рис. 1. Схематическое представление результатов оценки эффективности прогноза. Пояснения в тексте.

Поля таблицы, закрашенные серым цветом, обозначают, что для данного компонента в данном пункте наблюдения, значение коэффициента эффективности прогноза (J), оцененного по вариациям параметра нестационарности, получилось больше двух. Поля таблицы, закрашенные белым цветом, обозначают, соответственно, что в данном случае $J < 2$. Оценки эффективности прогноза с $J > 2$ получены для различных сочетаний параметров $C(T)$, W , TL^C и TL^ε , поскольку интерес представляла сама возможность использования экстремальных вариаций параметра

нестационарности в качестве прогнозных признаков сильных землетрясений, а также, не в последнюю очередь, пригодность описанного выше алгоритма для решения этой задачи. С этой точки зрения сознательное завышение диапазона изменения входных параметров представляется вполне оправданным. С другой стороны, вариабельность значений параметров $C(T)$, W , TL^C и TL^ε затрудняет интерпретацию результатов анализа. В частности, по полученным результатам сложно сделать точную оценку оптимального сочетания величин входных параметров $C(T)$, W , TL^C и TL^ε . Тем не менее, можно проанализировать, как зависит количество оценок эффективности прогноза с $J > 2$ от диапазона их изменения (рис. 2).

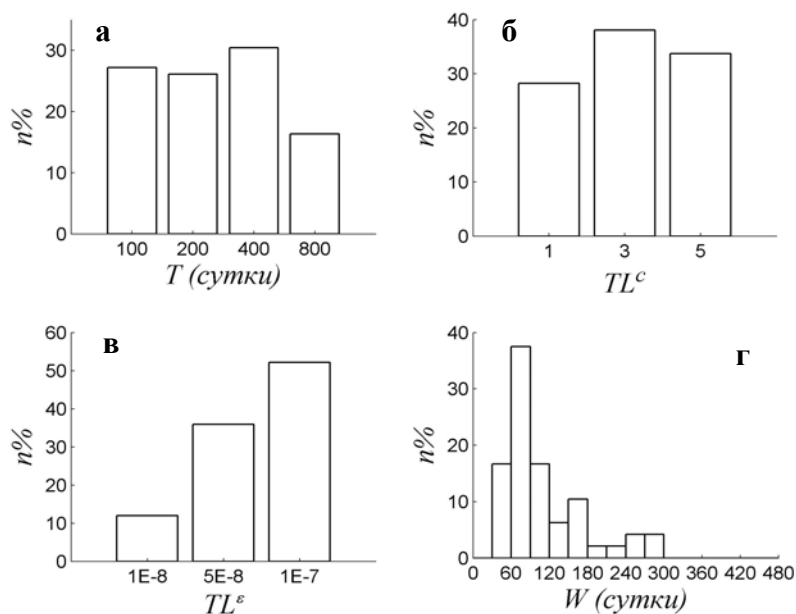


Рис. 2. Диаграммы распределения количества оценок $n\%$ (в процентах от общего количества) показателя эффективности прогноза $J > 2$ в зависимости от величин параметров: а – интервала усреднения T , б – порогового уровня для вариаций параметра нестационарности TL^C , в – порогового уровня для величин параметра Добровольского TL^ε , г – "времени тревоги" W . На последней диаграмме значения $n\%$ посчитаны только для значений параметра Добровольского $\varepsilon \geq 1 \cdot 10^{-7}$.

Из рисунка видно, что количество оценок эффективности прогноза ($J > 2$) слабо зависит от параметров T и TL^C и распределение $n\%$ для этих случаев практически равномерно. С другой стороны (диаграмма 2в), видно, что более 50% оценок эффективности прогноза получены для значений параметра $\varepsilon \geq 1 \cdot 10^{-7}$. Из этого можно сделать вывод о том, что по экстремальным вариациям параметра нестационарности, эффективный прогноз возможен только для наиболее сильных сейсмических событий. Из рисунка 2г следует, что наиболее вероятный интервал времени между экстремальным значением параметра нестационарности и сильным сейсмическим событием (время тревоги) составляет от 60 до 90 суток.

Вывод

Использование значимых вариаций параметра нестационарности в качестве прогнозных признаков, для ряда показателей химического и газового состава подземных вод, дает возможность прогноза в два раза большего количества сильных сейсмических событий, чем это можно было бы сделать "методом случайного угадывания", с упреждением до двух – трех месяцев.

Список литературы

1. Гусев А.А. Прогноз землетрясений по статистике сейсмичности // Сейсмичность и сейсмический прогноз, свойства верхней мантии и их связь с вулканизмом на Камчатке. Новосибирск: Наука, 1974. С. 109-119.
2. Добровольский И.П. Механика подготовки тектонического землетрясения. М.: ИФЗ, 1984. 188 с.
3. Кендалл М., Стюарт А. Статистические выводы и связи. М.: Наука, 1973. 899 с.
4. Копылова Г.Н. Изменения уровня воды в скважине Елизовская-1, Камчатка, вызванные сильными землетрясениями (по данным наблюдений в 1987 – 1998 гг.) // Вулканология и сейсмология. 2001. №2. С. 39-52.
5. Миркин Б.Г. Анализ качественных признаков и структур. М.: Статистика, 1980. 320 с.
6. Тимашев С.Ф. Фликкер-шумовая спектроскопия. Информация в хаотических сигналах. М.: Физматлит, 2007. 248 с.