УДК 52.17

01.00.00 Физико-математические науки

АДАПТИВНЫЕ МОДЕЛИ ВРЕМЕННОГО РЯДА УРОВНЯ ВОДЫ В РЕКЕ ГОРНОГО ТИПА

Титов Николай Георгиевич
аспирант

Кузякина Марина Викторовна
к.ф.-м.н.

Лебедев Константин Андреевич
д.ф.-м.н., проф.
Кубанский государственный университет, Краснодар, Россия

Предложена методика краткосрочного прогнозирования уровня воды в русле реки горного типа с использованием цепей Маркова

Ключевые слова: МАТЕМАТИЧЕСКОЕ МОДЕЛИРОВАНИЕ, АДАПТИВНЫЕ МОДЕЛИ, ПАВОДКОВАЯ СИТУАЦИЯ, ПРОГНОЗИРОВАНИЕ

UDC 52.17

Physics and Math

ADAPTIVE TIME SERIES MODELS OF A MOUNTAIN RIVER LEVEL

Titov Nikolay Georgievich
graduate

Kuzyakina Marina Viktorovna
Candidate of Physical and Mathematical Sciences

Lebedev Konstantin Andreevich
Doctor of Physical and Mathematical Sciences
Kuban state university, Krasnodar, Russia

The article presents a technique of short-term forecasting of water level in the river bed of a mountain type using Markov’s chains

Keywords: MATH MODELING, ADAPTIVE MODELS, FLOOD SITUATION, FORECASTING

Работа посвящена анализу и моделированию динамики уровня воды в горной реке Мзымта. Задача прогнозирования уровня горных рек является актуальной для Краснодарского края.

В настоящее время большое количество экономических исследований проводится с помощью методов корреляционного и регрессионного анализа, для решения поставленной задачи указанные методики применяются недавно [1].

Целью работы является проведение анализа рядов, составленных по ежедневным замерам уровня воды в горной реке Мзымта. Замеры проводились 2 раза в день в восемь утра и восемь вечера по Московскому времени. По этим данным требуется построить модели, адекватно описывающие динамику рядов, рассчитать точечные и интервальные прогнозы на семь дней, оценить точность построенных моделей, сравнивая прогнозные и фактические значения.

Представим графически динамику уровня воды в реке горного типа Мзымта за период январь-декабрь 2010 г. (рисунок 1). Отметим, что тренд за весь период – восходящий, он также отмечен на графике черной линией.

Наивысший уровень воды в реке Мзымта приходится на ноябрь. Как правило, для весенних и осенних месяцев наблюдается высокая вода, для зимних – низкая, что легко объясняется погодными условиями.

Рисунок 1 – Динамика уровня воды в реке Мзымта.

Проведем проверку гипотезы о наличии тренда с помощью критерия Фостера-Стюарта [2]. Если тест Фостера-Стюарта показывает наличие тренда во временном ряде, то подбирается наиболее подходящий тренд, который необходимо исключить для того, чтобы в последующем провести анализ и прогнозирование остатков выбранного ряда.

В методе Фостера-Стюарта гипотеза об отсутствии тренда проверяется с помощью вспомогательных функций. Проверка гипотезы осуществляется в несколько этапов:

1) Для начала определяют вспомогательные характеристики m_t и l_t:

$$m_t = \begin{cases} 1, \textit{если} \quad y_t > \textit{всех} \ \ \textit{предыдущих} \quad \textit{значений} \\ 0 \ \textit{в} \ \textit{противном} \quad \textit{случае} \end{cases}$$

$$l_t = \begin{cases} 1, \textit{если} \quad y_t < \textit{всех} \ \ \textit{предыдущих} \quad \textit{уровней} \\ 0 \ \textit{в} \ \textit{противном} \quad \textit{случае} \end{cases}$$

2) Вычисляют

$$d_t = m_t - l_t .$$

Эта величина может принимать значения -1,0,1.

$$D = \sum_{t=2}^{n} d_t = 8 .$$

3) Применяют критерий Стьюдента:

$$t_{наб} = \frac{D}{\sigma_D},$$

где σ_D вычисляется по формуле [3]

$$\sigma_D = \sqrt{2 \ln 730 - 0{,}8456} = 3{,}5129 ,$$

тогда

$$t_{наб} = \frac{8}{3{,}5129} = 2{,}27 ,$$

$$t_{кр}(\alpha; n-1) = t_{кр}(0{,}05; 729) = 1{,}96 .$$

Если $|t_{наб}| > t_{кр}$, то гипотеза об отсутствии тренда отвергается. Получаем, $|t_{наб}| > t_{кр}$, то есть $2{,}27 > 1{,}96$, следовательно гипотеза H_0 отвергается, тренд есть.

При добавлении стандартными средствами MS Excel линий тренда к графику исходных данных визуальный анализ показывает, что линейный тренд и полиномиальный тренд 5-ой степени соответствуют тенденции исследуемого ряда (рисунок 2). Были добавлены полиномиальные тренды от 1 до 6 степени включительно, но все, кроме 1 и 5 степени были отвергнуты на предварительном этапе анализа.

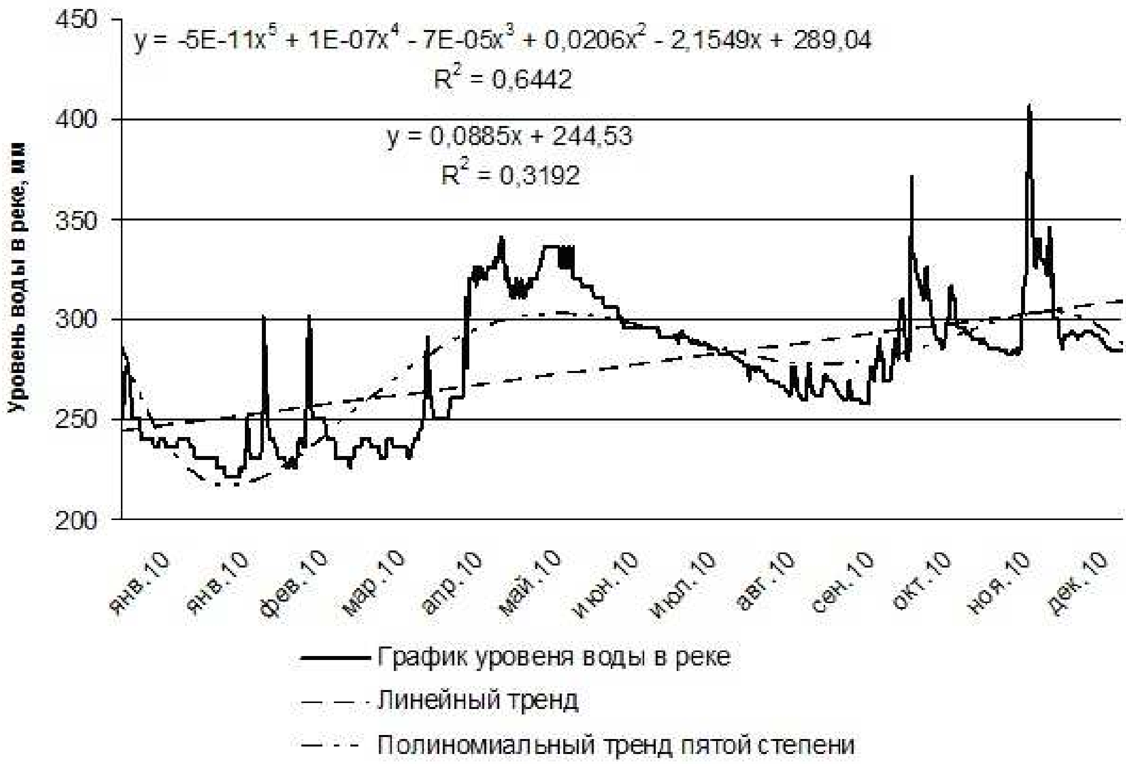

Рисунок 2 – График линейной и полиномиальной моделей для уровня воды в реке Мзымта

Линейная регрессионная модель имеет вид

$$y = 0{,}0885x + 244{,}53 \quad (1)$$

Полиномиальная регрессионная модель 5-го порядка имеет следующий вид

$$y = -5 \cdot 10^{-11} x^5 + 1 \cdot 10^{-7} x^4 - 7 \cdot 10^{-5} x^3 + 0{,}0206x^2 - 2{,}1549x + 289{,}04 \quad (2)$$

Тренды с соответствующими уравнениями и коэффициентами детерминации были рассчитаны с помощью стандартных средств MS Excel.

Коэффициент детерминации – это доля дисперсии зависимой переменной, объясняемая рассматриваемой моделью. При оценке регрессионных моделей это интерпретируется как соответствие модели данным. Для приемлемых моделей предполагается, что коэффициент детерминации должен быть хотя бы не меньше 50% (в этом случае коэффициент множественной корреляции превышает по модулю 70%). Равенство коэффициента детерминации единице означает, что объясняемая переменная в точности описывается рассматриваемой моделью.

Коэффициенты детерминации для моделей (1) и (2) равны соответственно $R^2 = 0{,}32$ и $R^2 = 0{,}64$, т.е. обе модели не очень хорошо ($R^2 < 0{,}75$) описывают ряд, но полиномиальная модель пятого порядка имеет гораздо большую точность. Здесь и далее R^2 – коэффициент детерминации Таким образом, будем рассматривать модель (2).

После построения модели необходимо проверить 5 предпосылок регрессионного анализа [4]. При выполнении всех пяти предпосылок оценки коэффициентов регрессии будут обладать свойствами несмещенности, эффективности и состоятельности. Проверим пять следующих предпосылок: случайный характер остатков модели, равенство нулю математического ожидания остатков, отсутствие автокорреляционной зависимости в остатках, гомоскедастичность (однородная вариативность значений наблюдений, выражающаяся в относительной стабильности, гомогенности дисперсии случайной ошибки регрессионной модели; явление, противоположное гетероскедастичности) дисперсии остатков, подчинение остатков нормальному закону распределения.

График остатков представлен на рисунке 3.

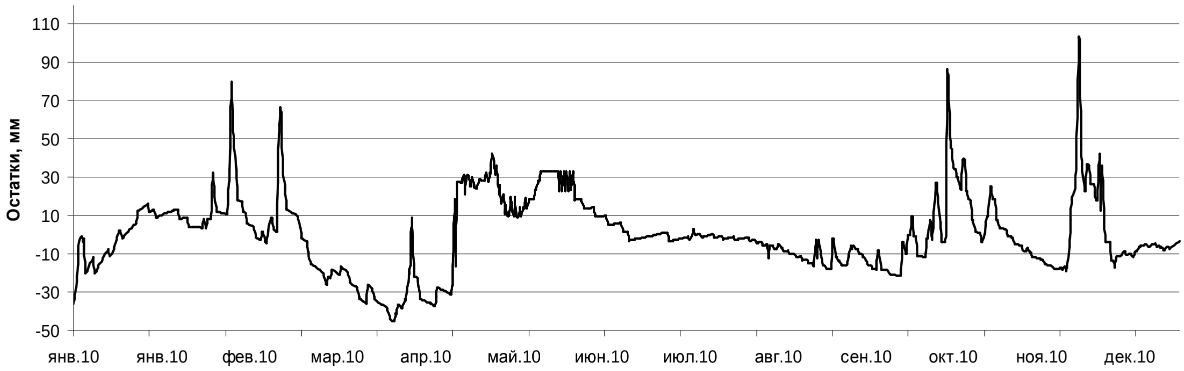

Рисунок 3 – График остатков модели

Математическое ожидание остатков имеет значение $M(E) = -4{,}7 \cdot 10^{-13}$ близко к нулю. Отличие от нуля обусловлено погрешностью вычислений.

Остатки принадлежат интервалу $[-3S; 3S] = [-102; 102]$, где S – стандартная ошибка регрессии [5].

Вычислим коэффициенты асимметрии

$$Ass = \frac{M\left(E^3\right)}{S^3} = 0{,}14$$

и эксцесса

$$Ex = \frac{M\left(E^4\right)}{S^4} = 0{,}58,$$

и воспользуемся статистикой Жака-Бера, которая выражается следующей формулой [5]:

$$W = n\left(\frac{Ass^2}{6} + \frac{(Ex-3)^2}{24}\right) = 180{,}59,$$

где n – объем выборки.

Сама статистика Жака-Бера, рассчитанная по выборке, является случайной величиной распределенной как Хи-квадрат $\chi^2(2)$ с двумя степенями свободы. Другими словами, статистика W подчиняется распределению $\chi^2(2)$ при справедливости гипотезы о нормальности распределения.

Критерий Жака-Бера используется для проверки гипотезы о том, что исследуемая выборка является выборкой нормально распределенной случайной величины с неизвестным математическим ожиданием и дисперсией. Как правило, этот критерий применяется перед тем, как использовать методы параметрической статистики, требующие нормальности исследуемых случайных величин.

Значение статистики принимает значение $W = 180{,}59$, что больше квантили распределения $\chi^2(2) = 5{,}99$, следовательно гипотезу о нормальном распределении остатков нужно отвергнуть.

Проверим распределение остатков в пакете прикладных программ Statistica [6], т. к. иногда бывает, в связи с большим объёмом выборки, что критерий Жака-Бера оказывается некорректным. Будем иметь явно выраженное нормальное распределение (рисунок 4).

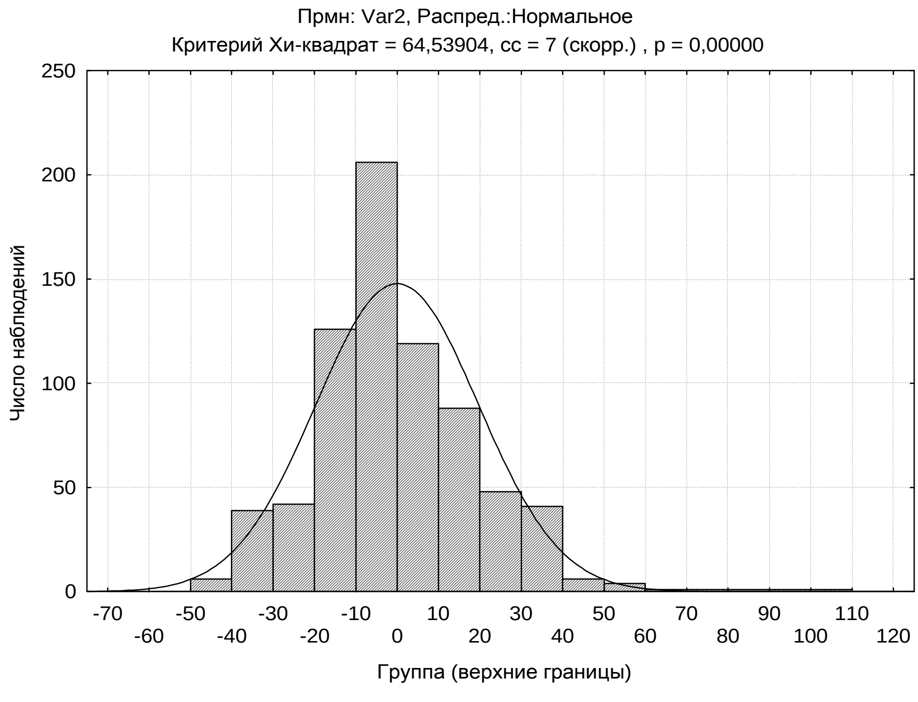

Рисунок 4 – Закон распределения остатков модели

Для проверки остатков на случайность используем критерий «восходящих и нисходящих серий» [7]. Он состоит в проверке следующих двух условий:

$$\begin{cases} \nu(n) > \left[\frac{1}{3}(2n-1) - 1{,}96\sqrt{\frac{16n-29}{90}}\right], \\ \tau_{\max}(n) < \tau_0(n), \end{cases}$$

$$\tau_0(n) = \begin{cases} 5, \textit{ если } n \le 26, \\ 6, \textit{ если } 26 < n \le 153, \\ 7, \textit{ если } 153 < n \le 1170. \end{cases}$$

(3)

где n – длина ряда, $\nu(n)$ – число серий, $\tau_{\max}(n)$ – максимальная длина серии.

Тогда (3) примет следующий вид:

$$\begin{cases} 225 > 464 - \textit{не выполняется}, \\ 31 < 7 - \textit{не выполняется}. \end{cases}$$

Оба условия из (3) не выполняются, следовательно, выборка остатков неслучайна.

Гетероскедастичность (неоднородность наблюдений, выражающаяся в неодинаковой (непостоянной) дисперсии случайной ошибки регрессионной модели) остатков модели регрессии может привести к негативным последствиям:

1) оценки неизвестных коэффициентов нормальной линейной модели регрессии являются несмещёнными и состоятельными, но при этом теряется свойство эффективности;

2) существует большая вероятность того, что оценки стандартных ошибок коэффициентов модели регрессии будут рассчитаны неверно, что конечном итоге может привести к утверждению неверной гипотезы о значимости коэффициентов регрессии и значимости модели регрессии в целом.

Для проверки наличия гетероскедастичности используем ранговый коэффициент корреляции Спирмена, который рассчитывается по следующей формуле [8]:

$$\rho = 1 - \frac{6\sum d^2}{n\left(n^2 - 1\right)},$$

где d – абсолютная разность между рангами значений Y_t и $|E_t|$, n – длина выборки.

Для рассматриваемого ряда остатков $\rho = 0{,}1427$. При количестве наблюдений n более 30 можно рассчитать критические значения с помощью t-критерия Стьюдента. Оценим статистическую значимость ρ с помощью t-критерия:

$$t_\rho = \frac{\rho\sqrt{n-2}}{\sqrt{1-\rho^2}} \approx 3{,}88 \cdot$$

Сравним эту величину с табличной $t_\lambda \approx 1{,}97$ при уровне значимости $\lambda = 0{,}05$. Получаем $3{,}88 > 1{,}97$, следовательно, гипотеза об отсутствии гетероскедастичности остатков отвергается.

Стационарный ряд – это ряд, в котором вероятностные характеристики (параметры случайной величины) постоянны, не зависят от времени, т. е. на свойства не влияет изменение начала отсчёта времени. Определить, стационарен ли ряд, можно по виду автокорреляционной функции и частной автокорреляционной функции и путем проведения теста Дики-Фуллера [9].

Графики автокорреляционной функции (АКФ) и частной автокорреляционной функции (ЧАКФ) ряда остатков, полученные с помощью ППП Statistica [5], представлены на рисунках 5 и 6 соответственно.

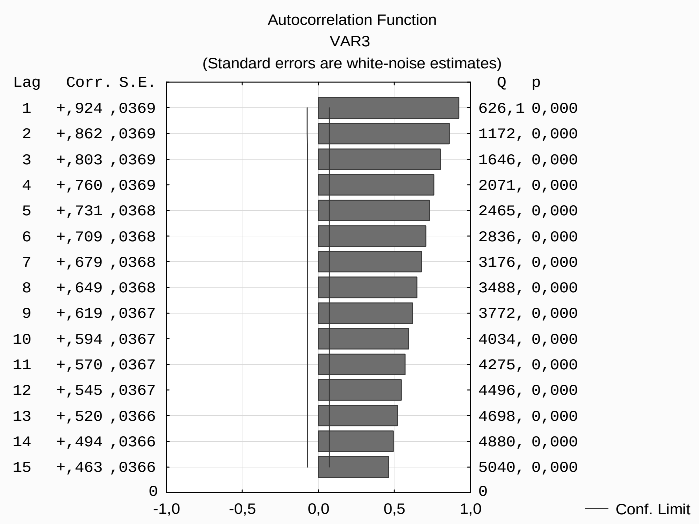

Рисунок 5 – График автокорреляционной функции

АКФ убывает и имеет положительные значения, которые можно считать белым шумом.

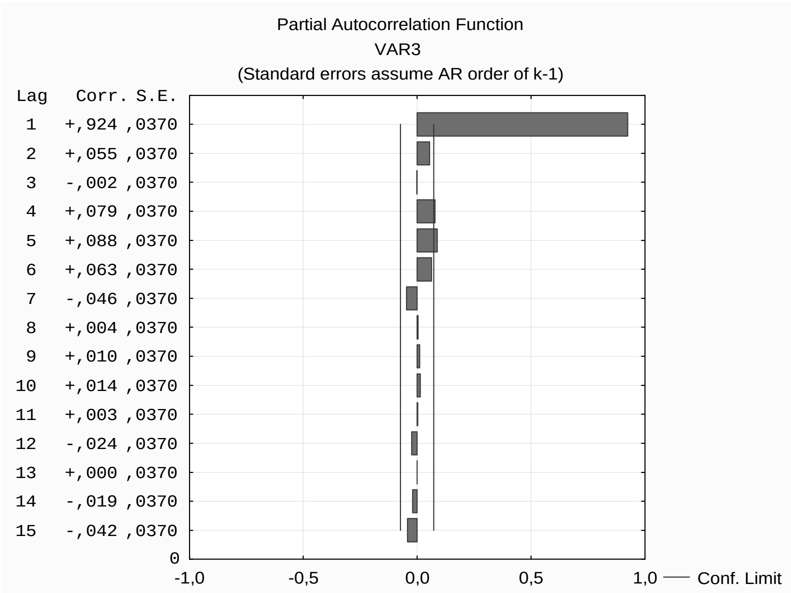

Рисунок 6 – График частной автокорреляционной функции

Из графика ЧАКФ видно, что значимым является лишь значение ЧАКФ при лаге 1.

Суть критерия Дики-Фуллера состоит в том, что необходимо проверять нулевую гипотезу о стационарности ряда [9]:

$$y_t = \alpha y_{t-1} + \varepsilon_t .$$

Есть альтернативная гипотеза: $\alpha < 1$. Взяв первую разность, можно получить следующее уравнение:

$$y_t - y_{t-1} = (\alpha - 1) y_{t-1} + \varepsilon_t ,$$

или

$$\Delta y_t = \beta y_{t-1} + \varepsilon_t ,$$

тогда

$$H_0 : \beta = 0 ,\ H_1 : \beta < 0 .$$

Если H_0 выполняется, то ряд стационарный.

$$\Delta y_t = -0{,}07 y_{t-1} - 0{,}04 .$$

После нахождения оценки $\hat{\beta} = -0{,}07$, вычисляют статистику

$$t_{набл} = \frac{\hat{\beta}}{S(\hat{\beta})} \approx -7 .$$

Гипотеза не принимается и ряд признается нестационарным, если $t_{набл} > t_{крит}$. $-7 < -2{,}58$, т. е. $t_{набл} < t_{крит}$, следовательно, нулевая гипотеза принимается и ряд признается стационарным.

Определим вид модели авторегрессии проинтегрированного скользящего среднего (АРПСС). Для этого необходимо оценить модель АРПСС (p, d, q) (Ps, Ds, Qs), где *p* – порядок авторегрессии, d – порядок разности, q – порядок скользящего среднего, Ps – порядок сезонной авторегрессии, Ds – порядок сезонной разности, Qs – сезонный параметр скользящего среднего [6].

На основе проведённого анализа была выбрана АРПСС (0,0,2) (0,0,3), параметры которой статистически значимы (рисунок 7).

Paramet.	Input: VAR3 (Таблица остатков 2010.sta) Transformations: none Model:(0,0,2)(0,0,3) Seasonal lag: 12 MS Residual= 95,755					
	Param.	Asympt. Std.Err.	Asympt. t(725)	p	Lower 95% Conf	Upper 95% Conf
q(1)	-0,848813	0,038597	-21,9915	0,000000	-0,924588	-0,773037
q(2)	-0,514080	0,032428	-15,8531	0,000000	-0,577744	-0,450417
Qs(1)	-0,307653	0,038994	-7,8897	0,000000	-0,384208	-0,231098
Qs(2)	-0,149195	0,036535	-4,0836	0,000049	-0,220923	-0,077467
Qs(3)	-0,114423	0,039206	-2,9185	0,003626	-0,191393	-0,037453

Рисунок 7 – Параметры модели АРПСС

На рисунках 8, 9 изображены исходные данные и прогнозные значения согласно выбранной модели АРПСС (0,0,2) (0,0,3).

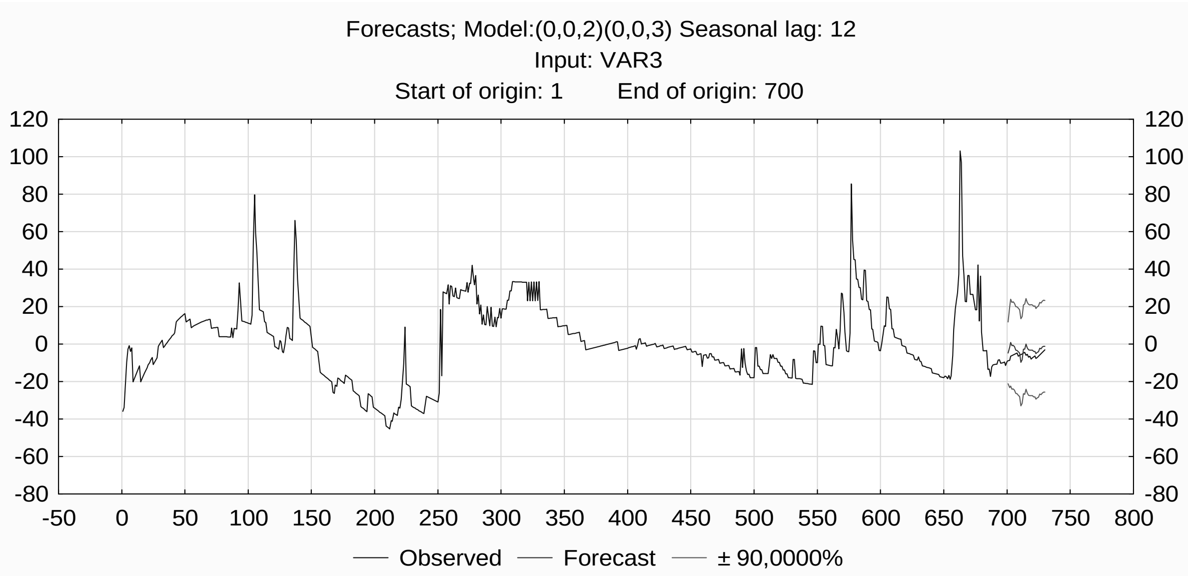

Рисунок 8 – График исходных и прогнозных значений

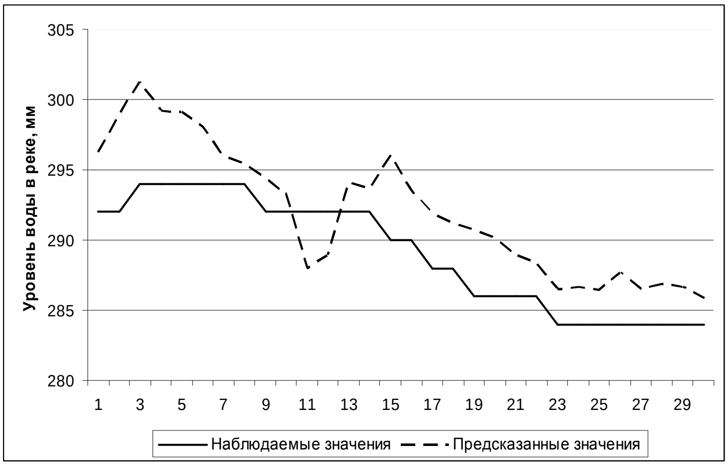

Рисунок 9 – График наблюдаемых и предсказанных значений

Таким образом, в ходе исследования была построена модель АРПСС (0,0,2) (0,0,3). Здесь 2 – параметр скользящего среднего, 3 – сезонный параметр скользящего среднего, авторегерессия не учитывается. Были выявлены стационарность и сравнительно наибольшая эффективность. Прогнозные значения отличаются от наблюдаемых в среднем на 3,43 мм, но не превышают 7,23 мм. Значения были предсказаны на 2 недели вперед. Отклонения составляют 2%. Можно считать прогноз адекватным, а предложенную методику – состоятельной.

Данная статья выполнена при поддержке гранта РФФИ № 13-01-96518 р_юг_а и является продолжением цикла исследований, посвящённых математическому моделированию состояния уровня воды в горной реке [10–12].

Список использованных источников

1. *Титов, Н.Г.* Сравнительный анализ методов математического моделирования уровня воды в реке горного типа (на примере реки Мзымта) /Н.Г. Титов, Е.А.Семенчин, М.В. Кузякина, К.А. Лебедев // Фундаментальные исследования, ISSN: 1812-7339, № 12 за 2014 год (часть 5), С. 952-957

2. *Четыркин Е.М.* Статистические методы прогнози- рования. - М.: Статистика, 1977.

3. *Кобзарь А.И.* Прикладная математическая статистика. — М.: Физматлит, 2006. — 816 с.

4. *Елисеева И.И.* Эконометрика. — М.: Финансы и статистика, 2007. — 576 с.

5. *Кендэл М.*. Временные ряды. — М.: Финансы и статистика, 1981. — 191 с.

6. *Халафян А.А.* Статистический анализ данных. STATISTICA 6.0. – Краснодар: изд-во "КубГУ", 2003, 191с.

7. *Айвазян С. А.* Прикладная статистика: Основы моделирования и первичная обработка данных. Справочное изд. / С. А. Айвазян, И. С. Енюков, Л. Д. Мешалкин. — М.: Финансы и статистика, 1983. — 471 с.

8. *Айвазян С.А., Мхитарян В.С.* Прикладная статистика и основы эконометрики. – Москва: «ЮНИТИ», 1998, 1006 с.

9. *Магнус Я. Р., Катышев П. К., Пересецкий А. А.* Эконометрика. Начальный курс. – М.: Дело, 2007. – 504 с. – ISBN 978-5-7749-0473-0

10. *Титов Н.Г.*, Прогноз уровня воды в реке с крутым падением водотока, основанное на фильтрации Кальмана-Бьюси / Н.Г. Титов, М.В. Кузякина, К.А. Лебедев // Политематический сетевой электронный научный журнал Кубанского государственного аграрного университета 2014, №104

11. *Титов Н.Г.*, Сравнительный анализ методов математического моделирования уровня воды в реке горного типа (на примере реки Мзымта) /Н.Г. Титов, Е.А.Семенчин,

М.В. Кузякина, К.А. Лебедев // Фундаментальные исследования, ISSN: 1812-7339, № 12 за 2014 год (часть 5), С. 952-957

12. Titov N.G., Kuzyakina M.V., Lebedev K.A. Su uno della metodologia per la valutazione del danno economico inflitto diluvio fiume tipo mountain trama regione. // Italian Science Review. 2014; 12(21), ISSN: 2308-832X. PP. 234-236

References

1. Titov, N.G. Sravnitel'nyj analiz metodov matematicheskogo modelirovanija urovnja vody v reke gornogo tipa (na primere reki Mzymta) /N.G. Titov, E.A.Semenchin, M.V. Kuzjakina, K.A. Lebedev // Fundamental'nye issledovanija, ISSN: 1812-7339, № 12 za 2014 god (chast' 5), S. 952-957

2. Chetyrkin E.M. Statisticheskie metody prognozirovanija. - M.: Statistika, 1977.

3. Kobzar' A.I. Prikladnaja matematicheskaja statistika. — M.: Fizmatlit, 2006. — 816 s.

4. Eliseeva I.I. Jekonometrika. — M.: Finansy i statistika, 2007. — 576 s.

5. Kendjel M.. Vremennye rjady. — M.: Finansy i statistika, 1981. — 191 s.

6. Halafjan A.A. Statisticheskij analiz dannyh. STATISTICA 6.0. – Krasnodar: izd-vo "KubGU", 2003, 191s.

7. Ajvazjan S. A. Prikladnaja statistika: Osnovy modelirovanija i pervichnaja obrabotka dannyh. Spravochnoe izd. / S. A. Ajvazjan, I. S. Enjukov, L. D. Meshalkin. — M.: Finansy i statistika, 1983. — 471 s.

8. Ajvazjan S.A., Mhitarjan V.S. Prikladnaja statistika i osnovy jekonometriki. – Moskva: «JuNITI», 1998, 1006 s.

9. Magnus Ja. R., Katyshev P. K., Pereseckij A. A. Jekonometrika. Nachal'nyj kurs. – M.: Delo, 2007. – 504 s. – ISBN 978-5-7749-0473-0

10. Titov N.G., Prognoz urovnja vody v reke s krutym padeniem vodotoka, osnovannoe na fil'tracii Kal'mana-B'jusi / N.G. Titov, M.V. Kuzjakina, K.A. Lebedev // Politematicheskij setevoj jelektronnyj nauchnyj zhurnal Kubanskogo gosudarstvennogo agrarnogo universiteta 2014, №104

11. Titov N.G., Sravnitel'nyj analiz metodov matematicheskogo modelirovanija urovnja vody v reke gornogo tipa (na primere reki Mzymta) /N.G. Titov, E.A.Semenchin, M.V. Kuzjakina, K.A. Lebedev // Fundamental'nye issledovanija, ISSN: 1812-7339, № 12 za 2014 god (chast' 5), S. 952-957

12. Titov N.G., Kuzyakina M.V., Lebedev K.A. Su uno della metodologia per la valutazione del danno economico inflitto diluvio fiume tipo mountain trama regione. // Italian Science Review. 2014; 12(21), ISSN: 2308-832X. PP. 234-236