

**ПРОГНОЗИРОВАНИЕ ВРЕМЕННЫХ РЯДОВ  
В ГИДРОЛОГИЧЕСКИХ ЗАДАЧАХ  
НА ОСНОВЕ ВАРИАТИВНОГО МОДЕЛИРОВАНИЯ \***

**В. В. Губарев, О. К. Альсова**

*Новосибирский государственный технический университет, г. Новосибирск  
E-mail: gubarev@vt.cs.nstu.ru*

Рассматриваются основные результаты комплексного исследования возможных путей решения задачи прогнозирования годового притока реки Обь в створе Новосибирской ГЭС с использованием идей и методов вариативного моделирования. Построен набор вероятностных моделей, позволяющих осуществить прогнозирование притока с разной точностью и заблаговременностью (период между датой выдачи прогноза и началом прогнозируемого периода). Намечены пути дальнейших исследований.

**1. Постановка задачи.** Анализ и прогнозирование притока реки Обь в створе Новосибирской ГЭС – актуальная задача, решение которой послужит основой для разработки оптимальной методики управления уровнем воды в водохранилище и ее сброски [1]. Данные по притоку представляют собой гидрологические временные ряды (ВР) расхода притока ( $\text{м}^3/\text{с}$ ) реки Обь в створе Новосибирской ГЭС разной дискретности (декада, месяц, квартал, год) за 1894–2005 гг.

В данной работе представлены результаты долгосрочного прогнозирования притока (на год и более). На рис. 1 приведен исходный временной ряд среднегодового расхода притока реки Обь за 1894–2005 гг.

С точки зрения прогнозирования изменение притока во времени представляет собой многофакторный процесс, имеющий стохастическую природу. Поэтому для идентификации и прогнозирования речного притока обоснованным представляется применение вариативного моделирования, которое предполагает построение и одновременное использование для решения поставленной задачи нескольких моделей исследуемого объекта [2]. В рамках этого метода исследуется пригодность различных классов вероятностно-

---

\* Работа выполнена при поддержке Совета по грантам Президента РФ (грант «Поддержка молодых ученых кандидатов наук и их научных руководителей» № МК2198.2006.9).

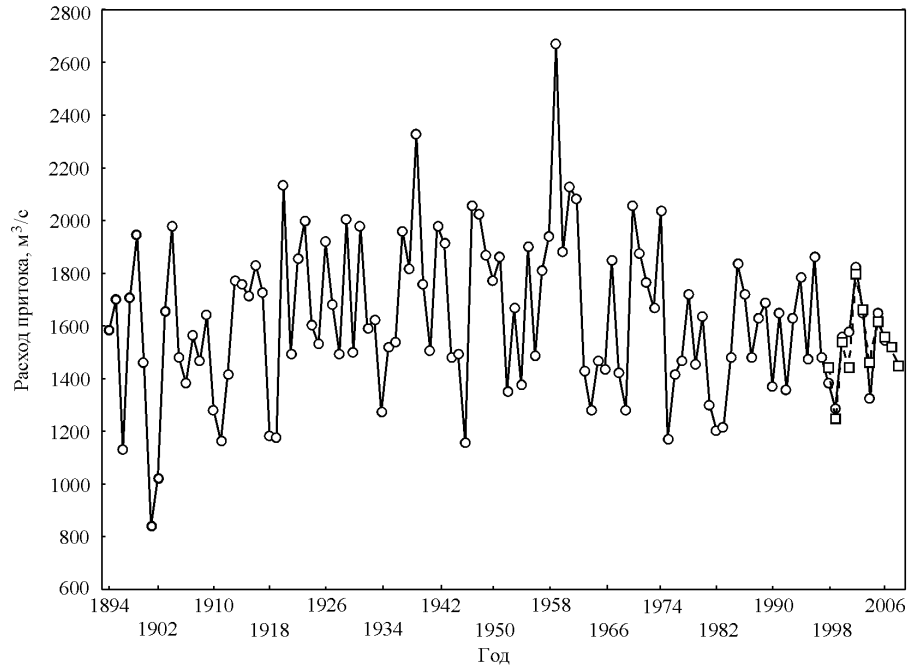


Рис. 1. Временной ряд среднегодового расхода притока за 1894-2005 гг. с наложением прогнозной кривой

статистических моделей для прогнозирования среднегодового притока и строится набор взаимодополняющих и уточняющих друг друга моделей, которые отражают влияние различных факторов на формирование притока.

**2. Прогнозирование среднегодового притока на основе метода главных компонент.** Исследовалась возможность прогнозирования притока с помощью метода главных компонент (МГК). Главные компоненты представляют собой новое множество исследуемых признаков, полученных в результате некоторой линейной комбинации исходных признаков, непосредственно измеренных на объектах. В случае применения МГК для обработки одного ВР в качестве исходных признаков рассматриваются исследуемый временной ряд и этот же временной ряд, но сдвинутый на  $1, 2, \dots, N$  моментов времени. Кроме того, корректное применение МГК требует выполнения условия стационарности ВР [3].

Временной ряд среднегодового расхода притока квазистационарен, поэтому использование МГК для разложения ВР правомерно в данном случае для каждого участка стационарности.

Главные компоненты  $Z_j$  образуются по правилам:

$$Z_j = \{z_{\tau j}\}_{(i-N) \times 1}, \quad j=1, 2, \dots, N+1, \quad \tau = i, i-1, \dots, N+1;$$

$$z_{\tau j} = \sum_{k=0}^N \alpha_{kj} y_{\tau-k}, \quad \sum_{k=0}^N \alpha_{kj}^2 = 1, \quad (1)$$

где  $i$  – номер последнего наблюдения в исследуемом (исходном) ВР;  $\alpha_{kj}$  (вклад  $k$ -го признака в  $j$ -ю компоненту) выбирается так, чтобы  $Z_j$  были центрированы, не коррелированы и ранжированы по дисперсии

$$D(Z_1) \geq D(Z_2) \geq \dots \geq D(Z_{N+1}).$$

При этом сумма дисперсий главных компонент устанавливается равной сумме дисперсий исходных признаков.

В работе [4] приведены некоторые результаты исследования структуры гидрологических временных рядов притока реки Обь с помощью МГК.

Главные компоненты представляют собой ВР, которые изменяются более закономерно по сравнению с исходным ВР и могут быть спрогнозированы. Предлагается применить один из возможных алгоритмов прогнозирования на основе использования МГК – прогнозирование первой главной компоненты ВР, а затем обратный переход от значений главных компонент к значениям исходного ВР [5]. В качестве прогнозной модели ВР компоненты  $Z_1(t)$  используется модель квазигармонической функции:

$$Z_1(t) = f(t) + \sum_{i=1}^{p/2} [A_i \sin ix + B_i \cos ix], \quad (2)$$

где  $x = \frac{2\pi}{p}t$ ;  $f(t)$  – функция тренда;  $i$  – номер гармоники;  $A_i, B_i$  – параметры модели;  $p$  – период колебаний ВР.

Были построены две модели гармонической функции по ВР первой главной компоненты ( $p = 8$ ): по исходным данным за 1987–2004 гг. (рис. 2, два полных периода предыстории) и по исходным данным за 1994–2004 гг. (один полный период предыстории).

На основе этих моделей было реализовано прогнозирование первой главной компоненты ВР и исходного ВР на 2005–2007 гг. Результаты прогнозирования представлены в табл. 1.

В табл. 2 приведены результаты прогнозирования притока на 1997–2005 гг. Анализ таблицы позволяет сделать следующее заключение: с помощью МГК был получен сравнительно точный прогноз среднегодового расхода притока на 1997–2004 гг. Как выяснилось, по истечении 1997–2004 гг. ошибка прогноза составила от 0,8 до 10,6 % (см. рис. 1); все прогнозные значения попадают в 95-процентный доверительный интервал. Более точный прогноз на 2005 г. соответствует модели, построенной на основе данных за 1994–2004 гг., ошибка прогноза 1,3 %. Для модели, построенной по данным за 1987–2004 гг., ошибка прогноза составляет 11,5 %.

Для реализации более точного прогнозирования притока каждый год модель (2) параметрически и структурно адаптировалась с учетом поступления новых данных о притоке. Также адаптировался интервал наблюдений (количество полных циклов) ВР компоненты, на основе которого идентифицируется прогнозная модель, так как оптимальный интервал (позволяющий построить наиболее точную прогнозную модель) с течением времени может измениться. Так, при прогнозе среднегодового притока на 2002–2004 гг. оптимальным интервалом был период, учитывающий два периода предыстории (ошибка прогнозирования в 2002 г. составила 1,3 % от наблюдаемых значений притока; в 2003 г. – 10,6 %; в 2004 г. – 1,8 %), а при прогнозе притока на

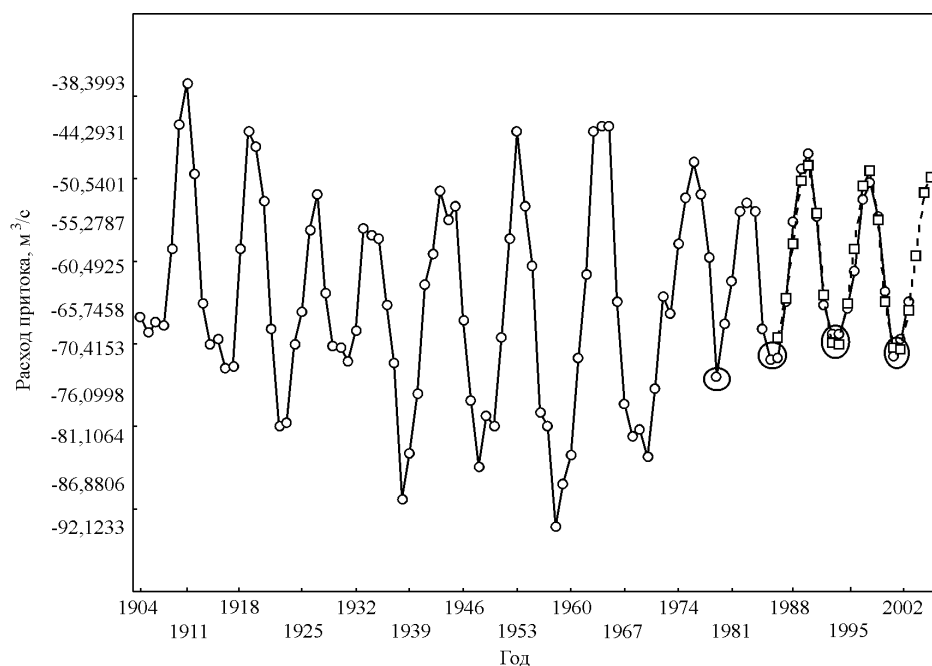


Рис. 2. Первая компонента ВР расхода притока с наложением прогнозной кривой

2005 г. – период, учитывающий один период предыстории. В табл. 2 приведены уточненные прогнозы на 2006–2007 гг. с учетом данных за 2005 г.

Отметим, что при исследовании декадных притоков были выявлены годы с нехарактерным распределением притока по декадам марта–мая по сравнению с общей картиной водности. Эти годы (1979, 1987, 1988, 1995, 2002) соответствуют минимальным значениям первой главной компоненты ВР (см. рис. 2, отмечены большими кружками). Временной ряд первой компоненты среднегодового расхода притока последние 28 лет имеет периодичность, равную 7–8 годам, следовательно, можно предположить, что 2009 г.

Т а б л и ц а 1

**Результаты прогнозирования первой главной компоненты ВР и ВР среднегодового расхода притока на 2005–2007 гг.**

Год	Прогнозное значение первой главной компоненты ВР по данным		Прогнозное значение расхода притока ( $\text{м}^3/\text{с}$ ) по данным	
	1987–2004 гг.	1994–2004 гг.	1987–2004 гг.	1994–2004 гг.
2005	-59,2	-62,0	1365,0	1561,4
2006	-51,8	-53,5	1569,0	1499,2
2007	-50,5	-51,6	1390,8	1438,7

## Результаты прогнозирования среднегодового расхода притока

Год	Наблюдаемый расход притока, м <sup>3</sup> /с	Прогнозное значение расхода притока, м <sup>3</sup> /с	Ошибка прогноза расхода притока		Доверительный интервал						
			м <sup>3</sup> /с	%	±σ (68 %)			±1,965σ (95 %)			
					нижняя граница	верхняя граница	нижняя граница	верхняя граница	нижняя граница	верхняя граница	
1997	1420	1439,0	-19,0	1,3	1305,07	1572,93	1175,83	1702,17	1702,17		
1998	1280	1247,1	32,9	2,6	1113,17	1381,03	983,93	1510,27	1510,27		
1999	1550	1537,9	12,1	0,8	1403,97	1671,83	1274,73	1801,07	1801,07		
2000	1570	1438,2	131,8	8,4	1304,27	1572,13	1175,03	1701,37	1701,37		
2001	1820	1800,0	20,0	1,1	1666,07	1933,93	1536,83	2063,17	2063,17		
2002	1640	1661,0	-21,0	1,3	1527,07	1794,93	1397,83	1924,17	1924,17		
2003	1320	1459,3	-139,3	10,6	1325,37	1593,23	1196,13	1722,47	1722,47		
2004	1645	1615,3	29,7	1,8	1481,37	1749,23	1352,13	1878,47	1878,47		
2005 (1) <sup>1</sup>	1542	1365,0	177,0	11,5	1231,07	1498,93	1101,83	1628,17	1628,17		
2006 (1)	–	1518,0	–	–	1384,07	1651,93	1254,83	1781,17	1781,17		
2007 (1)	–	1447,6	–	–	1313,67	1581,53	1184,43	1710,77	1710,77		
2005 (2) <sup>2</sup>	1542	1561,4	-19,4	1,3	1427,47	1695,33	1298,23	1824,57	1824,57		
2006 (2)	–	1480,5	–	–	1346,57	1614,43	1217,33	1743,67	1743,67		
2007 (2)	–	1436,2	–	–	1302,27	1570,13	1173,03	1699,37	1699,37		

Примечание. 1 – модель построена по данным 1987–2004 гг.; 2 – модель построена по данным 1994–2005 гг.

также будет соответствовать минимальному значению на ВР первой компоненты и будет аномальным.

**3. Скользящее прогнозирование среднегодового притока на основе множественных линейно-регрессионных моделей.** Как уже упоминалось в разд. 1, нами ставилась задача применения для прогнозирования идей и методов вариативного моделирования. Поэтому исследовалась возможность применения, помимо моделей, полученных методом главных компонент, линейных регрессионных моделей для скользящего прогнозирования среднегодового расхода притока по данным предшествующих месяцев этого же года.

Общий вид модели:

$$\bar{Y}_x = a + \sum_{j=1}^k b^j x^j + \varepsilon. \quad (3)$$

Здесь  $k$  – количество объясняющих переменных;  $x^j = (x_1^j, x_2^j, \dots, x_N^j)$  – вектор значений  $j$ -й объясняющей переменной;  $N$  – количество наблюдений;  $x_i^j$  –  $i$ -е наблюдение  $j$ -й объясняющей переменной;  $a, b^1, b^2, \dots, b^k$  – параметры модели;  $\varepsilon$  – ошибка прогноза (остаток);  $Y = (y_1, y_2, \dots, y_N)'$  – вектор-столбец наблюдаемых значений прогнозируемой переменной;  $y_i$  – значение  $i$ -го наблюдения прогнозируемой переменной;  $\bar{Y}_x$  – среднее значение  $Y$  при конкретных  $x$ .

В табл. 3 приведены множественные линейные регрессионные (корреляционно-регрессионные) модели, пригодные для осуществления прогнозирования среднегодового расхода притока по итогам предшествующих месяцев ( $V$  – объем расхода притока). Модели построены методом пошагового рег-

Т а б л и ц а 3

Статистические модели для прогнозирования среднегодового расхода притока

Прошедшие месяцы	Прогнозное уравнение	Коэффициент детерминации $d$	Остаток (min, max)	СКО остатков
Январь–май	$V_{\text{год}} = 640,8 + 0,1V_{\text{апр}} + 0,18V_{\text{май}}$	0,57	-286,8 332,8	178,6
Январь–июнь	$V_{\text{год}} = 218,2 + 0,86V_{\text{февр}} + 0,07V_{\text{апр}} + 0,07V_{\text{май}} + 0,18V_{\text{июнь}}$	0,83	-256,0 277,3	112,3
Январь–июль	$V_{\text{год}} = 445,5 + 0,08V_{\text{апр}} + 0,07V_{\text{май}} + 0,08V_{\text{июнь}} + 0,14V_{\text{июль}}$	0,89	-186,9 201,4	90,5
Январь–август	$V_{\text{год}} = 237,7 + 0,09V_{\text{апр}} + 0,08V_{\text{май}} + 0,12V_{\text{июнь}} + 0,06V_{\text{июль}} + 0,15V_{\text{авг}}$	0,96	-115,5 112,8	55,6
Январь–сентябрь	$V_{\text{год}} = 190,5 + 0,08V_{\text{апр}} + 0,08V_{\text{май}} + 0,1V_{\text{июнь}} + 0,07V_{\text{июль}} + 0,08V_{\text{авг}} + 0,16V_{\text{сен}}$	0,99	-58,7 93,4	28,9

рессионного анализа. Критерий качества аппроксимации – минимум средне-квадратичных отклонений реально наблюдаемых расходов притока от значений, полученных в результате расчета с помощью прогнозной модели.

**Пример прогнозирования.** Имея данные за апрель–май 2005 г., необходимо спрогнозировать значение притока на весь 2005 г. Прогнозное значение расхода притока, рассчитанное на основе модели 1 в табл. 3, составило  $1557,7 \text{ м}^3/\text{с}$ . Как выяснилось, по истечении 2005 г. ошибка прогноза составила 1,02 % от значения реально наблюдаемого притока ( $1542 \text{ м}^3/\text{с}$ ). По мере поступления новых данных по притоку (за июнь, июль, ...) этот прогноз может быть уточнен с помощью других моделей в табл. 3. Таким образом реализуется скользящее прогнозирование притока.

Отметим, что прогнозные значения на 2005 г., рассчитанные с помощью регрессионных моделей, близки ( $1514,0\text{--}1568,1 \text{ м}^3/\text{с}$ ; ошибка прогноза 1,01–1,7 %). Также необходимо отметить, что этот прогноз близок к прогнозу, построенному с помощью МГК и модели 2 гармонической функции (прогноз расхода притока  $1561,4 \text{ м}^3/\text{с}$ ). Следовательно, применяя две модели одновременно, по истечении апреля и мая текущего года прогноз среднегодового расхода притока, рассчитанный на основе МГК, может быть уточнен и скорректирован.

**4. Другие методы прогнозирования.** Исследовалась возможность прогнозирования притока на основе использования моделей авторегрессии и скользящего среднего разных порядков; применялись различные варианты прогнозирования притока на основе выделения года-аналога, а также альтернативные методы анализа данных – нейросетевые методы. Однако получить удовлетворительные результаты в точности прогнозов на основе этих методов не удалось.

**Заключение.** Полученные результаты подтверждают в целом перспективность использования для прогнозирования притока построенного набора уточняющих друг друга моделей. Прогноз на основе МГК выдается на начало прогнозируемого года и корректируется по мере поступления новых данных (по прошествии месяцев прогнозируемого года) с помощью множественных линейных регрессионных моделей.

В данной работе для скользящего прогнозирования притока использовались линейные вероятностно-статистические модели. Для прогнозирования притока перспективным представляется построение и использование набора нелинейных регрессионных моделей. Основная проблема заключается в выборе структуры (структур) нелинейных моделей. Эта проблема может быть решена, в частности, на основе применения алгоритмов и методов эволюционного моделирования.

Достоинство созданных моделей состоит в том, что при прогнозировании они основываются только на данных по притоку и не требуют дополнительной информации (например, об изменении метеорологических и климатических факторов).

Основные направления дальнейших исследований связаны с поиском новых и модернизацией описанных методов в целях повышения точности прогнозирования среднегодовых притоков реки Обь в створе Новосибирской ГЭС.

Усовершенствование методики прогнозирования годового притока реки Обь на основе МГК возможно в следующих направлениях:

- разработка правил выбора оптимального интервала для идентификации прогнозной модели компоненты;
  - исследование возможности прогнозирования притока на основе прогноза совокупности нескольких главных компонент ВР;
  - исследование применимости других моделей (помимо моделей гармонических функций) для идентификации ВР главных компонент.
- Одно из перспективных направлений дальнейших исследований – выделение и учет при построении прогнозных моделей факторов, определяющих величину притока: количества выпавших осадков в верховьях реки, динамики изменения температуры воздуха, тенденции изменения климата и ряд других.

#### СПИСОК ЛИТЕРАТУРЫ

1. **Губарев В. В., Альсова О. К., Беленький А. И. и др.** Управление Новосибирским водохранилищем на основе прогнозирования притока // Водное хозяйство России. Проблемы. Технологии. Управление. 2000. **2**, № 5. С. 484.
2. **Губарев В. В.** Интеллектуальный анализ данных и вариативное моделирование в экспериментальных исследованиях // Информационные системы и технологии (ИСТ, 2001): Сб. науч. статей. Новосибирск: НГТУ, 2001. С. 5.
3. **Пугачев В. С.** Основы общей теории случайных функций. М.: Изд-во Акад. арт. наук, 1952.
4. **Губарев В. В., Альсова О. К.** Исследование гидрологических временных рядов притока реки Обь в створе Новосибирской ГЭС методом главных компонент // Науч. вестн. НГТУ. 2000. №2 (9). С. 3.
5. **Ефимов В. М., Галактионов Ю. К., Шушпанова Н. Ф.** Анализ и прогнозирование временных рядов методом главных компонент. Новосибирск: Наука, 1988.

*Поступила в редакцию 11 апреля 2006 г.*