

УДК 551.461

ИСПОЛЬЗОВАНИЕ ЭКСПОНЕНЦИАЛЬНОЙ РЕГРЕССИИ ДЛЯ МОДЕЛИРОВАНИЯ КОЛЕБАНИЙ УРОВНЯ КАСПИЙСКОГО МОРЯ

© 2012 г. И.А. Кожевникова*, В.И. Швейкина**

* *Московский государственный университет*

** *Институт водных проблем Российской академии наук*

В статье исследованы данные по колебаниям уровня Каспийского моря, дополненные по апрель 2010 г. Статистический анализ этих наблюдений, и в частности эмпирическая гистограмма, дает основание считать, что в колебаниях уровня моря появилось третье устойчивое состояние на отметке –27 м абс. В предыдущих работах с участием авторов в колебаниях уровня моря до 1995 г. были установлены два устойчивых и одно неустойчивое состояния. Дополненные данные показали, что ранее разработанные модели, основанные на полиномиальной регрессии, моделируют колебания уровня с переходами между двумя определенными состояниями. В статье предложена усовершенствованная модель, основанная на экспоненциальной регрессии, решения которой демонстрируют переходы уровня моря между тремя устойчивыми состояниями.

Введение. Мы живем во время, когда во многие области науки (физика элементарных частиц, астрофизика, биология, химия, медицина, климатология, метеорология и многие др.) прочно вошли понятия физики неустойчивых систем: неустойчивость и множественность режимов, самоорганизация и детерминированный хаос, автоколебания и странный аттрактор. Динамическая неустойчивость присуща и колебаниям уровня Каспийского моря. По утверждению И.И. Пригожина [9], одной из существенных особенностей сложного поведения системы – это способность осуществлять переходы между различными режимами. В полной мере это относится и к уровню Каспийского моря. За время существования моря колебания его уровня происходили в очень широких пределах: от –140 до 25 м абс. За последние 700 лет уровень колебался в интервале от минус 30 до минус 25 м абс. [2]. Установлено, что при существенно нелинейной зависимости площади водоема от уровня возможно существование нескольких состояний равновесия водоема с окружающей средой [7]. Не только среднемесячные наблюдения по в/п г. Махачкалы (1900–2010 гг.), которые исследованы в статье (рис. 1), но и среднегодовые по г. Баку (1837–2005 гг.) [8] обнаруживают третье равновесное состояние уровня моря, около которого уровень находится последние 20 лет. На рис. 2 приведены гистограммы данных обоих пунктов наблюдения.

В работе авторов [6] с помощью вычисленных хаотических характеристик показано, что колебания уровня Каспийского моря имеют характер детерминированного хаоса. Колебания подобного типа могут непредсказуемо бурно реагировать на незначительные изменения в окружающей среде, из чего следует ограниченная предсказуемость их состояния на всех уровнях наблюдения. Согласно современному мировоззрению, прогнозы уровня Каспийского моря могут быть только вероятностными, для расчета которых необходима исключительно нелинейная модель. В статье предложена усовершенствованная стохастическая модель, решения которой демонстрируют переходы уровня моря между тремя равновесными состояниями. Настоящее время характеризуется небольшими флуктуациями уровня около отметки –27 м абс. Это состояние уровня моря соответствует определенной конфигурации поверхности моря, определенной площади побережья, свободного от воды. Все это – важная информация для народнохозяйственной деятельности. Однако необходимо быть готовым к быстрым изменениям сегодняшнего состояния.

Модель колебаний уровня моря можно получить разными способами. Первый способ, разработанный В.И. Найденовым совместно с одним из авторов [5], основан на полиномиальной регрессии приращений уровня как функции уровня. Было установлено, что полиномиальная регрес-

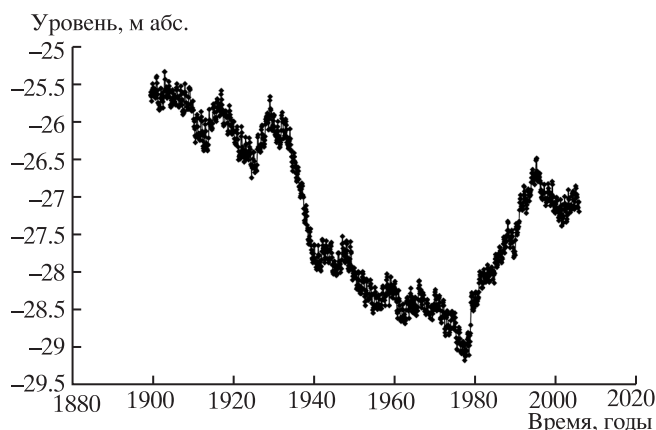


Рис. 1. Колебания уровня Каспийского моря по данным наблюдений в/п г. Махачкала.

сия 5-го порядка как оценка коэффициента сноса стохастического дифференциального уравнения (СДУ), на основе которой строится модель колебаний уровня (среднемесячных данных наблюдений уровня (в/п г. Махачкала, 1900–2010 гг.), фиксирует только два устойчивых уровня. Отчетливо видный на рисунке колебаний уровня моря (рис. 1) 3-й, вновь появившийся, уровень полиномиальная регрессия 5-го порядка игнорирует. Естественно, стационарное решение уравнения Фоккера-Планка-Колмогорова (ФПК), соответствующее этому СДУ, имеет плотность распределения вероятности с двумя модами, что получается в соответствии с приближающей регрессией. Таким образом, полиномиальная регрессия как самое простое приближение не реагирует на наблюдения последних лет, на основании которых можно предполагать изменения в колебаниях моря.

В связи с этим возникает задача построить регрессию, фиксирующую все три устойчивых уровня в соответствии с рис. 1. Авторами разработан новый метод, основанный на аппроксимации коэффициента сноса СДУ приращений уровня экспоненциальными функциями, который дает четкое выделение трех устойчивых состояний. Установлено также, что соответствующая модель обладает меньшей остаточной суммой квадратов по сравнению с полиномиальной регрессией. В статье излагаются математический аппарат, лежащий в основе данного метода, и его применение к анализу колебаний уровня моря.

Постановка задачи. Для построения непрерывной модели в виде диффузионного процесса, используется СДУ:

$$dX_t = \Phi(X)dt + \bar{\sigma}dW_t, \quad (1)$$

где $\Phi(X)$ – функциональная зависимость, достаточно хорошо фиксирующая связь приращений уровня от уровня (коэффициент сноса), $\bar{\sigma} > 0$ – коэффициент диффузии, $\Phi(X)$ – стандартный винеровский процесс. В фазовом пространстве данные по колебаниям моря могут быть представлены в виде точек (регрессионное облако), координатами которых являются уровень моря по оси ОХ и приращение уровня по оси ОУ. Эти точки, соединенные линией, представляют собой фазовый портрет (рис. 3) и описывают траекторию, которая связывает функциональной зависимостью скорость изменения уровня с его величиной. Точки скопления на фазовом портрете показывают наиболее часто встречающиеся ситуации в колебаниях уровня. Почти такую же информацию можно получить из оценки плотности распределения вероятностей, гистограммы, при построе-

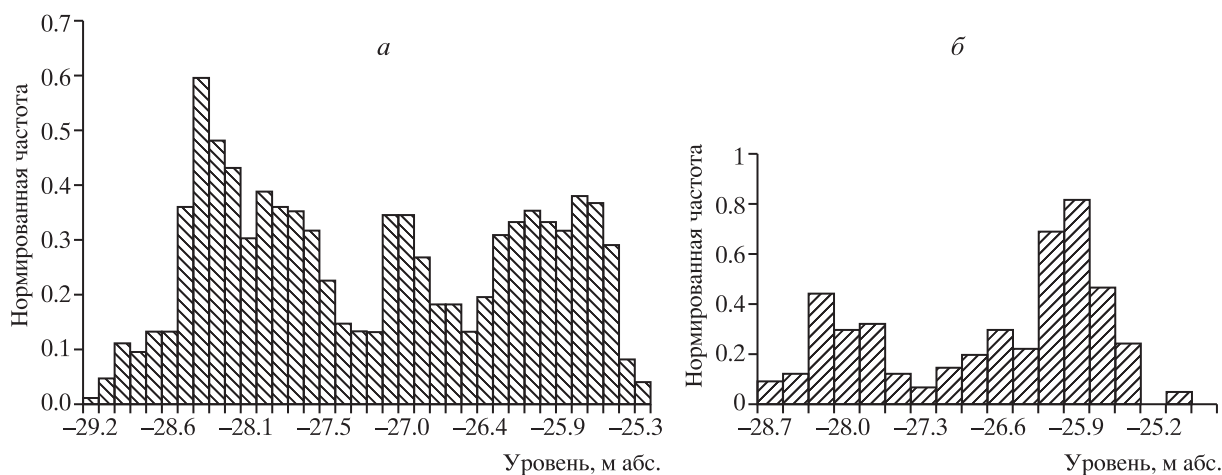


Рис. 2. Оценка плотности распределения вероятностей, построенная по данным колебаний уровня Каспийского моря: среднемесячным (а) и по среднегодовым (б).

нии которой используются группированные данные. Именно из-за группировки, которая зависит от квалификации и задач исполнителя, часть информации теряется. Тем не менее как на фазовом портрете, так и на гистограмме исходных данных четко выделяются три устойчивых состояния.

Известно [1], что зависимость $\Phi(X)$ является производной от потенциала с обратным знаком:

$$\Phi(X) = -dU/dt.$$

Следовательно, при известном потенциале – известна зависимость $\Phi(X)$. Потенциал связан со стационарной плотностью распределения $p(x)$ [1], которая получается в результате решения уравнения ФПК, соответствующего диффузионному процессу (1):

$$\frac{\partial p(y, t|x)}{\partial t} = - \frac{\partial[\Phi(y)p(y, t|x)]}{\partial y} + \frac{\bar{\sigma}^2}{2} \frac{\partial^2(p(y, t|x))}{\partial y^2}.$$

В этом уравнении $\Phi(X)$ – коэффициент сноса дифференциального уравнения (1), $\frac{\bar{\sigma}^2}{2}$ – коэффициент диффузии, который мы принимаем равным константе. Стационарная плотность вероятностей уровня при $\partial p/\partial t = 0$ выражается следующим образом:

$$p(x) = \frac{C}{\bar{\sigma}^2} \exp \left[\frac{2}{\bar{\sigma}^2} \int_{-\infty}^x \Phi(u) du \right], \quad (2)$$

$$X_{\min} \leq x \leq X_{\max},$$

где C – нормирующий множитель.

Как правило, стационарная плотность значительно отличается от оценки плотности вероятностей, в качестве которой рассматривается гистограмма. Приблизим исходную оценку плотности вероятностей подходящей функцией, и эту функцию будем считать стационарной плотностью вероятностей. В результате можно будет рассчитать потенциал, который сразу приводит к регрессионной зависимости приращений уровня от уровня. Таким образом, будет получена стохастическая модель колебаний уровня Каспийского моря в виде диффузионного процесса вида (1).

Задача статистического разделения смесей вероятностных распределений. В качестве функции, приближающей оценку плотности вероятностей, рассмотрим смесь гауссовских распределений. Анализируя исходный временной ряд, выделим три периода, когда колебания уровня происходят около одного из устойчивых состояний. Первый период, включающий 37 лет

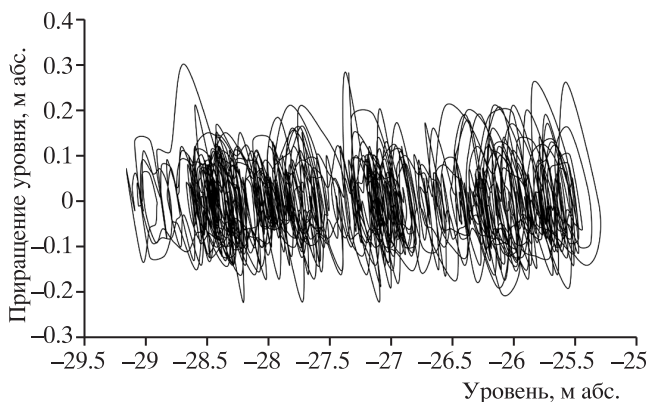


Рис. 3. Фазовый портрет исходных данных по среднемесячным колебаниям уровня Каспийского моря.

наблюдений, приходится на начало ряда, когда море тяготело к уровню -26 м абс., второй период включает резкое падение уровня моря и его стабилизацию около -28 м абс., и третий, содержащий наблюдения с 1986 г., характеризуется колебанием уровня около -27 м абс. Выбор гауссовских распределений основан на проверке гипотезы о нормальном распределении всех трех выборок, соответствующих выделенным трем периодам.

Наиболее универсальным критерием, применяемым во многих приложениях для проверки гипотезы о распределении, является критерий хи-квадрат. Этот критерий использует только группированные данные, что ведет к потере информации, поэтому он может быть рекомендован только для выборок больших длин и только для тех распределений, для которых нет специализированных критериев.

В 2002 г. был введен отечественный стандарт ГОСТ Р ИСО 5479-2002 “Статистические методы. Проверка отклонения распределения вероятностей от нормального распределения”. Подчеркивается, что критерии согласия при малых объемах выборок проигрывают по мощности специальным критериям проверки на отклонение от нормальности. В стандарте рассматривается графический метод проверки на нормальность с использованием вероятностной бумаги, являющийся прикидочным методом, критерии проверки на симметричность и на значение эксцесса, а также критерии Шапиро-Уилка [4], основанные на регрессионном анализе порядковых статистик и предназначенные для проверки гипотезы о нормальном (гауссовском) распределении. Критерии Шапиро-Уилка очень сложны в употреблении и предназначены для анализа выборок малых длин, не более 50. Существует также приближенный критерий Шапиро-Уилка, он рекомендуется для

Таблица 1. Значения критических величин по Д’Агустино для выборок колебаний уровня Каспийского моря

Среднемесячные наблюдения в/п г. Махачкала (1900–2010)				
Годы	Длина выборки	Уровень Каспия	Значение критерия	Уровень значимости критерия
1900–1930	361	верхний	2.290	0.0001
1940–1984	540	нижний	1.006	0.05
1990–2010	182	средний	-1.505	0.05
Среднегодовые наблюдения в/п г. Баку (1837–2005)				
1837–1930	94	верхний	-0.516	0.05
1940–1984	45	нижний	-1.592	0.05
1990–2007	18	средний	-1.997	0.05

анализа выборок, длина которых не превышает 100. При применении этих двух вариантов критериев Шапиро-Уилка для выборок длины больше 100 эти критерии дают неверные результаты, причем чем длиннее выборка, тем хуже и фантастичнее результат. Этот факт легко устанавливается с помощью широко известного пакета STATISTICA, не содержащего ограничителя длины выборки. Критерий Д’Агустино [3] является специализированным для проверки гипотезы нормальности на основе выборки и применяется для выборок больших и малых длин. Этот критерий помимо своего основного назначения позволяет определить направление отклонений асимметрии и эксцесса от нормального, что очень существенно при отклонении гипотезы от нормальности. Именно критерий Д’Агустино рекомендуется отечественным стандартом для проверки гипотезы нормальности.

В [3] приведены таблицы, дающие критические значения для различных уровней значимости в случае выборок больших и малых длин. В табл. 1 приведены вычисленные значения критерия для трех выборок, соответствующих верхнему, среднему и нижнему уровням. Все три выборки для среднемесячных значений уровня моря имеют большие длины. Для подтверждения значимости полученных результатов были вычислены подобные характеристики и приведены в той же таблице для ряда среднегодовых уровней Каспийского моря [8], для которого рассматриваемые выборки имеют малые длины. И в том, и в другом случаях принимается гипотеза о нормальности распределения выборок. Таким образом, задача аппроксимации истинного распределения сведена к статистическому определению смеси гауссовских распределений, которое имеет неизвестные параметры.

Если компоненты смешанного распределения имеют гауссовские распределения, то теоретиче-

ские моменты смеси выражаются через первый и второй моменты каждой компоненты следующим образом:

$$f(x) = p_1 f_1(x; \mu_1, \sigma_1^2) + p_2 f_2(x; \mu_2, \sigma_2^2) + (1 - p_1 - p_2) f_3(x; \mu_3, \sigma_3^2), \quad -\infty < x < \infty,$$

где $f_i(x; \mu_i, \sigma_i^2)$ – плотность i -компоненты смешанного распределения; p_i, μ_i, σ_i^2 – соответственно вес, математическое ожидание и дисперсия i -компоненты.

Чтобы оценить параметры каждой компоненты и веса компонент, используем модифицированный метод моментов [4]. В случае смеси трех гауссовских распределений необходимо вычислить моменты смешанного распределения вплоть до восьмого порядка, так как имеется восемь неизвестных параметров. Приведем формулы для моментов гауссовского распределения $N(\mu, \sigma^2)$ [11]:

$$m_1 = \mu, \\ m_j = \mu^j + \sum_{k=1}^{\lfloor \frac{j}{2} \rfloor} C_j^{2k} \sigma^{2k} \mu^{j-2k} (2k-1)!!, \quad j \geq 2,$$

где j – порядок момента гауссовского распределения, $(2k-1)!!$ – произведение нечетных чисел от единицы до $(2k-1)$ включительно. Тогда формулы для вычисления моментов смешанного распределения с компонентами имеют вид

$$m_{q,1} = \sum_{i=1}^q p_i \mu_i, \quad q = 3, \\ m_j = \mu^j + \sum_{k=1}^{\lfloor \frac{j}{2} \rfloor} C_j^{2k} \sigma^{2k} \mu^{j-2k} (2k-1)!!, \quad j \geq 2, \\ \sum_{i=1}^q p_i = 1.$$

Начальные моменты в левой части уравнений оцениваем по выборке, применяя формулы

$$C_i = \frac{1}{n} \sum_{k=1}^n x_k^i,$$

i – номер выборочного момента, n – длина выборки.

Все вычисления для получения оценок параметров смешанного распределения получены на основании минимума функции

$$Q = \sum_{j=1}^8 \left(\frac{m_j - C_j}{C_j} \right)^2.$$

С помощью итерационных методов решения системы алгебраических уравнений смешанного распределения вероятностей вычислены значения оценок параметров, которые приведены в табл. 2. Полученные значения оценок параметров позволяют построить смешанное распределение. Таким образом, в предположении, что стационарная плотность распределения вероятностей уровня имеет три максимума, предложена ее форма в виде суммы трех гауссовых распределений и найдены оценки ее параметров. Принятие предположения о гауссовости распределений, входящих в смесь, опирается на применение критерия Д'Агустино. Далее решается система из восьми уравнений, для решения которой используются выборочные оценки моментов смешанного распределения. Ясно, что точность получаемых оценок параметров распределения вероятностей уровня Каспия невелика, поскольку выборочные оценки даже второго момента дают достаточно широкие доверительные интервалы, включающие в себя истинное значение параметров распределения. Тем не менее использование вычисленных оценок параметров приводит к неплохому согласованию эмпирической оценки плотности распределения с теоретической, построенной на основании смеси трех гауссовских распределений (рис. 4). Полученная таким образом стационарная плотность имеет три максимума, которые связаны с равновесными состояниями уровня.

Стохастическая модель колебаний уровня Каспийского моря. Рассмотрим формулу для стационарной плотности вероятностей (2) и подставим в нее выражение $\Phi(X) = -dU/dt$. Считая, что $p_s(x) = f(x)$, получим связь между смешанным распределением и вероятностным потенциалом:

$$f(x) = p(x) = \frac{C}{\bar{\sigma}^2} \exp\left(-\frac{2}{\bar{\sigma}^2} U(x)\right).$$

Логарифмируя последнее выражение, получаем

$$\frac{2}{\bar{\sigma}^2} U(x) = \ln \frac{C}{\bar{\sigma}^2} - \ln f(x).$$

Таблица. 2. Оценки параметров q компонент смешанного распределения

	$q = 1$	$q = 2$	$q = 3$
μ_q	$\mu_1 = -25.86$	$\mu_2 = -28.27$	$\mu_3 = -26.97$
σ_q	$\sigma_1 = 0.548$	$\sigma_2 = 0.439$	$\sigma_3 = 0.192$
p_q	$p_1 = 0.39$	$p_2 = 0.48$	$p_3 = 0.13$

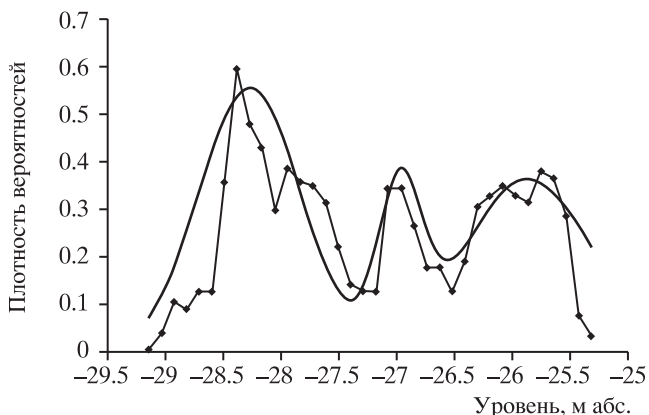


Рис. 4. Плотность смешанного распределения (сплошной линией обозначены данные эмпирической гистограммы, маркером – смесь трех нормальных распределений).

Теперь коэффициент сноса или регрессионная зависимость между приращением уровня моря и его уровнем имеет вид

$$\Phi(x) = -\frac{dU}{dx} = \frac{\bar{\sigma}^2}{2} \frac{f'(x)}{f(x)}.$$

Зависимость $\Phi(X)$ показана на рис. 5. Используя эту зависимость, получаем модель колебаний уровня Каспийского моря, выраженную следующим стохастическим уравнением:

$$dX_i = \frac{\bar{\sigma}^2}{2} \frac{f'(x)}{f(x)} dt + \bar{\sigma} dW_i,$$

где

$$f(X) = \frac{1}{\sqrt{2\pi}} \sum_{k=1}^3 \frac{p_k}{\bar{\sigma}^2} \exp\left(-\frac{(X - \mu_k)^2}{2\bar{\sigma}_k^2}\right),$$

$$f'(X) = -\frac{1}{\sqrt{2\pi}} \sum_{k=1}^3 \frac{p_k(X - \mu_k)}{\bar{\sigma}_k^3} \exp\left(-\frac{(X - \mu_k)^2}{2\bar{\sigma}_k^2}\right).$$

В ранее предложенной авторами модели [10] коэффициент сноса выражался полиномом 5-й степени, теперь он выражен с помощью экспоненциальной регрессии (рис. 5) и достаточно хорошо описывает регрессионное облако. Характерная реализация этой модели (рис. 6) демонстрирует переходы уровня от одного равновесного

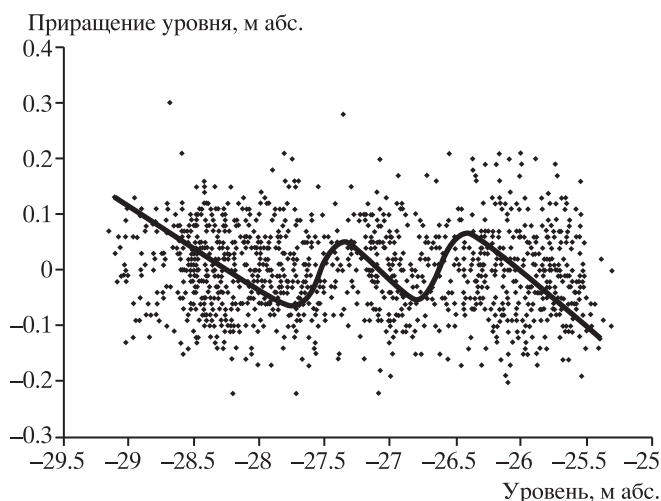


Рис. 5. Экспоненциальная регрессия, аппроксимирующая исходные данные и использованная для построения модели колебаний уровня Каспийского моря.

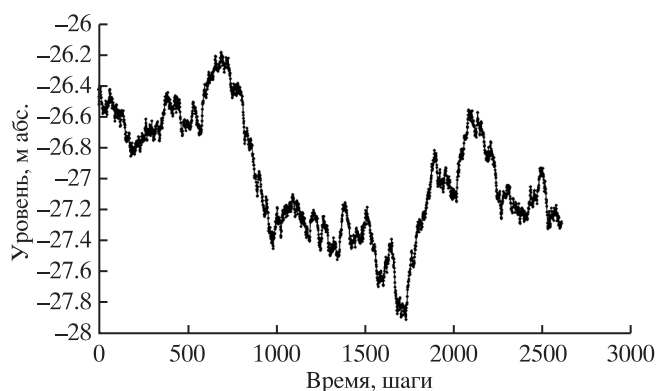


Рис. 6. Смоделированная с помощью экспоненциальной регрессии характерная траектория колебаний уровня Каспийского моря.

состояния к другому, качественно похожие на то, что наблюдается в действительности.

Для доказательства адекватности полученной модели исходным данным была рассчитана остаточная сумма квадратов отклонений экспоненциальной регрессии от приращений наблюдаемых данных. Получена величина 9.87, и, следовательно, остаточная дисперсия будет равна $9.87/(1272-6) \cong 9.81 \cdot 10^{-3}$. Остаточная дисперсия регрессионной модели, использующей в качестве коэффициента сноса полиномиальную регрессию 5-го порядка, получается равной 11.81. Таким образом, переход к экспоненциальной регрессии хотя и незначительно улучшает количественный показатель адекватности, но вновь полученная модель является более точной, так как реагирует на изменения в колебаниях уровня моря.

Заключение. В статье четко изложено, как, используя смешанное распределение, получить стохастическую модель, моделирующую сложный природный процесс с неустойчивыми равновесными состояниями. В работе для построения смешанного распределения использованы нормальные распределения, причем для проверки гипотезы о нормальном распределении применен специализированный критерий. Смешанное распределение может быть и многомодальным, и одномодальным. Его составляющими могут быть самые разные распределения: гамма, Парето, смешанное, равномерное и т.д. Во всех случаях, как только есть смешанное распределение, можно получить стохастическую модель, адекватно описывающую реальный процесс.

Подобная модель, основанная на полиномиальной регрессии, была разработана ранее в исследованиях с участием авторов. Казалось бы, если полиномиальная регрессия не подошла по каким-либо причинам, можно было бы применить экспоненциальную регрессию без привлечения уравнения ФПК, как это было выполнено в работе [12]. Применение этого метода требует дополнительного исследования. Основной вопрос и цель настоящей работы заключается в том, чтобы получить переходы с одного устойчивого состояния в другое с заданной стационарной плотностью распределения вероятностей. По мнению авторов, это проще всего выполнить на основании уравнения ФПК. Предлагаемая модель, основанная на стационарной плотности вероятностей, показывает такие переходы. На протяжении последних 180 лет Каспий вел себя именно так: долгое стояние около равновесных состояний и быстрые переходы от одного состояния к другому. Именно это и демонстрирует полученная в предлагаемой статье модель. Ее характерная реализация подчеркивает адекватность модели исходным данным. Кроме этого получена и численная оценка адекватности модели.

Из вида модели следует, что для колебаний уровня Каспийского моря характерна не только внешняя непредсказуемость, создаваемая климатическими изменениями, но и внутренняя, обусловленная неустойчивой динамикой водного баланса.

СПИСОК ЛИТЕРАТУРЫ

1. *Гардинер К.В.* Стохастические методы в естественных науках. М.: Мир, 1986. 526 с.
2. *Гидрометеорология и гидрохимия морей. Каспийское море.* СПб.: Гидрометеоздат, 1992. Т. VI. 538 с.

3. Кобзарь А.И. Прикладная математическая статистика. М.: Физматлит, 2006. 813 с.
4. Кожевникова И.А. Исследование статистических свойств оценок параметров смеси распределений с примерами из гидрологии // Обозрение прикладной и промышленной математики. М.: Научное изд-во ТВП, 1997. Т. 4. Вып. 3. С.483–498.
5. Кожевникова И.А., Найденов В.И. Нелинейная стохастическая модель колебаний уровня Каспийского моря // Вод. ресурсы. 1998. Т. 25. № 6. С. 661–670.
6. Кожевникова И.А., Швейкина В.И. Нелинейная динамика колебаний уровня Каспийского моря // Вод. ресурсы. 2008. Т. 35. № 3. С. 313–320.
7. Найденов В.И. Нелинейная динамика поверхностных вод суши. М.: Наука. 2004. 317 с.
8. Панин Г.Н., Мамедов Р.М., Митрофанов И.В. Современное состояние Каспийского моря. М.: Наука, 2005. 356 с.
9. Пригожин И.И., Стенгерс И. Время, хаос, квант. К решению парадокса времени. М.: Эдиториал, 2000. 239 с.
10. Швейкина В.И., Кожевникова И.А. О вероятностях перехода уровня Каспия из одного состояния в другое на основе решения обратного уравнения Колмогорова // Вод. ресурсы. 2008. Т. 35. № 1. С. 45–52.
11. Ширяев А.Н. Вероятность. М.: Наука, 2004. 418 с.
12. Ozaki T. Non-linear threshold autoregressive models for non-linear random vibrations // J. Appl. Prob. 1981. V. 18. P. 443–451.

Use of Exponential Regression for Modeling of Caspian Sea Level Fluctuation

I.A. Kozhevnikova*, V.I. Shveikina**

* *Moscow State University*

** *Institute of Water Problems, RAS*

Data on Caspian Sea level fluctuation, added up to April 2010 is examined in the article. Statistic analyses of these observations and in particular empiric histogram permits to consider that in sea level fluctuation the third stable condition appears at the level of -27 m. In previous works in sea level fluctuations up to 1995 two stable and one unstable condition were fixed. Supplemented data shows that early developed models, based on polynomial regression are modeling level fluctuations with the transition of two defined conditions. A new improved model, based on exponential regression is proposed in the article, the solution of which demonstrates the transition of the sea level between three stable conditions