

УДК 550.34

АНАЛИЗ ДЛИТЕЛЬНЫХ НАБЛЮДЕНИЙ ЗА УРОВНЕМ ПОДЗЕМНЫХ ВОД В АСЕЙСМИЧЕСКОМ РЕГИОНЕ

© 2019 г. А. А. Любушин^{1*}, О. С. Казанцева^{1**}, А. Б. Манукин^{1***}

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

*E-mail: lyubushin@yandex.ru

**E-mail: o.kazantseva@yandex.ru

***E-mail: amanukin@yandex.ru

Поступила в редакцию 27.02.2018 г.

Принята в печать 19.09.2018 г.

В статье приводятся результаты анализа непрерывных прецизионных временных рядов наблюдений за атмосферным давлением и уровнем подземных вод в скважине, пробуренной на глубину 400 м и расположенной на территории г. Москвы. Эти наблюдения примечательны своей длительностью — более 22 лет (с 02.02.1993 по 04.04.2015 гг.) и временным интервалом опроса 10 минут. Такая большая длительность наблюдений позволяет исследовать вопрос о том, насколько стационарны свойства гидрогеологических временных рядов в сейсмически спокойном регионе, что представляет методический интерес при интерпретации этих наблюдений, проводимых в сейсмически активных регионах для целей прогноза землетрясений. В результате применения факторного и кластерного анализа к последовательности многомерных векторов статистических свойств ряда наблюдений за уровнем подземных вод в последовательных временных окнах длиной 10 суток, после адаптивной компенсации влияния атмосферного давления, было выделено 5 различных статистически значимых состояний временного ряда, между которыми происходят переходы. Делается попытка геофизической интерпретации выделенных состояний. Спектральный анализ последовательности моментов времен переходов между кластерами выделил 2 значимых периода 46 и 275 суток.

Ключевые слова: уровень подземных вод, временные ряды, передаточные функции, энтропия, мультифракталы, индекс предсказуемости, факторный анализ, кластерный анализ

DOI: <https://doi.org/10.31857/S0002-33372019247-67>

ВВЕДЕНИЕ

Изменения уровня подземных вод несут информацию об объемных деформациях в земной коре. Если скважина пробурена во флюидонасыщенный горизонт достаточной мощности, то вариации уровня будут характеризовать объемные деформации, определяющие региональные процессы в земной коре. Анализ данных о вариациях уровня воды в скважине позволяет исследовать лунно-солнечные приливные процессы в твердой земле [Bredenhoft, 1967; Rojstaczer, Agnew, 1989; Багмет и др., 1989; Любушин и др., 1997; Виноградов и др., 2011], а также получать данные об изменчивости свойств земной коры в данном районе по отклику изменения уровня на изменения атмосферного давления [Любушин, Малугин, 1993; Любушин, Лежнев, 1995; Копылова и др., 2009]. Долговременные

наблюдения в сейсмически активном районе могут дать информацию о предвестниках землетрясений [Roeloffs, 1988; Roeloffs et al., 1989; Igarashi, Wakita, 1991; Копылова и др., 2000; Копылова, 2001; 2006; Болдина, Копылова, 2017], а информация, полученная в асейсмических, платформенных областях, необходима для выявления путем сравнительного анализа развития аномальных процессов в сейсмоактивных районах [Киссин, Гумен, 1994].

Параллельные измерения на нескольких скважинах, пробуренных в разные флюидонасыщенные горизонты, могут дать информацию о гидродинамических процессах в данном регионе [Любушин и др., 1999; Беседина и др., 2015]. Наконец, анализ результатов измерений поможет выявить гидродинамические режимы, обусловленные техногенными воздействиями: отбор воды из данного или другого водонасыщенного

горизонта, закачка жидкости, изменение нагрузки на почву в связи с проведением строительных работ и т. д. Ценность информации о геофизических и гидродинамических процессах возрастает при условии проведения многолетних непрерывных наблюдений с высокой чувствительностью и относительно высокой частотой съема данных. При этих условиях удается выявить тонкие эффекты в широком диапазоне периодов.

Аппаратура, входящая в состав уровнемерно-го комплекса для параллельного измерения вариаций уровня воды в скважине и атмосферного давления, характеризующаяся высоким разрешением и большим динамическим диапазоном, достаточной долговременной стабильностью, надежностью и автономностью, позволила решить проблему организации и проведения многолетних непрерывных наблюдений.

ИСХОДНЫЕ ДАННЫЕ

Непрерывные измерения вариаций уровня подземных вод проводились с 02.02.1993 по 04.04.2015 гг. в скважинах, расположенных на северо-западе Москвы на территории Центрального института травматологии и ортопедии (ЦИТО), координаты (55.8202° с.ш., 37.5292° в.д.). Пробуренная на глубину 404 м скважина вскрывает верхнефаменский водоносный горизонт [Гидрогеология..., 1966], сложенный трещиноватыми доломитами и известняками. Общая мощность обводненных пород — свыше 200 м.

Скважина была пробурена в 1969–1970 гг. и предназначалась для бальнеологического лечения в ЦИТО, однако с начала 1980-х годов она по своему назначению не использовалась и была законсервирована. Был проведен химический анализ и определена минерализация воды, которая составляет 5 г/л. Специальный пункт наблюдений был оборудован в отдельном помещении площадью около 10 м^2 , расположенном рядом с устьем скважины.

Одновременно с измерениями вариаций уровня подземных вод проводились измерения вариаций атмосферного давления. Датчики уровня и атмосферного давления построены по манометрическому типу с емкостным частотным преобразователем перемещения чувствительного элемента — мембраны в выходной сигнал. Чувствительность датчиков: для уровня — 0.1 мм водяного столба, для давления — 10 мкбар. Датчики уровня и атмосферного давления установлены в скважине на тросах. Высота столба

воды над датчиком уровня составляет 2.5–3.0 м. Датчик атмосферного давления опущен на глубину около 10 м. Регистрация данных производится с дискретностью отсчетов 10 мин в цифровой форме с помощью специального интерфейсного устройства на блоки твердотельной памяти. На рис. 1 представлены графики временных рядов исходных данных. Полная длина записей равна $1'165'862$ 10-минутных отсчетов. В записях имеются перерывы, связанные со сбоями системы регистрации.

ПЕРЕДАТОЧНЫЕ ФУНКЦИИ И СПЕКТРЫ МОЩНОСТИ

Как показывают оценки спектров когерентности, основным внешним метеорологическим фактором, влияющим на вариации уровня подземных вод, является изменение атмосферного давления. Наличие длительных рядов наблюдений (22 года) позволяет статистически значимо оценить амплитудные частотные передаточные функции от атмосферного давления к вариациям уровней для очень широкого диапазона периодов — от 20 мин (частоты Найквиста для данных временных рядов) до 1 года. Сама возможность получения такой оценки уникальна и обеспечивается сочетанием благоприятствующих факторов, перечисленных во Введении.

Оценки передаточных функций строились классическим непараметрическим методом, основанным на преобразовании Фурье «входных» (атмосферное давление) и «выходных» (вариации уровня) сигналов, вычислении их взаимной спектральной функции и спектра мощности «входа» (с помощью усреднения периодограмм и кросс-периодограммы) и отношения кросс-спектра к спектру мощности [Brillinger, 1975]. Большой диапазон периодов создавал технические трудности для получения оценки сразу по всей выборке. Поэтому оценки строились сшиванием трех различных оценок. Для периодов от 20 мин до 10 ч оценка строилась путем усреднения для каждого значения частоты кросс-периодограмм исходных 10-минутных временных рядов, полученных в скользящих временных окнах длиной 2048 отсчетов (с предварительным удалением локальных линейных трендов в каждом окне и последующим переходом к приращениям) и завершающим усреднением по частоте скользящим частотным окном радиуса 20 значений частоты ($9.7 \cdot 10^{-4} \text{ мин}^{-1}$). Для периодов от 10 до 100 ч оценка строилась аналогично, но для временных рядов после перехода к 1-часовым отсчетам (путем усреднения

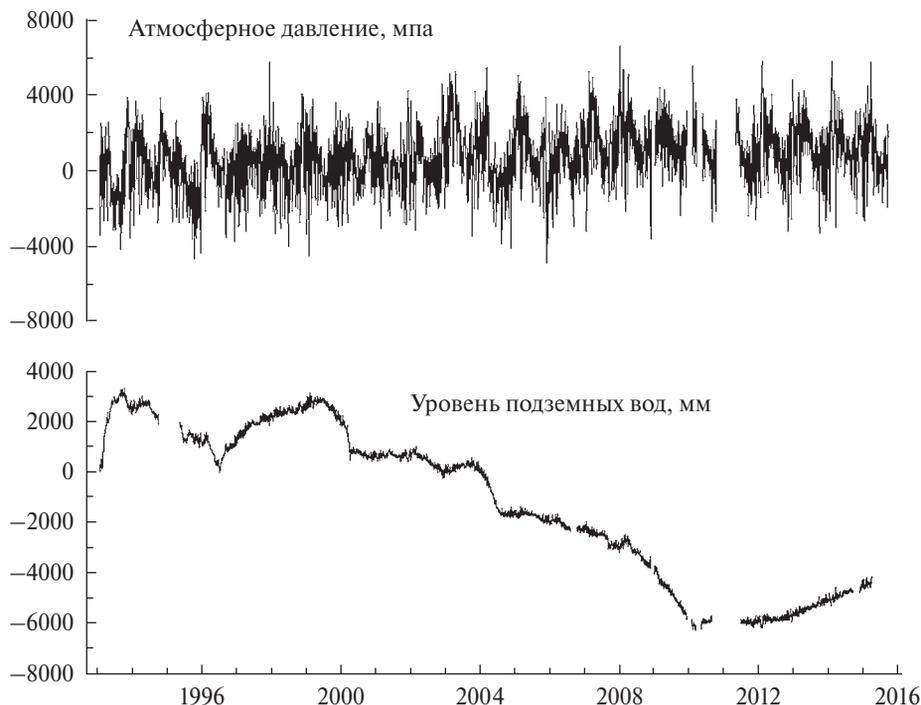


Рис. 1. Графики исходных данных: вариаций атмосферного давления и уровня подземных вод, 02.02.1993–04.04.2015 гг., шаг по времени 10 минут.

и прореживания в 6 раз исходных рядов), при этом радиус окна частотного усреднения стал равен $9.7 \cdot 10^{-3} \text{ ч}^{-1}$. Наконец, для оценки для периодов от 100 до 10000 часов осуществлялся предварительный переход (после усреднения и прореживания исходных данных в 144 раза) к 1-суточным отсчетам, оценка строилась аналогично, но при этом длина скользящего временного окна бралась 1024 отсчета, а радиус частотного усреднения – 10 значений частоты ($9.7 \cdot 10^{-3} \text{ сут}^{-1}$). Во всех случаях смещение временных окон бралось равным 1/8 от длины окна.

Учет пропущенных значений производился следующим образом. Сначала все пропуски восполнялись искусственными данными, которые строились по поведению настоящих данных слева и справа от интервала пропусков на временных фрагментах той же длины, что и длина пропуска. Далее, поскольку оценки производились путем усреднения периодограмм от скользящих временных окон, то в обработку принимались лишь те окна, в которых доля искусственных (восполненных) значений не превышала 2% от длины временного окна.

На рис. 2 представлены графики оценок квадрата модуля спектра когерентности (частотно-зависимого квадрата коэффициента корреляции) и амплитудной частотной передаточной функции.

Следует иметь в виду, что дисперсия непараметрической оценки передаточной функции (асимптотическое выражение для достаточно большого числа отсчетов) имеет вид [Brillinger, 1975] $c \cdot (1 - \gamma_{xy}^2(\omega)) \cdot (S_{yy}(\omega) / S_{xx}(\omega))$. Здесь: ω – частота; $S_{yy}(\omega)$ – спектр мощности выходного сигнала (уровня); $S_{xx}(\omega)$ – спектр мощности входного сигнала (атмосферное давление); $\gamma_{xy}^2(\omega)$ – их квадрат модуля спектра когерентности; c – постоянная, зависящая от способа усреднения периодограмм, в частности, $c \sim (k(2m+1))^{-1}$, где m – радиус частотного усреднения периодограмм в числе значений частоты (20 либо 10 в нашем случае), а k – число независимых временных интервалов, по которым проводилось усреднение. Что касается дисперсии оценки квадрата модуля спектра когерентности, то она равна $2c \cdot (1 - \gamma_{xy}^2(\omega))^2 \cdot \gamma_{xy}^2(\omega)$. Поэтому для тех периодов, где оценка $\gamma_{xy}^2(\omega)$ мала, следует ожидать широкого доверительного интервала для оценки передаточной функции. Однако использование очень глубокого усреднения периодограмм как по временным интервалам, так и по частоте (большие значения k и m), делает полученную оценку достаточно устойчивой к шуму.

Атмосферное давление может быть использовано как зондирующий сигнал. По изменению частотной функции барометрического отклика

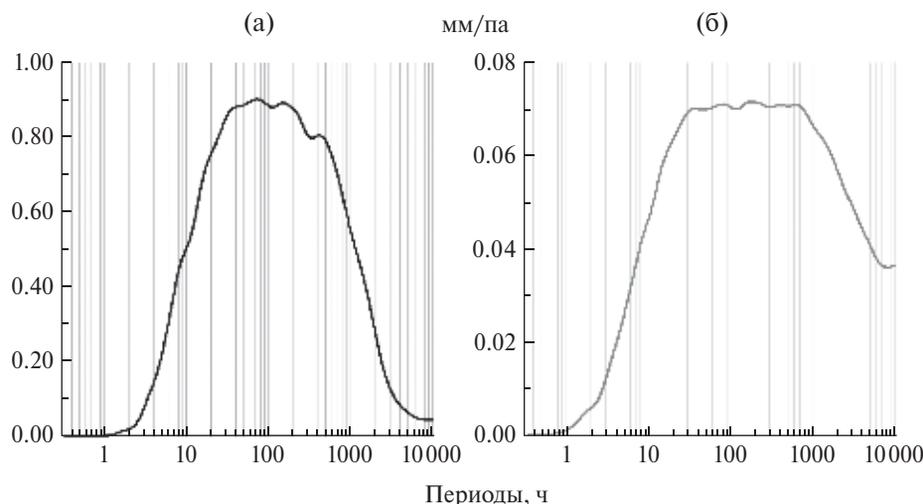


Рис. 2. Графики оценок квадратичного спектра когерентности (а) и амплитудной частотной передаточной функции (б).

уровня подземных вод можно выделять временные интервалы и частотные полосы аномального поведения [Любушин, Малугин, 1993; Любушин, Лежнев, 1995; Копылова и др., 2000]. Ниже оценки изменчивости частотной функции отклика были проведены после перехода к 1-часовым отсчетам путем вычисления средних значений по 6 смежным 10-минутным значениям. Было взято скользящее временное окно длиной 672 ч (28 суток, или лунный месяц, — это естественный масштаб времени в анализе геофизических низкочастотных фоновых процессов в земной коре). Для каждого положения окна были проведены оценки частотной передаточной функции от вариаций атмосферного давления с использованием параметрической регрессионной модели:

$$x(t) = \sum_{j=0}^p b_j \cdot u(t-j) + d + e(t). \quad (1)$$

Здесь: $x(t)$ — приращения временного ряда уровня подземных вод; $u(t)$ — приращения атмосферного давления; p — порядок регрессии; $(b_j, j = 0, \dots, p, d)$ — параметры модели, которые определялись в каждом временном окне из условия минимума суммы квадратов остатков идентификации $\sum e^2(t) \rightarrow \min$. Взаимное смещение окон бралось равным 14 суток. Кроме того, перед идентификацией параметров модели (1), для увеличения устойчивости оценки к наличию выбросов, проводилась так называемая «винзоризация» выборок [Huber, 1981], означающая итеративную срезку выбросов за уровень ± 3 выборочных стандартных отклонений с последующим пересчетом среднего значения и стандартного отклонения. После определения параметров модели (1) частотная функция

отклика уровня подземных вод на вариации атмосферного давления в зависимости от частоты ω определялась по формуле:

$$H(\omega) = \sum_{j=0}^p b_j \cdot \exp(-i\omega j). \quad (2)$$

В наших расчетах порядок регрессии в модели (1) был выбран $p = 5$ как минимальное значение, после дальнейшего увеличения которого результат оценки функций отклика менялся незначительно.

На рис. 3 представлены результаты оценки амплитудной передаточной функции отклика $A(\omega) = |H(\omega)|$. Рис. 3а представляет график максимального значения $\max_{\omega} A(\omega)$ для каждого временного окна, а рис. 3б — частотно-временную диаграмму амплитудной частотной передаточной функции. Заштрихованные прямоугольники соответствуют временным интервалам, в которых доля искусственных (восполненных) значений временных рядов превышает 2% от длины окна.

Из рис. 3 видно, что в целом функция отклика стационарна, но содержит короткие аномальные временные интервалы, в течение которых максимум частотной амплитудной функции отклика претерпевает резкие изменения как в сторону увеличения, так и уменьшения. Таким образом, даже в стабильной платформенной области, тем не менее, происходит скрытая «тектоническая жизнь». Существование таких короткоживущих аномалий было отмечено в работе [Любушин и др., 1999], где они были названы «медленными событиями» в асейсмическом регионе.

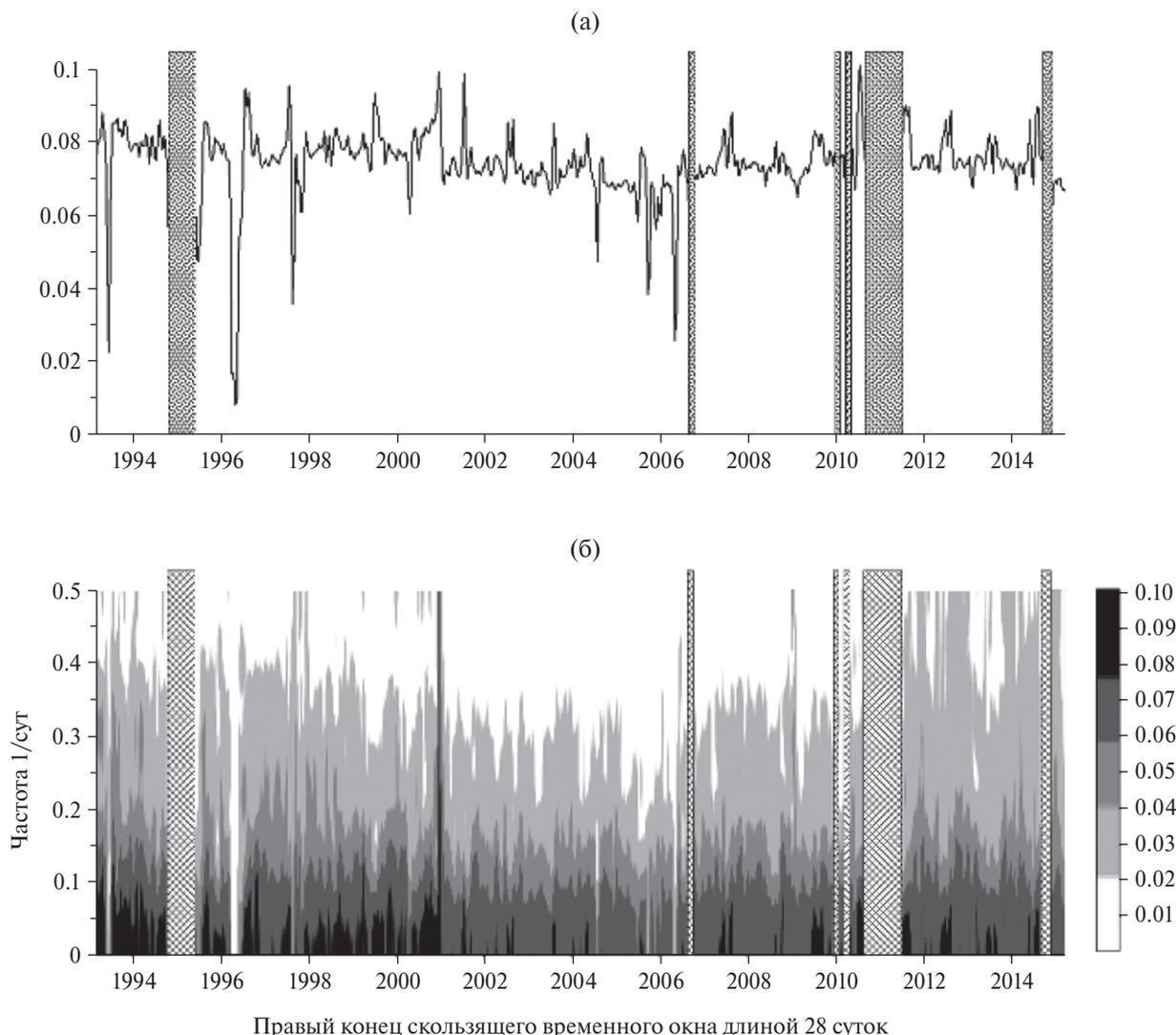


Рис. 3. (а) – График максимального значения $\max_{\omega} A(\omega)$ для каждого временного окна; (б) – частотно-временная диаграмма амплитудной частотной передаточной функции. Заштрихованные прямоугольники соответствуют временным интервалам, в которых доля искусственных (восполненных) значений временных рядов превышает 2% от длины окна. Минимальная частота на диаграмме (б) соответствует максимальному периоду, равному длине окна, то есть 28 суткам.

Для дальнейшего анализа нам необходимо получить временной ряд вариаций уровня подземных вод, «очищенный» от влияния атмосферного давления. Для этой операции был применен компенсирующий адаптивный частотный фильтр [Любушин, 2007]. В скользящих временных окнах длиной 28 суток или 4032 отсчетов с шагом 10 минут, взятых с минимальным смещением 1 отсчет, вычислялись спектр мощности $S_{uu}(\omega)$ давления и комплексный кросс-спектр $S_{xu}(\omega)$ между уровнем подземных вод и давлением. Эти оценки получают путем сглаживания периодограмм и кросс-периодограмм частотным окном длиной 1/32 от длины окна. Далее в каждом окне вычислялась частотная

передаточная функция $H_{xu} = S_{xu}(\omega) / S_{uu}(\omega)$. При этом перед сглаживанием периодограмм подавлялись приливные частотные полосы $[1/11, 1/13]$ и $[1/23, 1/27]$ ч^{-1} , а оценки внутри этих частотных полос получались интерполяцией оценок от соседних значений частот. Результат компенсации $\tilde{E}_x(\omega)$ в частотной области внутри каждого временного окна вычислялся по формуле $\tilde{E}_x(\omega) = \tilde{X}(\omega) - H_{xu}(\omega) \cdot \tilde{U}(\omega)$, где $(\tilde{X}(\omega), \tilde{U}(\omega))$ – дискретные преобразования Фурье от уровня подземных вод и атмосферного давления внутри текущего окна. Результат компенсации $e_x(t)$ во временной области внутри каждого временного окна определяется в результате обратного дискретного преобразования Фурье от $\tilde{E}_x(\omega)$.

Заключительной операцией получения компенсированного сигнала является сшивка внутриоконных результатов компенсации в один сигнал. Для первого окна вкладом в такой единый сигнал является содержимое части временного окна, соответствующее первой половине окна, а для последнего окна — соответствующей второй половине. Что же касается прочих, «промежуточных» временных окон, то из них в единый компенсированный сигнал берется лишь один отсчет, соответствующий центру окна.

На рис. 4 представлены графики исходного и компенсированного уровня подземных вод для 2-месячного временного фрагмента для начала 1999 г. Видно, насколько более четко выделяются приливные вариации уровня в компенсированном сигнале по сравнению с исходными данными, где они «тонут» в вариациях, обусловленных влиянием атмосферного давления.

На рис. 5 приведены оценки спектра мощности вариаций уровней для исходных временных рядов и после компенсации влияния

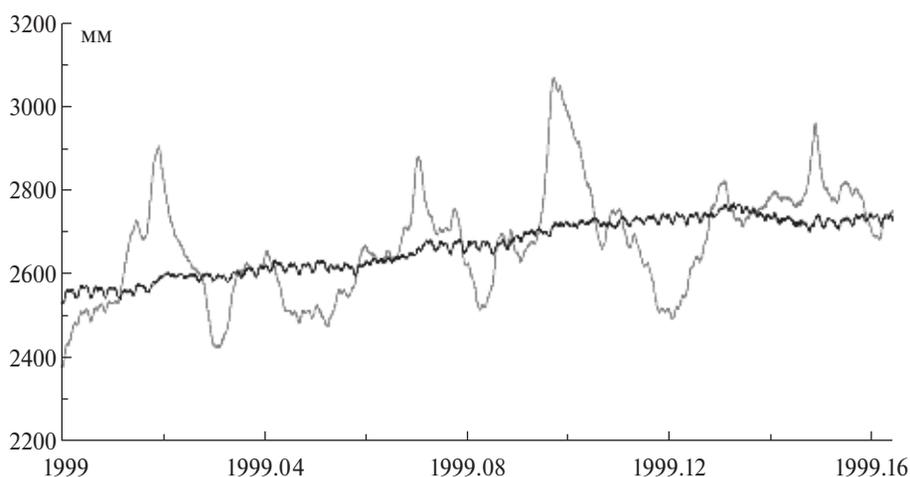


Рис. 4. Сравнение исходных данных уровня подземных вод (серая линия) и результата адаптивной компенсации влияния атмосферного давления в скользящем окне длиной 28 суток (4032 отсчетов с шагом по времени 10 мин — черная линия) для временного фрагмента 2-х первых месяцев 1999 г.

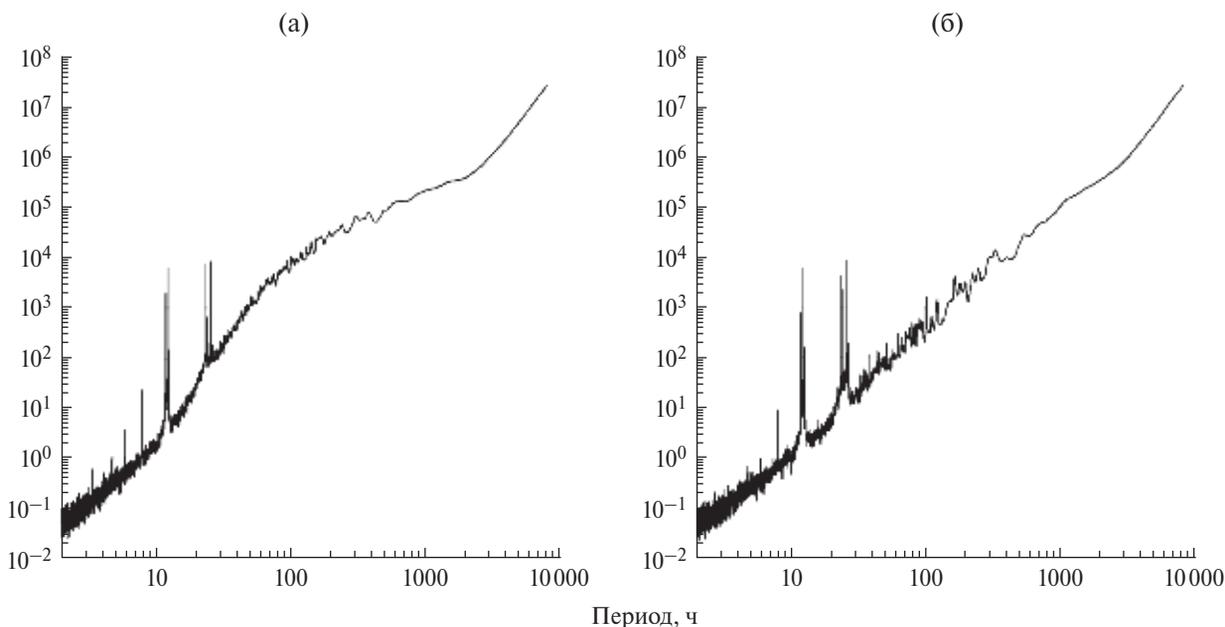


Рис. 5. Сравнение спектров мощности исходных данных уровня подземных вод (а) и после адаптивной компенсации влияния атмосферного давления в скользящем окне длиной 28 суток (б).

атмосферного давления. На рис. 5б видно, что очень надежно выделяются 12-и 24-часовые группы приливных гармоник, причем спектральные пики на периодах, соответствующих высшим обертонам суточного колебания — 6, 4 и т. д. часов, присутствующие на спектре мощности исходного сигнала, после компенсации исчезли. Это свидетельствует о том, что они обусловлены влиянием негармонической формы суточного колебания атмосферного давления. Кроме того, следует подчеркнуть выпрямление графика спектра мощности (в двойном логарифмическом масштабе) после компенсации влияния давления — полностью исчез «горб» на спектре мощности для периодов от 10 до 3000 ч.

СТАТИСТИЧЕСКИЕ СВОЙСТВА ВРЕМЕННОГО РЯДА КОМПЕНСИРОВАННОГО УРОВНЯ ПОДЗЕМНЫХ ВОД

План дальнейшего анализа состоит в том, чтобы оценить набор статистических свойств, описывающих поведение временного ряда уровня подземных вод после компенсации влияния атмосферного давления в последовательных временных фрагментах и использовать их значения для задачи выделения различных состояний множества длительных наблюдений за динамикой водовмещающего горизонта. Длина временного фрагмента была выбрана $N = 1440$ отсчетов, что составляет 10 суток при шаге по времени 10 минут. Далее в этом разделе статьи дается краткое описание используемых статистик, их общее число равно 10. Все оценки производились для временного ряда $x(t)$ приращений уровня подземных вод после компенсации влияния атмосферного давления.

Минимальная нормализованная энтропия вейвлет-коэффициентов En

Пусть $x(t)$ — конечная выборка некоторого случайного сигнала, $t = 1, \dots, N$ — индекс, нумерующий последовательные отсчеты (дискретное время). Определим нормализованную энтропию конечной выборки формулой:

$$En = - \sum_{k=1}^N p_k \cdot \log(p_k) / \log(N),$$

$$p_k = c_k^2 / \sum_{j=1}^N c_j^2, \quad 0 \leq En \leq 1. \quad (3)$$

Здесь $c_k, k = 1, N$ — коэффициенты ортогонального вейвлет-разложения с некоторым

базисом. Ниже использовались 17 ортогональных вейвлетов Добеши: 10 обычных базисов с минимальным носителем с числом обнуляемых от 1 до 10 и 7 так называемых симлетов Добеши [Малла, 2005], с числом обнуляемых моментов от 4 до 10. Для каждого из базисов вычислялась нормализованная энтропия распределения квадратов коэффициентов (3) и находился базис, обеспечивающий минимум величине (3). Заметим, что в силу ортогональности вейвлет-преобразования сумма квадратов коэффициентов равна дисперсии (энергии) сигнала $x(t)$. Таким образом, величина (3) вычисляет энтропию распределения энергии колебаний на различных частотных и временных масштабах. Статистика En использовалась в работах [Lyubushin, 2012; 2014; 2018] при исследовании прогностических свойств сейсмического шума на Японских островах.

Индекс Донохо-Джонстона γ

После того, как вейвлет-базис определен для данного сигнала из условия минимума энтропии, мы можем определить набор вейвлет-коэффициентов, которые являются наименьшим по модулю. В вейвлет-фильтрации эти вейвлет-коэффициенты могут быть обнулены перед обратным вейвлет-преобразованием с целью «уменьшения шума» [Donoho, Johnstone, 1995; Mallat, 1999]. Мы предполагаем, что шум концентрируется в основном в вариациях на первом уровне детальности. Из-за ортогональности вейвлет-преобразования, дисперсия вейвлет-коэффициентов равна дисперсии исходного сигнала. Таким образом, мы оцениваем стандартное отклонение шума как стандартное отклонение вейвлет-коэффициентов на первом уровне детализации. Эта оценка должна быть устойчивой, т. е. нечувствительна к выбросам в значениях вейвлет-коэффициентов на первом уровне. Для этого мы можем использовать робастную медианную оценку стандартного отклонения для нормальной случайной величины:

$$\sigma = \text{med}\{|c_k^{(1)}|, k = 1, \dots, N/2\} / 0.6745, \quad (4)$$

где: $c_k^{(1)}$ — вейвлет-коэффициенты на первом уровне детальности, $N/2$ — число таких коэффициентов. Оценка стандартного отклонения σ из формулы (2) определяет величину $\sigma\sqrt{2 \cdot \ln N}$ как «естественный» порог для выделения шумовых вейвлет-коэффициентов. Величина $\sigma\sqrt{2 \cdot \ln N}$ известна в вейвлет-анализе как порог Донохо-Джонстона, а само выражение для этой величины основано на формуле для

асимптотической вероятности максимальных уклонений гауссовского белого шума [Mallat, 1999]. В результате можно определить безразмерную характеристику сигнала γ , $0 < \gamma < 1$ как отношение числа наиболее информативных вейвлет-коэффициентов, для которых выполнено неравенство $|c_k| > \sigma\sqrt{2 \cdot \ln N}$, к общему числу N всех вейвлет-коэффициентов. Формально чем больше индекс γ , тем более информативным (менее «шумовым») является сигнал.

Модель авторегрессии

Далее широко будет использоваться модель авторегрессии [Box, Jenkins, 1970; Kashyap, Rao, 1976] для временного ряда $x(t)$. Запишем ее в общей форме:

$$x(t) + \sum_{k=1}^p a_k^{(p)} x(t-k) = e^{(p)}(t) + d^{(p)}. \quad (5)$$

Здесь целое число $p \geq 1$ – порядок авторегрессии, вектор $c = (a_1^{(p)}, \dots, a_p^{(p)}, d^{(p)})^T$ является вектором неизвестных параметров. Верхний индекс (p) в формуле (5) подчеркивает, что используется модель авторегрессии p -го порядка. Здесь $a_k^{(p)}$ – коэффициенты авторегрессии, $d^{(p)}$ – параметр статического смещения, $e^{(p)}(t)$ – остаточный сигнал с нулевым средним и дисперсией σ_p^2 . Модель (5) можно записать в компактной форме:

$$x(t) = c^T Y(t) + e^{(p)}(t), \\ Y(t) = (-x(t-1), \dots, -x(t-p), 1)^T. \quad (6)$$

Пусть имеется конечная выборка $\{x(t), t=1, \dots, N\}$. Тогда оценка вектора параметров c из условия минимума суммы квадратов остатков $\sum_{t=p+1}^N (e^{(p)}(t))^2 \rightarrow \min_c$ сводится к решению системы нормальных уравнений с симметричной положительно определенной матрицей A :

$$Ac = R, \quad A = \sum_{t=p+1}^N Y(t)Y^T(t), \quad R = \sum_{t=p+1}^N x(t)Y(t). \quad (7)$$

Полный вектор параметров модели (5) есть $\theta^{(p)} = (c^T, \sigma_p)^T$.

Далее в качестве характеристик фрагментов временного ряда, наряду с прочими параметрами, будут использованы значения коэффициента $a_1^{(1)}$ модели авторегрессии первого порядка и логарифм дисперсии остатка в этой модели: $\lg \sigma_1^2$.

Индекс линейной предсказуемости c_{Pred}

Индекс линейной предсказуемости был введен в работе [Любушин, 2010; Lyubushin, 2012]. Рассмотрим величину: $c_{\text{Pred}} = V_0 / V_{AR} - 1$. Здесь V_0 – дисперсия ошибки $\varepsilon_0(t+1)$ тривиального прогноза $\hat{x}_0(t+1)$ на 1 шаг вперед для сигнала $x(t)$, который равен среднему по предыдущему «малому» временному окну длиной n отсчетов: $\hat{x}_0(t+1) = \sum_{s=t-n+1}^t x(s) / n$.

Таким образом, $\varepsilon_0(t+1) = x(t+1) - \hat{x}_0(t+1)$,

а $V_0 = \sum_{t=n+1}^N \varepsilon_0^2(t) / (N-n)$, где $N > n$ – число отсчетов в последовательных «больших» временных фрагментах. Величина V_{AR} вычисляется по аналогичной формуле $V_{AR} = \sum_{t=n+1}^N \varepsilon_{AR}^2(t) / (N-n)$,

в которой $\varepsilon_{AR}(t+1) = x(t+1) - \hat{x}_{AR}(t+1)$ – ошибка линейного прогноза $\hat{x}_{AR}(t+1)$ на 1 шаг вперед с помощью модели авторегрессии 2-го порядка, коэффициенты которой оцениваются также по предыдущему «малому» временному окну длиной n отсчетов.

Выбор 2-го порядка авторегрессии обусловлен тем, что этот порядок – минимальный для AR -модели, при котором описывается колебательное движение и допускается положение максимума спектральной плотности модели авторегрессии в значениях частот между частотой Найквиста и нулевой. AR -прогноз использует свойство коррелированности соседних значений, и если коррелированность имеет место, то $V_{AR} < V_0$ и $c_{\text{Pred}} > 0$. При длине «большого» временного окна $N = 1440$ длина малого окна была равной $n = 144$.

Выбор 2-го порядка авторегрессии обусловлен тем, что этот порядок – минимальный для AR -модели, при котором описывается колебательное движение и допускается положение максимума спектральной плотности модели авторегрессии в значениях частот между частотой Найквиста и нулевой. AR -прогноз использует свойство коррелированности соседних значений, и если коррелированность имеет место, то $V_{AR} < V_0$ и $c_{\text{Pred}} > 0$. При длине «большого» временного окна $N = 1440$ длина малого окна была равной $n = 144$.

Авторегрессионная мера нестационарности сигнала R^2

Пусть $x(t)$ – изучаемый сигнал, n – половина длины скользящего временного окна, которое далее будем называть «коротким». Пусть τ – центр двойного скользящего временного окна, в которое, тем самым, входят временные отсчеты t , удовлетворяющие условию $\tau - n \leq t \leq \tau + n$. Для левой и правой половин короткого окна построим скалярную авторегрессионную модель (5) порядка $p=2$ сигнала $x(t)$. Оценивая модель независимо по выборкам, попавшим в левую и правую половины двойного скользящего временного окна, получим два вектора параметров $\theta_1^{(p)}$ и $\theta_2^{(p)}$, соответственно. Обозначим $\Delta\theta = \theta_2^{(p)} - \theta_1^{(p)}$ – разницу между векторами

оценок на правой и левой половинах скользящего временного окна.

Если поведение изучаемого сигнала на левой и правой половинах сильно различаются, то будет увеличиваться разница $\Delta\theta$. Для «взвешивания» вектора $\Delta\theta$ в качестве метрической матрицы логично использовать матрицу Фишера, поскольку она определяет скорость изменения логарифмической функции правдоподобия в окрестности точки максимума по параметрам:

$$B = -\frac{\partial^2 \ln(\Phi)}{\partial \theta \partial \theta},$$

$$\ln(\Phi) = -(n-p)\ln(\sigma_p) - \frac{1}{2\sigma_p^2} \sum_t (x(t) - c^T Y(t))^2 \quad (8)$$

– матрицы вторых производных по параметрам от условной логарифмической функции правдоподобия авторегрессионной модели. Обозначим через $B^{(1)}$ и $B^{(2)}$ матрицы, вычисленные по левой и правой половинам скользящего окна, соответственно. Тогда мерой нестационарности поведения процесса $x(t)$ в симметричной окрестности точки τ будет величина:

$$r^2(\tau) = (\Delta\theta^T B^{(1)} \Delta\theta + \Delta\theta^T B^{(2)} \Delta\theta) / (2(n-p)). \quad (9)$$

В формуле (9) полусумма длин вектора разности параметров $\Delta\theta$, измеряемых с помощью метрических матриц $B^{(1)}$ и $B^{(2)}$, делится на $(n-p)$ – число отсчетов в левой и правой частях скользящего окна за вычетом числа авторегрессионных параметров. Такая метрика обеспечивает естественную безразмерную меру нестационарности поведения исследуемого сигнала. Проведя несложные выкладки, нетрудно получить следующее выражение:

$$\Delta\theta^T B \Delta\theta = \frac{2(\Delta\sigma_p)^2}{\sigma_p^2} + \frac{\Delta c^T \left(\sum_t Y(t) Y^T(t) \right) \Delta c}{\sigma_p^2 (n-p)} + \frac{4\Delta c^T \Delta\sigma_p \sum_t e^{(p)}(t) Y(t)}{\sigma_p^3 (n-p)}, \quad (10)$$

полезное при вычислении значения меры нестационарности (9). Мера нестационарного поведения (9), (10) была введена в работе [Любушин и др., 1999; Любушин, 2007].

Используя формулы (9), (10) можно определить другую, более устойчивую меру нестационарного поведения исследуемого сигнала внутри «длинного» временного интервала, состоящего

из N последовательных отсчетов. Для этого возьмем «короткое» окно радиуса n отсчетов, $2n+1 < N$, и вычислим меру нестационарного поведения $r^2(\tau)$ для всех возможных положений центральной точки τ внутри «длинного» окна, при которых «короткое» окно целиком лежит внутри «длинного». Нетрудно посчитать, что число таких допустимых положений центральной точки τ равно $N-2n$. Определим интегральную меру нестационарности R^2 для «длинного» окна как медиану значений $r^2(\tau)$ для всех допустимых положений центральной точки τ «короткого» окна внутри «длинного». В вычислениях мы использовали длины окон $N=1440$ и $n=144$. В дальнейшем мы будем рассматривать логарифм меры нестационарности $\lg(R^2)$.

Мультифрактальные параметры $\Delta\alpha$, α^* и α_{\min}

Рассмотрим некоторое случайное колебание $x(t)$ на интервале времени $[t-\delta/2, t+\delta/2]$ длиной δ с центром во временной точке t . Рассмотрим размах $\mu(t, \delta)$ случайного колебания на этом интервале, то есть разницу между максимальным и минимальным значениям:

$$\mu(t, \delta) = \max_{t-\delta/2 \leq s \leq t+\delta/2} x(s) - \min_{t-\delta/2 \leq s \leq t+\delta/2} x(s). \quad (11)$$

Если устремить $\delta \rightarrow 0$, то $\mu(t, \delta)$ будет также стремиться к нулю, но здесь важна скорость этого убывания. Если скорость определяется законом $\delta^{h(t)}$: $\mu(t, \delta) \sim \delta^{h(t)}$ или если существует предел $h(t) = \lim_{\delta \rightarrow 0} \frac{\log(\mu(t, \delta))}{\log(\delta)}$, то величина $h(t)$

называется экспонентой Гельдера-Липшица. Если величина $h(t)$ не зависит от момента времени t : $h(t) = \text{const} = H$, то случайное колебание $x(t)$ называется монофрактальным, а величина H – экспонентой Херста. Если же экспоненты Гельдера-Липшица $h(t)$ различаются для разных моментов времени t , то случайное колебание называется мультифракталом и для него можно определить понятие спектра сингулярности $F(\alpha)$ [Feder, 1991]. Для этого выделим множество $C(\alpha)$ таких моментов времени t , которые имеют одно и то же значение α экспоненты Гельдера-Липшица: $h(t) = \alpha$. Множества $C(\alpha)$ не являются пустыми не для всех значений α , то есть существуют некоторые минимальное α_{\min} и максимальное α_{\max} , такие что лишь для $\alpha_{\min} < \alpha < \alpha_{\max}$ множества $C(\alpha)$ содержат некоторые элементы. Мультифрактальный спектр сингулярности $F(\alpha)$ – это фрактальная размерность множества точек $C(\alpha)$. Параметр $\Delta\alpha = \alpha_{\max} - \alpha_{\min}$, называемый шириной носителя спектра сингулярности, представляется

важной мультифрактальной характеристикой. Кроме того, значительный интерес представляет аргумент α^* , доставляющий максимум спектру сингулярности: $F(\alpha^*) = \max_{\alpha_{\max} \leq \alpha \leq \alpha_{\min}} F(\alpha)$, называемый обобщенным показателем Херста. Максимум спектра сингулярности не может превосходить 1 – размерности вмещающего множества или оси времени, $0 < F(\alpha^*) \leq 1$, обычно $F(\alpha^*) = 1$. Заметим, что для монофрактального сигнала $\Delta\alpha = 0$, $\alpha^* = H$.

Ниже для оценки мультифрактальных характеристик сигналов использовался метод, основанный на анализе флуктуаций после устранения масштабозависимых трендов [Kantelhardt et al., 2002]. В работах [Lyubushin, 2010; 2012; 2013; 2014; 2018] мультифрактальные параметры $\Delta\alpha$ и α^* были использованы при оценке сейсмической опасности по свойствам сейсмического шума в Японии (см. также [Любушин, 2009; 2012; 2014]).

Коэффициент эксцесса κ определяется формулой $\kappa = M(x^4)/(M(x^2))^2$ [Cramer, 1999]. Коэффициент эксцесса характеризует «остроту» графика плотности вероятности распределения случайной величины x с нулевым средним и дает меру отклонения плотности вероятности от нормального закона, для которого $\kappa = 3$. Здесь операция $M(\dots)$ означает вычисление математического ожидания, в данном случае просто выборочного среднего случайной величины. Обычно под коэффициентом эксцесса понимается введенная выше величина κ , из которой вычитается 3, чтобы для нормального закона эксцесс был равен нулю. Однако далее мы будем рассматривать логарифм эксцесса $\lg(\kappa)$ – поэтому значение 3 не вычитается, чтобы гарантировать положительные значения κ .

Итак для каждого временного окна имеется 10 параметров, характеризующих статистические свойства временного ряда внутри этого окна: минимальная нормированная энтропия вейвлет-коэффициентов En , индекс Донохона-Джонстона γ , коэффициент $a_1^{(1)}$ и логарифм дисперсии $\lg \sigma_1^2$ в модели авторегрессии 1-го порядка, индекс линейной предсказуемости c_{Pred} и логарифм меры нестационарности $\lg(R^2)$, основанные на использовании модели авторегрессии 2-го порядка, мультифрактальные параметры $\Delta\alpha$, α^* , α_{\min} и логарифм коэффициента эксцесса $\lg(\kappa)$.

Десятимерный вектор параметров, характеризующих статистические свойства временного ряда внутри последовательных временных фраг-

ментов длиной 10 суток (1440 отсчетов с шагом по времени 10 мин) обозначим:

$$\zeta = (En, \gamma, a_1^{(1)}, \lg \sigma_1^2, c_{Pred}, \lg(R^2), \Delta\alpha, \alpha^*, \alpha_{\min}, \lg(\kappa)). \quad (12)$$

На рис. 6 представлены графики изменения компонент 10-мерного вектора ζ свойств временного ряда в зависимости от положения правого конца последовательных временных окон длиной 10 суток.

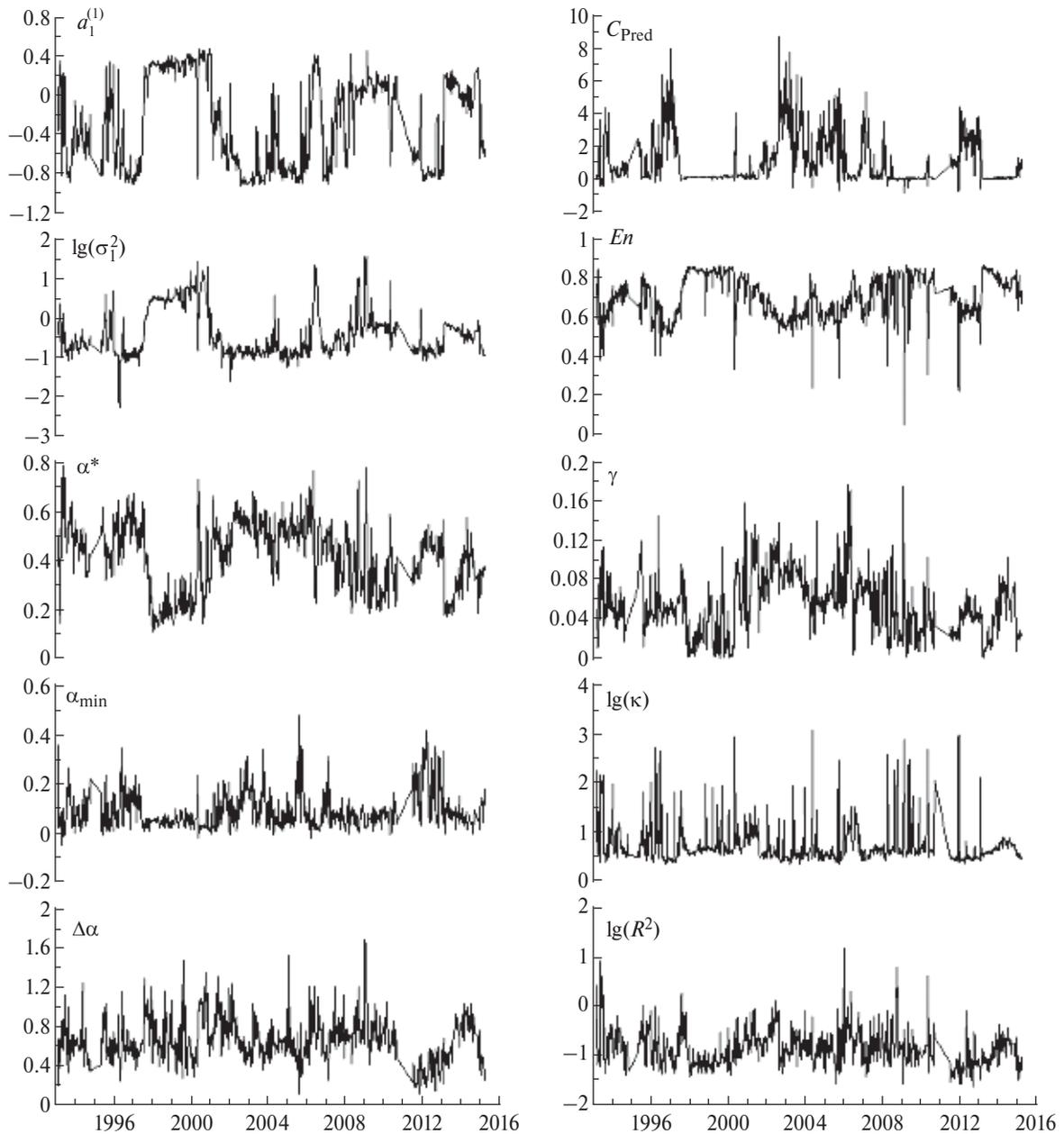
ФАКТОРНЫЙ АНАЛИЗ ВЕКТОРА СВОЙСТВ ВРЕМЕННОГО РЯДА

Сделаем попытку выделить различные состояния в 22-летней истории временного ряда наблюдений за уровнем подземных вод, используя кластерный анализ 10-мерного вектора признаков (12). Следует отметить, что отдельные компоненты вектора ζ очевидно демонстрируют, что всю историю наблюдений можно разбить на несколько интервалов с различным поведением, например, $a_1^{(1)}, \lg \sigma_1^2, c_{Pred}, \alpha^*, En$. Для формального разбиения полученного облака векторов ζ на кластеры выполним предварительную операцию понижения размерности с помощью факторного анализа. Модель факторного анализа [Hartman, 1967] в рассматриваемом случае описывается формулой:

$$z = \Lambda \cdot f + e, \quad (13)$$

где 10-мерный вектор z получается из вектора ζ операцией нормировки, которая заключается в удалении выборочного среднего и делении на выборочную оценку стандартного отклонения для каждой компоненты вектора ζ . После выполнения операции нормировки вычисляется корреляционная матрица R_{zz} .

В формуле (13) f – вектор размерности $q < p = 10$, состоящий из скрытых факторов – некоторых случайных векторов, которые «управляют» значениями скалярных компонент многомерного вектора z посредством умножения на матрицу факторных нагрузок Λ размером p строк на q столбцов. Элементы матрицы $\Lambda = (\lambda_{j\alpha})$, $j = 1, \dots, p$; $\alpha = 1, \dots, q$ являются неизвестными параметрами модели, которые необходимо определить, имея выборочную оценку R_{zz} корреляционной матрицы исходных данных. Пока будем считать, что число скрытых параметров q известно. Относительно свойств случайного вектора f предполагается, что его среднее равно нулю, $M\{f\} = 0$, а его ковариационная матрица единична $M\{f \cdot f^T\} = I_q$, где I_q – q -мерная единичная матрица.



Правый конец временного окна длиной 10 суток

Рис. 6. Графики 10 свойств временного ряда уровня подземных вод после компенсации влияния атмосферного давления в последовательных окнах длиной 10 суток.

Это условие означает ортогональность факторов (в гауссовском случае — их независимость). Условие равенства единице дисперсий ортогональных факторов является своего рода нормировкой, так как в противном случае этого можно добиться масштабированием элементов матрицы Λ . Вектор e в формуле (13) имеет ту же размерность, что и исходный вектор z и состоит из случайных величин, описывающих шум по каждой из компонент вектора z ,

то есть, не несущих полезной информации. Так как шумы по различным компонентам должны быть независимы, то относительно вектора e предполагается, что он центрирован и его ковариационная матрица имеет диагональный вид: $M\{e \cdot e^T\} = \Psi^2 = \text{diag}(\psi_1^2, \dots, \psi_p^2)$, где $\psi_j^2, j = 1, \dots, p$ — т.н. остаточные дисперсии или дисперсии шумов. Элементы диагональной матрицы Ψ^2 также являются параметрами модели (13).

Существуют несколько способов идентификации параметров модели (13), но среди них наиболее надежным и простым является метод минимальных остатков [Harman, 1967]. Из условий диагональности ковариационных матриц векторов f и e нетрудно получить, что в силу модели (13) ковариационная матрица вектора z равна:

$$\Sigma = M\{zz^T\} = \Lambda\Lambda^T + \Psi^2. \quad (14)$$

Метод минимальных остатков заключается в определении элементов матрицы Λ из условия минимума суммы квадратов разностей между выборочными оценками и теоретическими значениями попарных коэффициентов корреляции. Тем самым, критерием близости модели к данным является близость всех теоретических коэффициентов корреляции их выборочным оценкам. Обозначим через r_{ij} элементы матрицы R_{zz} . Тогда необходимо минимизировать следующую функцию от элементов матрицы факторных нагрузок:

$$\Phi(\Lambda) = \sum_{j=1}^{p-1} \sum_{i=j+1}^p (r_{ij} - \sum_{\alpha=1}^q \lambda_{j\alpha} \lambda_{i\alpha})^2 \rightarrow \min_{\Lambda}. \quad (15)$$

При этом на элементы матрицы Λ должны быть наложены ограничения:

$$\sum_{\alpha=1}^q \lambda_{j\alpha}^2 \leq 1, \quad j = 1, \dots, p, \quad (16)$$

вытекающие из условия равенства единице диагональных элементов теоретической матрицы корреляций (14). Стоит отметить, что задача определения матрицы Λ независима от определения диагональной матрицы остаточных дисперсий Ψ^2 . После того, как задача на минимум (15) при ограничениях (16) решена, остаточные дисперсии находятся автоматически:

$$\psi_j^2 = 1 - \sum_{\alpha=1}^q \lambda_{j\alpha}^2, \quad j = 1, \dots, p. \quad (17)$$

После определения матрицы факторных нагрузок финальным шагом анализа является вычисление собственно реализаций ортогональных факторов — облака q -мерных векторов f . Наиболее простая оценка следует из условия, что вектор шумов e распределен согласно p -мерному нормальному распределению с ковариационной матрицей Ψ^2 . В этом случае оценкой максимального правдоподобия будет оценка взвешенного метода наименьших квадратов:

$$e^T \Psi^{-2} e \rightarrow \min_f, \quad f = (\Lambda^T \Psi^{-2} \Lambda)^{-1} \Lambda^T \Psi^{-2} z. \quad (18)$$

Однако оценка (18) дает вектор общих факторов с недиагональной ковариационной

матрицей. Для того, чтобы компоненты вектора факторов были ортогональны, следует применить модификацию оценки (18), предложенную в работе [Anderson, Rubin, 1956]:

$$f = (\Lambda^T \Psi^{-2} \Sigma \Psi^{-2} \Lambda)^{-1/2} \Lambda^T \Psi^{-2} z, \quad \Sigma = \Lambda \Lambda^T + \Psi^2. \quad (19)$$

Целью получения реализаций вектора общих факторов f является снижение размерности задачи [Айвазян и др., 1989]. Вопрос о выборе числа q общих факторов — размерности вектора f является самым трудным в факторном анализе. Для его решения был предложен критерий Риппе [Lawley, Maxwell, 1971], который опирается на предположение о нормальности распределения векторов z . Однако этот критерий показал очень сильную чувствительность к малым отклонениям от нормальности, что делает его практически неприменимым. Если отсутствует априорная информация о числе q , то можно получить оценку максимально допустимого числа общих факторов, начав решать задачу с минимального $q = 1$ и постепенно увеличивать q на единицу до тех пор, пока модель факторного анализа не вырождается (общее число параметров станет избыточным). После этого можно взять в качестве q последнее максимальное значение перед вырождением задачи. Вырождение задачи факторного анализа называется «случаем Хейвуда» [Harman, 1967] и заключается в обнулении остаточной дисперсии ψ_j^2 для одного или нескольких компонент вектора z . Практически вместо обнуления наблюдается резкое уменьшение остаточной дисперсии для той или иной компоненты по сравнению с прочими (на несколько порядков).

Именно этот метод выбора q использовался в нашем случае, и он определил максимальное допустимое значение числа общих ортогональных факторов $q = 4$. На рис. 7 представлены графики 4-х общих ортогональных факторов для множества 10-мерных векторов свойств временного ряда уровня подземных вод после компенсации влияния атмосферного давления.

КЛАСТЕРНЫЙ АНАЛИЗ ОРТОГОНАЛЬНЫХ ОБЩИХ ФАКТОРОВ

После понижения размерности множества векторов статических свойств последовательности временных интервалов временного ряда путем перехода к рассмотрению 4 ортогональных общих факторов f , выделим кластеры в пространстве общих факторов, применив популярный метод k -средних (от же ISODATA) [Айвазян

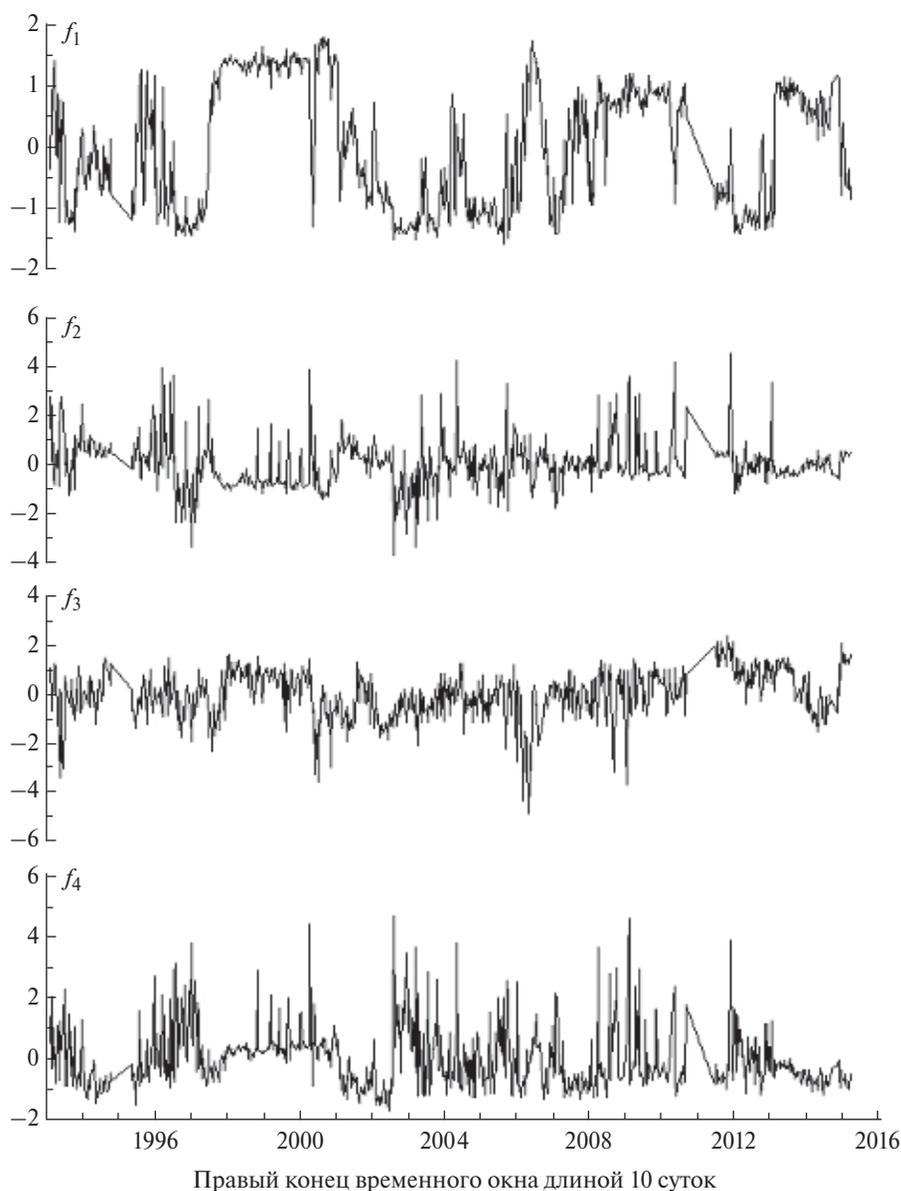


Рис. 7. Графики 4 ортогональных общих факторов набора 10 свойств временного ряда уровня подземных вод после компенсации влияния атмосферного давления в последовательных окнах длиной 10 суток.

и др., 1989; Duda, Hart, 1973]. В нашем случае объекты классификации представляют собой точки в 4-мерном евклидовом пространстве, причем каждая компонента этих векторов имеет нулевое среднее и единичное стандартное отклонение. Поэтому логично ввести между векторами обычное евклидово расстояние. Рассмотрим облако 4-мерных векторов f общих ортогональных факторов. Внутри минимального параллелепипеда, содержащего классифицируемые точки f , случайным образом располагаются центры пробных кластеров, причем число $q \geq 2$ таких кластеров фиксировано. Обозначим символом Γ начальное случайное положение пробных

кластеров. Для заданного расположения центров кластеров производится пробное разбиение множества точек по принципу минимума расстояния до того или иного центра. Пусть $c_k, k = 1, \dots, q$ – вектора центров кластеров, n_k – число точек в k -ом кластере, $\sum_{k=1}^q n_k = M$ – общее число точек в разбиваемом множестве. В нашем случае $M = 746$, что соответствует числу последовательных временных интервалов длиной 10 суток, которые содержат не более 2% восполненных данных на местах пропусков. Пусть B_k – множество векторов, принадлежащих

k -му кластеру. Вычислим вектора центров тяжести получившихся кластеров: $r_k = \sum_{f \in B_k} \xi / n_k$.

Если для всех $c_k = r_k$, то разбиение завершается. В противном случае вектора центров кластеров c_k перемещаются в центры тяжести r_k , производится новое разбиение на кластеры, вычисляются новые центры тяжести, проверяется условие завершения разбиения и т. д. Процедура быстро сходится. Однако то разбиение на кластеры, которое получилось по завершению итераций, зависит от случайных положений центров пробных кластеров Γ в самом начале итераций. Качество финального разбиения оценивается критерием компактности кластеров:

$$J(q|\Gamma) = \sum_{k=1}^q \sum_{f \in B_k} |f - c_k|^2. \quad (20)$$

Для заданного числа кластеров q естественно искать такое случайное начальное расположение Γ , для которого величина (20) минимальна. Это достигается методом Монте-Карло: случайные эксперименты по вбрасыванию центров пробных кластеров внутрь облака точек повторяются большое число раз (ниже, при анализе конкретных данных, мы использовали 10^4 попыток), после чего выбирается то разбиение, для которого реализовался минимум по Γ .

Далее возникает задача определения оптимального числа кластеров, на которое следует разбить множество признаков. Пусть $J_0(q) = \min_{\Gamma} J(q|\Gamma)$. Если последовательно уменьшать число пробных кластеров q от некоторого достаточно большого значения до минимального $q = 2$, то величина $J_0(q)$ будет монотонно увеличиваться, однако при оптимальном значении кластеров (если такое существует) она будет претерпевать излом. Более эффективный метод выявления оптимального числа кластеров состоит в использовании псевдо- F -статистики [Vogel, Wong, 1979], заимствованного из дисперсионного анализа:

$$PFS(q) = (M - q) \cdot \sum_{k=1}^q n_k \cdot |c_k - r_0|^2 / ((q - 1) \cdot J_0(q)), \quad (21)$$

где $r_0 = \sum f / M$ – общий центр тяжести всего классифицируемого множества точек. Оптимальному значению числа кластеров соответствует точка максимума функции (21).

На рис. 8 представлена зависимость псевдо- F -статистики от числа пробных кластеров, из которой видно, что оптимальное число кластеров в пространстве ортогональных общих факторов равно 5.

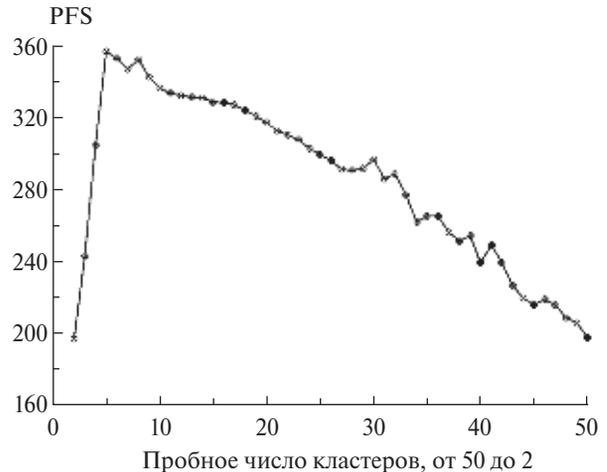


Рис. 8. График псевдо- F -статистики для кластеризации 4 ортогональных общих факторов.

СПЕКТРАЛЬНЫЙ АНАЛИЗ ПОСЛЕДОВАТЕЛЬНОСТИ ПЕРЕХОДОВ МЕЖДУ КЛАСТЕРАМИ

На рис. 9а, 9б показаны последовательность переходов между выделенными 5 состояниями и последовательность 279 событий смены номера кластера. Исследуем вопрос о наличии периодических составляющих интенсивности этих переходов. Ниже используется метод выделения периодических компонент точечного процесса, основанный на вычислении разности между максимальными значениями логарифмических функций правдоподобия для двух моделей: пуассоновского процесса с постоянной интенсивностью и процессом, имеющим периодическую составляющую, предложенный в работе [Любушин и др., 1998].

Пусть $t_i, i = 1, \dots, N$ – времена последовательности событий, наблюдаемых на интервале $(0, T]$. Рассмотрим следующую модель интенсивности, содержащую периодическую компоненту:

$$\lambda(t) = \mu(1 + a \cos(\omega t + \varphi)), \quad (22)$$

где: частота ω , амплитуда $a, 0 \leq a \leq 1$, фазовый угол $\varphi, \varphi \in [0, 2\pi]$ и множитель $\mu > 0$ (описывающий пуассоновскую часть интенсивности) являются параметрами модели. Таким образом, пуассоновская часть интенсивности модулируется гармоническим колебанием.

Зафиксируем какое-то значение частоты ω . Логарифмическая функция правдоподобия модели (22) в этом случае для серии наблюдаемых событий равна:

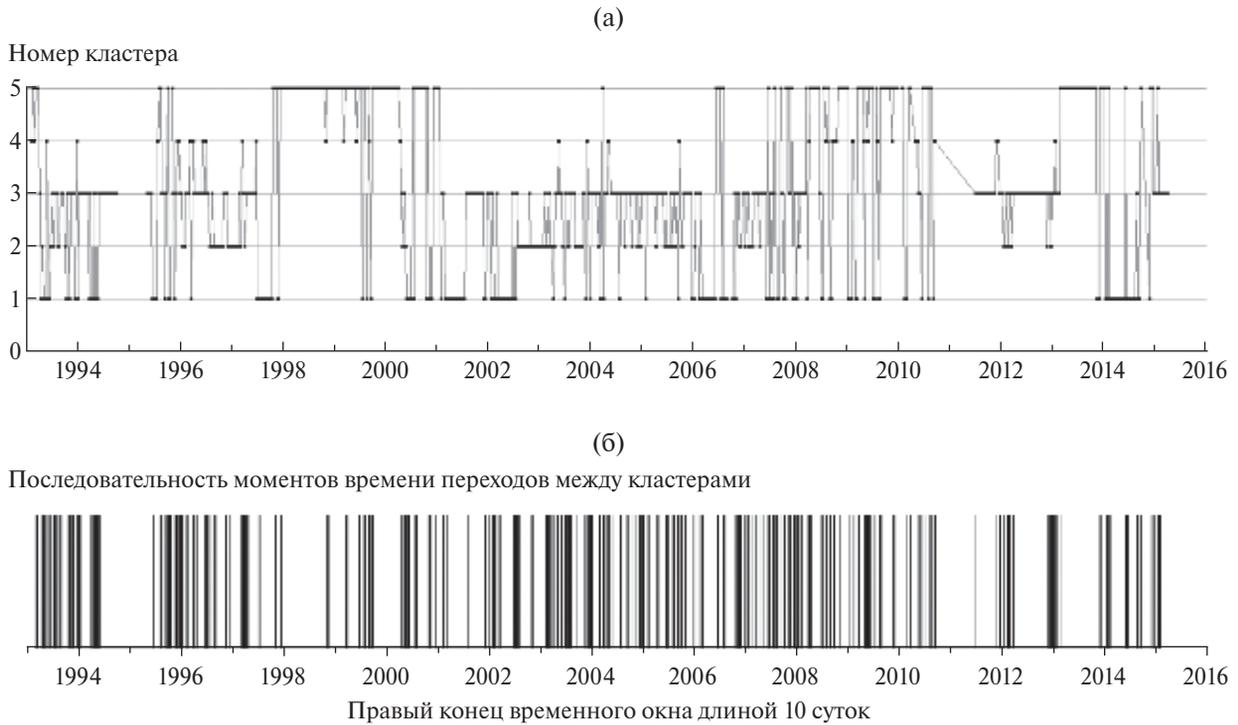


Рис. 9. (а) – Последовательность 5 кластеров в пространстве 4-х общих факторов статистических свойств временных фрагментов длиной 10 суток временного ряда уровня подземных вод после компенсации влияния атмосферного давления (рис. 7); (б) – последовательность 279 переходов между кластерами.

$$\begin{aligned} \ln L(\mu, a, \varphi | \omega) &= \sum_{t_i} \ln(\lambda(t_i)) - \int_0^T \lambda(u) du = \\ &= N \ln(\mu) + \sum_{t_i} \ln(1 + a \cos(\omega t_i + \varphi)) - \\ &- \mu T - \frac{\mu a}{\omega} [\sin(\omega T + \varphi) - \sin(\varphi)]. \end{aligned} \quad (23)$$

Взяв максимум выражения (23) по отношению к параметру μ нетрудно найти что:

$$\mu = \hat{\mu}(a, \varphi | \omega) = \frac{N}{T + a(\sin(\omega T + \varphi) - \sin(\varphi)) / \omega}. \quad (24)$$

Подставляя (24) в формулу (23) получаем:

$$\begin{aligned} \ln(L(\hat{\mu}, a, \varphi | \omega)) &= \sum_{t_i} \ln(1 + a \cos(\omega t_i + \varphi)) + \\ &+ N \ln(\hat{\mu}(a, \varphi | \omega)) - N. \end{aligned} \quad (25)$$

Приращение логарифмической функции правдоподобия вследствие рассмотрения более богатой, чем для чисто случайного потока событий, модели интенсивности с гармонической компонентой с заданной частотой ω равно:

$$\begin{aligned} \Delta \ln L(a, \varphi | \omega) &= \sum_{t_i} \ln(1 + a \cos(\omega t_i + \varphi)) + \\ &+ N \ln(\hat{\mu}(a, \varphi | \omega) / \hat{\mu}_0). \end{aligned} \quad (26)$$

Пусть

$$R(\omega) = \max_{a, \varphi} \Delta \ln L(a, \varphi | \omega), \quad 0 \leq a \leq 1, \varphi \in [0, 2\pi]. \quad (27)$$

Функция (27) может рассматриваться как обобщение спектра для последовательности событий. График этой функции показывает насколько «более выгодна» периодическая модель интенсивности по сравнению с чисто случайной моделью. Максимальные значения функции (27) выделяют частоты, присутствующие в потоке событий. Как показано в работе [Любушин и др., 1998]:

$$P\{R(\omega) < x\} = 1 - e^{-x}, \quad N \rightarrow \infty. \quad (28)$$

Асимптотическое распределение (28) можно использовать для оценки статистической значимости максимумов функции (27). Из формулы (28) следует, что вероятность превышения величины (27) порога 4 равна 0.02, то есть асимптотически значимость частот ω , для которых $R(\omega) > 4$, равна 98%.

Непосредственное применение изложенного метода спектрального анализа последовательности событий, представленной на рис. 9б затрудняется наличием достаточно длинных пропусков данных в 1995 г и в конце 2010 г. Такие пропуски данных могут привести к появлению ложных низкочастотных компонент. Поэтому рассмотрим последовательность 217 переходов между кластерами лишь для интервала времени от 22.06.1995 до 10.09.2010 гг., когда не было длинных пропусков регистрации. На рис. 10 представлен график функции (27), вычисленной для 2000 пробных значений периодов в равномерной логарифмической шкале от 20 до 1000 суток. Из этого графика следует, что у переходов между кластерами есть 2 значимых периода (с вероятностью принятия гипотезы о существовании таких периодичностей не менее 98%) с периодами 46 и 275 суток.

ПРОВЕРКА ГИПОТЕЗЫ СВЯЗИ С СИЛЬНЕЙШИМИ ЗЕМЛЕТРЯСЕНИЯМИ

Частотные оценки вероятностей нахождения в каждом из выделенных состояний, представленных на рис. 9а, равны 0.217, 0.120, 0.341, 0.055, 0.267, соответственно, для состояний 1–5. Таким образом, 4-е состояние можно назвать аномальным, поскольку вероятность нахождения в этом кластере равна 5.5%. Проверим гипотезу о том, что пребывание свойств временного ряда в этом аномальном кластере может быть как-то связано с последовательностью 27 сильнейших землетрясений с магнитудой не менее 8, происшедших в течение 22 лет наблюдений, 1993–2015 гг. На рис. 11а, 11б вертикальными штрихами представлены 2 последовательности событий: 34 момента времени, соответствующие правому концу 10-суточного временного окна, после которого свойства ряда наблюдений переходят в 4-й кластер, рис. 11а и 27 сильнейших землетрясений с магнитудой не менее 8, происшедших в течение 22 лет наблюдений, 1993–2015 гг. (рис. 11б).

Для проверки гипотезы применим метод матриц влияния, основанный на линейной модели взаимодействия между потоками событий. Этот метод был предложен в работе [Любушин, Писаренко, 1993; Любушин, 2007]), где рассматривалась модель взаимодействия последовательностей сейсмических событий из нескольких регионов и использования функции влияния события, затухающей по степенному закону, используемому в модели ETAS (epidemic-type aftershock sequence) [Ogata, Katsura, 1993],

причем показатель степенного закона затухания входит в список определяемых параметров модели. Ниже метод упрощен и изложен для частного случая 2-х последовательностей событий и использования экспоненциального закона затухания функции влияния события.

Пусть $t_j^{(\alpha)}$, $j = 1, \dots, N_\alpha$; $\alpha = 1, 2$ представляют собой моменты времен 2-х потоков событий. Представим интенсивность какого-нибудь процесса в виде:

$$\lambda^{(\alpha)}(t) = b_0^{(\alpha)} + \sum_{\beta=1}^2 b_\beta^{(\alpha)} g^{(\beta)}(t), \quad (29)$$

где: $b_0^{(\alpha)} \geq 0$, $b_\beta^{(\alpha)} \geq 0$ – параметры; $g^{(\beta)}(t)$ – функция влияния событий потока с номером β . Представим функцию влияния события в виде:

$$g^{(\beta)}(t) = \sum_{t_j^{(\beta)} < t} \exp(-(t - t_j^{(\beta)})/\tau), \quad (30)$$

где $\tau > 0$ является характерным временным масштабом рассмотрения взаимодействия между потоками событий. Таким образом, в соответствии с формулой (30), вес события с номером j становится ненулевым для времен $t > t_j^{(\beta)}$ и экспоненциально затухает с характерным временем τ по мере возрастания текущего времени t . Сумма всех таких затухающих экспонент образует функцию влияния $g^{(\beta)}(t)$ потока с номером β . Параметр $b_\beta^{(\alpha)}$ является масштабирующим множителем и именно он определяет степень влияния потока β на поток α : $\beta \rightarrow \alpha$. Параметр $b_\alpha^{(\alpha)}$ определяет степень влияния потока α на самого себя (самовозбуждение), а параметр $b_0^{(\alpha)}$ отражает чисто случайную компоненту интенсивности, для которой функция влияния постоянна и тождественно равна 1.

Зафиксируем параметр τ и рассмотрим задачу определения параметров $b_0^{(\alpha)}$, $b_\beta^{(\alpha)}$. Логарифмическая функция правдоподобия для нестационарного пуассоновского процесса равна [Cox, Lewis, 1966]:

$$\ln(L_\alpha) = \sum_{j=1}^{N_\alpha} \ln(\lambda^{(\alpha)}(t_j^{(\alpha)})) - \int_0^T \lambda^{(\alpha)}(s) ds, \quad \alpha = 1, 2, \quad (31)$$

где $[0, T]$ есть интервал наблюдения. Таким образом, необходимо найти максимум функции (31) по отношению к параметрам $b_0^{(\alpha)}$, $b_\beta^{(\alpha)}$. Принимая во внимание формулу (29) и используя правило дифференцирования сложной функции, нетрудно получить следующее выражение:

$$b_0^{(\alpha)} \frac{\partial \ln(L_\alpha)}{\partial b_0^{(\alpha)}} + \sum_{\beta=1}^2 b_\beta^{(\alpha)} \frac{\partial \ln(L_\alpha)}{\partial b_\beta^{(\alpha)}} = N_\alpha - \int_0^T \lambda^{(\alpha)}(s) ds. \quad (32)$$

Поскольку параметры $b_0^{(\alpha)}, b_\beta^{(\alpha)}$ должны быть неотрицательными, то каждый член в левой части формулы (32) равен нулю в точке максимума функции (31) – либо в силу необходимых условий экстремума (если параметры положительны), либо, если максимума достигается на границе, то сами параметры равны нулю. Следовательно, в точке максимума логарифмической функции правдоподобия (31) выполняется равенство:

$$\int_0^T \lambda^{(\alpha)}(s) ds = N_\alpha. \quad (33)$$

Подставим выражение (29) в (33) и разделим на длину интервала наблюдения. Тогда получим другой вид формулы (33):

$$b_0^{(\alpha)} + \sum_{\beta=1}^m b_\beta^{(\alpha)} \cdot \bar{g}^{(\beta)} = \lambda_0^{(\alpha)} \equiv \frac{N_\alpha}{T}, \quad (34)$$

где $\bar{g}^{(\beta)} = \frac{1}{T} \int_0^T g^{(\beta)}(s) ds$ – среднее значение функции влияния. Подставляя $b_0^{(\alpha)}$ из (34) в (31), получим следующую задачу на максимум, эквивалентную задаче максимизации (31):

$$\Phi^{(\alpha)}(b_1^{(\alpha)}, b_2^{(\alpha)}) = \sum_{j=1}^{N_\alpha} \ln(\lambda_0^{(\alpha)} + \sum_{\beta=1}^2 b_\beta^{(\alpha)} \Delta g^{(\beta)}(t_j^{(\alpha)})) \rightarrow \max, \quad (35)$$

где $\Delta g^{(\beta)}(t) = g^{(\beta)}(t) - \bar{g}^{(\beta)}$, при ограничениях:

$$b_1^{(\alpha)} \geq 0, b_2^{(\alpha)} \geq 0, \sum_{\beta=1}^2 b_\beta^{(\alpha)} \bar{g}^{(\beta)} \leq \lambda_0^{(\alpha)}. \quad (36)$$

Функция (35) является выпуклой с отрицательно определенным гессианом [Любушин, Писаренко, 1993] и, следовательно, задача (35),

(36) имеет единственное решение. Эта задача решается численно методом проекции градиента [Моисеев и др., 1978].

Решив задачу (35), (36) для заданного τ , можно ввести доли интенсивности $\mu_\beta^{(\alpha)}, \alpha = 1, 2; \beta = 0, 1, 2$ согласно формулам:

$$\kappa_0^{(\alpha)} = \frac{b_0^{(\alpha)}}{\lambda_0^{(\alpha)}} \geq 0, \quad \kappa_\beta^{(\alpha)} = \frac{b_\beta^{(\alpha)} \cdot \bar{g}^{(\beta)}}{\lambda_0^{(\alpha)}} \geq 0, \quad (37)$$

которые можно назвать элементами матрицы влияния.

Интерпретация этих величин вполне естественна: $\kappa_0^{(\alpha)}$ является частью средней интенсивности процесса с номером α , являющейся чисто стохастической, часть $\kappa_\alpha^{(\alpha)}$ вызвана влиянием самовозбуждения $\alpha \rightarrow \alpha$ и $\kappa_\beta^{(\alpha)}, \beta \neq \alpha$ обусловлена внешним влиянием $\beta \rightarrow \alpha$. Из формулы (33) вытекает условие нормировки:

$$\kappa_0^{(\alpha)} + \sum_{\beta=1}^2 \kappa_\beta^{(\alpha)} = 1, \quad \alpha = 1, 2. \quad (38)$$

В таблице представлены результаты расчета элементов матриц влияния для двух значений времен затухания, 10 и 100 суток. Видно, что рассматриваемые последовательности событий являются пуассоновскими для времени затухания 10 суток, но для 100 суток в последовательности моментов времени перехода в 4-й кластер появляется небольшая самовозбуждающаяся компонента. Однако для всех рассматриваемых времен затухания связь между потоками событий отсутствует.

ВЫВОДЫ

Конечным результатом предложенного многофакторного анализа свойств временного ряда наблюдений за уровнем подземных вод является диаграмма переходов между статистически значимыми кластерами, представленная на рис. 9а.

Матрицы влияния для анализа связи между сильнейшими землетрясениями и временами перехода в аномальный кластер для двух времен затухания экспоненциальной функции влияния событий, 10 и 100 суток

Время затухания τ , сутки	Последовательность событий	Пуассоновская часть	Самовозбуждение	Влияние другой последовательности
10	Землетрясения	0.954	0.046	0.000
	Переход в 4-й кластер	1.000	0.000	0.000
100	Землетрясения	1.000	0.000	0.000
	Переход в 4-й кластер	0.784	0.216	0.000

Спектральный анализ точечного процесса времен перехода между различными кластерами (рис. 9б) позволил выделить значимые периодические компоненты интенсивности таких переходов с периодами 46 и 275 суток (рис. 10). Наличие таких периодов, возможно, отражает как региональные, так и глобальные причины, воздействующие на обширный подземный горизонт. Проверка связи переходов в аномальный кластер и сильнейших землетрясений мира (рис. 11) дала отрицательный результат — такая

связь для наблюдений в асейсмическом регионе отсутствует (таблица).

Дальнейшая интерпретация полученных результатов нуждается в привлечении информации о других геофизических полях в рассмотренном регионе. Тем не менее, кое-какие общие соображения о наличии нескольких состояний временного ряда длительных наблюдений могут быть приведены.

В асейсмических регионах, к которым относится территория Москвы, разрядка потока

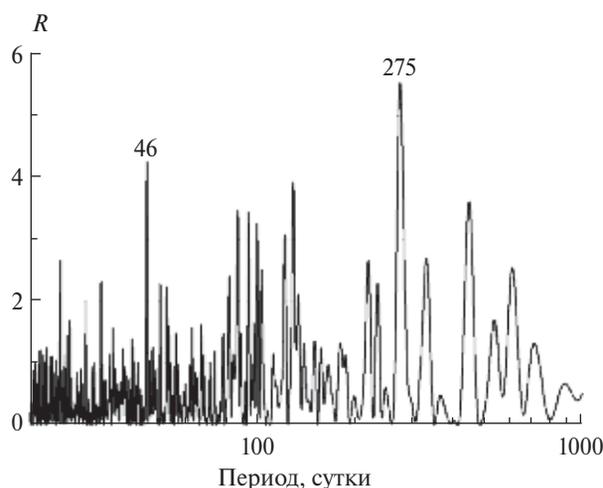
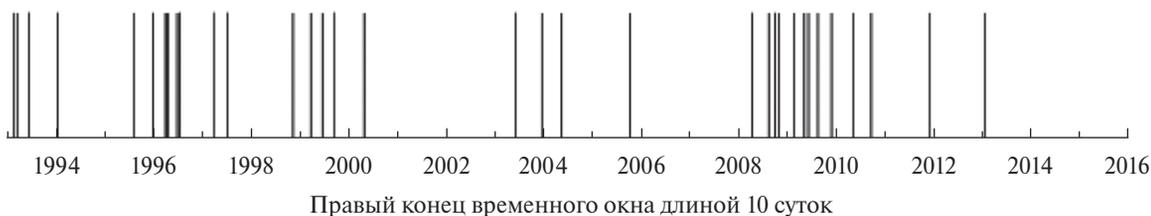


Рис. 10. График максимума разности логарифмических функций правдоподобия в зависимости от периода для последовательности времен переходов между 5 кластерами (рис. 9б) в интервале времени 22.06.1995–10.09.2010 гг. (217 событий) для пробных периодов от 20 до 1000 суток. Отмечены 2 периода со значимостью 98% со значениями 46 и 275 суток.

(a)

Последовательность временных интервалов переходов в 4-й кластер



(б)

Последовательность сильнейших землетрясений, $M \geq 8$, 1993–2015 гг.

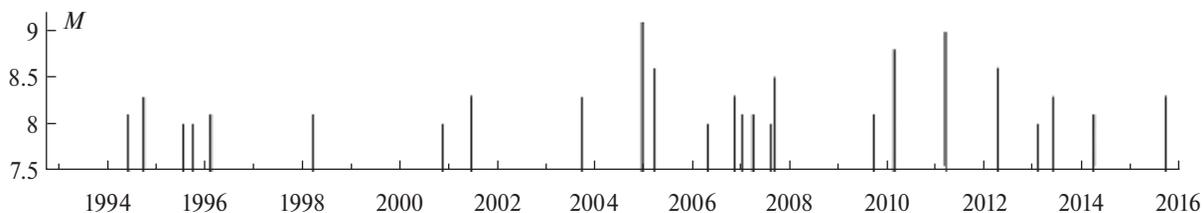


Рис. 11. (а) — Последовательность временных интервалов длиной 10 суток, когда происходит переход в 4-й «аномальный» кластер; (б) — последовательность сильнейших землетрясений магнитудой не менее 8.

энергии из земных недр носит плавный характер и имеет вид не резкого возбуждения сейсмических волн, а увеличения «медленных» движений земной коры, происходящих вдоль сравнительно узких зон локализации деформаций — линеаментов. Интенсификация подобного рода движений земной коры может приводить к усилению миграции подземных вод, обладающих способностью разупрочнять блоки земной коры, усиливать карстово-суффозионные процессы. В свою очередь, последнее приводит к увеличению вероятности оползней, растрескивания фундаментов больших зданий, прорыва подземных тоннелей, коррозии путей метрополитена. Таким образом, выделенная последовательность временных фрагментов с отличающимися статистическими характеристиками может быть использована для корреляции временных интервалов из различных кластеров с интенсивностью техногенных происшествий на территории мегаполиса.

Работа выполнена при поддержке Российского фонда фундаментальных исследований, грант № 18-05-00133.

СПИСОК ЛИТЕРАТУРЫ

- Айвазян С.А., Бухштабер В.М., Енюков И.С., Мешалкин Л.Д.* Прикладная статистика. Классификация и снижение размерности. М.: Финансы и статистика. 1989.
- Багмет А.Л., Багмет М.И., Барабанов и др.* Исследование земноприливных колебаний уровня подземных вод на скважине «Обнинск» // Изв. АН СССР Сер. Физика Земли. 1989. № 11. С. 84–95.
- Беседина А.Н., Виноградов Е.А., Горбунова Э.М., Кабыченко Н.В., Свинцов И.С., Пигулевский П.И., Свистун В.К., Щербина С.В.* Отклик флюидонасыщенных коллекторов на лунно-солнечные приливы. Часть 1. Фоновые параметры приливных компонент в смещении грунта и уровне подземных вод // Физика Земли. 2015. № 1. С. 73–82
- Болдина С.В., Копылова Г.Н.* Эффекты Жупановского землетрясения 30 января 2016 г., $M_w = 7.2$, в изменениях уровня воды в скважинах ЮЗ-5 и Е-1, Камчатка // Геодинамика и тектонофизика. 2017. Т. 8. № 4. С. 863–880.
- Виноградов Е.А., Горбунова Э.М., Кабыченко Н.В., Кочарян Г. Г, Павлов Д.В., Свинцов И.С.* Мониторинг уровня подземных вод по данным прецизионных измерений // Геоэкология. Инженерная геология. Гидрогеология. Геокриология. 2011. № 5. С. 439–449.
- Гидрогеология СССР. М.: Недра. 1966. Т. 1. С. 423
- Киссин И.Г., Гумен А.М.* Гидрогеологические индикаторы современных движений земной коры в асейсмическом регионе // Докл. РАН. 1994. Т. 334. № 6. С. 768–772.
- 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.
- Копылова Г.Н.* Изменения уровня воды в скважине ЮЗ-5, Камчатка, вызванные землетрясениями // Вулканология и сейсмология. 2006. № 6. С. 52–64.
- Копылова Г.Н., Любушин А.А., Малугин В.А., Смирнов А.А., Таранова Л.Н.* Гидродинамические наблюдения на Петропавловском полигоне, Камчатка // Вулканология и сейсмология. 2000. № 4. С. 69–79.
- Копылова Г. Н, Горбунова Э.М., Болдина С.В., Павлов Д.В.* Оценка деформационных свойств системы «пласт-скважина» на основе анализа барометрического и приливного откликов уровня воды в скважине // Физика Земли. 2009. № 10. С. 69–78.
- Любушин А.А.* Анализ данных систем геофизического и экологического мониторинга. М.: Наука. 2007. 228 с.
- Любушин А.А.* Тренды и ритмы синхронизации мультифрактальных параметров поля низкочастотных микросейсм // Физика Земли. 2009. № 5. С. 15–28.
- Любушин А.А.* Статистики временных фрагментов низкочастотных микросейсм: их тренды и синхронизация // Физика Земли. 2010. № 6. С. 86–96.
- Любушин А.А.* Прогноз Великого Японского землетрясения // Природа. 2012. № 8. С. 23–33.
- Любушин А.А.* Прогностические свойства случайных флуктуаций геофизических характеристик // Биосфера. 2014. № 4. С. 319–338.
- Любушин А.А., Писаренко В.Ф.* Исследование сейсмического режима с помощью линейной модели интенсивности взаимодействующих точечных процессов // Физика Земли. 1993. № 12. С. 81–87.
- Любушин А.А., Малугин В.А.* Статистический анализ отклика уровня подземных вод на вариации атмосферного давления // Физика Земли. 1993. № 12. С. 74–80.
- Любушин А.А., Лежнев М.Ю.* Анализ изменчивости функции отклика уровня подземных вод на баровариации на Южных Курилах (о. Шикотан) // Физика Земли. 1995. № 8. С. 79–84.
- Любушин А.А., Малугин В.А., Казанцева О.С.* Мониторинг приливных вариаций уровня подземных вод в группе водоносных горизонтов // Физика Земли. 1997. № 4. С. 52–64.
- Любушин А.А., Писаренко В.Ф., Ружич В.В., Буддо В.Ю.* Выделение периодичностей в сейсмическом режиме // Вулканология и сейсмология. 1998. № 1. С. 62–76.
- Любушин А.А., Малугин В.А., Казанцева О.С.* Выделение «медленных событий» в асейсмическом регионе // Физика Земли. 1999. № 3. С. 35–44.
- Моисеев Н.Н., Иванюков Ю.П., Столярова Е.М.* Методы оптимизации. М.: Наука. 1978. 352 с.
- Anderson T.W., Rubin H.* Statistical inference in factor analysis, Proc. 3rd Berkley Symp. on Math. Statistics and Probability. 1956. V. 5. P. 111–150.

- Box G.E.P., Jenkins G.M.* Time series analysis. Forecasting and control. Holden-Day. San Francisco, Cambridge, London, Amsterdam. 1970. 406 с. (Русский перевод: *Бокс Дж., Дженкинс Г.* Анализ временных рядов. Прогноз и управление. 1970. М.: Мир. (В 2-х выпусках). 1974. 197 с.).
- Bredehoft J.D.* Response of a Well-Aquifer System to Earth Tides // *J. Geophys. Res.* 1967. V. 12. P. 3075–3087.
- Brillinger D.R.* Time series. Data analysis and theory. Holt, Rinehart Winston, Inc., N.Y., Chicago, San Francisco. 1975. (Русский перевод: *Бриллинджер Д.* (1980) Временные ряды. Обработка данных и теория. М.: Мир. 536 с.)
- Cox D.R., Lewis P.A.W.* The statistical analysis of series of events. London, Methuen. 1966. (Русский перевод: *Кокс Д., П. Льюис* Статистический анализ последовательностей событий. М.: Мир. 1969. 312 с.)
- Cramer H.* Mathematical Methods of Statistics. 1999. Princeton University Press.
- Donoho D.L. and Johnstone I.M.* Adapting to Unknown Smoothness via Wavelet Shrinkage // *Journal of the American Statistical Association.* 1995. V. 90. Iss. 432. P. 1200–1224.
- Duda R.O., P.E. Hart* Pattern classification and scene analysis, John Wiley & Sons. N.Y., London, Sydney, Toronto (1973) (Русский перевод: *Дуда Р., Харп П.* Распознавание образов и анализ сцен. М.: Мир. 1976. 511 с.)
- Feder J.* Fractals. 1988. Plenum Press, New York, London (Русский перевод: *Федер Е.* Фракталы. М.: Мир. 1991. 254 с.)
- Harman, H.H.* Modern factor analysis. Chicago: University of Chicago Press. 1967. (Русский перевод: *Харман Г.* Современный факторный анализ. М.: Финансы и статистика. 1972. 420 с.).
- Huber P.J.* Robust statistics. John Wiley and Sons. New York; Chichester; Brisbane; Toronto. 1981 (Рус. пер.: *Хьюбер П.* Робастность в статистике. М.: Мир. 1984. 303 с.).
- Igarashi G., Wakita H.* Tidal response and earthquake-related changes in the water level of deep wells // *J. Geophys. Res.* 1991. V. 96. P. 4269–4278.
- Kantelhardt J.W., Zschiegner S.A., Koncsienly-Bunde E., Havlin S., Bunde A., and Stanley H.E.* Multifractal detrended fluctuation analysis of nonstationary time series // *Physica A.* 2002. V. 316. P. 87–114.
- Kashyap R.L., Rao A.R.* Dynamic stochastic models from empirical data. 1976. New York; San Francisco; London: Acad. Press (Русский перевод: *Кашьяп Р.Л., Рао А.Р.* Построение динамических стохастических моделей по экспериментальным данным. М.: Наука. 1983. 384 с.)
- Lawley D.N., Maxwell A.E.* Factor analysis as a statistical method. New York: American Elsevier. 1971. (Русский перевод 1-го издания: *Лоули Д., Максвелл А.* Факторный анализ как статистический метод. М.: Мир. 1967).
- Lyubushin A.* Multifractal Parameters of Low-Frequency Microseisms/V. de Rubeis et al. (eds.), Synchronization and Triggering: from Fracture to Earthquake Processes, GeoPlanet: Earth and Planetary Sciences 1, Springer-Verlag Berlin Heidelberg. 2010. Chapter 15. P. 253–272.
- Lyubushin A.* Prognostic properties of low-frequency seismic noise // *Natural Science.* 2012. V. 4. Iss. 8 A. P. 659–666.
- Lyubushin, A.* How soon would the next mega-earthquake occur in Japan? // *Natural Science.* 2013. V. 5. Iss. (8 A1). P. 1–7.
- Lyubushin A.A.* Dynamic estimate of seismic danger based on multifractal properties of low-frequency seismic noise // *Natural Hazards.* 2014. V. 70. Iss. 1. P. 471–483.
- Lyubushin A.* Synchronization of Geophysical Fields Fluctuations / Tamaz Chelidze, Luciano Telesca, Filippos Vallianatos (eds.), Complexity of Seismic Time Series: Measurement and Applications, Elsevier 2018, Amsterdam, Oxford, Cambridge. Chapter 6. P. 161–197. DOI: <https://doi.org/10.1016/B978-0-12-813138-1.00006-7>.
- Mallat S.* A Wavelet Tour of Signal Processing. Second edition. Academic Press. San Diego, London, Boston, New York, Sydney, Tokyo, Toronto. 1999. (Русский перевод: *Малла С.* Вейвлеты в обработке сигналов. М.: Мир. 2005).
- Ogata Y., Katsura K.* Analysis of temporal and spatial heterogeneity of magnitude frequency distribution inferred from earthquake catalogues // *Geophys. J. Int.*, 1993. V. 113. P. 727–738.
- Rojstaczer S., Agnew D.C.* The influence of formation material properties on the response of water levels in wells to Earth tides and atmospheric loading // *J. Geophys. Res.* 1989. V. 94. P. 12403–12411.
- Roeloffs E.* Hydrologic precursors to earthquakes: A review // *PAGEOPH.* 1988. V. 126. № 2/4. P. 177–209.
- Roeloffs E., Burford S., Riley F.* Records A. Hydrologic effects on water level changes associated with episodic fault creep near Parkfield, California // *J. Geophys. Res.* 1989. V. 94. № B9. P. 12387–12402.
- Vogel M.A. and Wong, A.K.C.* (1979) PFS Clustering method // *IEEE Transactions on Pattern Analysis and Machine Intelligence.* № 1. P. 237–245. doi:10.1109/TPAMI.1979.4766919

Analysis of Long-Term Observations of the Groundwater Level in an Aseismic Region

A. A. Lyubushin^{a,*}, O. S. Kazantseva^{a,**}, and A. B. Manukin^{a,***}

^a*Schmidt Institute of Physics of the Earth, Russian Academy of Sciences, Moscow, 123242 Russia*

^{*}*E-mail: lyubushin@yandex.ru*

^{**}*E-mail: o.kazantseva@yandex.ru*

^{***}*E-mail: amanukin@yandex.ru*

Received February 27, 2018

The results of the analysis of continuous precise time series of atmospheric pressure and groundwater level fluctuations in a well drilled to a depth of 400 m in the territory of Moscow are presented. The observations are remarkable in terms of their duration of more than 22 years (from February 2, 1993 to April 4, 2015) and by the sampling interval of 10 min. These long observations are suitable for exploring the stationarity of the properties of hydrogeological time series in a seismically quiet region, which is important from the methodological standpoint for interpreting the similar observations in seismically active regions aimed at earthquake prediction. Factor and cluster analysis applied to the sequence of multivariate vectors of the statistical properties of groundwater level time series in the successive 10-day windows after adaptive compensation for atmospheric pressure effects distinguish five different statistically significant states of the time series with the transitions between them. An attempt to geophysically interpret the revealed states is made. Two significant periods – 46 and 275 days – are established by spectral analysis of the sequence of the transitions times between the clusters.

Keywords: groundwater level, time series, transfer functions, entropy, multifractals, predictability index, factor analysis, cluster analysis