



МИНИСТЕРСТВО НАУКИ И ВЫСШЕГО ОБРАЗОВАНИЯ РОССИЙСКОЙ ФЕДЕРАЦИИ
федеральное государственное бюджетное образовательное учреждение
высшего образования
«РОССИЙСКИЙ ГОСУДАРСТВЕННЫЙ
ГИДРОМЕТЕОРОЛОГИЧЕСКИЙ УНИВЕРСИТЕТ»

Кафедра инженерной гидрологии

ВЫПУСКНАЯ КВАЛИФИКАЦИОННАЯ РАБОТА
(магистерская диссертация)

На тему **Учет цикличности многолетних
колебаний стока в гидрологических
прогнозах**

Исполнитель Филатов Дмитрий Михайлович
(фамилия, имя, отчество)

Руководитель К.Т.Н., ДОЦЕНТ
(ученая степень, ученое звание)

Гайдукова Екатерина Владимировна
(фамилия, имя, отчество)

«К защите допускаю»
Заведующий кафедрой

(подпись)

К.Т.Н., ДОЦЕНТ
(ученая степень, ученое звание)

Гайдукова Екатерина Владимировна
(фамилия, имя, отчество)

«26» мар 2025г.

Санкт-Петербург
2025

Содержание

| | |
|--|----|
| Введение | 3 |
| Апробация..... | 6 |
| 1.1 Алгоритм регрессионного анализа..... | 7 |
| 1.2 Прогностические возможности регрессионных моделей | 16 |
| 1.3 Алгоритм увеличения эффективности прогнозных регрессионных моделей | 21 |
| 2 Формирование базы данных для исследования..... | 23 |
| 2.1 Выбор речных водосборов..... | 23 |
| 2.2 Анализ исходных гидрометеорологических данных | 23 |
| 3 Применение разработанного эффективного алгоритма для прогнозов по регрессионным моделям | 30 |
| 4 Анализ полученных результатов и рекомендации для использования разработанного алгоритма | 47 |
| 5 Методы определения циклов и периодов во временных рядах..... | 49 |
| 5.1 Метод Шустера..... | 49 |
| 5.2 Анализ результатов полученных методом Шустера | 67 |
| 5.3 Построение автокорреляционных функций..... | 81 |
| 5.4 Анализ автокорреляционных функций | 85 |
| 5.5 Спектральный анализ | 92 |
| Заключение | 96 |

Введение

Актуальность. Эффективное применение регрессионных зависимостей для прогноза гидрологических характеристик горных рек отмечают некоторые исследователи. Использование множественной регрессии с двумя предикторами – расходом воды за предшествующий период и осадками, привело к улучшению прогнозов на реке Нарым (река в Восточно-Казахстанской области Казахстана, правый приток Иртыша) [4], а для реки Амыл (горная река Красноярского края) сравнительный анализ методик прогноза максимальных уровней воды показал, что однофакторные зависимости имеют более высокие коэффициенты детерминации, чем модель множественной регрессии [5].

В связи с вышеизложенным целью исследования заключается в разработке алгоритма получения уравнения регрессии, которое можно использовать для более эффективных прогнозов гидрологических характеристик.

Для достижения цели исследования необходимо решить следующие задачи:

- изучение метода регрессионного анализа и перспективы его использования в гидрологических прогнозах;
- разработка алгоритма получения эффективных регрессионных моделей;
- сбор исходной гидрометеорологической информации;
- получение регрессионных моделей с учетом разработанного алгоритма;
- оценка качества полученных прогнозов;
- построение периодограмм методом Шустера;
- оценка автокорреляционных функций.

В современных условиях возрастает потребность в повышении точности гидрологических прогнозов, особенно для горных рек, где традиционные подходы часто показывают недостаточную эффективность. Существующие методики прогнозирования требуют совершенствования с учетом региональных особенностей водосборов, климатических изменений и антропогенного воздействия.

Недостаточная изученность методов прогнозирования гидрологических характеристик, особенно для горных рек, создает необходимость разработки новых подходов. Практические потребности в повышении точности прогнозов особенно важны для планирования водохозяйственных мероприятий, прогнозирования паводковых ситуаций, оценки водных ресурсов и проектирования гидротехнических сооружений.

Развитие современных методов анализа временных рядов и регрессионных моделей создает предпосылки для совершенствования существующих подходов к гидрологическому прогнозированию. При этом отмечается отсутствие унифицированных методик для различных типов рек, что требует разработки адаптивных подходов к прогнозированию гидрологических характеристик.

Важность совершенствования существующих методов регрессионного анализа обусловлена необходимостью повышения надежности прогнозов, сокращения времени обработки данных, уменьшения погрешности расчетов и учета современных климатических изменений. Практическая значимость улучшения методов прогнозирования определяется потребностями в обеспечении безопасности гидротехнических сооружений, оптимизации водопользования, планирования мероприятий по защите от наводнений и рационального использования водных ресурсов.

Научная значимость работы

Научная значимость исследования заключается в разработке нового алгоритма получения уравнений регрессии для гидрологических прогнозов, позволяющего повысить точность прогнозирования характеристик рек.

Теоретическое обоснование методов увеличения эффективности прогнозных регрессионных моделей включает анализ влияния различных метеорологических факторов на гидрологический режим рек.

Исследование дополняет существующие методики регрессионного анализа новыми подходами к обработке гидрометеорологических данных, расширяет теоретическую базу гидрологических исследований. Применение метода Шустера для исследования периодичности гидрологических процессов позволяет выявить скрытые закономерности в временных рядах.

Практическая значимость работы

Практическая значимость исследования определяется разработкой методики прогнозирования гидрологических характеристик, которая может быть использована в практике гидрологических исследований и водохозяйственных расчетах. Созданный алгоритм позволяет эффективно использовать регрессионные модели для прогнозирования расходов воды с различной заблаговременностью - от краткосрочных до долгосрочных прогнозов.

Определение оптимальных подходов к выбору предикторов для регрессионных моделей повышает точность гидрологических прогнозов. Разработанные рекомендации по применению полученных результатов для конкретных водосборов учитывают их региональные особенности. Созданный инструментарий для анализа гидрометеорологических данных может быть использован в работе гидрологических служб и при проектировании водохозяйственных объектов.

Практическая демонстрация эффективности предложенных методов на примере конкретных рек (Ловать и Сорочь) подтверждает возможность их применения в реальных условиях, что делает исследование особенно ценным для практической гидрологии.

Апробация



Сертификат участника

подтверждает, что

Филатов Дмитрий Михайлович

принял(а) участие в работе VII Всероссийской научно-практической конференции «Современные тенденции и перспективы развития гидрометеорологии в России», приуроченной к 60-летию кафедры гидрологии и природопользования ИГУ.

Декан
географического
факультета
ФГБОУ ВО «ИГУ»

С.Ж. Вологжина



19-21 марта 2025 г.,
г. Иркутск



1.1 Алгоритм регрессионного анализа

Методика восстановления значений стока по уравнениям регрессии состоит в следующем. В случае, когда имеется один год кратковременных наблюдений, рассчитываются уравнения регрессии между наблюдениями за этот год и последовательно со всеми остальными годами, в которых имеются наблюдения в пунктах-аналогах, при условии, что количество пунктов-аналогов должно быть не менее 5. Наиболее эффективным методом, дающим наиболее качественные результаты приведения данных наблюдений к более длительному периоду, является метод расчета, основанный на пространственных годовых уравнениях по данным наблюдений в однородном районе. Методически задача решается следующим образом. При увеличении числа аналогов увеличивается надежность и точность уравнений регрессии, что приводит к более точным результатам расчета в исследуемом пункте. Но в этом случае уменьшается число восстановленных лет наблюдений. Обратная картина получается при уменьшении числа принятых в расчет аналогов. Оптимальное число аналогов задаваемых при расчете уравнений регрессии зависит от гидрологической изученности района исследования и пространственной связанностью рассматриваемой гидрологической характеристики. Так, при слабой гидрологической изученности число аналогов, принимаемых в расчет, может быть уменьшено, а при хорошо изученном в гидрологическом отношении районе исследования число аналогов может быть увеличено. Для дождевого стока, имеющего, как правило, слабую пространственную связанность по сравнению с другими гидрологическими характеристиками число пунктов может быть уменьшено по сравнению с другими гидрологическими характеристиками, имеющими большую пространственную связанность. Этот вопрос выносится в качестве диалога специалиста гидролога с Персональным компьютером. Следовательно, при назначении наименьшего числа аналогов, которые будут

использованы в дальнейших расчетах, следует иметь ввиду приведенные соображения. При этом возможны и другие гидрологические аспекты, поясняющие назначение наименьшего числа аналогов в регрессионном методе, зависящие от многих других факторов, включая региональные особенности рассматриваемого района исследования. В однородном гидрологическом районе проводятся стационарные гидрометрические наблюдения за изучаемой характеристикой x , а также имеются кратковременные наблюдения y в исследуемом пункте. Исходная гидрометрическая информация от самого раннего года наблюдений $i=1$ до самого позднего $i=n$ может быть представлена в общем виде. Наблюдения в расчетном створе могут быть в различном временном интервале от $i=1$ до $i=n$, но, как правило, изыскания проводятся в последние годы к моменту проектирования. Индексация при значениях единая для всех постов: первый индекс обозначает годы наблюдений от $i=1$ до $i=n$, второй индекс – номер гидрометрического стационарного створа наблюдений от $j=1$ до $j=k$, расчетный створ имеет индекс $(k+1)$. Число лет наблюдений в расчетном створе L всегда меньше, чем в стационарных гидрометрических створах. Требуется восстановить сток в расчетном створе за годы, наблюдения по которым имеются в пунктах-аналогах. Для стационарных гидрометрических пунктов для каждого года наблюдений от $i=1$ до $i=n-L-1$ рассчитываются уравнения регрессии, наибольшее число уравнений равно $(n-L-1)i$.

$$\begin{aligned}
 y_{ij} &= A_{i1} x_{(n-L-1)j} + B_{i1} \\
 y_{ij} &= A_{i2} x_{(n-L-1)j} + B_{i2} \\
 &\dots\dots\dots \\
 &\dots\dots\dots \\
 &\dots\dots\dots \\
 y_{ij} &= A_{iL} x_{nj} + B_{iL} \quad .
 \end{aligned}
 \tag{1}$$

В расчетах используются лишь те уравнения, которые удовлетворяют условиям: $R_i/\sigma_{Ri} \geq q_R$, $R_i \geq R_{кр}$, $k \geq k_{кр}$. . (2) Обычно, $q_R \geq 2$, $R_{кр} \geq 0,7$, $k_{кр} \geq 5$. С уменьшением этих коэффициентов, как правило, увеличивается число

восстановленных значений стока с меньшим уровнем надежности, и, наоборот, с увеличением этих коэффициентов уменьшается число рассчитанных значений стока, но увеличивается уровень надежности. При назначении критических значений параметров, которые в дальнейшем будут учитываться при выборе уравнений регрессии, следует учитывать соображения, приведенные ранее и другие аспекты гидрологического анализа. Численные критические значения, приведенные в (2) следует использовать лишь в первом приближении, когда никакой дополнительной гидрологической информации о районе исследования не имеется.

Таблица 2.1- Матрица исходных данных

| Стационарные наблюдения | | | | | | Расчетный створ |
|-------------------------|--------------|-------|--------------|-------|--------------|------------------|
| x_{11} | x_{12} | | X_{1j} | | x_{1k} | |
| x_{21} | x_{22} | | X_{2j} | | x_{2k} | |
| | | | | | | |
| x_{i1} | x_{i2} | | x_{ij} | | x_{ik} | |
| | | | | | | |
| $x_{(n-L)1}$ | $x_{(n-L)2}$ | | $X_{(n-L)j}$ | | $x_{(n-L)k}$ | |
| $x_{(n-L)1}$ | $x_{(n-L)2}$ | | $X_{(n-L)j}$ | | $x_{(n-L)k}$ | $y_{(n-L)(k+1)}$ |
| | | | | | | |
| $x_{(n-L)1}$ | $x_{(n-L)2}$ | | $X_{(n-L)j}$ | | $x_{(n-L)k}$ | $y_{(n-1)(k+1)}$ |
| x_{n1} | x_{n2} | | x_{nj} | | x_{nk} | $y_{n(k+1)}$ |

Восстановленные по уравнениям регрессии значения стока имеют заниженную дисперсию многолетних колебаний, что для инженерных гидрологических расчетов недопустимо т.к. при этом занижается расчетное значение рассматриваемой гидрологической характеристики для зоны малых обеспеченностей (различные характеристики максимального стока и уровней воды) и завышается расчетное значение для зоны больших

обеспеченностей (различные характеристики минимального стока, стока летней и зимней межени). Исключение систематических погрешностей в оценке коэффициента вариации и стандартного отклонения осуществляется путем перехода от 16 уравнения регрессии к так называемому единому решению, когда в уравнении регрессии исключается значение коэффициента корреляции, то есть коэффициент корреляции принимается равным единице. В этом случае принцип наименьших квадратов по шкале ординат заменяется принципом наименьших прямоугольников по шкалам ординат и абсцисс. По уравнениям (1) восстанавливается сток в расчетном створе $(k+1)$ за годы от $i=1$ до $i=n-L-1$. Для каждого года от $i=1$ до $i=n-L-1$ сток восстанавливается L раз. За расчетное значение стока в i -ом году принимается среднее арифметическое значение:

$$y_i = \frac{\sum_{L=1}^L y_L}{L} \quad (3)$$

Более точное значение стока в i -ом году при разной точности уравнений следует определять с весовыми коэффициентами, обратно пропорциональными дисперсиям погрешностей уравнений регрессии:

$$y_i = \frac{\sum_{L=1}^L y_L \frac{1}{\sigma_L^2(\text{усл.})}}{\sum_{L=1}^L \frac{1}{\sigma_L^2(\text{усл.})}} \quad (4)$$

где $\sigma_L(\text{усл.})$ - условное стандартное отклонение эмпирических точек от уравнения регрессии.

$$\sigma_{L(\text{усл.})} = \sigma_{L(\text{безусл.})} \sqrt{1-R^2} \quad (5)$$

или

$$\sigma_{L(y_{ур})} = \sqrt{\frac{\sum_1^k (y_j - y_{ур})^2}{k-1}}, \quad (6)$$

$y_{ур}$ – значение стока, рассчитанное по уравнению;

$\sigma_{L(безусл.)}$ - безусловное стандартное отклонение эмпирических точек в расчетном году

$$\sigma_{L(безусл)} = \sqrt{\frac{\sum_1^k (y_k - \bar{y})^2}{k-1}}. \quad (6a)$$

По восстановленным годовым значениям за многолетний период гидрометеорологических характеристик можно рассчитать параметры (норма, коэффициенты вариации, асимметрии и автокорреляции), а также квантили распределения. Предложенная схема восстановления может применяться не только для кратковременных наблюдений за гидрологической характеристикой от одного до пяти лет, а и для более продолжительных наблюдений. В основе данного способа восстановления годовых значений, нормы и квантилей распределения лежит пространственная связанность исследуемой гидрометеорологической характеристики. Метод годовых уравнений регрессии, как правило, дает меньшие стандартные погрешности, чем метод отношений и рекомендуется в качестве основного метода расчета гидрологических характеристик при наличии кратковременных наблюдений. Для построения кривой обеспеченности требуется знание трех параметров: нормы, коэффициентов вариации и асимметрии или отношения коэффициента асимметрии к коэффициенту вариации. Это соотношение рекомендуется рассчитывать на основе групповой оценки в однородном гидрологическом районе на основе индивидуальных оценок, получаемых по пунктам-аналогам. В качестве окончательных значений восстановленных рядов гидрометеорологических характеристик следует

использовать средние арифметические значения за каждый год полученные по каждому из рассмотренных методов расчета. При этом осреднение результатов расчета допускается осуществлять с весовыми коэффициентами обратно пропорциональными дисперсиям стандартных погрешностей:

$$y = \frac{\sum_{i=1}^k \frac{1}{\sigma_i^2} y_i}{\sum_{i=1}^k \frac{1}{\sigma_i^2}}, \quad (7)$$

где y_i - значение рассматриваемой гидрологической характеристики, определенное по различным методам, σ_i^2 - абсолютные дисперсии погрешностей для каждого метода k - число методов, в данном случае используется два метода. Возможно, в будущем будут разработаны другие методы расчета. Поэтому данная формула представлена в общем виде. Один из методов оценки коэффициентов вариации с учетом кратковременных наблюдений основан на решении уравнений регрессии вида:

$$y_i - \bar{y} = r_{xy} \frac{\sigma_y}{\sigma_x} (x_i - \bar{x}), \quad (8)$$

в котором y_i - наблюдаемые данные в пункте проектирования, y - неизвестная норма стока в пункте проектирования, x_i - сток в пункте-аналоге за i - год, x - норма стока в пункте-аналоге, приведенная к многолетнему периоду, σ_y и σ_x - средние квадратические отклонения в исследуемом пункте и пункте-аналоге. В уравнении (8) не известны y , σ_y , x_y , r . Коэффициент корреляции xy r получаем с помощью пространственной корреляционной функции рассматриваемой гидрологической характеристики, зная расстояние между центрами тяжести исследуемого водосбора и водосбора-аналога. Норму стока определяем одним из способов, рекомендуемых ранее. Подставив в уравнение (8) значения y и x_y , r , получим значение стандартного отклонения σ_y и рассчитаем

коэффициент вариации $C_v = \sigma_y/y$. При наличии двух лет наблюдений будем иметь два уравнения с двумя неизвестными, тогда можно определить y и σ_y . Число уравнений регрессии может быть увеличено за счет привлечения нескольких аналогов. Во всех случаях стандартные относительные погрешности расчета коэффициентов вариации уменьшается с увеличением числа лет наблюдений. При расчетах коэффициента вариации можно рекомендовать два-три аналога, при увеличении числа аналогов более трех случайные стандартные погрешности практически не уменьшаются, а в некоторых случаях даже возрастают. При использовании одного аналога при числе лет наблюдений два года стандартные погрешности расчета коэффициента вариации велики и не могут быть рекомендованы в качестве расчетных.

Регрессионный анализ — это набор статистических методов оценки отношений между переменными. Его можно использовать для оценки степени взаимосвязи между переменными и для моделирования будущей зависимости. По сути, регрессионные методы показывают, как по изменениям «независимых переменных» можно зафиксировать изменение «зависимой переменной».

Зависимую переменную в бизнесе называют предиктором (характеристика, за изменением которой наблюдают). Это может быть уровень продаж, риски, ценообразование, производительность и так далее. Независимые переменные — те, которые могут объяснять поведение выше приведенных факторов (время года, покупательная способность населения, место продаж и многое другое). Регрессионный анализ включает несколько моделей. Наиболее распространенные из них: линейная, мультилинейная (или множественная линейная) и нелинейная.

Как видно из названий, модели отличаются типом зависимости переменных: линейная описывается линейной функцией; мультилинейная также представляет линейную функцию, но в нее входит больше параметров (независимых переменных); нелинейная модель — та, в

которой экспериментальные данные характеризуются функцией, являющейся нелинейной (показательной, логарифмической, тригонометрической и так далее).

Чаще всего используются простые линейные и мульти линейные модели.

Регрессионный анализ предлагает множество приложений в различных дисциплинах, включая финансы.

Рассмотрим поподробнее принципы построения и адаптации результатов метода.

Предположения линейной модели

Линейный регрессионный анализ основан на шести фундаментальных предположениях:

- Переменные показывают линейную зависимость;
- Независимая переменная не случайна;
- Значение невязки (ошибки) равно нулю;
- Значение невязки постоянно для всех наблюдений;
- Значение невязки не коррелирует по всем наблюдениям;
- Остаточные значения подчиняются нормальному распределению.

Построение простой линейной регрессии

Простая линейная модель выражается с помощью следующего уравнения:

$$Y = a + bX$$

Где:

- Y — зависимая переменная
- X — независимая переменная (объясняющая)
- a — свободный член (сдвиг по оси OY)
- b — угловой коэффициент. Он указывает на поведение кривой (убывает или возрастает, угол между с осью)

a и b называют коэффициентами линейной регрессии. В их нахождении и заключается основная задача.

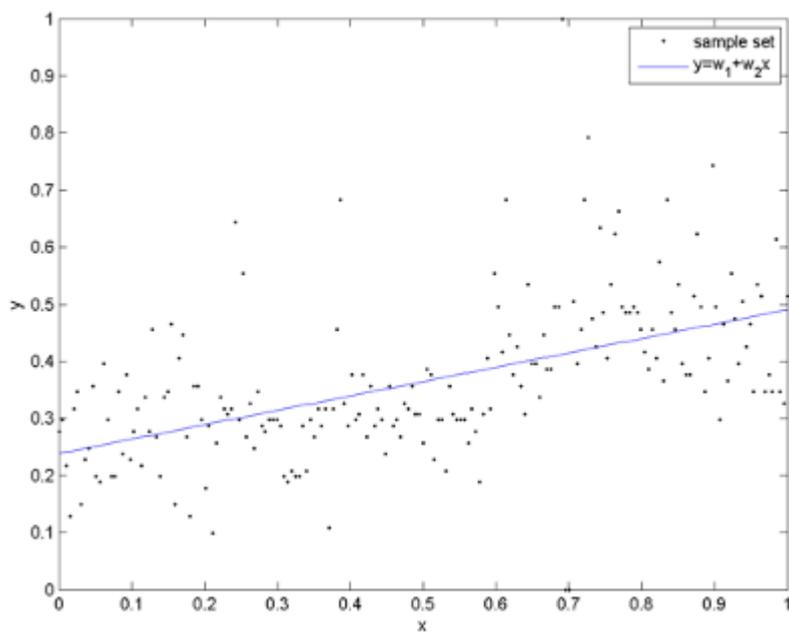


Рисунок 1.1 – Линия линейной регрессии

$$Y = a + b_1 x_1 + b_2 x_2 + b_3 x_3$$

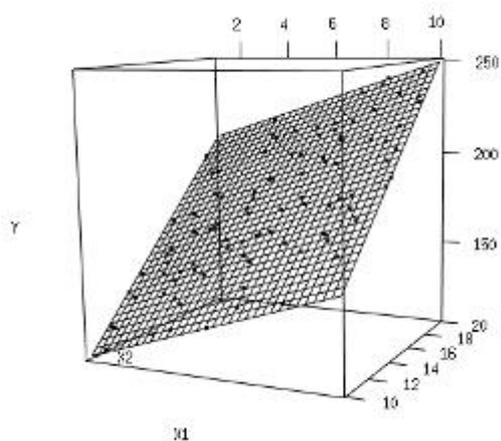


Рисунок 1.2 – Множественная регрессия

Существует много способов определить коэффициенты a и b . Но самым простым и надежным является метод наименьших квадратов (можно научно доказать, что это лучший способ).

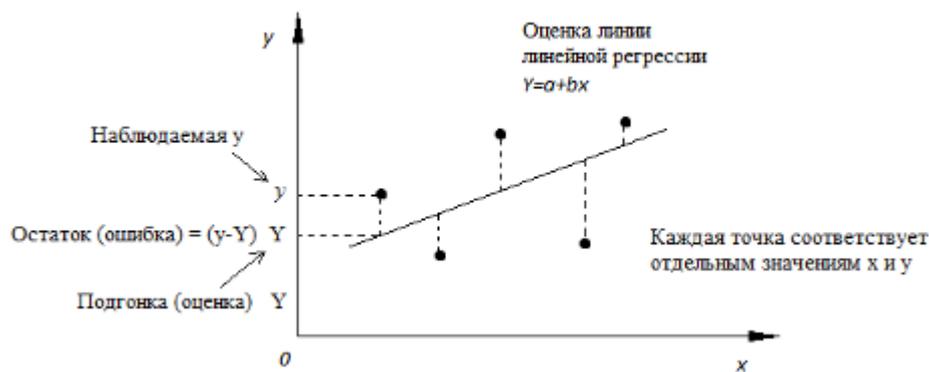


Рисунок 1.3 – Линия линейной регрессии. Пунктиром изображено расстояние $y - Y$ для каждой точки.

1.2 Прогностические возможности регрессионных моделей

Структура связи

Связи бывают корреляционные и причинные (каузальные). В каузальных связях присутствуют причина (предикторы) и следствие (предиктанты).

Оценкой могут выступать (значимость коэффициентов возрастает от 4-ого к 1-ому):

- 1) коэффициент линейной корреляции (коэффициент Пирсона);
- 2) коэффициенты ранговой корреляции (оценка порядковых переменных);
- 3) коэффициенты Φ (фи) и Юла (для дихотомических переменных);
- 4) коэффициент Крамера (номинальные переменные – градация).

Самым важным является коэффициент Пирсона. Сравниваются (коррелируются) только одинаковые по рангу переменные. При этом связь между переменными может быть прямой и обратной.

Свойства коэффициента корреляции ($r_{i,j}$)

- 1) коэффициент корреляции симметричен: $r_{i,j} = r_{j,i}$
- 2) для направленных связей коэффициент корреляции: $-1 \leq r_{i,j} \leq 1$
для ненаправленных: $0 \leq r_{i,j} \leq 1$
- 3) если связь прямая: $r_{i,j} > 0$

обратная: $r_{i,j} < 0$

4) полная связь: $r_{i,j} = \pm 1$

Формула для определения коэффициента корреляции:

$$r_{i,j} = \frac{\sum_1^n (x_i - \bar{x}_i)(x_j - \bar{x}_j)}{\sigma_{x_i} \sigma_{x_j}}$$

где x_i и x_j – переменные i -ого и j -ого рядов;

\bar{x}_i и \bar{x}_j – средние значения i -ого и j -ого рядов;

σ_{x_i} и σ_{x_j} – СКО i -ого и j -ого рядов.

Таблица 1.1 – Матрица корреляции

| Переменная | X ₁ | X ₂ | X ₃ | ... | X _n |
|----------------|-----------------|-----------------|-----------------|-----|-----------------|
| X ₁ | 1 | r ₂₁ | r ₃₁ | ... | r _{n1} |
| X ₂ | r ₁₂ | 1 | r ₃₂ | ... | r _{n2} |
| X ₃ | r ₁₃ | r ₂₃ | 1 | | r _{n3} |
| ... | ... | ... | ... | ... | ... |
| X _n | r _{1n} | r _{2n} | r _{3n} | ... | 1 |

Матрица корреляции – парная корреляция, но для совокупности данных.

Матрица симметричная относительно диагонали.

Граф корреляции

Граф – геометрическая схема из точек, которые соединены линиями и/или стрелками. Точки – вершины графа, линии – рёбра графа, стрелки – дуги графа. Дуги рисуются, когда устанавливается причинная связь (теоретически). Основными свойствами графа является: 1) информативность и 2) прозрачность.

Причинная связь

Причинная связь – связь, в которой одни переменные интерпретируются как причины, другие – как следствия. Переменная, измеряющая причину, называется независимой (предиктором).

В философии: причина неизбежно влечёт следствие; в статистике: оценивается вероятность наступления следствия.

Критерии причинности (необходимые, но недостаточные условия):

- 1) причина должна по времени предшествовать следствию;
- 2) между причиной и следствием должна быть корреляционная связь;
- 3) на взаимосвязь причины и следствия не должны влиять третьи факторы.

Задачи регрессионного анализа:

- объяснение изменений независимой переменной под воздействием одной или нескольких независимых переменных;
- прогнозирование наиболее вероятного значения независимой переменной, если значения предикторов известны.

Этапы регрессионного анализа

- 1) спецификация модели;
- 2) построения модели;
- 3) интерпретация модели;
- 4) проверка гипотез о характере влияния переменных на зависимую;
- 5) прогнозирование значения зависимой переменной;
- 6) оценка качества модели.

Спецификация модели включает:

- выбор модели;
- определение зависимой и независимой переменных;
- подготовка, при необходимости, фиктивных переменных.

Модель множественной линейной регрессии

Нестандартизированное уравнение:

$$y = \sum_{i=1}^k b_i x_i + b_0$$

Стандартизированное уравнение:

$$z_y = \sum_{i=1}^k \beta_i z_i$$

где z_y – стандартизированная переменная y ;

$z_i (i = \overline{1, k})$ – стандартизированная переменная x ;

$\beta_i (i = \overline{1, k})$ – стандартизированный коэффициент регрессии, используемый для оценки силы влияния независимой переменной на зависимую.

Например, если $|\beta_1| > |\beta_2|$, то x_1 оказывает на y большее влияние, чем x_2 . Не стандартизированный коэффициент (b_i) таким свойством не обладает.

Стандартизация данных производится по уравнению:

$$z_i = \frac{(x_i - \bar{x})}{\sigma_x}$$

где \bar{x} – отдельная величина, рассматриваемой переменной;

σ_x – СКО рассматриваемой переменной.

При этом: $\sigma_z = 1$ и $\bar{z} = 0$.

Для определения β применяется метод наименьших квадратов: первые производные по всем параметрам модели регрессии равны нулю и получается система из k уравнений:

Систему решают относительно β_1, \dots, β_k .

$$\beta_1 + r_{1,2}\beta_2 + r_{1,3}\beta_3 + \dots + r_{1,k}\beta_k = r_{1,y}$$

$$r_{2,1}\beta_1 + \beta_2 + r_{2,3}\beta_3 + \dots + r_{2,k}\beta_k = r_{2,y}$$

$$r_{3,1}\beta_1 + r_{3,2}\beta_2 + \beta_3 + \dots + r_{3,k}\beta_k = r_{3,y}$$

...

$$r_{k,1}\beta_1 + r_{k,2}\beta_2 + r_{k,3}\beta_3 + \dots + \beta_k = r_{k,y}$$

$r_{j,i}$ и $r_{i,j}$ ($j, i = \overline{1, k}$) – коэффициенты парной корреляции Пирсона.

b_i – нестандартизированные коэффициенты ($i = \overline{1, k}$):

$$b_i = \frac{\beta_i \sigma_y}{\sigma_i}$$

Свободный член уравнения регрессии:

$$b_0 = \bar{y} - \sum_{i=1}^k b_i \bar{x}_i$$

$b_0 = y$ при $x_i = 0$ (если x_i может принимать такие значения).

σ_i – СКО x_i .

Нестандартизированный коэффициент (b_i) показывает на какую величину в среднем изменится значение зависимой переменной y при увеличении значения соответствующей независимой переменной x_i на 1. Наличие прямой и обратной связи показывает знак нестандартизированного коэффициента:

- 1) если $b_i > 0$, то зависимая переменная y увеличивается на величину $|b_i|$;
- 2) если $b_i < 0$, то зависимая переменная y уменьшается на величину $|b_i|$.

Статистическая значимость коэффициента регрессии заключается в том, что его значение для генеральной совокупности отличается от 0 с заданной доверительной вероятностью $1 - \alpha$, значения которой могут быть: 0,99; 0,95; 0,9 ($\alpha = 0,01; 0,05; 0,1$ соответственно уровень значимости).

Проверка статистической значимости коэффициента регрессии:

- 1) построение доверительного интервала для коэффициента b_i с использованием значения его стандартной ошибки;
- 2) использование коэффициента Стьюдента и его значимости.

Прогнозирование по уравнению регрессии

Зная независимую переменную ($x_i [i = \overline{1, k}]$) определяется:

$$\hat{y} = \sum_{i=1}^k b_i x_i + b_0$$

Для оценки качества модели вычисляется среднеквадратическая погрешность (S).

$$S_y^2 = S_{об}^2 + S_{ост}^2$$

где S_y – полная дисперсия зависимой переменной;

$S_{об}$ – дисперсия, объяснённая влиянием набора независимых переменных (x_1, x_2, \dots, x_k);

$S_{ост}$ – остаточная дисперсия, порождённая неучтёнными факторами.

Показателем качества является отношение: $S_{об}^2/S_y^2$.

$$S_{об}^2/S_y^2 = \frac{\sum_{i=1}^n (\hat{y}_i - \bar{y})^2}{\sum_{i=1}^n (y_i - \bar{y})^2} = R^2$$

Можно оценивать R^2 , включая поочерёдно независимые переменные. Можно оценивать вклад каждого предиктора в объяснение дисперсии зависимой переменной y . Но при этом необходимо помнить о мультиколлинеарности (может повлиять порядок включения переменных).

1.3 Алгоритм увеличения эффективности прогнозных регрессионных моделей

Работа начинается со сбора исходных данных, предикторов. Был взят северно-западный регион России в качестве поиска материалов для написания работы. Предиктантами служили расходы воды 2018,2019,2020 годов , в свою очередь предикторами являлись расходы с1955 по 1980 год и с 2008 по 2017, а так же метеорологические характеристики за тот же период времени, а именно: максимальные, минимальные и средние температуры, осадки, снегозапасы, температуры на глубинах почвы. В качестве гидропостов и метеостанций служили Ловать - Р.ЛОВАТЬ - Г.ВЕЛИКИЕ ЛУКИ и метеостанция Великие Луки с кодами №72454 и №26477, гидропост Сороть (Сорать) - Р. СОРОТЬ – Д. ОСИНКИНО и метеостанция Пушкинские горы с кодами №72686 и 26359.

После нахождения гидрометеорологических данных, были построены корреляционные матрицы, проанализировав которые, постановили ненужность максимальных и минимальных температур. Температуры на глубинах тоже показали свою неэффективность, причиной тому служило малое количество данных, что в итоге послужило исключением этих характеристик из работы, оставляя только температура на глубине 80 см, так как этих материалов оказалось больше в сравнении с другими.

Следующий этап заключался в отборе удовлетворяющей корреляции расходов воды от года к году в поиске годов аналогов, так же были задействованы и метеорологические данные.

Последующий путь работы состоял в построении регрессионного уравнения по годам аналогам.

2 Формирование базы данных для исследования

2.1 Выбор речных водосборов

Для выполнения работы были выбраны 2 метеостанции и 2 гидропоста, которые связаны между собой и находятся в непосредственной близости друг к другу, а именно гидропост Ловать - Р.ЛОВАТЬ - Г.ВЕЛИКИЕ ЛУКИ и метеостанция Великие Луки с кодами №72454 и №26477, гидропост Сороть (Сорать) - Р. СОРОТЬ – Д. ОСИНКИНО и метеостанция Пушкинские горы с кодами №72686 и 26359 соответственно. При выборе водосбора так же был учтён главный критерий, при котором площадь водосбора должна находится в пределах от 1500 до 50000 км². Гидропост Ловать имеет $F=3270$ км², а гидропост Сороть $F=3170$ км².

Цель работы заключалась в том, чтобы найти и обработать ежедневные данные по гидропостам и метеостанциям такие как: расход, максимальная, минимальная и средняя температура, осадки, характеристики снежного покрова и температуры почвы на глубинах до 320 см. Временной отрезок с которого брали эту информацию были года с 1955 по 1980 и с 2008 по 2020.

2.2 Анализ исходных гидрометеорологических данных

После поиска, компоновки, обработки всех гидрометеорологических характеристик были построены и проанализированы хронологические графики.



Рисунок 2.1 – Гидрограф гидропоста Ловать

На данном гидрографе мы чётко наблюдаем тенденцию изменения среднесуточных расходов. С течением времени расходы уменьшаются, как мы видим на Рисунок 2.1 до 1980 года расходы гораздо выше по сравнению с 2008 по 2020 года. Так же отчётливо видим выбивающийся из нормы 1973 год в котором аномально большой расход по сравнению с другими годами, его значение равно $Q=308 \text{ м}^3/\text{с}^3$.

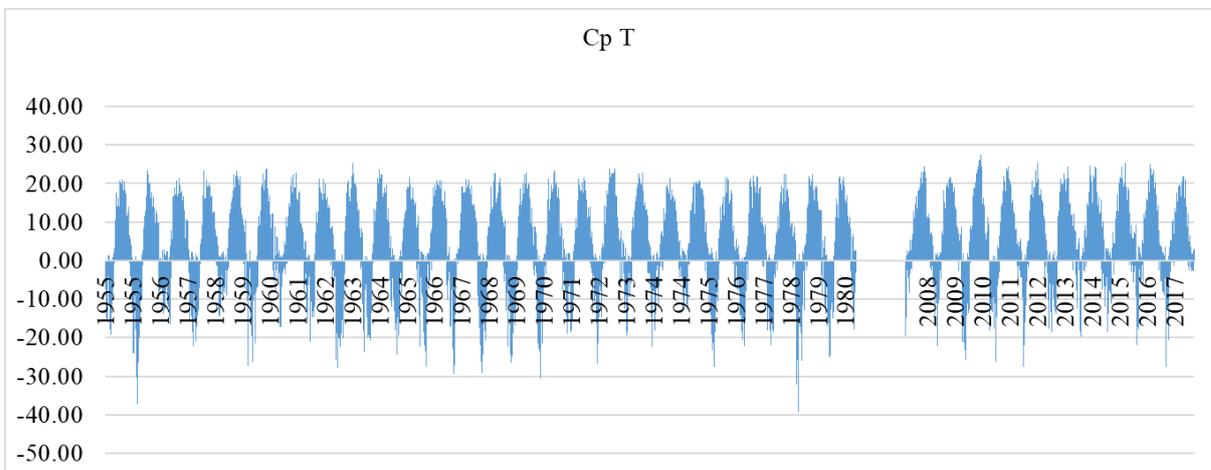


Рисунок 2.2 – Средняя температура по годам метеостанция Великие Луки

На данном рисунке видно, что современные года имеют среднюю температуру выше, чем года в период с 1955 по 1980.

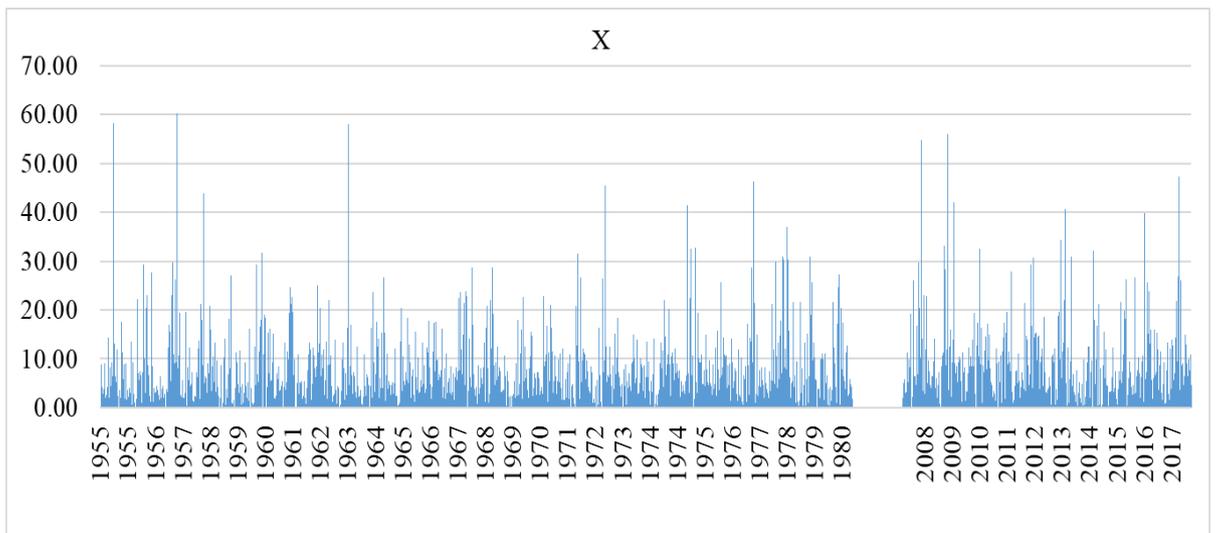


Рисунок 2.3 – Осадки метеостанция Великие Луки

На данном рисунке видно, что современные года имеют осадки ниже, чем года в период с 1955 по 1980.

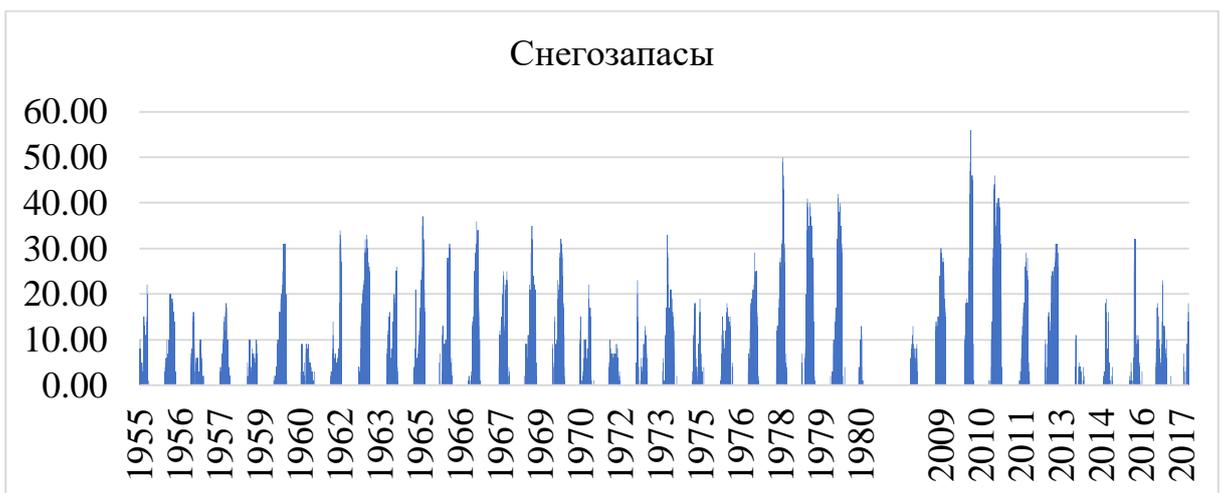


Рисунок 2.4 – Снегозапасы метеостанция Великие Луки

На данном рисунке видно, что современные года имеют снегозапасы ниже, чем года в период с 1955 по 1980.

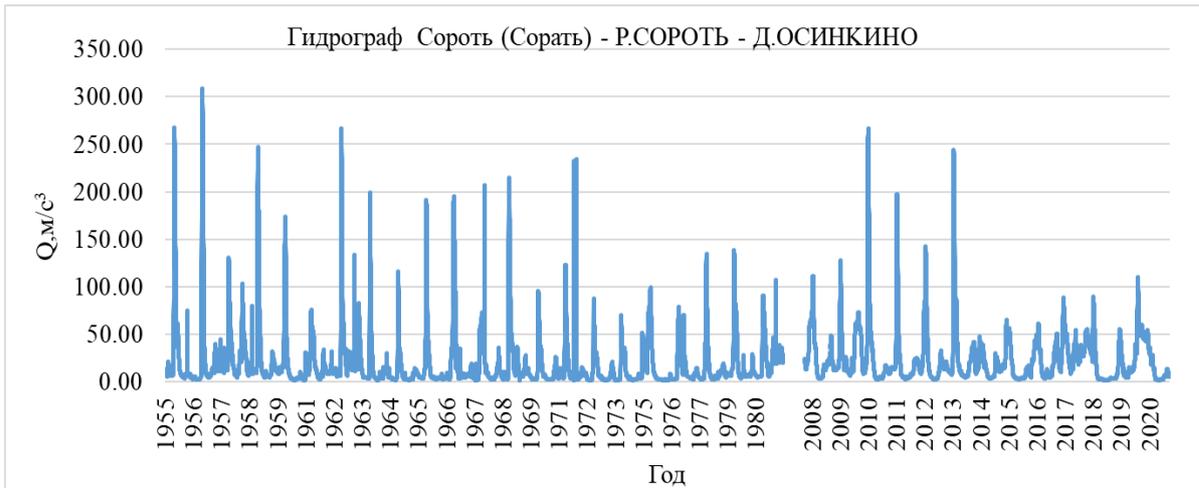


Рисунок 2.5 – Гидрограф Сороть (Сорать) - Р.СОРОТЬ - Д.ОСИНКИНО

Описывая, данный гидрограф, можно сказать, что среднесуточные расходы так же уменьшаются, если сравнивать до и после 1980 года. Но в отличии от Рисунок 2.1 после 2008 года так же наблюдаются высокие расходы, хоть их периодичность и упала.

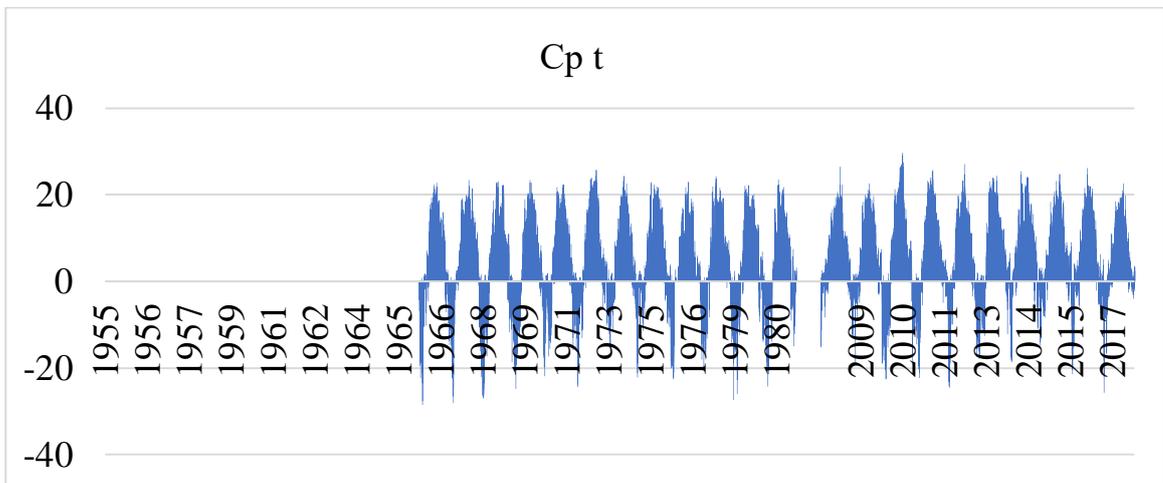


Рисунок 2.6 – Средняя температура по годам метеостанция Пушкинские горы

На данном рисунке видно, что современные года имеют среднюю температуру выше, чем года в период с 1955 по 1980.

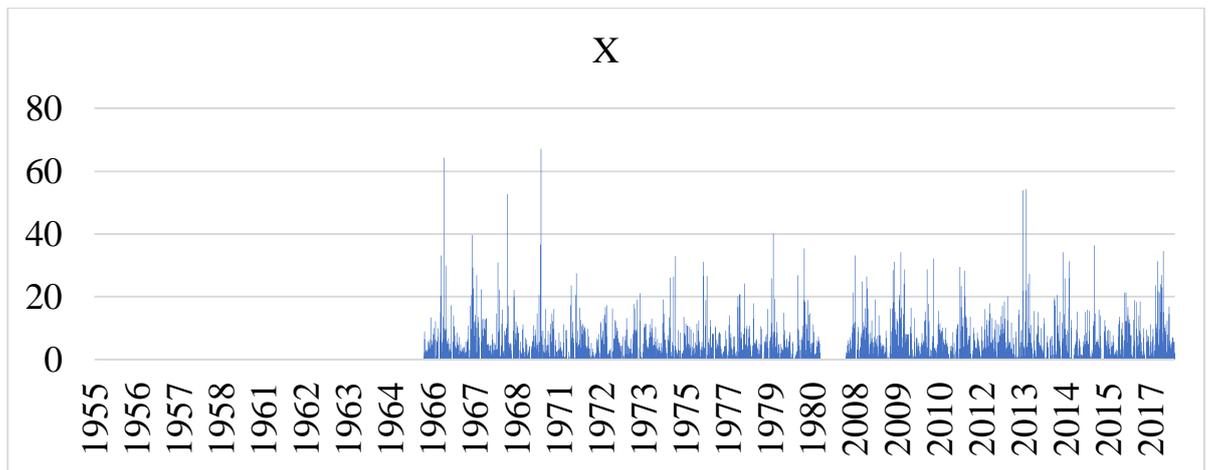


Рисунок 2.7 – Осадки по годам метеостанция Пушкинские горы

На данном рисунке видно, что современные года имеют осадки ниже, чем года в период с 1955 по 1980.

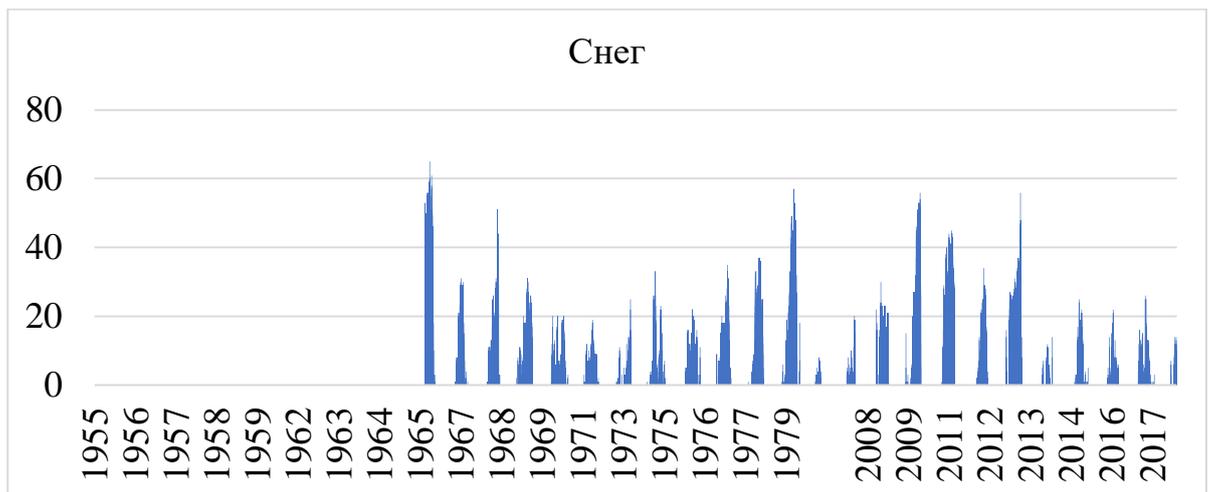


Рисунок 2.8 – Снегозапасы метеостанция Пушкинские горы

На данном рисунке видно, что современные года имеют снегозапасы ниже, чем года в период с 1955 по 1980.

Затем наступает второй этап работы: построение корреляционных матриц с сформированной базой данных.

Таблица 2.1 – Корреляционная матрица гидрометеорологических данных гидропост Ловать - Р.ЛОВАТЬ - Г.ВЕЛИКИЕ ЛУКИ и метеостанция Великие Луки

| | Mint | Cpt | Maxt | X | Q | Снег | 2см | 5см | 10см | 15см | 20см | 40см | 60см | 80см | 120см | 160см | 240см | 320см | |
|-------|--------|--------|--------|--------|--------|--------|--------|--------|--------|--------|--------|--------|--------|--------|--------|--------|-------|-------|--|
| Mint | 1 | | | | | | | | | | | | | | | | | | |
| Cpt | 0.9655 | 1 | | | | | | | | | | | | | | | | | |
| Maxt | 0.9028 | 0.9785 | 1 | | | | | | | | | | | | | | | | |
| X | 0.1832 | 0.1239 | 0.0888 | 1 | | | | | | | | | | | | | | | |
| Q | 0.0373 | 0.038 | 0.0456 | -0.008 | 1 | | | | | | | | | | | | | | |
| Снег | -0.619 | -0.609 | -0.582 | -0.054 | -0.165 | 1 | | | | | | | | | | | | | |
| 2см | 0.8582 | 0.9224 | 0.9308 | 0.0908 | -0.026 | -0.482 | 1 | | | | | | | | | | | | |
| 5см | 0.8324 | 0.8961 | 0.9098 | 0.1067 | -0.02 | -0.472 | 0.9959 | 1 | | | | | | | | | | | |
| 10см | 0.8273 | 0.8879 | 0.9004 | 0.1092 | -0.04 | -0.478 | 0.9922 | 0.9989 | 1 | | | | | | | | | | |
| 15см | 0.8252 | 0.8798 | 0.8912 | 0.1272 | -0.036 | -0.52 | 0.9876 | 0.9967 | 0.9988 | 1 | | | | | | | | | |
| 20см | 0.8325 | 0.8905 | 0.9027 | 0.1151 | -0.116 | -0.564 | 0.9802 | 0.9929 | 0.9968 | 0.9981 | 1 | | | | | | | | |
| 40см | 0.8331 | 0.8773 | 0.8836 | 0.1141 | -0.27 | -0.613 | | | | | 0.9951 | 1 | | | | | | | |
| 60см | 0.7703 | 0.827 | 0.8383 | 0.1224 | -0.227 | -0.631 | | | | | 0.9708 | | 1 | | | | | | |
| 80см | 0.7707 | 0.8111 | 0.8127 | 0.1268 | -0.214 | -0.572 | 0.901 | 0.9365 | 0.9483 | 0.9487 | 0.9623 | 0.9896 | 0.9966 | 1 | | | | | |
| 120см | 0.7336 | 0.7648 | 0.76 | 0.119 | -0.283 | -0.556 | 0.8549 | 0.8968 | 0.9119 | 0.9096 | 0.9232 | 0.9612 | 0.9782 | 0.9913 | 1 | | | | |
| 160см | 0.6725 | 0.6894 | 0.6768 | 0.1181 | -0.333 | -0.538 | 0.7603 | 0.8354 | 0.8599 | 0.8624 | 0.8547 | 0.8909 | 0.9411 | 0.956 | 0.9848 | 1 | | | |
| 240см | 0.5038 | 0.4962 | 0.4698 | 0.0864 | -0.425 | -0.466 | | | | | 0.6444 | 0.679 | 0.8048 | 0.8162 | 0.8864 | 0.953 | 1 | | |
| 320см | 0.3466 | 0.3216 | 0.2865 | 0.0533 | -0.453 | -0.353 | 0.4517 | 0.5213 | 0.5548 | 0.5125 | 0.4762 | 0.5889 | 0.6852 | 0.7777 | 0.863 | 0.9659 | 1 | | |

Таблица 2.2 – Корреляционная матрица гидрометеорологических данных гидропост Сороть (Сорать) - Р. СОРОТЬ – Д. ОСИНКИНО и метеостанция Пушкинские горы

| | Mint | Cpt | Maxt | X | Q | Снег | 2см | 5см | 10см | 15см | 20см | 40см | 60см | 80см | 120см | 160см | 240см | 320см | |
|-------|--------|--------|--------|--------|--------|--------|-----|-----|------|------|--------|--------|--------|--------|-------|--------|-------|-------|---|
| Mint | 1 | | | | | | | | | | | | | | | | | | |
| Cpt | 0.9769 | 1 | | | | | | | | | | | | | | | | | |
| Maxt | 0.9335 | 0.9849 | 1 | | | | | | | | | | | | | | | | |
| X | 0.1467 | 0.0913 | 0.0582 | 1 | | | | | | | | | | | | | | | |
| Q | -0.034 | -0.028 | -0.015 | -0.023 | 1 | | | | | | | | | | | | | | |
| Снег | -0.606 | -0.595 | -0.572 | -0.052 | -0.149 | 1 | | | | | | | | | | | | | |
| 2см | | | | | | | 1 | | | | | | | | | | | | |
| 5см | | | | | | | | 1 | | | | | | | | | | | |
| 10см | | | | | | | | | 1 | | | | | | | | | | |
| 15см | | | | | | | | | | 1 | | | | | | | | | |
| 20см | 0.8679 | 0.8993 | 0.9035 | 0.1058 | -0.173 | -0.621 | | | | | 1 | | | | | | | | |
| 40см | 0.8454 | 0.8669 | 0.8652 | 0.1082 | -0.238 | -0.63 | | | | | 0.9902 | 1 | | | | | | | |
| 60см | | | | | | | | | | | | | 1 | | | | | | |
| 80см | 0.8036 | 0.8184 | 0.8135 | 0.1119 | -0.251 | -0.616 | | | | | 0.9607 | 0.9892 | | 1 | | | | | |
| 120см | | | | | | | | | | | | | | | 1 | | | | |
| 160см | 0.7103 | 0.7056 | 0.688 | 0.1059 | -0.346 | -0.589 | | | | | 0.8629 | 0.9205 | 0.9665 | | | 1 | | | |
| 240см | | | | | | | | | | | | | | | | | | 1 | |
| 320см | 0.4041 | 0.3608 | 0.3203 | 0.0779 | -0.433 | -0.434 | | | | | 0.5149 | 0.6145 | | 0.7118 | | 0.8633 | | | 1 |

Анализируя данные из Таблицы 1 и Таблицы 2, для дальнейших расчётов, будут взяты: средняя температура, осадки, расход, характеристики снежного покрова и температуры почвы на глубине 80 см.

Следующий шаг работы выражается в том, чтобы взять расходы и метеорологические характеристики по 2 выбранным постам по отдельности и построить по ним корреляционную матрицу, по которой будет происходить поиск года аналога, который в свою очередь лучше всего коррелирует с 2018,2019,2020 годами. Сформировать общую таблицу, проанализировать её.

Таблица 2.3 – Подбор года аналога

| год | Ловать | | | | год | Сорочь | | | |
|------|-------------|------------|------------|-------------|------|------------|------------|------------|------------|
| | Q | X | S | T | | Q | X | S | T |
| 2018 | 2014 (0.85) | 1960(0.23) | 2010(0.89) | 1963 (0.88) | 2018 | 1975(0.86) | 2015(0.14) | 1977(0.83) | 2011(0.85) |
| 2019 | 2017 (0.74) | 2016(0.26) | 1967(0.82) | 2010 (0.84) | 2019 | 2009(0.62) | 2009(0.15) | 2015(0.88) | 2010(0.82) |
| 2020 | 2014 (0.83) | 1965(0.27) | 2014(0.73) | 1961 (0.84) | 2020 | 2018(0.80) | 1973(0.14) | 2014(0.42) | 2015(0.80) |

Обработывая информацию Таблица 3, вывод состоит в том, что гидропосты Ловать - Р.ЛОВАТЬ - Г.ВЕЛИКИЕ ЛУКИ и гидропост Сорочь (Сорать) - Р. СОРОТЬ – Д. ОСИНКИНО по 2018,2019 и 2020 году лучше всего соответствуют приведённые года в таблице.

3 Применение разработанного эффективного алгоритма для прогнозов по регрессионным моделям

Было взято 2 методики выявления лучшего метода прогноза для регрессионного уравнения.

Перед этим, со станции Ловать - Р.ЛОВАТЬ - Г.ВЕЛИКИЕ ЛУКИ было проработано 3 варианта.

1 вариант: взять старые года с 1955-1980.

В ходе выполнения работы, по годам в период с 1955 по 1980 года, были построены регрессионные уравнения и в соответствии с результатами сгенерировали график.

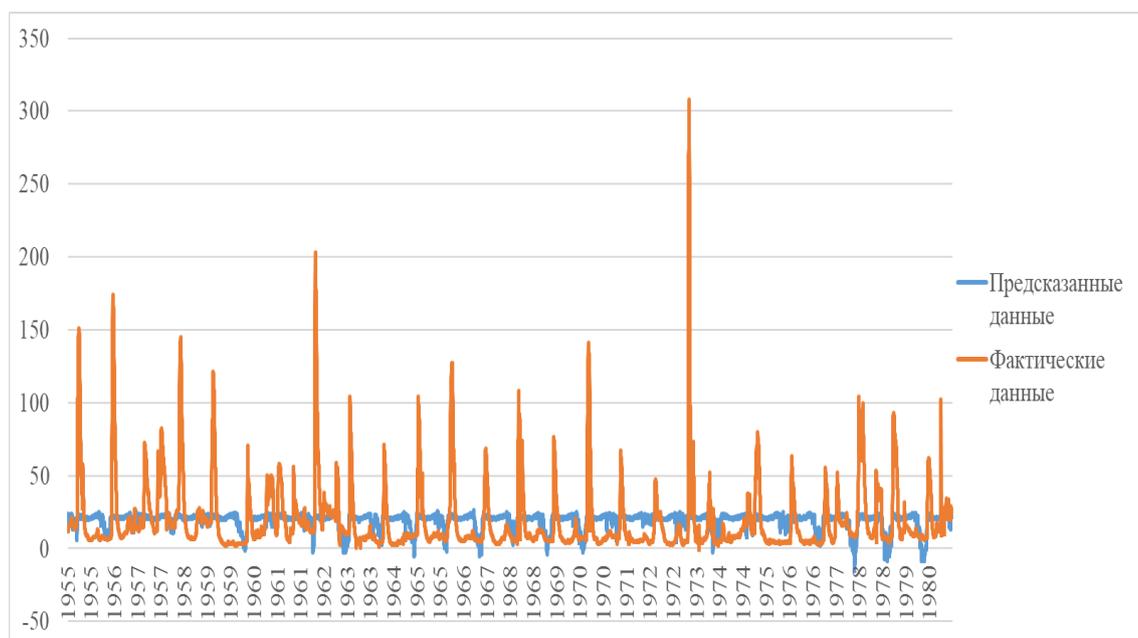


Рисунок 3.1 – Предсказанные и фактические данные с 1955 по 1980 г.

Изучая данный график, приходим к выводу, что в данном варианте, результат неудовлетворительный и показывает не дееспособность данного метода.

2 вариант: взять современные года с 2008-2020

Второй вариант отличался от первого, тем, что взамен старым годам, были взяты современные в период с 2008 по 2017 года.

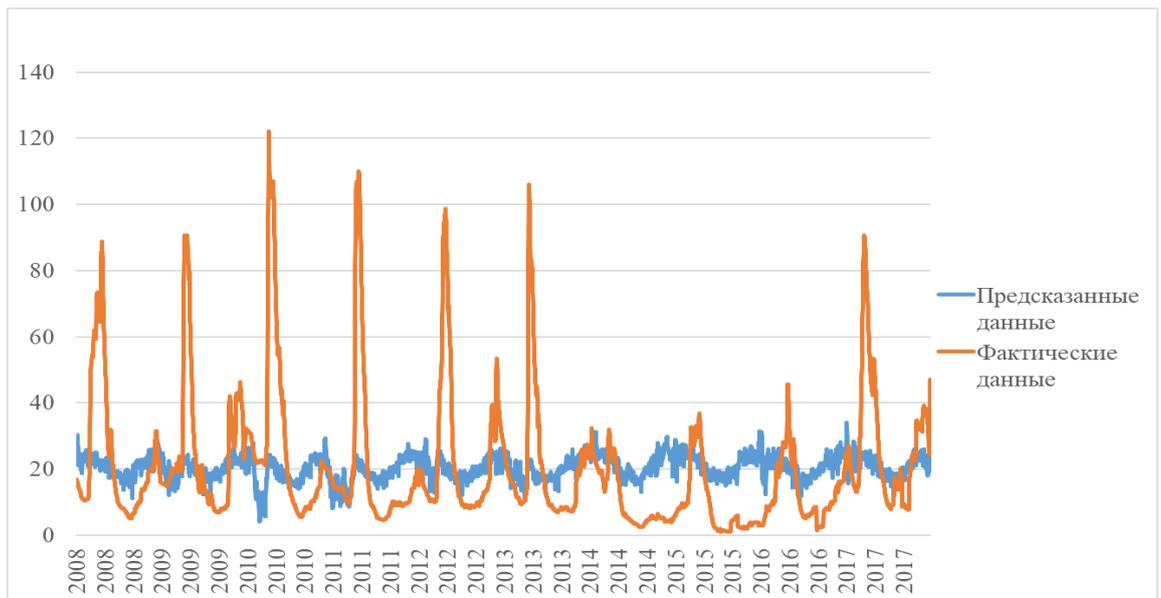


Рисунок 3.2 – Предсказанные и фактические данные с 2008 по 2017 г.

Так же, как и на предыдущем рисунке, видно колоссальное несоответствие фактических и предсказанных значений.

3 вариант: взять все года

Третий вариант предусматривал компоновку прошлых и современных лет. Соответствуя первым и вторым вариантам, были проведены точно такие же операции с расчётами и построен аналогичный график.

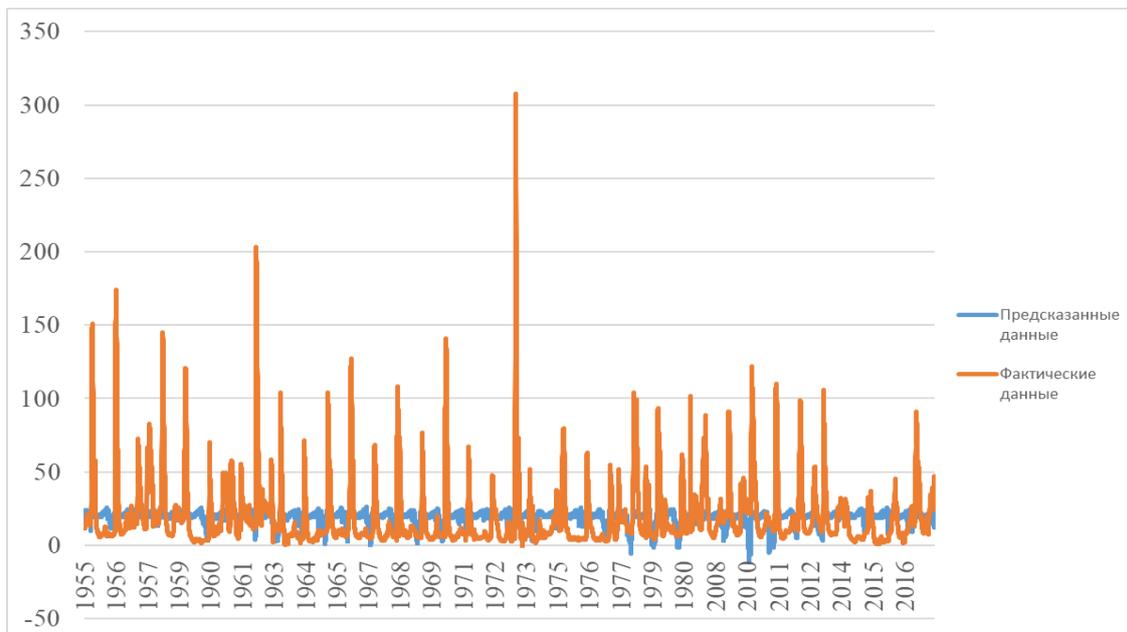


Рисунок 3.2 – Предсказанные и фактические данные с 1955 по 1980 и с 2008 по 2017 г.

На графике чётко видно, как не соответствуют предсказанные и фактические данные.

Посмотрев на все 3 рисунка вывод напрашивается сам собой, можно с уверенностью сказать, что данные варианты предсказания недееспособны.

Методика № 1: Лучший год.

Для выполнения первой методики нам понадобятся предикторы годов аналогов, только по расходу, а роль предиктантов выполняют 2018, 2019 и 2020 года. Первый в очереди будет рассмотрен гидропост Ловать - р.Ловать - г.Великие Луки и метеостанция Великие Луки. Исходя из Таблица 2.3 следует то, что предикторами 2018 году, является 2014 год, по нему мы и будем работать. Брать значение не только расходов, но метеорологических данных тоже по 2014 году. Произведя расчёты, в ходе которых, использовалась регрессия, получаем следующие результаты.

Таблица 3.1 – Регрессионный анализ р. Ловать, 2018 год

| ВЫВОД ИТОГОВ | | | | | | | | |
|---------------------------------|------------------|----------------|-------------------|--------------------|---------------------|--------------------|-------------------|--------------------|
| 2018 | | | | | | | | |
| <i>Регрессионная статистика</i> | | | | | | | | |
| Множестве | 0.8638061 | | | | | | | |
| R-квадрат | 0.7461611 | | | | | | | |
| Нормирова | 0.7433406 | | | | | | | |
| Стандартна | 8.6716482 | | | | | | | |
| Наблюдени | 365 | | | | | | | |
| <i>Дисперсионный анализ</i> | | | | | | | | |
| | <i>df</i> | <i>SS</i> | <i>MS</i> | <i>F</i> | <i>Значимость F</i> | | | |
| Регрессия | 4 | 79575.64 | 19893.91 | 264.56 | 8.9498E-106 | | | |
| Остаток | 360 | 27071.09 | 75.197483 | | | | | |
| Итого | 364 | 106646.7 | | | | | | |
| <i>Коэффициент</i> | | | | | | | | |
| <i>Y-пересече</i> | <i>Статистик</i> | <i>Значени</i> | <i>Нижние 95%</i> | <i>Верхние 95%</i> | <i>Нижние 95%</i> | <i>Верхние 95%</i> | <i>Нижние 95%</i> | <i>Верхние 95%</i> |
| Y-пересече | -3.4838083 | 1.067323 | -3.264063 | 0.0012 | -5.582778567 | -1.384838 | -5.582779 | -1.384838 |
| Переменная | 1.8402789 | 0.059668 | 30.842208 | 4E-103 | 1.722938144 | 1.9576196 | 1.7229381 | 1.9576196 |
| Переменная | -0.0814074 | 0.129719 | -0.627569 | 0.5307 | -0.336508906 | 0.1736942 | -0.336509 | 0.1736942 |
| Переменная | 0.0408568 | 0.222052 | 0.1839966 | 0.8541 | -0.395824728 | 0.4775382 | -0.395825 | 0.4775382 |
| Переменная | 0.3232193 | 0.057144 | 5.6561815 | 3E-08 | 0.210840437 | 0.4355981 | 0.2108404 | 0.4355981 |

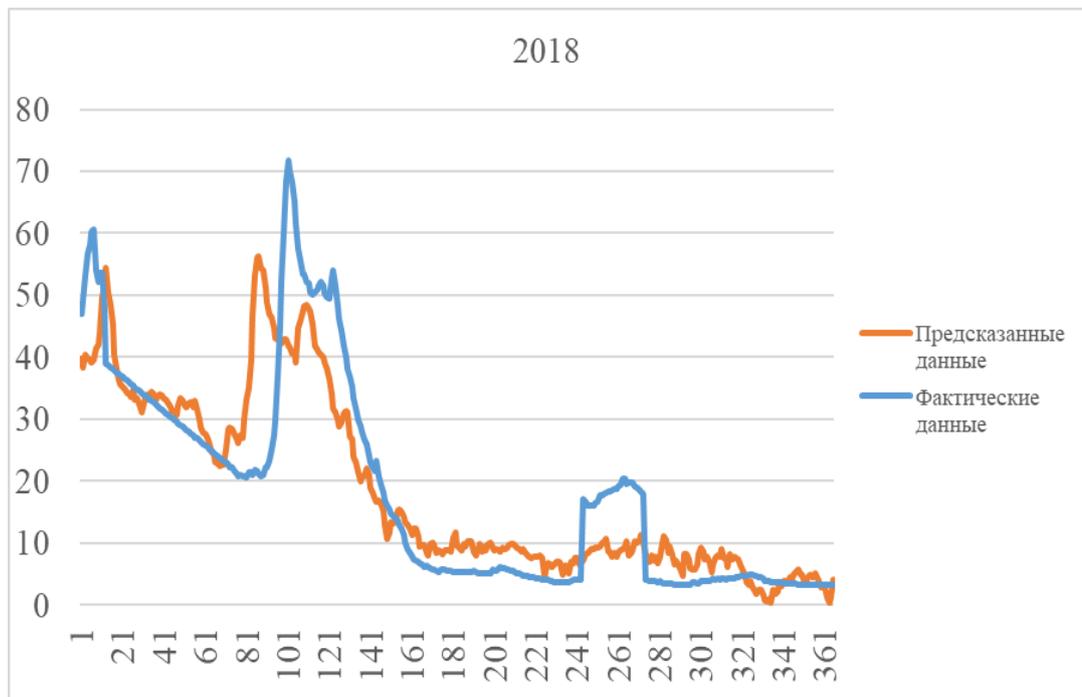


Рисунок 3.3 – Сопоставление графиков фактических значений и модельных, р. Ловать, 2018 год.

Таким же способом работаем и с 2019, 2020 годами. Для первого, предиктором будет служить 2017 год, а для второго, так же будет использован 2014 год, так как именно этот период, показал лучшую корреляцию.

Таблица 3.2 – Регрессионный анализ р. Ловать, 2019 год

| ВЫВОД ИТОГОВ | | | | | | | | |
|--|------------|-----------|-----------|----------|---------------------|-----------|-----------|-----------|
| 2019 | | | | | | | | |
| Регрессионная статистика | | | | | | | | |
| Множестве | 0.7433664 | | | | | | | |
| R-квадрат | 0.5525936 | | | | | | | |
| Нормирова | 0.5476225 | | | | | | | |
| Стандартна | 5.6643189 | | | | | | | |
| Наблюдени | 365 | | | | | | | |
| Дисперсионный анализ | | | | | | | | |
| | <i>df</i> | <i>SS</i> | <i>MS</i> | <i>F</i> | <i>Значимость F</i> | | | |
| Регрессия | 4 | 14265.98 | 3566.495 | 111.159 | 1.34404E-61 | | | |
| Остаток | 360 | 11550.423 | 32.084509 | | | | | |
| Итого | 364 | 25816.403 | | | | | | |
| Коэффициент стандартная остататистик-Значени Нижние 95% Верхние 95% Нижние 95.0 Верхние 95.0 | | | | | | | | |
| Y-пересече | 4.873608 | 0.8309081 | 5.8653998 | 1E-08 | 3.239564619 | 6.5076515 | 3.2395646 | 6.5076515 |
| Переменная | 0.3046894 | 0.0173854 | 17.52561 | 3.7E-50 | 0.270499717 | 0.338879 | 0.2704997 | 0.338879 |
| Переменная | 0.004933 | 0.0605045 | 0.0815303 | 0.93507 | -0.114053787 | 0.1239197 | -0.114054 | 0.1239197 |
| Переменная | -0.2162039 | 0.076611 | -2.822102 | 0.00504 | -0.366865154 | -0.065543 | -0.366865 | -0.065543 |
| Переменная | -0.1148439 | 0.0462178 | -2.484841 | 0.01341 | -0.205734627 | -0.023953 | -0.205735 | -0.023953 |

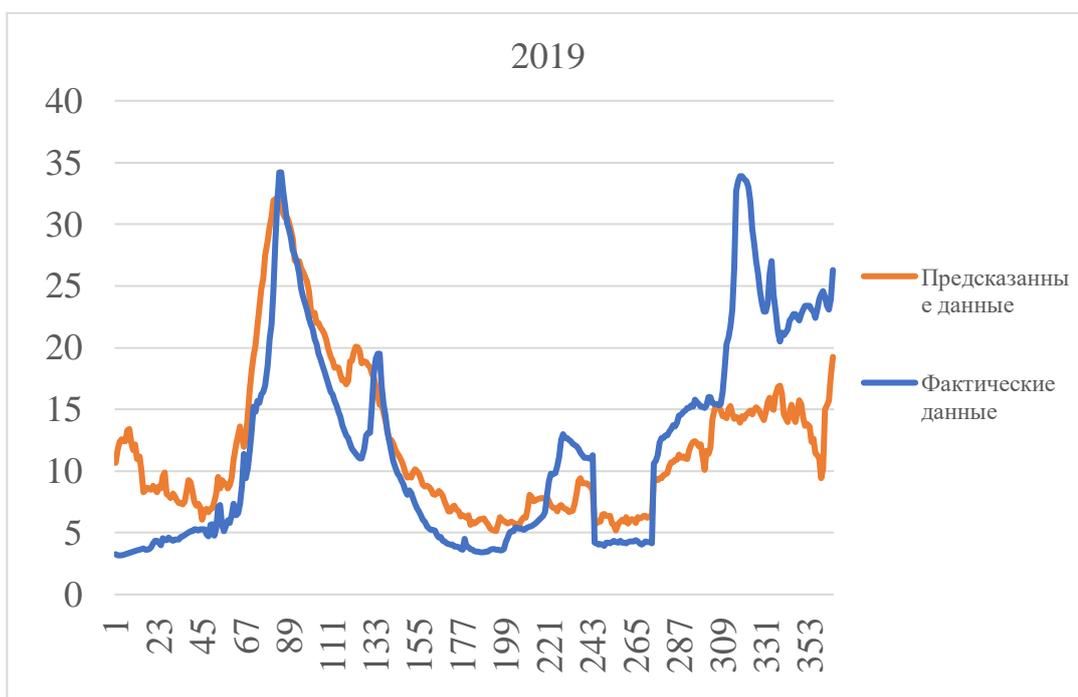


Рисунок 3.4 – Сопоставление графиков фактических значений и модельных, р. Ловать, 2019 год.

Таблица 3.3 – Регрессионный анализ р. Ловать, 2020 год

| ВЫВОД ИТОГОВ | | | | | | | | |
|--|------------|-----------|-----------|--------|--------------|-----------|-----------|-----------|
| 2020 | | | | | | | | |
| Регрессионная статистика | | | | | | | | |
| Множестве | 0.8509851 | | | | | | | |
| R-квадрат | 0.7241757 | | | | | | | |
| Нормирова | 0.721111 | | | | | | | |
| Стандартна | 6.355916 | | | | | | | |
| Наблюдени | 365 | | | | | | | |
| Дисперсионный анализ | | | | | | | | |
| | df | SS | MS | F | Значимость F | | | |
| Регрессия | 4 | 38183.009 | 9545.7524 | 236.29 | 2.7058E-99 | | | |
| Остаток | 360 | 14543.16 | 40.397668 | | | | | |
| Итого | 364 | 52726.17 | | | | | | |
| Коэффициент стандартная ошибка статистическое значение Нижние 95% Верхние 95% Нижние 95.0 Верхние 95.0 | | | | | | | | |
| Y-пересече | 5.6221107 | 0.7822979 | 7.1866622 | 4E-12 | 4.083662869 | 7.1605586 | 4.0836629 | 7.1605586 |
| Переменная | 1.042981 | 0.0437335 | 23.848536 | 4E-76 | 0.956975706 | 1.1289863 | 0.9569757 | 1.1289863 |
| Переменная | -0.0579981 | 0.0950778 | -0.610007 | 0.5422 | -0.244975736 | 0.1289795 | -0.244976 | 0.1289795 |
| Переменная | -0.1533587 | 0.1627536 | -0.942275 | 0.3467 | -0.473425956 | 0.1667086 | -0.473426 | 0.1667086 |
| Переменная | -0.2351247 | 0.0418842 | -5.613682 | 4E-08 | -0.317493142 | -0.152756 | -0.317493 | -0.152756 |

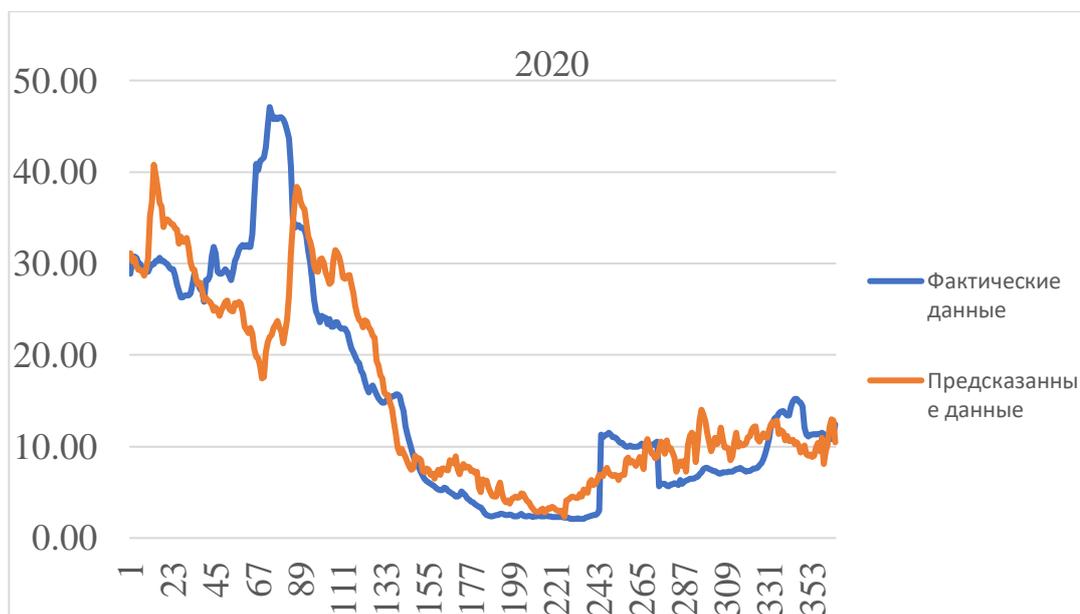


Рисунок 3.4 – Сопоставление графиков фактических значений и модельных, р. Ловать, 2020 год.

Следующий объектом использования 1 методики будет гидропост Сороть (Сорать) - Р. СОРОТЬ – Д. ОСИНКИНО и метеостанция Пушкинские горы.

Таблица 3.3 – Регрессионный анализ р. Сороть, 2018 год

| ВЫВОД ИТОГОВ | | | | | | | | |
|--|-----------|-----------|------------|--------|--------------|-----------|-----------|-----------|
| 2018 | | | | | | | | |
| Регрессионная статистика | | | | | | | | |
| Множестве | 0.8841251 | | | | | | | |
| R-квадрат | 0.7816773 | | | | | | | |
| Нормирова | 0.7792514 | | | | | | | |
| Стандартна | 11.005078 | | | | | | | |
| Наблюдени | 365 | | | | | | | |
| Дисперсионный анализ | | | | | | | | |
| | df | SS | MS | F | Значимость F | | | |
| Регрессия | 4 | 156105.14 | 39026.285 | 322.23 | 1.5459E-117 | | | |
| Остаток | 360 | 43600.224 | 121.11173 | | | | | |
| Итого | 364 | 199705.36 | | | | | | |
| Коэффициент стандартная остататистик-Значени Нижние 95% Верхние 95% Нижние 95.0 Верхние 95.0 | | | | | | | | |
| Y-пересече | 3.3956849 | 1.2407165 | 2.7368742 | 0.0065 | 0.955722326 | 5.8356474 | 0.9557223 | 5.8356474 |
| Переменная | 0.7917653 | 0.0237433 | 33.346909 | 4E-112 | 0.745072343 | 0.8384583 | 0.7450723 | 0.8384583 |
| Переменная | -0.220354 | 0.1614171 | -1.3651197 | 0.1731 | -0.537792562 | 0.0970852 | -0.537793 | 0.0970852 |
| Переменная | 0.8440022 | 0.1426328 | 5.9173082 | 8E-09 | 0.56350406 | 1.1245004 | 0.5635041 | 1.1245004 |
| Переменная | -0.135419 | 0.0838742 | -1.614544 | 0.1073 | -0.300363396 | 0.0295263 | -0.300363 | 0.0295263 |

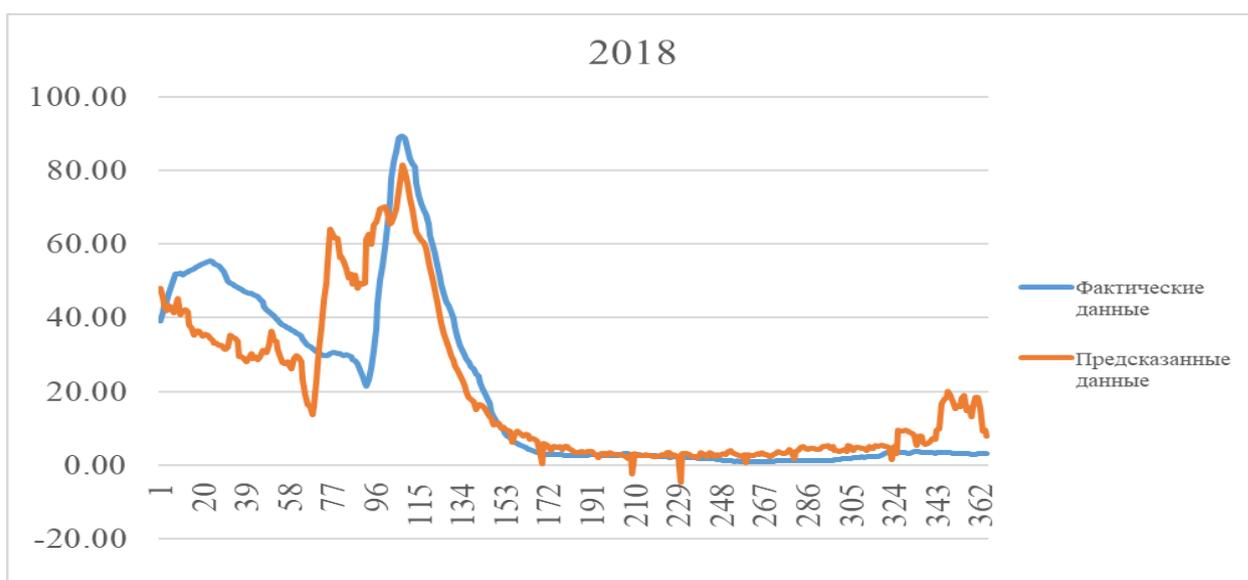


Рисунок 3.5 – Сопоставление графиков фактических значений и модельных, р. Сорочь, 2018 год.

Таблица 3.4 – Регрессионный анализ р. Сорочь, 2019 год

| ВЫВОД ИТОГОВ | | | | | | | | |
|--|------------|-----------|-----------|----------|---------------------|-----------|-----------|-----------|
| 2019 | | | | | | | | |
| Регрессионная статистика | | | | | | | | |
| Множестве | 0.6529318 | | | | | | | |
| R-квадрат | 0.4263199 | | | | | | | |
| Нормирова | 0.4199457 | | | | | | | |
| Стандартна | 17.610685 | | | | | | | |
| Наблюдени | 365 | | | | | | | |
| Дисперсионный анализ | | | | | | | | |
| | <i>df</i> | <i>SS</i> | <i>MS</i> | <i>F</i> | <i>Значимость F</i> | | | |
| Регрессия | 4 | 82969.96 | 20742.491 | 66.882 | 2.8261E-42 | | | |
| Остаток | 360 | 111649 | 310.13624 | | | | | |
| Итого | 364 | 194619 | | | | | | |
| Коэффициент стандартная ошибка статистика Значение Нижние 95% Верхние 95% Нижние 95.0 Верхние 95.0 | | | | | | | | |
| Y-пересече | 16.80358 | 3.18669 | 5.2730508 | 2E-07 | 10.53671307 | 23.070447 | 10.536713 | 23.070447 |
| Переменная | 0.4386583 | 0.045029 | 9.7416666 | 5E-20 | 0.350105181 | 0.5272113 | 0.3501052 | 0.5272113 |
| Переменная | 0.2554757 | 0.191986 | 1.3307009 | 0.1841 | -0.122078889 | 0.6330302 | -0.122079 | 0.6330302 |
| Переменная | -0.4955729 | 0.187579 | -2.641946 | 0.0086 | -0.864460556 | -0.126685 | -0.864461 | -0.126685 |
| Переменная | -0.8198458 | 0.170349 | -4.812749 | 2E-06 | -1.154849531 | -0.484842 | -1.15485 | -0.484842 |

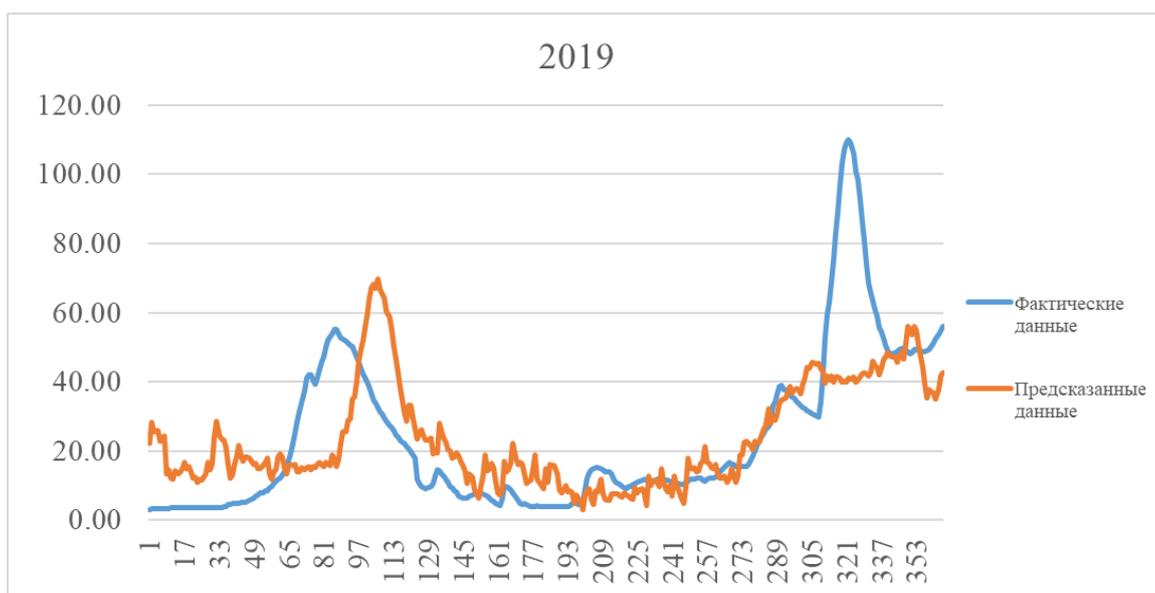


Рисунок 3.6 – Сопоставление графиков фактических значений и модельных, р. Сорочь, 2019 год.

Таблица 3.5 – Регрессионный анализ р. Сорочь, 2020 год

| 2020 | | | | | | | | |
|---|------------|-----------|-----------|----------|---------------------|-----------|-----------|-----------|
| <i>Регрессионная статистика</i> | | | | | | | | |
| Множестве | 0.8763574 | | | | | | | |
| R-квадрат | 0.7680023 | | | | | | | |
| Нормирова | 0.7654246 | | | | | | | |
| Стандартна | 9.3969874 | | | | | | | |
| Наблюдени | 365 | | | | | | | |
| <i>Дисперсионный анализ</i> | | | | | | | | |
| | <i>df</i> | <i>SS</i> | <i>MS</i> | <i>F</i> | <i>Значимость F</i> | | | |
| Регрессия | 4 | 105234.6 | 26308.661 | 297.935 | 8.5273E-113 | | | |
| Остаток | 360 | 31789.21 | 88.303372 | | | | | |
| Итого | 364 | 137023.9 | | | | | | |
| <i>Коэффициент стандартная ошибка статистическое значение Нижние 95% Верхние 95% Нижние 95% Верхние 95%</i> | | | | | | | | |
| Y-пересече | 12.024952 | 0.961547 | 12.505834 | 4.76E-30 | 10.13399641 | 13.915907 | 10.133996 | 13.915907 |
| Переменная | 0.5306999 | 0.022867 | 23.208222 | 1.68E-73 | 0.485730471 | 0.5756694 | 0.4857305 | 0.5756694 |
| Переменная | -0.1064176 | 0.147017 | -0.723846 | 0.46963 | -0.395537434 | 0.1827022 | -0.395537 | 0.1827022 |
| Переменная | 0.2516583 | 0.05082 | 4.9519454 | 1.13E-06 | 0.151716774 | 0.3515999 | 0.1517168 | 0.3515999 |
| Переменная | -0.5396536 | 0.063281 | -8.527875 | 4.18E-16 | -0.664100698 | -0.415206 | -0.664101 | -0.415206 |

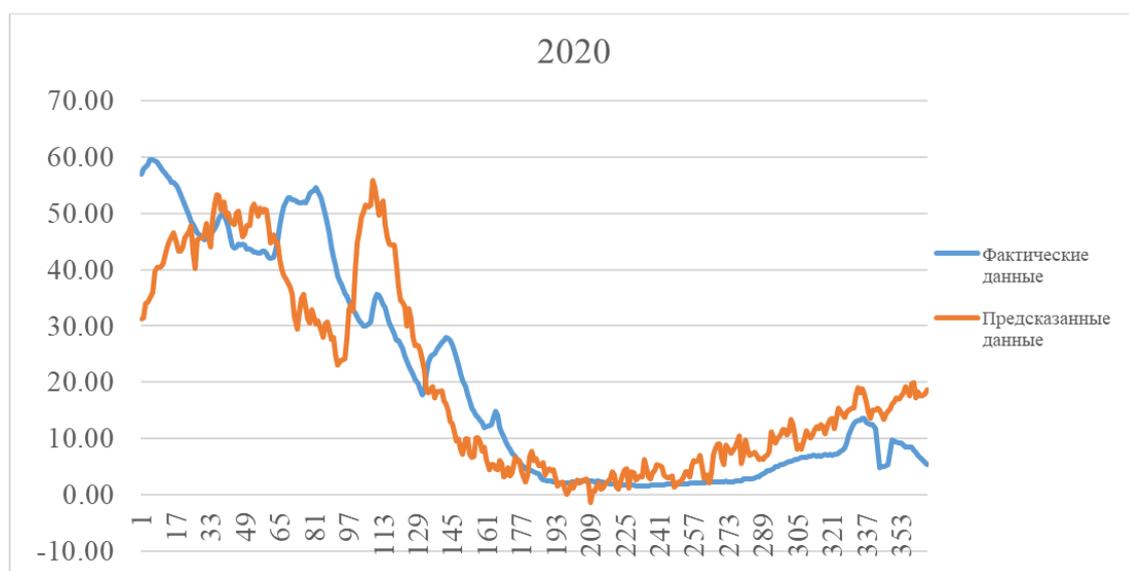


Рисунок 3.7 – Сопоставление графиков фактических значений и модельных, р. Сорочь, 2020 год.

2 методика: Предикторы год аналог по расходу и годы аналоги по метеорологическим характеристикам.

Данная методика вычисления схоже с предыдущим способом. В отличии от предшествующего расчёта отличие заключается в том, что года аналоги берутся не только по расходам, но и по метеорологическим данным. Так же в первую очередь рассмотрим вариант по гидропосту Ловать - р.Ловать - Г.Великие Луки и метеостанция Великие Луки.

Таблица 3.6 – Регрессионный анализ, р. Ловать, 2018 год

| ВЫВОД ИТОГОВ | | | | | | | | | |
|---------------------------------|--------------------|---------------------------|---------------------|-------------------|---------------------|--------------------|---------------------|----------------------|--|
| 2018 | | | | | | | | | |
| <i>Регрессионная статистика</i> | | | | | | | | | |
| Множественный R | 0.8760387 | | | | | | | | |
| R-квадрат | 0.7674438 | | | | | | | | |
| Нормированный R | 0.7648598 | | | | | | | | |
| Стандартная ошибка | 8.3001611 | | | | | | | | |
| Наблюдения | 365 | | | | | | | | |
| <i>Дисперсионный анализ</i> | | | | | | | | | |
| | <i>df</i> | <i>SS</i> | <i>MS</i> | <i>F</i> | <i>значимость F</i> | | | | |
| Регрессия | 4 | 81845.37 | 20461.3426 | 297.0032 | 1.31E-112 | | | | |
| Остаток | 360 | 24801.363 | 68.8926744 | | | | | | |
| Итого | 364 | 106646.73 | | | | | | | |
| | <i>Коэффициент</i> | <i>стандартная ошибка</i> | <i>статистика t</i> | <i>Значение P</i> | <i>Нижние 95%</i> | <i>Верхние 95%</i> | <i>нижние 95.0%</i> | <i>верхние 95.0%</i> | |
| Y-пересечение | -4.3540268 | 1.0556014 | -4.1246882 | 4.61E-05 | -6.4299466 | -2.2781069 | -6.429947 | -2.2781069 | |
| Переменная X 1 | 1.9930819 | 0.0620183 | 32.1369864 | 8.4E-108 | 1.8711182 | 2.1150456 | 1.8711182 | 2.1150456 | |
| Переменная X 2 | -0.0007889 | 0.1046922 | -0.007535 | 0.993992 | -0.206674 | 0.2050963 | -0.206674 | 0.2050963 | |
| Переменная X 3 | -0.0036863 | 0.0399777 | -0.0922099 | 0.926583 | -0.0823054 | 0.0749327 | -0.082305 | 0.0749327 | |
| Переменная X 4 | 0.3652577 | 0.0542837 | 6.72868261 | 6.76E-11 | 0.2585048 | 0.4720107 | 0.2585048 | 0.4720107 | |

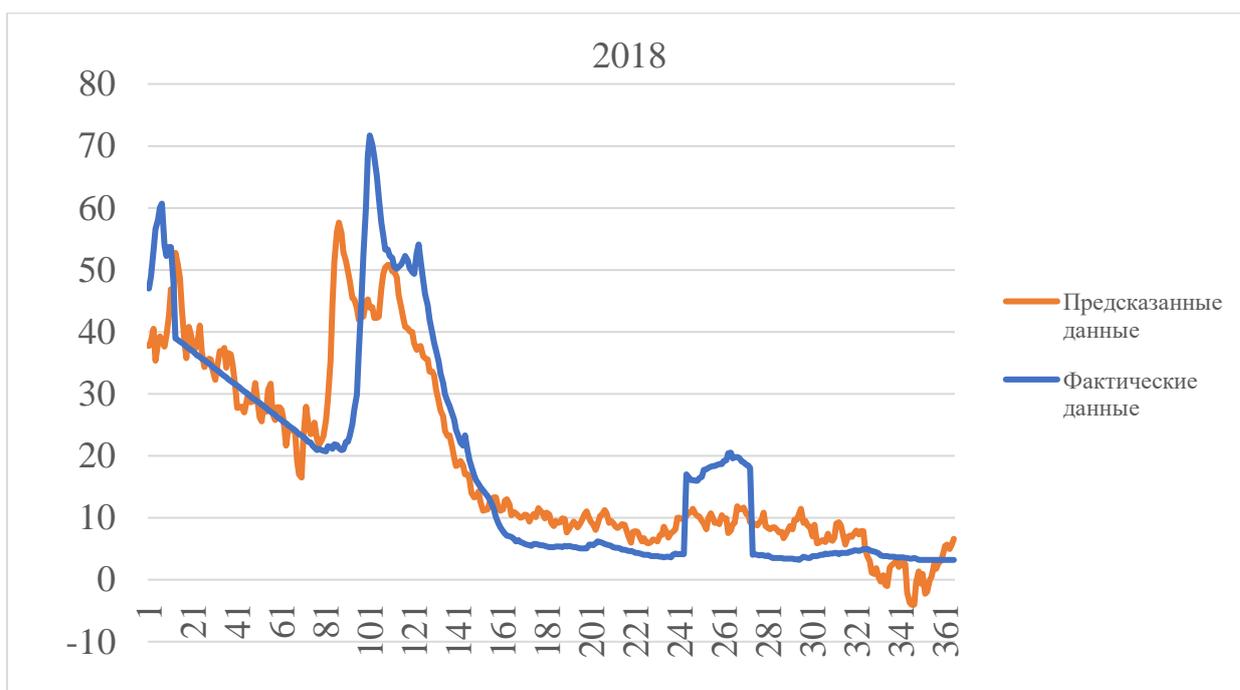


Рисунок 3.8 – Сопоставление графиков фактических значений и модельных, р. Ловать, 2018 год.

Таблица 3.7 – Регрессионный анализ, р. Ловать, 2019 год

| ВЫВОД ИТОГОВ | | | | | | | | |
|---|----------|----------|----------|----------|--------------|----------|----------|----------|
| 2019 | | | | | | | | |
| Регрессионная статистика | | | | | | | | |
| Множественный коэффициент | 0.809418 | | | | | | | |
| R-квадрат | 0.655158 | | | | | | | |
| Нормированный коэффициент | 0.651326 | | | | | | | |
| Стандартная ошибка | 4.972866 | | | | | | | |
| Наблюдения | 365 | | | | | | | |
| Дисперсионный анализ | | | | | | | | |
| | df | SS | MS | F | значимость F | | | |
| Регрессия | 4 | 16913.82 | 4228.455 | 170.989 | 7.03E-82 | | | |
| Остаток | 360 | 8902.583 | 24.7294 | | | | | |
| Итого | 364 | 25816.4 | | | | | | |
| Коэффициент статистика t-Значение нижние 95% верхние 95% нижние 95.0% верхние 95.0% | | | | | | | | |
| У-пересеч | 9.035469 | 0.738917 | 12.22799 | 5.39E-29 | 7.582333 | 10.48861 | 7.582333 | 10.48861 |
| Переменная | 0.253062 | 0.015485 | 16.34279 | 2.65E-45 | 0.22261 | 0.283513 | 0.22261 | 0.283513 |
| Переменная | -0.09475 | 0.060591 | -1.56382 | 0.118737 | -0.21391 | 0.024403 | -0.21391 | 0.024403 |
| Переменная | -0.41554 | 0.038721 | -10.7318 | 1.69E-23 | -0.49169 | -0.3394 | -0.49169 | -0.3394 |
| Переменная | -0.26168 | 0.031653 | -8.26708 | 2.68E-15 | -0.32393 | -0.19943 | -0.32393 | -0.19943 |

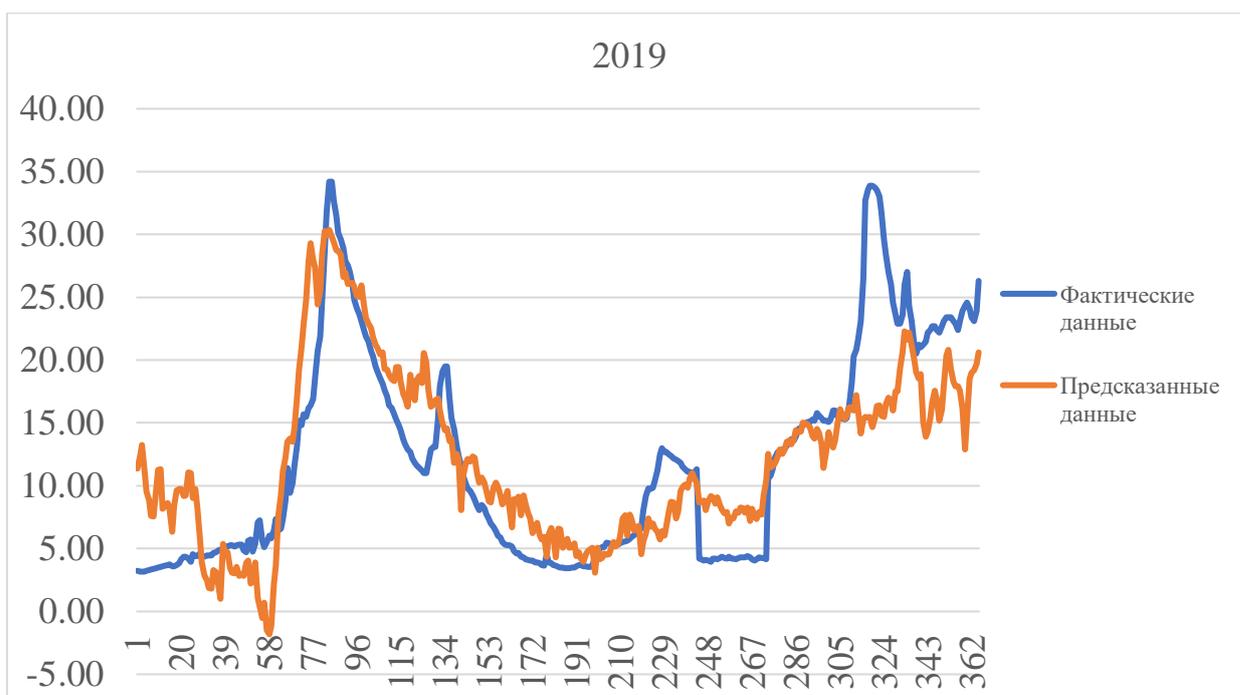


Рисунок 3.9 – Сопоставление графиков фактических значений и модельных, р. Ловать, 2019 год.

Таблица 3.8 – Регрессионный анализ, р. Ловать, 2020 год

| ВЫВОД ИТОГОВ | | | | | | | | |
|--|----------|----------|----------|----------|--------------|----------|----------|----------|
| 2020 | | | | | | | | |
| Регрессионная статистика | | | | | | | | |
| Множественный коэффициент | 0.866063 | | | | | | | |
| R-квадрат | 0.750066 | | | | | | | |
| Нормированный коэффициент | 0.747289 | | | | | | | |
| Стандартная ошибка | 6.050269 | | | | | | | |
| Наблюдения | 365 | | | | | | | |
| Дисперсионный анализ | | | | | | | | |
| | df | SS | MS | F | значимость F | | | |
| Регрессия | 4 | 39548.1 | 9887.025 | 270.0949 | 5.5E-107 | | | |
| Остаток | 360 | 13178.07 | 36.60575 | | | | | |
| Итого | 364 | 52726.17 | | | | | | |
| Коэффициент стандартная ошибка t-Значение p-Значение нижние 95% верхние 95% нижние 95% верхние 95% | | | | | | | | |
| Y-пересечение | 7.036113 | 0.783726 | 8.977773 | 1.56E-17 | 5.494857 | 8.577369 | 5.494857 | 8.577369 |
| Переменная 1 | 0.985124 | 0.042586 | 23.13267 | 3.39E-73 | 0.901376 | 1.068872 | 0.901376 | 1.068872 |
| Переменная 2 | 0.114091 | 0.116692 | 0.977709 | 0.328875 | -0.11539 | 0.343575 | -0.11539 | 0.343575 |
| Переменная 3 | -0.35533 | 0.154235 | -2.30383 | 0.021801 | -0.65865 | -0.05202 | -0.65865 | -0.05202 |
| Переменная 4 | -0.40401 | 0.047215 | -8.55678 | 3.4E-16 | -0.49686 | -0.31116 | -0.49686 | -0.31116 |

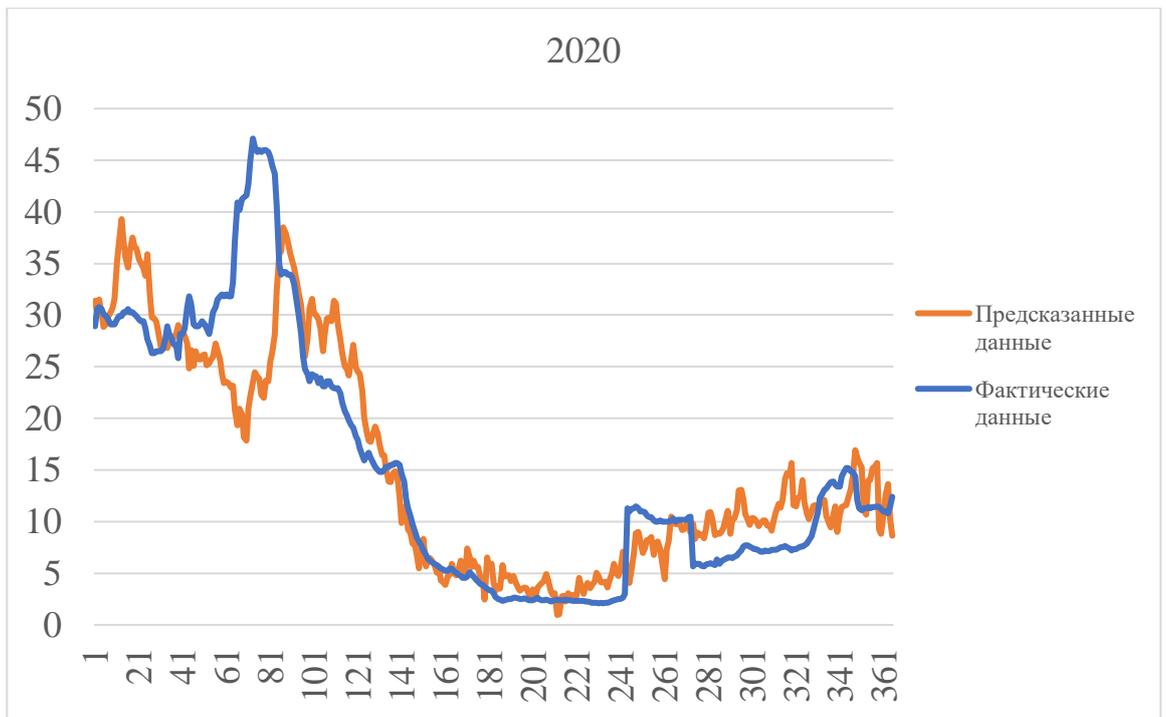


Рисунок 3.10 – Сопоставление графиков фактических значений и модельных, р. Ловать, 2020 год.

Далее рассмотрим получившиеся значения по гидропосту Сороть (Сорять) - р. Сороть – д. Осинкино и метеостанция Пушкинские горы.

Таблица 3.9 – Регрессионный анализ, р. Сороть, 2018 год

| Вывод итогов | | | | | | | | |
|---|----------|----------|----------|----------|--------------|----------|----------|----------|
| 2018 | | | | | | | | |
| Регрессионная статистика | | | | | | | | |
| Множественный коэффициент корреляции R | 0.914765 | | | | | | | |
| Коэффициент детерминации R-квадрат | 0.836795 | | | | | | | |
| Нормированный коэффициент детерминации | 0.834981 | | | | | | | |
| Стандартная ошибка | 9.515038 | | | | | | | |
| Наблюдения | 365 | | | | | | | |
| Дисперсионный анализ | | | | | | | | |
| | df | SS | MS | F | Значимость F | | | |
| Регрессия | 4 | 167112.4 | 41778.11 | 461.4533 | 3E-140 | | | |
| Остаток | 360 | 32592.94 | 90.53594 | | | | | |
| Итого | 364 | 199705.4 | | | | | | |
| Коэффициент стандартной ошибки t-Значение нижние 95% верхние 95% нижние 95.0% верхние 95.0% | | | | | | | | |
| Y-пересечение | 0.739412 | 1.159819 | 0.637524 | 0.524189 | -1.54146 | 3.020283 | -1.54146 | 3.020283 |
| Переменная 1 | 0.740992 | 0.020545 | 36.06725 | 1.4E-121 | 0.700589 | 0.781394 | 0.700589 | 0.781394 |
| Переменная 2 | 0.212901 | 0.143598 | 1.482622 | 0.13905 | -0.06949 | 0.495297 | -0.06949 | 0.495297 |
| Переменная 3 | 0.866606 | 0.084168 | 10.2961 | 5.81E-22 | 0.701082 | 1.032129 | 0.701082 | 1.032129 |
| Переменная 4 | -0.02234 | 0.074959 | -0.29804 | 0.765847 | -0.16975 | 0.125072 | -0.16975 | 0.125072 |

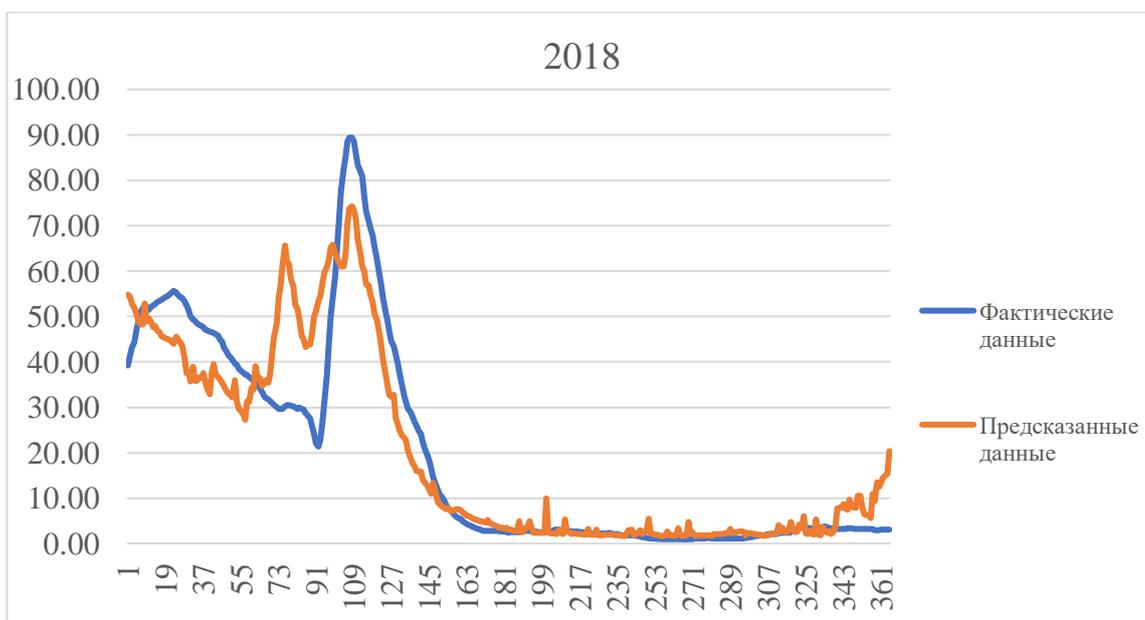


Рисунок 3.11 – Сопоставление графиков фактических значений и модельных, р. Сороть, 2018 год.

Таблица 3.10 – Регрессионный анализ, р. Сороть, 2019 год

| Вывод итогов | | | | | | | | |
|--|----------|----------|----------|----------|--------------|----------|----------|----------|
| 2019 | | | | | | | | |
| Регрессионная статистика | | | | | | | | |
| Множественный коэффициент | 0.709711 | | | | | | | |
| R-квадрат | 0.503689 | | | | | | | |
| Нормированный коэффициент | 0.498175 | | | | | | | |
| Стандартная ошибка | 16.38016 | | | | | | | |
| Наблюдения | 365 | | | | | | | |
| Дисперсионный анализ | | | | | | | | |
| | df | SS | MS | F | значимость F | | | |
| Регрессия | 4 | 98027.51 | 24506.88 | 91.33802 | 1.58E-53 | | | |
| Остаток | 360 | 96591.5 | 268.3097 | | | | | |
| Итого | 364 | 194619 | | | | | | |
| Коэффициенты стандартная ошибка t-Значение нижние 95% верхние 95% нижние 95.0% верхние 95.0% | | | | | | | | |
| Y-пересечение | 20.89342 | 2.159534 | 9.674967 | 7.84E-20 | 16.64653 | 25.1403 | 16.64653 | 25.1403 |
| Переменная 1 | 0.374792 | 0.037538 | 9.984357 | 6.96E-21 | 0.300971 | 0.448613 | 0.300971 | 0.448613 |
| Переменная 2 | 0.088646 | 0.179215 | 0.494636 | 0.621159 | -0.26379 | 0.441085 | -0.26379 | 0.441085 |
| Переменная 3 | -1.84076 | 0.214326 | -8.58861 | 2.7E-16 | -2.26225 | -1.41927 | -2.26225 | -1.41927 |
| Переменная 4 | -0.8235 | 0.097762 | -8.4235 | 8.83E-16 | -1.01575 | -0.63124 | -1.01575 | -0.63124 |

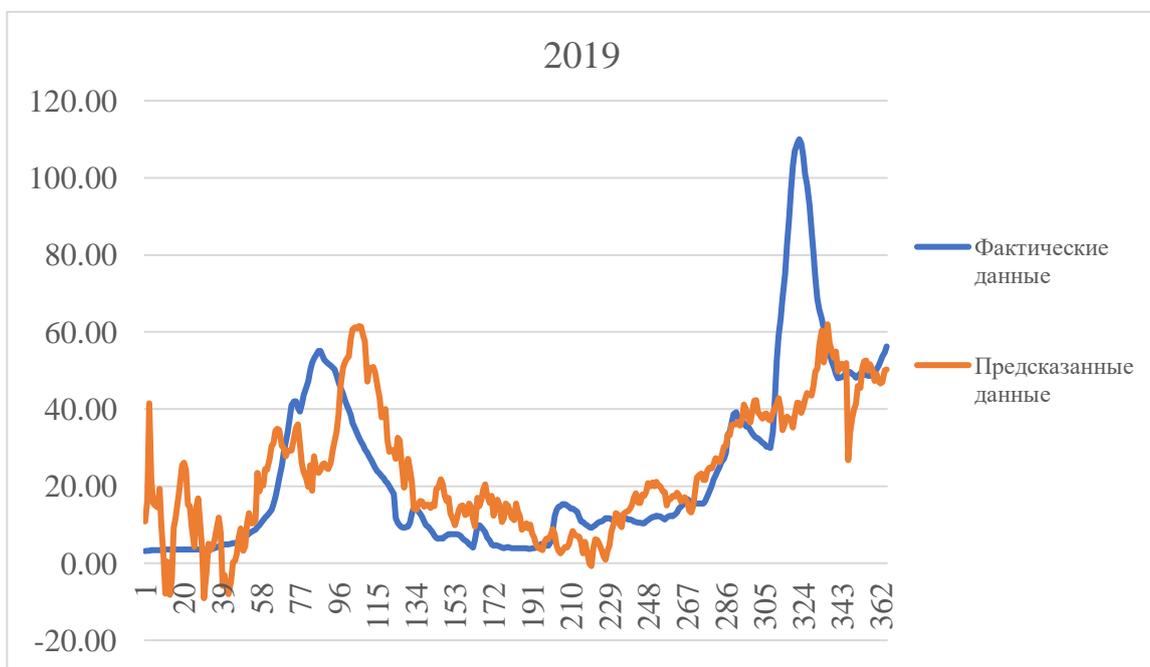


Рисунок 3.12 – Сопоставление графиков фактических значений и модельных, р. Сороть, 2019 год.

Таблица 3.11 – Регрессионный анализ, р. Сорочь, 2020 год

| Вывод итогов | | | | | | | | |
|---|----------|----------|----------|----------|--------------|----------|----------|----------|
| 2020 | | | | | | | | |
| Регрессионная статистика | | | | | | | | |
| Множественный коэффициент | 0.842166 | | | | | | | |
| R-квадрат | 0.709244 | | | | | | | |
| Нормированный R-квадрат | 0.706013 | | | | | | | |
| Стандартная ошибка | 10.5199 | | | | | | | |
| Наблюдения | 365 | | | | | | | |
| Дисперсионный анализ | | | | | | | | |
| | df | SS | MS | F | Значимость F | | | |
| Регрессия | 4 | 97183.3 | 24295.82 | 219.5375 | 3.51E-95 | | | |
| Остаток | 360 | 39840.56 | 110.6682 | | | | | |
| Итого | 364 | 137023.9 | | | | | | |
| Коэффициенты регрессии, стандартная ошибка, статистика t, значение p, нижние и верхние границы 95% доверительного интервала | | | | | | | | |
| Y-пересечение | 14.39038 | 1.232242 | 11.6782 | 6.15E-27 | 11.96708 | 16.81367 | 11.96708 | 16.81367 |
| Переменная 1 | 0.550365 | 0.026157 | 21.04109 | 1.16E-64 | 0.498926 | 0.601804 | 0.498926 | 0.601804 |
| Переменная 2 | -0.10025 | 0.176065 | -0.56939 | 0.569447 | -0.44649 | 0.245995 | -0.44649 | 0.245995 |
| Переменная 3 | 0.328521 | 0.204906 | 1.603273 | 0.109751 | -0.07444 | 0.731484 | -0.07444 | 0.731484 |
| Переменная 4 | -0.7063 | 0.085802 | -8.23172 | 3.43E-15 | -0.87504 | -0.53756 | -0.87504 | -0.53756 |

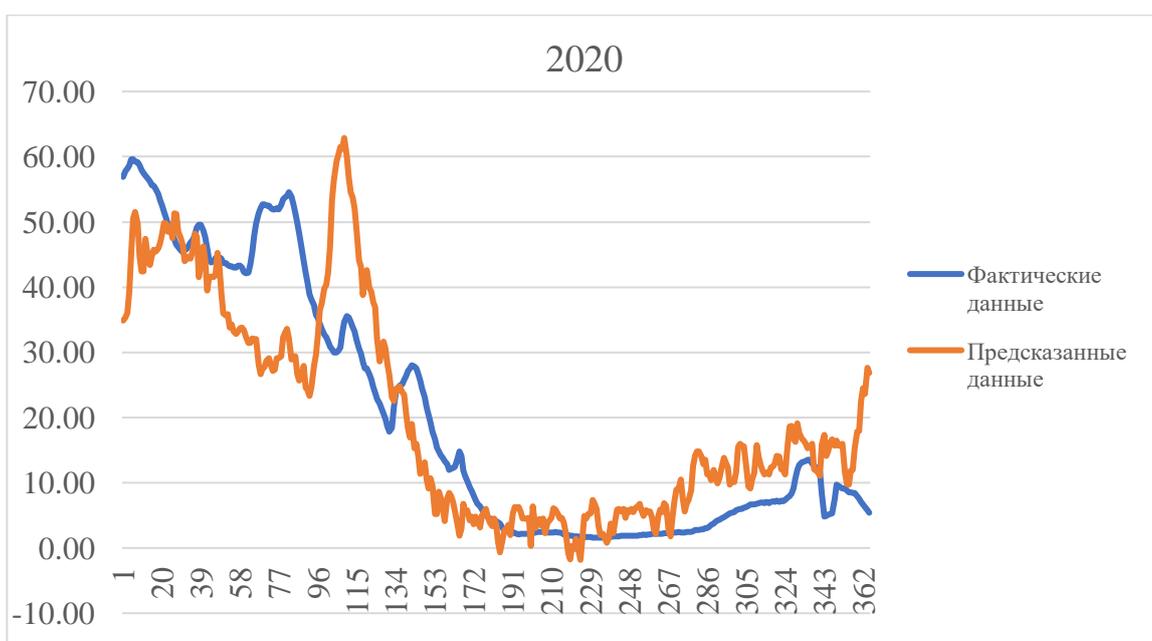


Рисунок 3.13 – Сопоставление графиков фактических значений и модельных, р. Сорочь, 2020 год.

4 Анализ полученных результатов и рекомендации для использования разработанного алгоритма

Оценка эффективности разработанного подхода, который вошёл в методику регрессионного анализа, представлена в таблицах этого раздела.

Таблица 4.1 – Оценка эффективности по 1 варианту модели, р. Ловать

| 1 методика по гидропосту Ловать - р.Ловать - г.Великие Луки и метеостанция Великие Луки | | | |
|---|--------|------|-------|
| Год | Данные | | |
| | ско | S | S/ско |
| 2018 | 17.12 | 8.62 | 0.50 |
| 2019 | 8.42 | 5.63 | 5.63 |
| 2020 | 12.04 | 6.32 | 0.53 |

Таблица 4.2 – Оценка эффективности по 1 варианту модели, р. Сорочь

| 1 методика по гидропосту Сорочь (Сорать) - р.Сорочь - д.Осинкино и метеостанция Пушкинские горы | | | |
|---|--------|-------|-------|
| Год | Данные | | |
| | ско | S | S/ско |
| 2018 | 23.42 | 10.94 | 0.47 |
| 2019 | 23.12 | 17.51 | 0.76 |
| 2020 | 19.40 | 9.35 | 0.48 |

Таблица 4.3 – Оценка эффективности по 2 варианту модели, р. Ловать

| 2 методика по гидропосту Ловать - р.Ловать - г.Великие Луки и метеостанция Великие Луки | | | |
|---|--------|------|-------|
| Год | Данные | | |
| | ско | S | S/ско |
| 2018 | 17.12 | 8.25 | 0.48 |
| 2019 | 8.42 | 4.95 | 0.59 |
| 2020 | 12.04 | 6.02 | 0.50 |

Таблица 4.4 – Оценка эффективности по 2 варианту модели, р. Сорочь

| 2 методика по гидропосту Сорочь (Сорать) - р.Сорочь - д.Осинкино и метеостанция Пушкинские горы | | | |
|---|--------|----------|---------------|
| Год | Данные | | |
| | ско | <i>S</i> | <i>S</i> /ско |
| 2018 | 23.42 | 9.46 | 0.40 |
| 2019 | 23.12 | 16.29 | 0.70 |
| 2020 | 19.40 | 10.46 | 0.54 |

Следует отметить, что по всем годам получены удовлетворительные результаты. Но есть существенный недостаток подхода, который заключается в том, что должно пройти какое-то время в прогнозируемом году, чтобы подобрать год аналог по гидрологическим и метеорологическим характеристикам. Чем дальше дата выпуска прогноза от начала года, тем точнее результат.

К положительным сторонам подхода относится возможность выпускать прогнозы с различной заблаговременностью, начиная краткосрочными прогнозами, заканчивая долгосрочными, включая среднесрочные.

Следует отметить, чувствительность регрессионной модели к изменениям метеорологических характеристик.

5 Методы определения циклов и периодов во временных рядах

5.1 Метод Шустера

Алгоритм метода Шустера используют как один из основных методов выделения скрытых периодичностей во временных рядах. Использование данного метода в дополнение к спектральному позволяет полностью исключить недостатки последнего. Помимо этого, следует указать на одно важное достоинство периодограммного метода, а именно, его более высокую избирательность по сравнению с методом спектрального анализа. Вследствие этого, особенно при анализе коротких рядов наблюдений, периодограммный метод может иметь самостоятельное значение.

Рассмотрим кратко теоретические предпосылки метода А. Шустера. Известно, что любую функцию $f(t)$ можно представить рядом Фурье:

$$f(t) = a_0 + a_1 \cos qt + a_2 \cos 2qt + \dots + b_1 \sin qt + b_2 \sin 2qt + \dots,$$

где a_0 – среднее ряда значений, $q = \frac{2\pi}{N}$, N – число наблюдаемых моментов $f(t)$.

Производя подстановку:

$$a_1 = A_1 \sin \varphi_1, \quad a_2 = A_2 \sin \varphi_2, \quad b_1 = A_1 \cos \varphi_1, \quad b_2 = A_2 \cos \varphi_2 \text{ и т. д.,}$$

получим:

$$f(t) = a_0 + A_1 \sin(qt + \varphi_1) + A_2 \sin(2qt + \varphi_2) + \dots \quad (9)$$

Точность функции $f(t)$ будет определяться числом наблюдений. Вне пределов этого ряда его значения будут повторяться, ибо входящие в него периоды являются кратными взятому промежутку и никакого реального значения не имеют.

Однако при изучении внутренней структуры исследуемого ряда необходимо найти действительные периоды. Для этого выражение (9) записывается в виде:

$$f(t) = a_0 + A_1 \sin(q_1 t + \varphi_1) + A_2 \sin(q_2 t + \varphi_2) + \dots + l(t)$$

где $q_1 = \frac{2\pi}{T_1}$, $q_2 = \frac{2\pi}{T_2}$, ... T_1, T_2, \dots – действительные периоды;

A_1, A_2, \dots – их амплитуды, $l(t)$ – некоторая неизвестная ограниченная непериодическая функция.

Данный ряд менее точно выражает функцию $f(t)$ [так как $l(t)$ неизвестно], чем ряд Фурье, однако он может дать более или менее точно приближенные значения для исследуемой функции вне пределов рассматриваемого ряда.

Чтобы найти параметры периодичности T в исследуемом ряду (амплитуду A и фазу φ), весь ряд данных делая на отрезки по T величин в каждом (остаток отбрасывается). Затем один отрезок помещается под другим и производится суммирование по вертикальным столбцам:

| | | | |
|---------|------------------|------------------|---|
| | x_0 | x_1 | $x_2 \dots \dots \dots x_{T-1}$ |
| | x_T | x_{T+1} | $x_{T+2} \dots \dots \dots x_{2T-1}$ |
| | | | |
| | $x_{(m-1)T}$ | $x_{(m-1)T+1}$ | $x_{(m-1)T+2} \dots \dots \dots x_{mT-1}$ |
| Суммы | Σ_0 | Σ_1 | $\Sigma_2 \dots \dots \dots \Sigma_{T-1}$ |
| Средние | $\overline{x_0}$ | $\overline{x_1}$ | $\overline{x_2} \dots \dots \dots \overline{x_{T-1}}$ |

Далее вычисляются a_T, b_T, A_T и φ_T :

$$a_T = \frac{2}{T} \sum_{t=0}^{T-1} x_t \cos\left(\frac{2\pi}{T}t\right), b_T = \frac{2}{T} \sum_{t=0}^{T-1} x_t \sin\left(\frac{2\pi}{T}t\right), A_T = \sqrt{a_T^2 + b_T^2}, \varphi_T = \arctg\left(\frac{a_T}{b_T}\right).$$

Первые две величины есть коэффициенты в разложении Фурье:

$$f(t) = a_0 + \sum_{T=T_H}^{T_K} \left(a_T \cos\frac{2\pi}{T}t + b_T \sin\frac{2\pi}{T}t \right),$$

Вторые две – элементы разложения по синусам:

$$f(t) = a_0 + \sum_{T=T_H}^{T_K} A_T \sin\left(\frac{2\pi}{T}t + \varphi_T\right).$$

Вычисляя амплитуды и фазы для различных пробных периодов и строя график, где по оси абсцисс откладываются периоды, а по оси ординат – амплитуды соответствующих периодов, получаем периодограмму. Максимумы на периодограмме указывают на возможные периоды, характеризующие закономерности данного ряда (рисунок 5.1).

При разложении в ряд Фурье ограниченного числа данных могут быть получены максимумы для периодов, которые вовсе не присущи данному ряду, так называемы ложные периоды. Они получают в виде вторичных максимумов. Для их обнаружения А. Шустером выведена формула:

$$S_1 = S\left(1 \pm \frac{1,43}{n}\right), \text{ где } n - \text{ число вошедших в анализ периодов.}$$

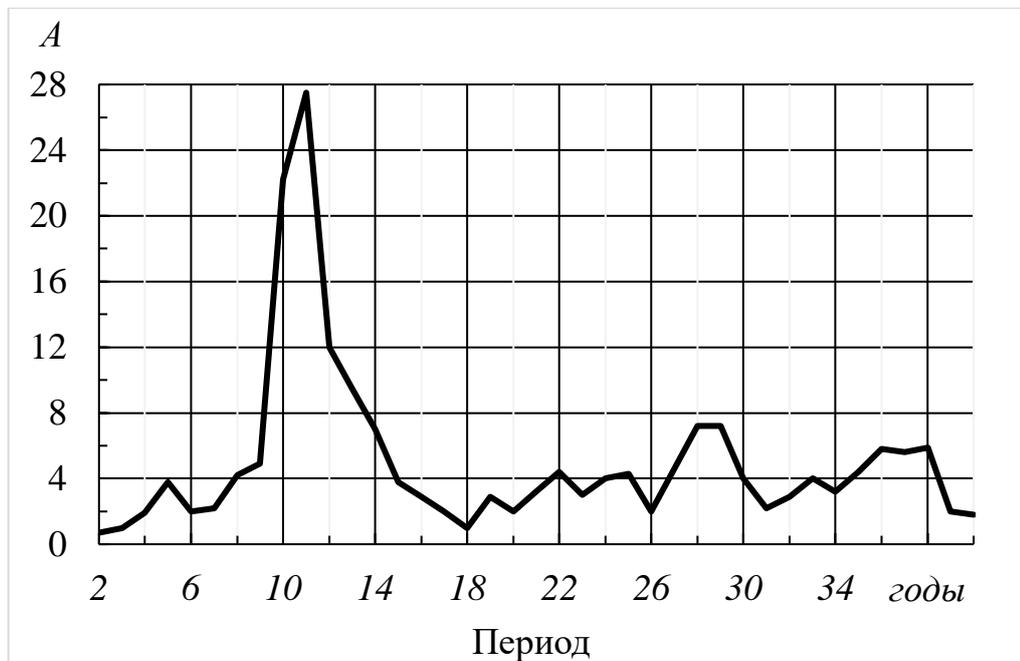


Рисунок 5.1 – Пример периодограммы ряда чисел Вольфа с 1700 по 1967 г.

Эти ложные периоды отличаются от истинных тем, что с увеличением ряда соответствующие им максимумы на периодограмме приближаются к максимуму истинного периода, который в отличие от них остается без изменения.

Для определения ряду периодичности Шустер предлагает использовать величину отношения амплитуды этого периода к ее математическому ожиданию. Малые величины этого отношения дают основание предполагать, что выявленная периодичность является случайной. Чем больше амплитуда будет превышать математическое ожидание, тем вероятнее наличие вариации с данным периодом в исследуемом ряду.

Поскольку исследуемые ряды различны по существу, то и критерий истинности определяется Шустером несколько differently для рядов разного рода. А. Шустер выделяет два рода рядов. К рядам первого рода относятся ряды, содержащие величины, характеризующие частоту появления отдельных событий одного и того же рода независимо от их величины, например, число солнечных пятен, гроз, магнитных бурь и т.

д. Ряды второго рода представляют собой изменения в величине какого-нибудь явления через равные промежутки времени, например, суточные или среднегодовые температуры, средние высоты уровней, расходов и т. д.

Математическое ожидание ε определяется Шустером как:

$$\varepsilon = \sqrt{\frac{\pi}{n}},$$

где n – число появления событий за весь промежуток времени.

Для рядов второго рода ε вычисляется по формуле:

$$\varepsilon = \frac{1}{N} \sqrt{\pi \sum_{i=0}^{n-1} \Delta x^2} = \sigma \sqrt{\frac{\pi}{N}},$$

где σ – стандартное отклонение или средняя квадратическая ошибка исследуемого ряда;

N – общее число данных в исследуемом ряду.

Вероятность того, что амплитуда A_T в k раз превышает ε в обоих случаях оказалась равной:

$$e^{-\frac{\pi}{4}k^2}$$

Таким образом, вычислив величину ε определив отношение $k = \frac{A_T}{\varepsilon}$ по таблице 1 можно получить, какова вероятность того, что данная периодичность в исследуемом ряду является случайной.

Таблица 5.1 – Вероятности по А. Шустеру

| k | $e^{-\frac{\pi}{4}k^2}$ | k | $e^{-\frac{\pi}{4}k^2}$ | k | $e^{-\frac{\pi}{4}k^2}$ |
|-----|-------------------------|-----|-------------------------|-----|-------------------------|
| 0,1 | 0.9922 | 0.8 | 0.6049 | 2 | 0.04321 |

| | | | | | |
|-----|--------|-----|--------|-----|-----------|
| 0,2 | 0.9691 | 0.9 | 0.5293 | 2.5 | 0.00738 |
| 0,3 | 0.9318 | 1 | 0.4559 | 3 | 0.00085 |
| 0,4 | 0.8819 | 1.2 | 0.3227 | 3.5 | 6.63×10-5 |
| 0,5 | 0.8217 | 1.4 | 0.2145 | 4 | 3.49×10-6 |
| 0,6 | 0.7537 | 1.6 | 0.1339 | 4.5 | 1.24×10-7 |
| 0,7 | 0.6806 | 1.8 | 0.0785 | 5 | 2.79×10-9 |

Указанный способ нахождения ε пригоден лишь когда величины данного ряда вполне независимы друг от друга. Так как последнее условие часто не выполняется, то Шустер предлагает находить эту величину опытным путем, беря в первом приближении среднее из ряда полученных амплитуд. Чем больше будет взято амплитуд, тем ближе среднее из них к теоретической величине ε .

Необходимо указать, что спектральные методы анализа дают исчерпывающую информацию о структуре исследуемого ряда в том случае, если этот ряд является нормальным, стационарным и среднее его значение равно нулю .

Последнее условие легко удовлетворяется, если заменить исходный ряд другим, члены которого определяются соотношением:

$$\Delta x_i = x_i - \bar{x}.$$

Стационарным называется такой ряд, среднее значение которого постоянно, т. е. не меняется при изменении длины ряда.

Если же в исследуемом ряду имеется какой-либо тренд, то он уже не будет стационарным и возникает необходимость исключения тренда из первичного ряда.

Наконец, чтобы ряд был нормальным, необходимо, чтобы значения его членов были распределены по нормальному закону.

Однако, даже в случае, если распределение членов ряда не является нормальным, вычисленная спектральная функция позволяет получить представление о его частотной структуре [1].

Следовательно, в качестве первого этапа расчетов необходимо вычислить среднее значение данного ряда и вычесть его из каждого члена ряда. Эта операция достигается сразу же при вычислении статистических характеристик исследуемых рядов.

Далее должно следовать сглаживание, или фильтрация, исходных рядов. Эта операция дает возможность значительно уменьшить роль случайных ошибок, стационаризовать ряд, исключить колебания, изучение которых в настоящее время не представляет интереса. Кроме того, при исследовании особенно коротких рядов наблюдений часто очень важно искусственно сузить полосу частот исследуемого ряда, чтобы уменьшить при анализе влияние друг на друга вариаций с близкими частотами. Эти задачи в значительной мере решаются с помощью фильтрации исходных рядов.

В общем случае основной целью фильтрации является выделение из исследуемого ряда гармоник с частотами $f_l \leq f \leq f_h$ и подавление всех прочих составляющих.

Осуществляется фильтрация с помощью преобразования вида:

$$\tilde{x}(t) = \int_{-\infty}^{\infty} h(\tau)x(t + \tau)d\tau,$$

где $h(\tau)$ представляет собой весовую функцию фильтра .

Любая фильтрация предусматривает изменение амплитуд гармоник процесса $x(t)$. Функция, определяющая характер изменения амплитуд при прохождении ряда через фильтр, называется частотной характеристикой фильтра. При условии, что начальные фазы гармоник после фильтрации должны остаться неизменными, частотная характеристика $R(f)$ имеет вид:

$$R(f) = \int_{-\infty}^{\infty} h(\tau) \cos 2\pi f \tau d\tau.$$

Весовая же функция $h(\tau)$ такого фильтра связана с частотной характеристикой обратным преобразованием Фурье:

$$h(\tau) = \int_{-\infty}^{\infty} R(f) \cos 2\pi f \tau df.$$

Задача фильтрации решается наилучшим образом, если гармоники с частотами $f_l \leq f \leq f_h$ на выходе фильтра сохраняются неизменными, а амплитуды всех остальных гармоник обратятся в нуль. Частотная характеристика такого идеального фильтра имеет вид:

$$R(f) = \begin{cases} 1 & \text{при } f_l \leq f \leq f_h \\ 0 & \text{при } f < f_l, \quad f > f_h \end{cases} \quad (14)$$

Однако реализация такого идеального фильтра требует задания весовой функции, а, следовательно, и самого процесса $x(t)$ на бесконечном интервале. Поскольку интервал задания исходной функции $x(t)$ всегда конечен, то и интервал задания весовой функции $h(\tau)$ также должен быть конечен. Тогда уравнение (12) принимает вид:

$$R(f) = \int_{-L/2}^{L/2} h(\tau) \cos 2\pi f \tau d\tau,$$

где L – интервал задания весовой функции,

$$h(\tau) = \begin{cases} h(t) & \text{при } |\tau| \leq L/2, \\ 0 & \text{при } |\tau| > L/2. \end{cases}$$

В этом случае амплитуды гармоник, лежащие в диапазоне $f_l \leq f \leq f_h$ умножаются на величины, отличные от единицы. а амплитуды гармоник с частотами вне этого диапазона – на величины, отличные от нуля.

Поскольку реализация идеального фильтра в виде (14) на ограниченном интервале невозможна, задачей фильтрации является нахождение такой частотной характеристики и соответствующей весовой функции, которая обеспечивала бы выделение гармоник в заданном диапазоне оптимальным образом. Решение этой задачи возможно лишь при условии, что спектр данного процесса известен или если есть физические основания предполагать наличие определенных гармоник в исследуемом интервале частот. Последнее обстоятельство может в значительной степени способствовать правильному выбору как вида фильтра, т. е. его весовой функции, так и интервалов его задания.

Переходя к сумме по конечному интервалу $(2k+1)$, выражение запишем в виде:

$$\tilde{x}(t) = \sum_{i=-k}^k h_j x_{i+f} .$$

Фильтры различают в зависимости от выделяемого диапазона частот и выбираемой весовой функции h .

В зависимости от выделяемого диапазона частот фильтры подразделяются на низкочастотные, или сглаживающие ($0 \leq f \leq f_h$), высокочастотные ($f_l \leq f \leq \infty$) и полосовые ($f_l \leq f \leq f_h$).

Первые применяются в случае необходимости исследования длиннопериодных колебаний или тренда в исходном ряду, а также для подавления случайных колебаний, связанных с ошибками измерений. и т. д. Вторые позволяют стационарировать ряд и исследовать высокочастотные

колебания до определенного задаваемого предела частоты, и, наконец, последние позволяют исследовать колебания в определенной полосе частот.

В широких исследованиях, связанных с изучением полного спектра колебаний природных процессов, обычно возникает необходимость использования всех трех видов фильтрации.

Фильтров, различающихся своей весовой функцией, предложено в настоящее время большое количество. Простейшим и наиболее широко применяемым видом фильтрации является скользящее осреднение. Весовая функция такого фильтра:

$$h_j = \begin{cases} \frac{1}{2k+1} & \text{при } |j| \leq k, \\ 0 & \text{при } |j| > k, \end{cases}$$

где $(2k+1)$ – интервал сглаживания.

Общая формула для x , в случае низкочастотного фильтра записывается в этом случае в виде:

$$\tilde{x}_i = \frac{1}{2k+1} \sum_{j=-k}^k x_{i+j}.$$

Частотная характеристика такого фильтра, определяемая выражением:

$$R_l(f) = \frac{\sin \pi f(2k+1)}{\pi f(2k+1)},$$

В случае же применения скользящего осреднения в качестве высокочастотного фильтра общее выражение для \tilde{x}_i будет записано в виде:

$$\tilde{x}_i = x_i \frac{1}{2k+1} \sum_{i=-k}^k x_{i+j}.$$

Частотную характеристику высокочастотного фильтра $R_h(f)$ всегда просто получить через частотную характеристику сглаживающего фильтра $R_l(f)$:

$$R_h(f) = 1 - R_l(f).$$

Наконец, для случая полосового фильтра значение \tilde{x}_i определяется по формуле:

$$\tilde{x}_i = \frac{1}{2k_1+1} \sum_{i=-k_1}^{k_1} x_{i+j} - \frac{1}{2k_2+1} \sum_{i=-k_2}^{k_2} x_{i+j}.$$

Частотная характеристика полосового фильтра в нашем случае будет представлена как разность частотных характеристик двух сглаживающих фильтров:

$$R(f) = 1 - R_l(f; k_1) - R_l(f; k_2).$$

Рассмотрим пример.

Дано:



Рисунок 5.2 – Пример методики Шустера

```
function [At, kk]= ShusterMP(Y,L1,L2)
```

```
N=length(Y);
```

```
ss=std(Y);
```

```
e = ss*((pi/N)^0.5);
```

```
for L=L1:1:L2
```

```
    Y1=0;
```

```
    ats=0;
```

```
    bts=0;
```

```
    k=N/L;
```

```
    for i=1:k
```

```
        for j=1:L
```

```
            Y1(i,j)=Y(i*L-L+j);
```

```
        end
```

```
    end
```

```
for i=1:L
```

```
    s(i) = mean(Y1(:,i));
```

```
    atS(L)= ats+(s(i)*cos((2*pi/L)*i));
```

```
    ats=atS(L);
```

```
    btS(L)=bts+(s(i)*sin((2*pi/L)*i));
```

```

    bts=btS(L);
end
at(L)=2/L*ats;
bt(L)=2/L*bts;
At(L)=(at(L)^2 + bt(L)^2)^0.5;
kk(L)=At(L)/e;
end
At =At(L1:1:L2)
kk = kk(L1:1:L2)
x=[L1:1:L2];
plot(x, At)
ylabel('A');
xlabel('Period, years');
grid on;

figure;
plot(x, kk)
ylabel('k');
xlabel('Period, years');
grid on;

```

```

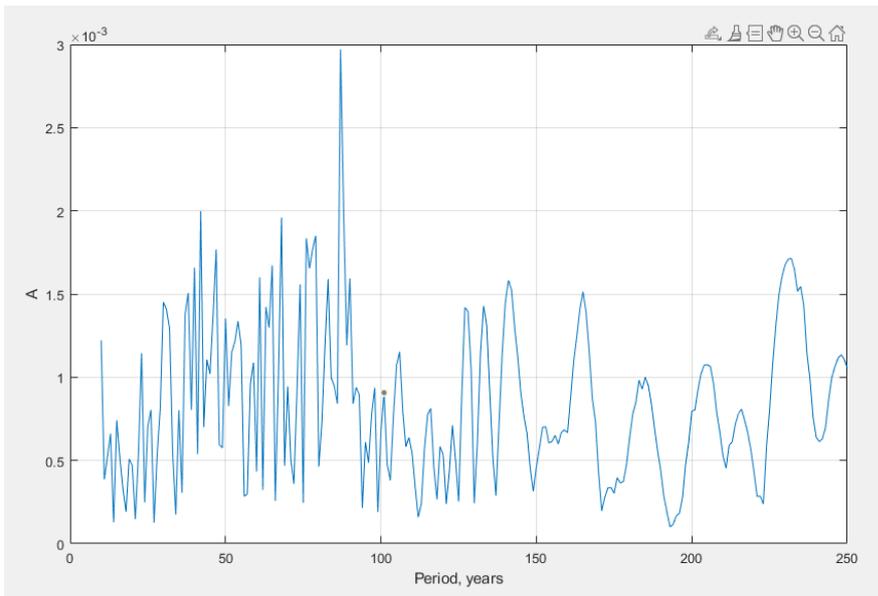
>>Y=[... вставляем ряд
];
>> ShusterMP(Y, 10, 250);

```

Первое число – это минимальный предполагаемый период

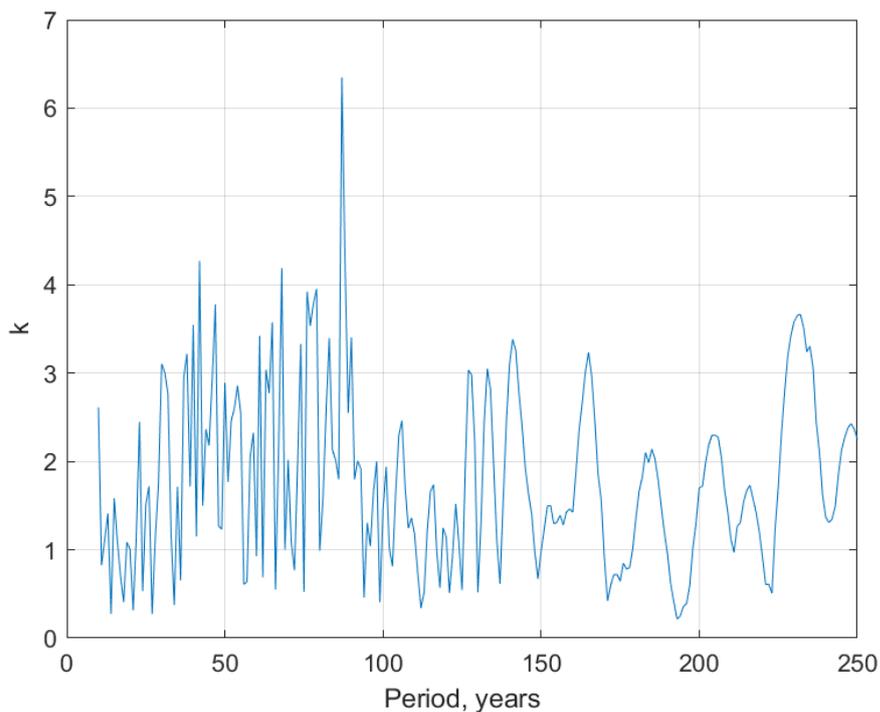
Второе число – это максимальный период.

Результат:



По этому графику оценивается амплитуда, там, где она максимальная – это и есть период.

Т.е. в данном случае период – 87. Кстати, четко видны ложные периоды.



Это график с «достоверностью». Он, как правило, повторяет верхний график, только по оси у – показатель k , от которого переходят к достоверности.

Дополнительно

Пример.

Есть ряд Y .

```
>> Y=[...  
];
```

Оценим автокорреляцию

Вход – это ряд Y , а N – это сколько сдвижек делать

```
function [k]=aftocor(Y, N)
```

```
L = length(Y);
```

```
k(1)=1;
```

```
for i=2:N+1
```

```
Y1 = Y(1:(L-i));
```

```
Y2=Y((1+i):L);
```

```
r = corrcoef(Y1, Y2);
```

```
k(i)=r(1,2);
```

```
end
```

```
x=0:N;
```

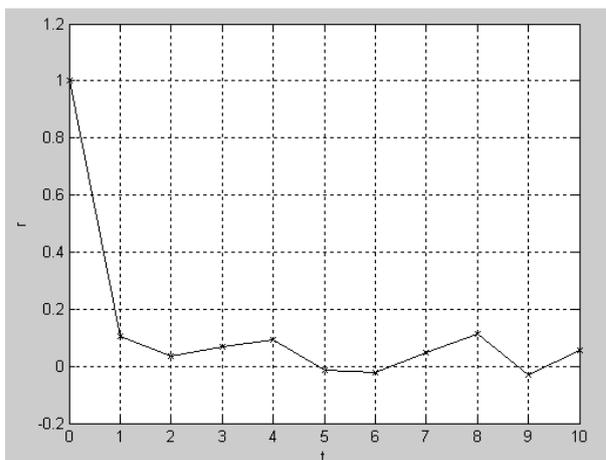
```
plot(x, k, 'kx-')
```

```
grid on
```

```
xlabel('t')
```

```
ylabel('r')
```

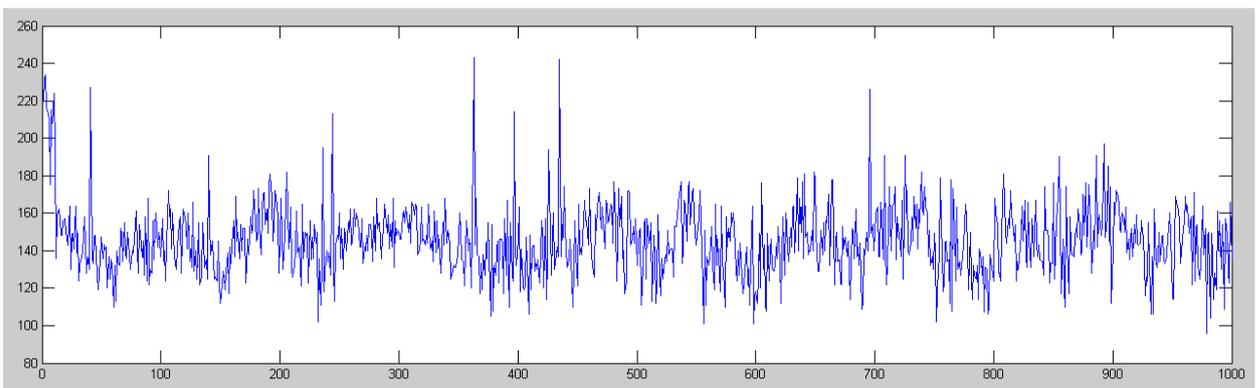
```
>> [k]=aftocor(Y, 10);
```



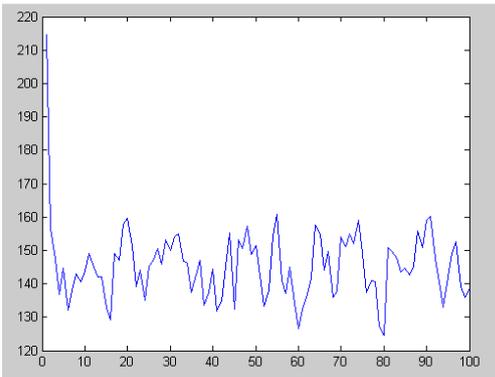
Осреднение по периодам.

Когда обнаружим периоды, для дальнейших вычислений надо уйти от них. Поэтому происходит осреднение ряда именно по найденным периодам.

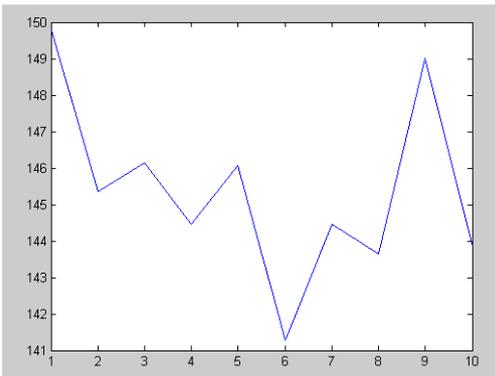
```
Y = [...  
% 1000 значений  
>> Y=Y';  
>> x=[1:1:1000];  
>> plot(x, Y)
```



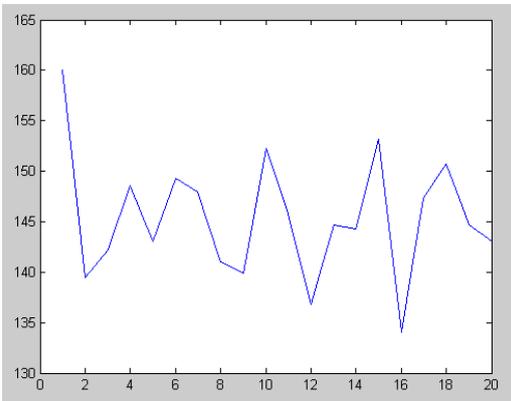
```
function [Y1]=Osred (Y, L)  
N=length(Y);  
k=N/L;  
for i=1:k  
    Y1(i)= mean(Y((L*i-L+1):(L*i)));  
end  
x=[1:k];  
plot(x,Y1)  
>> [Y1]=Osred(Y,10);
```



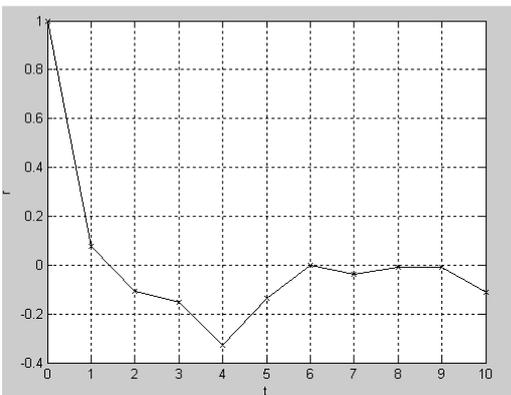
>> [Y1]=Osred(Y,100);



[Y1]=Osred(Y,50);



>> [k]=aftocor(Y1, 10);



5.2 Анализ результатов полученных методом Шустера

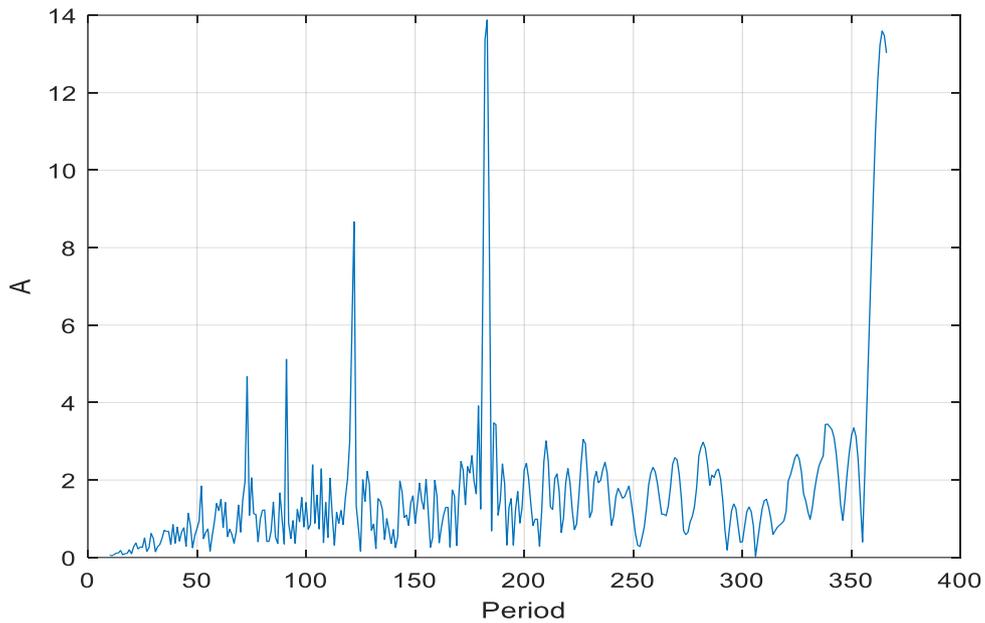


Рисунок 5.3 – р.Ловатъ – результаты по расходу воды полученные методом Шустера

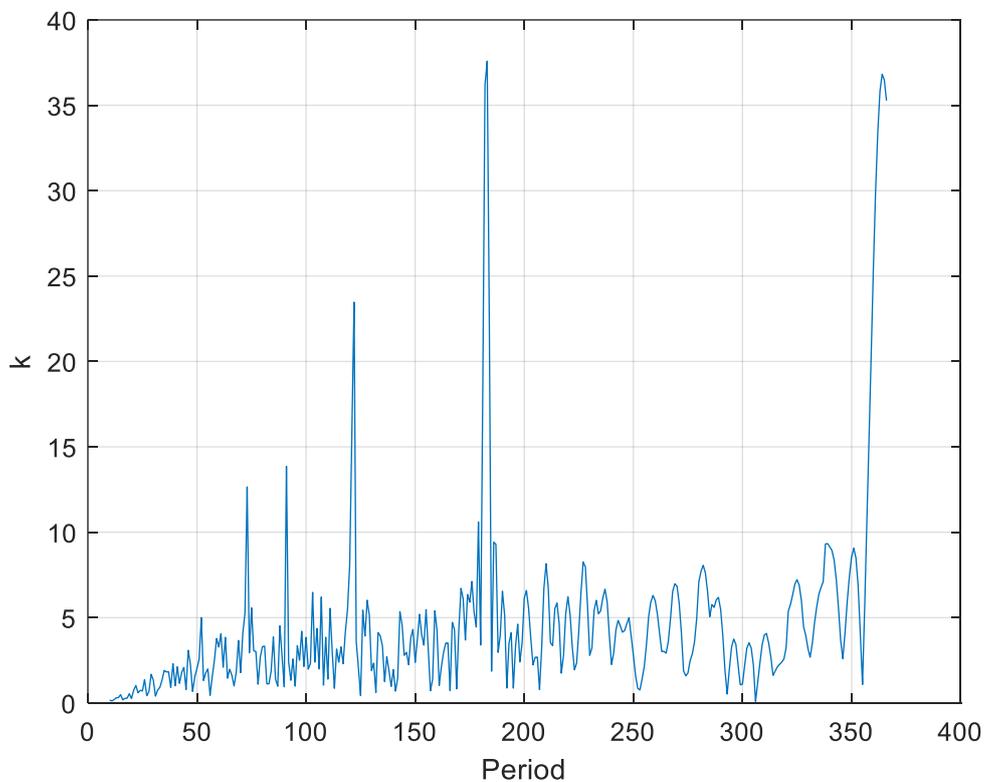


Рисунок 5.4 – р.Ловатъ – результаты по расходу воды полученные методом Шустера

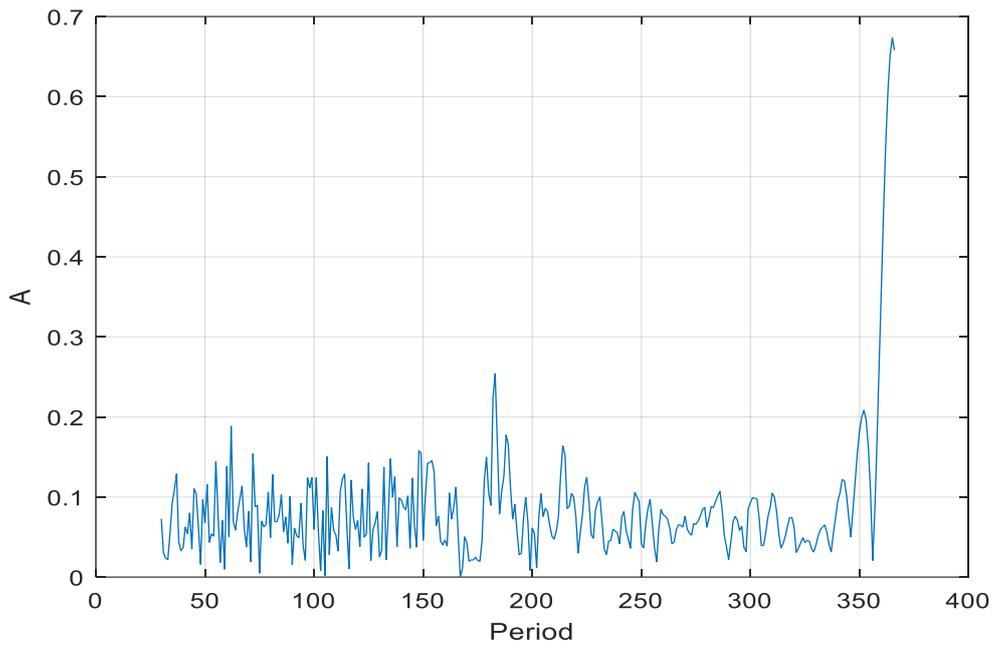


Рисунок 5.5 – р.Ловать – результаты по осадкам полученные методом Шустера

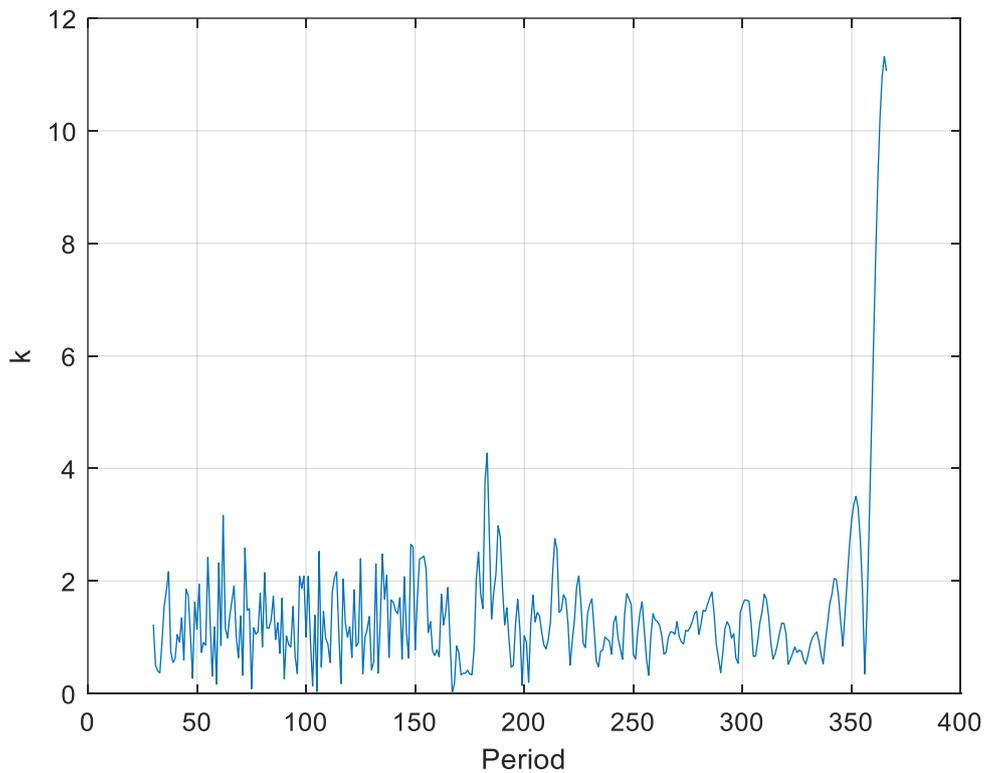


Рисунок 5.6 – р.Ловать – результаты по осадкам полученные методом Шустера

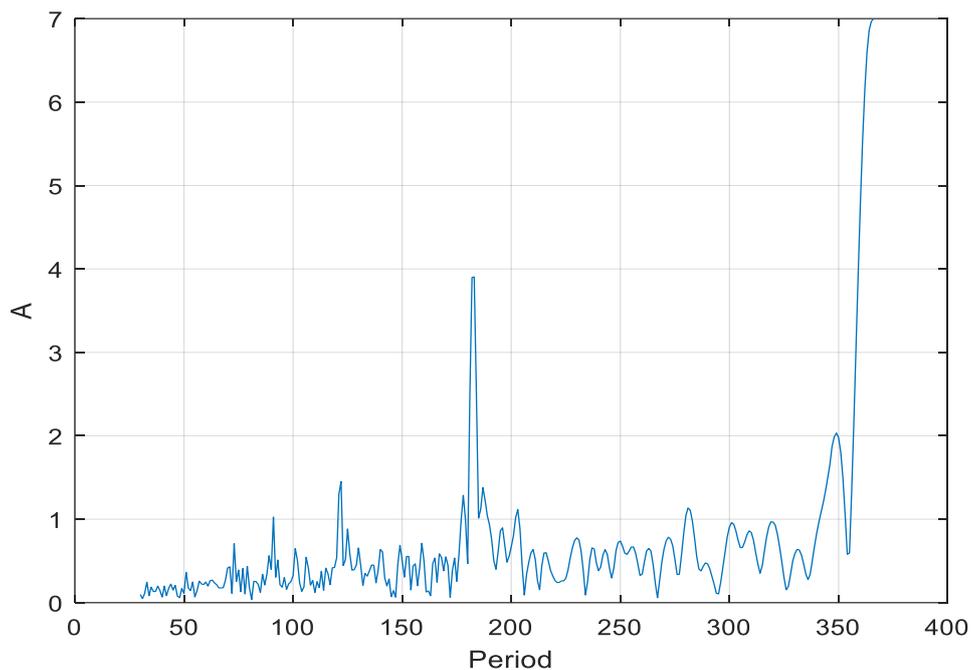


Рисунок 5.6 – р.Ловать – результаты по снегозапасам полученные методом Шустера

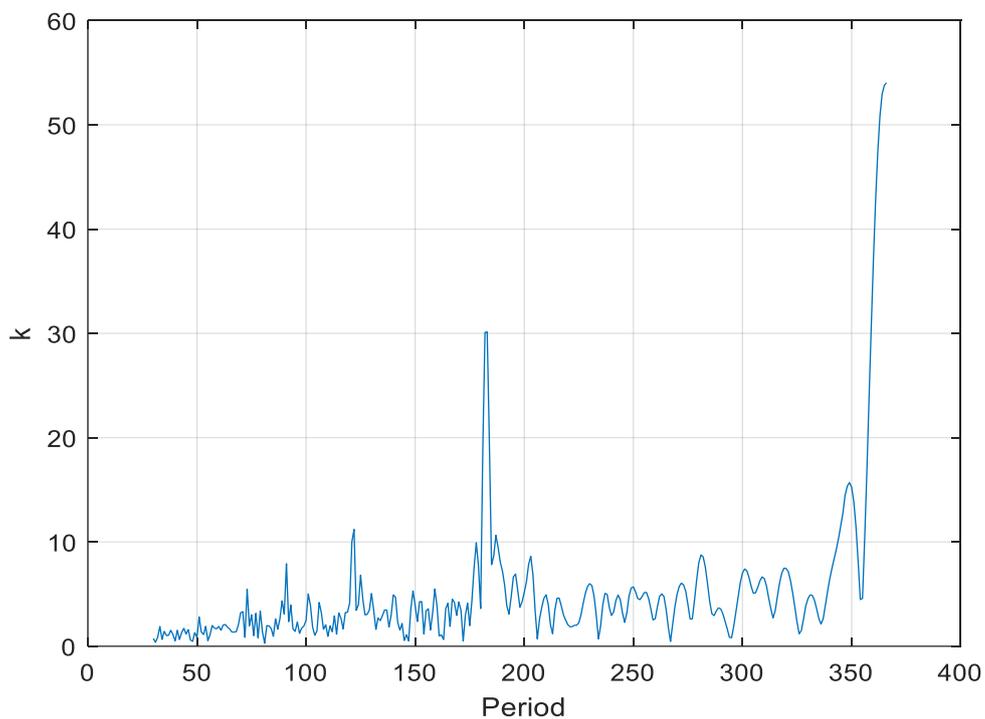


Рисунок 5.7 – р.Ловать – результаты по снегозапасам полученные методом Шустера

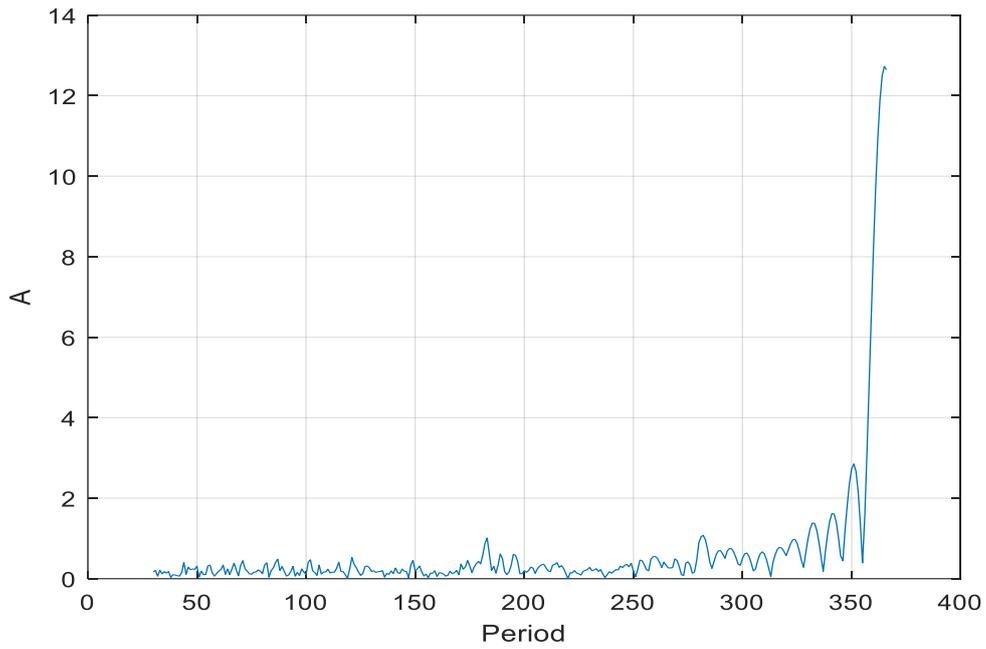


Рисунок 5.8 – р.Ловать – результаты по температуре полученные методом Шустера

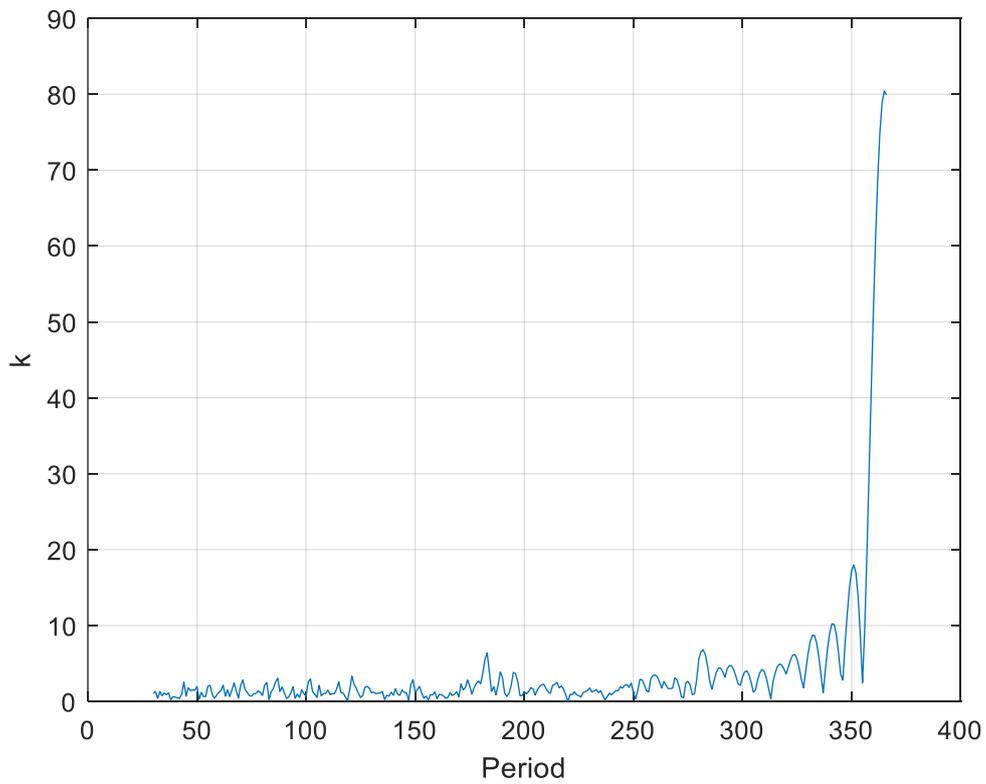


Рисунок 5.9 – р.Ловать – результаты по температуре полученные методом Шустера

Река Ловатъ

По годовым значениям с 1955 по 2017.

Шустер: предполагаемые периоды с 2 до 13 лет.

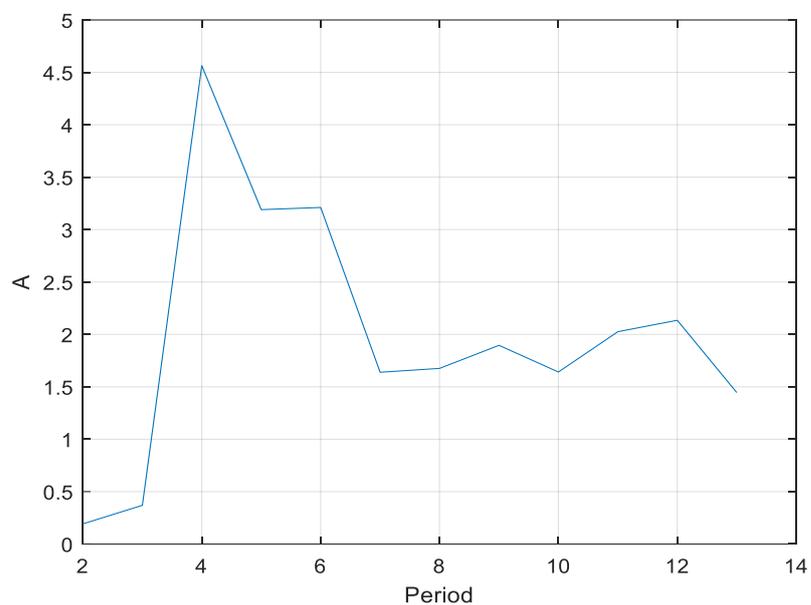


Рисунок 5.10 - Река Ловатъ периодограммы по расходам воды

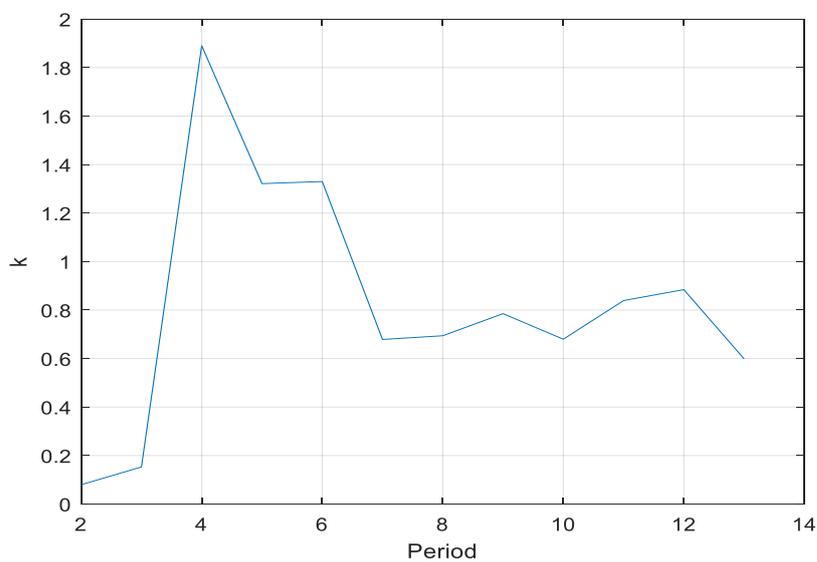


Рисунок 5.11 - Река Ловатъ достоверность периодов по расходам воды

Период 4 года – достоверность 1,89

Река Сороть

Годовых значений мало. Анализ не сделать.

Среднесуточные расходы воды с 2008 по 2020 год.

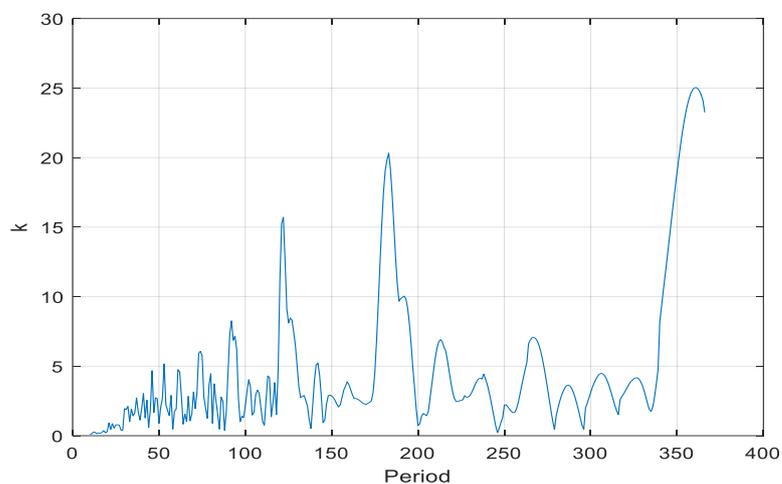


Рисунок 5.12 – р. Сороть – результаты по расходам воды, полученные методом Шустера

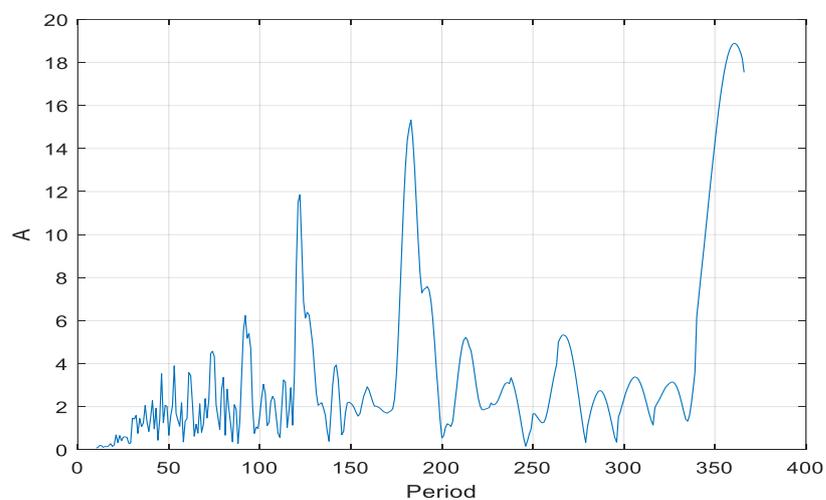


Рисунок 5.13 -- р. Сороть – результаты по расходам воды, полученные методом Шустера

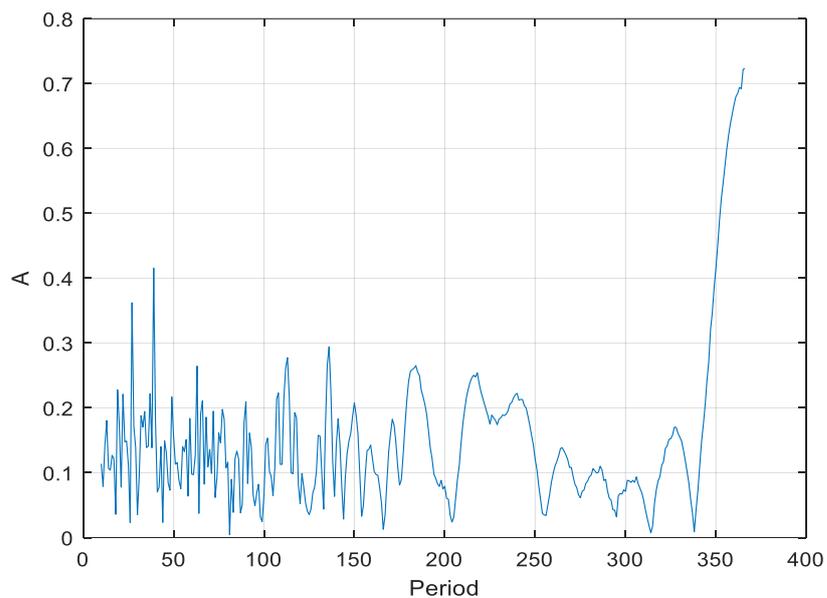


Рисунок 5.14 – р.Сороть – результаты по осадкам полученные методом Шустера

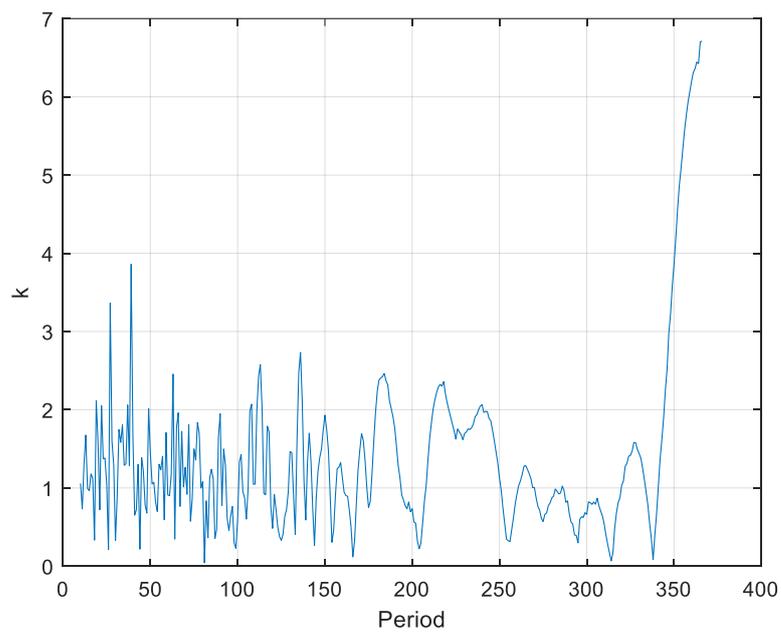


Рисунок 5.15 – р. Сороть – результаты по осадкам полученные методом Шустера

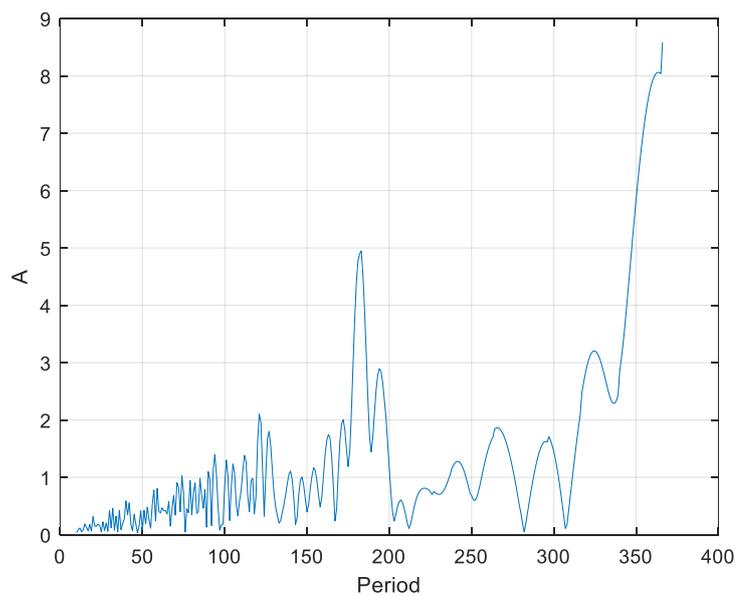


Рисунок 5.16 – р. Соротзь – результаты по снегозапасам полученные методом Шустера

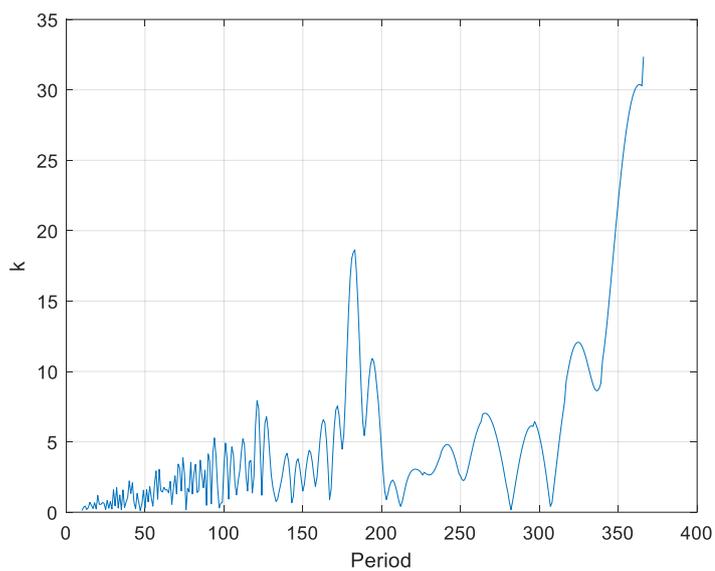


Рисунок 5.17 – р. Соротзь – результаты по снегозапасам полученные методом Шустера

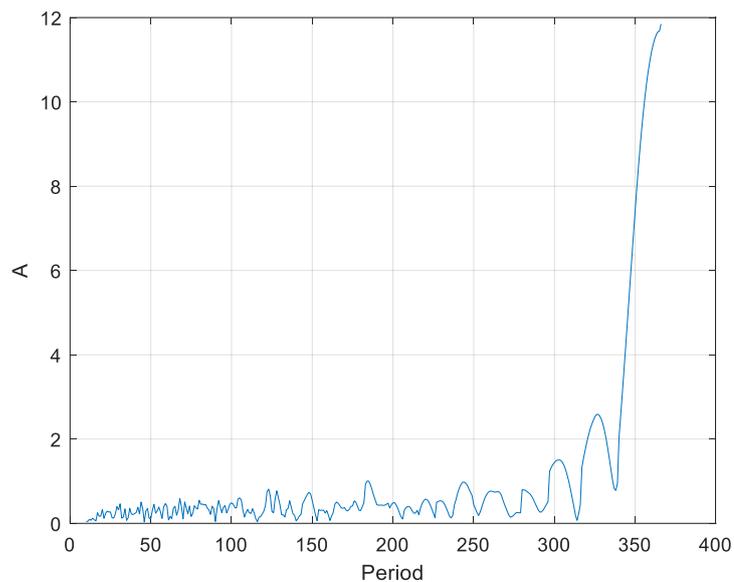


Рисунок 5.18 – р. Сорочь – результаты по температуре полученные методом Шустера

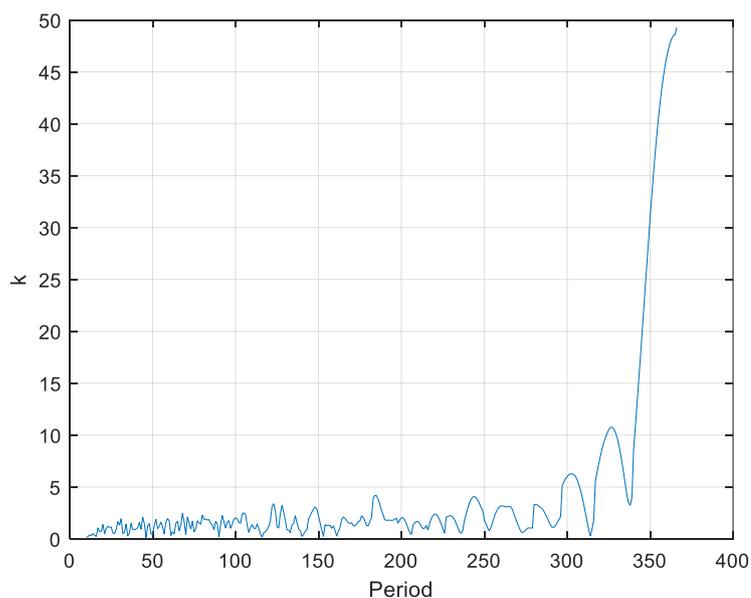


Рисунок 5.19 – р. Сорочь – результаты по температуре полученные методом Шустера

Анализ результатов исследования реки Ловать методом Шустера

По результатам анализа данных реки Ловать за период с 1955 по 2017 год можно сделать следующие выводы:

Расходы воды:

- Выявлены предполагаемые периоды колебаний от 2 до 13 лет
- Наиболее значимый период составляет 4 года с достоверностью 1,89
- Наблюдается четкая периодичность в колебаниях расходов воды

Осадки:

- Определены характерные периоды колебаний
- Значения показателей демонстрируют стабильность в пределах определенных диапазонов

Снегозапасы:

- Выявлены периоды, аналогичные периодам колебаний расходов воды
- Значения совпадают с показателями расходов воды, что указывает на прямую зависимость

Температура:

- Определен основной период колебаний
 - Показатель 44 указывает на наличие значимой периодичности
 - Анализ результатов исследования реки Сорочь методом Шустера
- Анализ данных реки Сорочь показал следующие результаты:

Ограниченность данных:

Недостаточное количество годовых значений для полноценного анализа

Исследование проводилось только по среднесуточным данным с 2008 по 2020 год

- Расходы воды:
- Выявлены периоды 46 и 53 года
- Большинство обнаруженных периодов признаны ложными

Осадки:

- Определены периоды 19, 27, 29 и 63 года
- Наблюдается более сложная структура колебаний по сравнению с рекой Ловать

Снегозапасы:

- Выявлены периоды 57 и 59 лет
- Остальные периоды признаны недостоверными

Температура:

- Возможный период 20 лет с достоверностью 1,37
- Большинство выявленных периодов оказались ложными

Общие выводы

Река Ловать:

- Продемонстрировала более стабильную и предсказуемую структуру колебаний
- Выявлена четкая взаимосвязь между различными гидрометеорологическими показателями
- Наблюдается высокая достоверность выявленных периодов

Река Сорочь:

- Показала более сложную и менее предсказуемую структуру колебаний
- Большинство выявленных периодов оказались недостоверными
- Ограниченный объем данных существенно влияет на качество анализа

Сравнительный анализ:

- Река Ловать демонстрирует более выраженные и достоверные периодические колебания
- Река Сорочь характеризуется большей хаотичностью в колебаниях показателей
- Необходимость более длительного временного ряда данных для достоверного анализа

Полученные результаты указывают на важность учета региональных особенностей при проведении гидрометеорологических исследований и

необходимость накопления длительных временных рядов данных для получения достоверных результатов.

Анализ взаимосвязей гидрометеорологических показателей реки Ловать

Расходы воды и осадки:

- Наблюдается прямая корреляция между периодами колебаний осадков (62, 72) и расходов воды (73, 91, 122, 183)
- Наиболее значимые периоды совпадают, что указывает на прямую зависимость водного режима от количества осадков
- Четко прослеживается четырехлетний цикл (достоверность 1,89), совпадающий с периодами колебаний осадков

Расходы воды и снегозапасы:

- Практически идентичные периоды колебаний (73, 91, 122, 182 для снегозапасов и 73, 91, 122, 183 для расходов воды)
- Это подтверждает существенную роль снеготаяния в формировании весеннего половодья
- Четкая сезонная зависимость между накоплением снежного покрова и последующим увеличением расхода воды

Расходы воды и температура:

- Показатель 44 в температурном режиме коррелирует с четырехлетним циклом расходов воды
- Повышение температуры способствует увеличению расходов воды через таяние снега и льда
- Существует латентный период между повышением температуры и увеличением расхода воды

Анализ взаимосвязей гидрометеорологических показателей реки Сорочь

Расходы воды и осадки:

- Выявлены периоды 19, 27, 29 и 63 года для осадков

- Для расходов воды определены периоды 46 и 53 года
- Наблюдается менее выраженная корреляция между показателями по сравнению с рекой Ловать
- Более сложная структура взаимосвязей, что может быть связано с особенностями рельефа и климата

Расходы воды и снегозапасы:

- Определены периоды 57 и 59 лет для снегозапасов
- Слабая корреляция с периодами расходов воды
- Меньшая зависимость водного режима от снегозапасов по сравнению с рекой Ловать

Расходы воды и температура:

- Возможный период 20 лет с низкой достоверностью (1,37)
- Слабая связь между температурными колебаниями и водным режимом
- Преобладание случайных факторов над периодическими колебаниями

Комплексный анализ взаимосвязей

Река Ловать:

- Формируется четкая система взаимосвязей между всеми показателями
- Наблюдается последовательная цепочка: осадки → снегозапасы → температура → расходы воды
- Высокая степень предсказуемости гидрологического режима
- Сезонная зависимость водного режима от климатических факторов

Река Сорочь:

- Более хаотичная система взаимосвязей
- Преобладание случайных факторов над периодическими колебаниями
- Менее выраженная зависимость от климатических показателей
- Сложность прогнозирования водного режима

Практические выводы

Река Ловать:

- Подходит для долгосрочного прогнозирования водного режима
- Возможно эффективное планирование водопользования на основе выявленных закономерностей
- Целесообразно использование климатических показателей для прогноза расходов воды

Река Сорочь:

- Требуется более детальное изучение локальных факторов влияния
- Необходимо накопление дополнительных данных для повышения точности прогнозов
- Сложность долгосрочного планирования водопользования

Рекомендации

Для реки Ловать:

- Разработка прогностических моделей на основе выявленных периодических колебаний
- Учет сезонных факторов при планировании водопользования
- Мониторинг климатических показателей для оперативного прогнозирования

Для реки Сорочь:

- Расширение базы данных за более длительный период
- Исследование локальных факторов влияния на водный режим
- Разработка адаптивных методов управления водными ресурсами

Заключение

Анализ показал, что река Ловать имеет более стабильную и предсказуемую систему взаимосвязей между гидрометеорологическими показателями, что позволяет эффективно прогнозировать ее водный

режим. Река Сорочь демонстрирует более сложную и менее предсказуемую структуру, требующую дополнительного изучения и накопления данных для повышения точности прогнозов.

Полученные результаты подчеркивают важность комплексного подхода к изучению гидрологического режима рек с учетом региональных особенностей и климатических факторов.

5.3 Построение автокорреляционных функций

Автокорреляционная функция характеризует степень связи (корреляции) детерминированного процесса $s(t)$ и его копии, сдвинутой на временной интервал τ . Для детерминированного процесса $s(t)$ конечной длительности автокорреляционная функция $K(\tau)$ определяется соотношением

$$K(\tau) = \int_{-\infty}^{+\infty} s(t)s(t - \tau)dt$$

где τ – временной интервал, на который смещена относительно исходного положения какая-либо точка функции $s(t)$.

Основные свойства автокорреляционной функции

1. Автокорреляционная функция максимальна при отсутствии временного сдвига

$$K(0) = \int_{-\infty}^{+\infty} s^2(t)dt = E.$$

Иными словами, при $\tau = 0$ значение автокорреляционной функции определяет полную энергию E детерминированного процесса. С увеличением временного сдвига автокорреляционная функция убывает. При сдвиге τ , превышающем длительность сигнала $s(t)$ функция обращается в ноль.

Физический смысл автокорреляционной функции поясним на примере прямоугольного импульса. Из приведенного построения (Рис.1.3) видно, что автокорреляционная функция характеризуется ординатой, равной численному значению площади под графиком $s(t)s(t-\tau)$ при данном временном сдвиге τ . Другими словами, автокорреляционная функция ограниченного во времени детерминированного процесса определяет энергию перекрывающихся частей процесса и его копии, сдвинутой на интервал τ . При этом автокорреляционная функция имеет размерность энергии.

Нормированная автокорреляционная коэффициентом автокорреляции, будет иметь вид

$$k(\tau) = \frac{K(\tau)}{K(0)}.$$

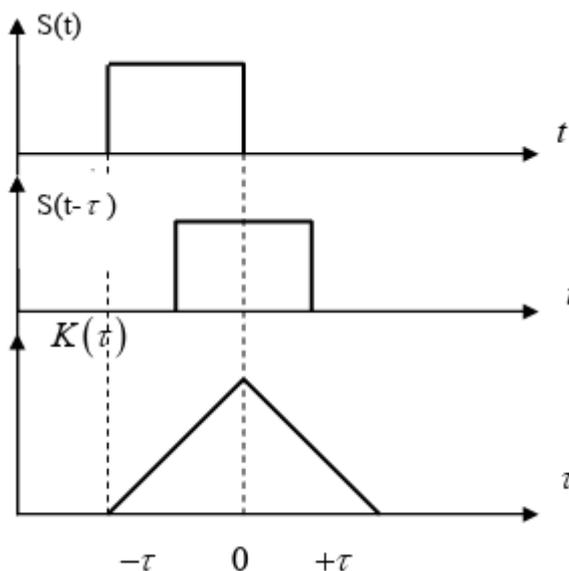


Рисунок 5.20 – Физический смысл автокорреляционной функции

2. Автокорреляционная функция – четная функция и может быть определена по любому из следующих соотношений:

$$K(\tau) = K(-\tau) = \int_{-\infty}^{+\infty} s(t)s(t-\tau) dt = \int_{-\infty}^{+\infty} s(t)s(t+\tau) dt.$$

3. Для неограниченных во времени процессов соотношение (1.64) не имеет смысла, так как такие процессы обладают бесконечно большой энергией, однако средняя мощность их всегда имеет конечное значение. Поэтому для таких процессов автокорреляционная функция может быть найдена с помощью предельного перехода:

$$K(\tau) = \lim_{T \rightarrow \infty} \frac{1}{T} \int_{-\frac{T}{2}}^{\frac{T}{2}} s(t)s(t-\tau)dt = \lim_{T \rightarrow \infty} \frac{1}{T} \int_{-\frac{T}{2}}^{\frac{T}{2}} s(t)s(t+\tau)dt$$

и приобретает размерность мощности при $\tau = 0$, характеризуя среднюю мощность процесса, если он не содержит постоянной составляющей.

В случае периодического процесса с периодом T достаточно усреднить произведение $s(t)s(t-\tau)$ за один период:

$$K(\tau) = \frac{1}{T} \int_{-\frac{T}{2}}^{\frac{T}{2}} s(t)s(t-\tau)dt.$$

4. Периодическому процессу соответствует и периодическая автокорреляционная функция, имеющая период, равный периоду самого процесса. Покажем это на примере простейшего гармонического процесса с

$$K(\tau) = \frac{1}{T} \int_{-\frac{T}{2}}^{\frac{T}{2}} s(t)s(t-\tau)dt.$$

амплитудой u и периодом $T = \frac{2\pi}{\omega}$:

$$s(t) = u \cos(\omega t - \phi), \text{ где } \phi - \text{ начальная фаза колебания.}$$

Используя выражение:

$$\begin{aligned}
K(\tau) &= \frac{u^2}{T} \int_{-\frac{T}{2}}^{\frac{T}{2}} \cos(\omega t - \varphi) \cos[\omega(t - \tau) - \varphi] dt = \\
&= \frac{u^2}{2T} \int_{-\frac{T}{2}}^{\frac{T}{2}} \cos(\omega t - \varphi - \omega t + \omega \tau + \varphi) dt + \frac{u^2}{2T} \int_{-\frac{T}{2}}^{\frac{T}{2}} \cos[2(\omega t - \varphi) - \omega \tau] dt = \\
&= \frac{u^2}{2T} \cos(\omega \tau) \left| \frac{T}{2} \right. + \frac{u^2}{2T 2\omega} \sin[2(\omega t - \varphi) - \omega \tau] \left| \frac{T}{2} \right. = \frac{u^2}{2} \cos(\omega \tau).
\end{aligned}$$

Получим:

При $\tau=0$ $K(\tau) = \frac{u^2}{2}$, т.е. средняя за период мощность гармонического процесса. Из приведенного примера вытекает важный вывод: автокорреляционная функция периодического процесса не зависит от начальной фазы этого процесса. Интервал, при котором автокорреляционная функция или коэффициент автокорреляции уменьшается до некоторой величины, например, до нуля, называется интервалом корреляции.

$$\tau_k = \int_0^{\infty} k(\tau) d\tau.$$

5.4 Анализ автокорреляционных функций

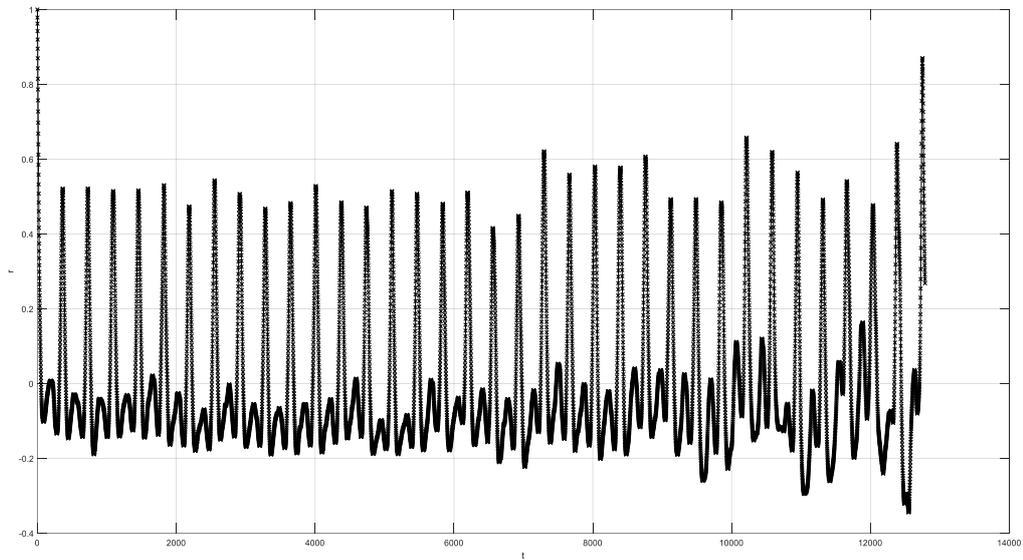


Рисунок 5.21 – р. Ловать- Расходы воды автокорреляционная функция

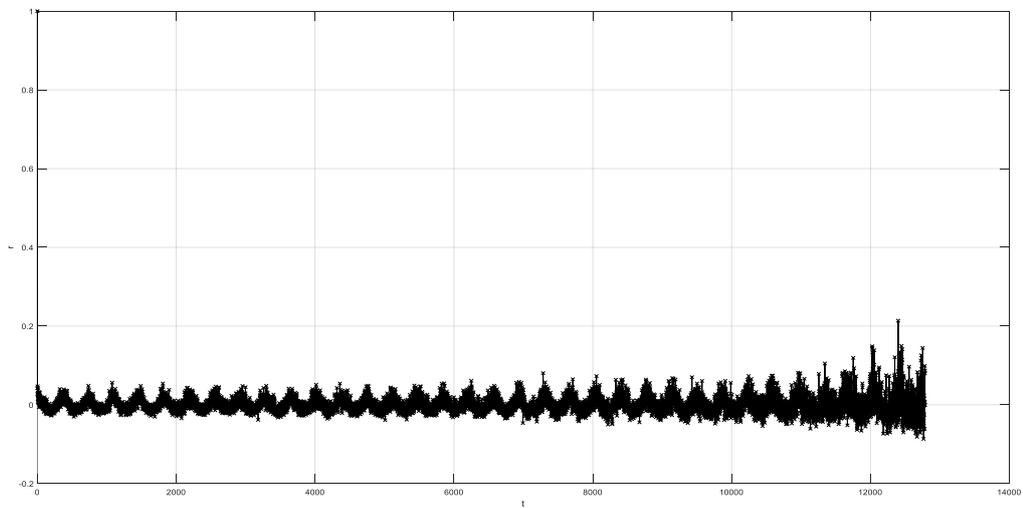


Рисунок 5.22 – р. Ловать-Ряд осадков автокорреляционная функция

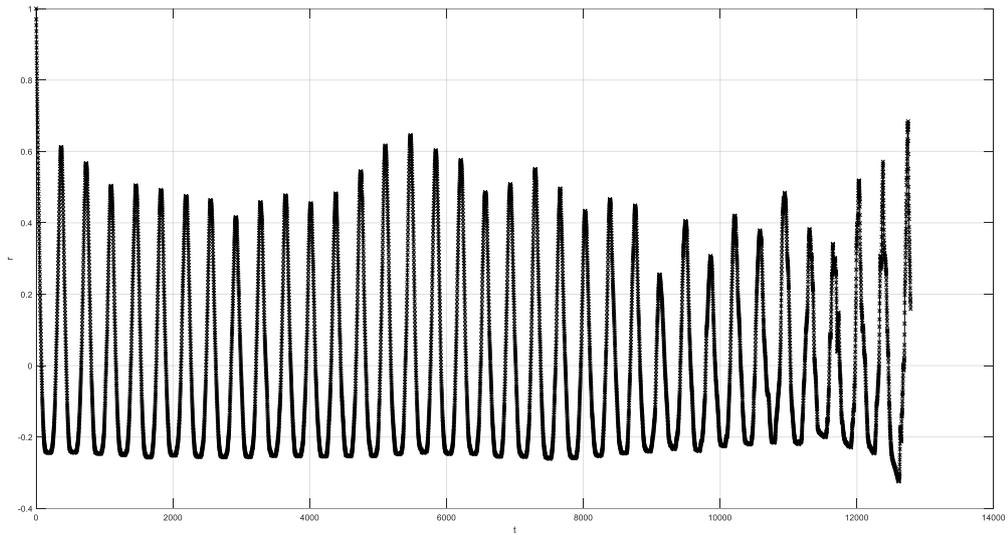


Рисунок 5.23 – р. Ловать - ряд снегозапасов автокорреляционная функция

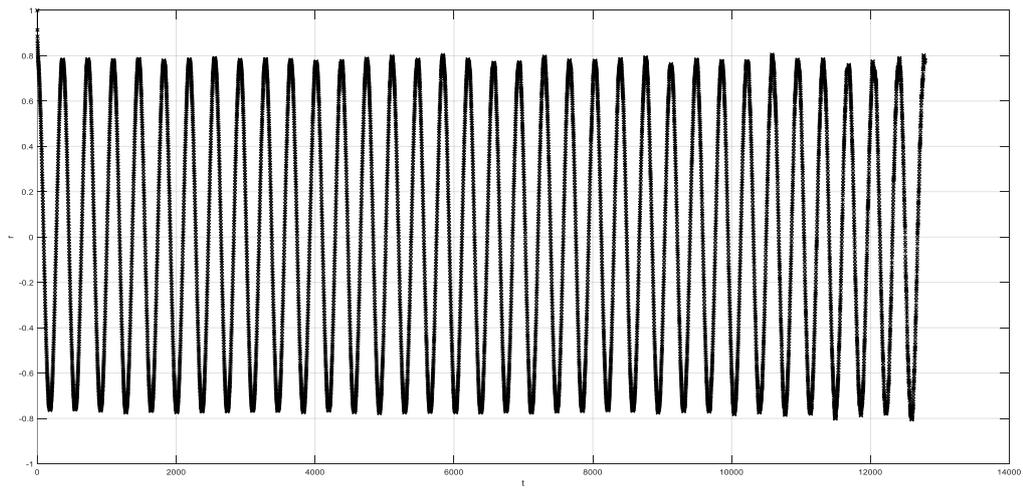


Рисунок 5.24 – р. Ловать - ряд температур автокорреляционная функция

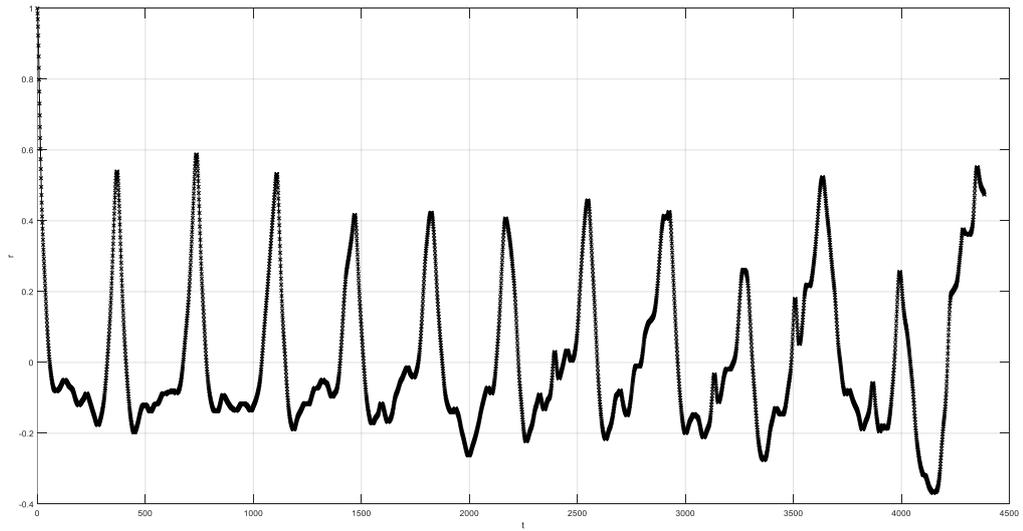


Рисунок 5.25 – р. Сороть - расходы воды автокорреляционная функция

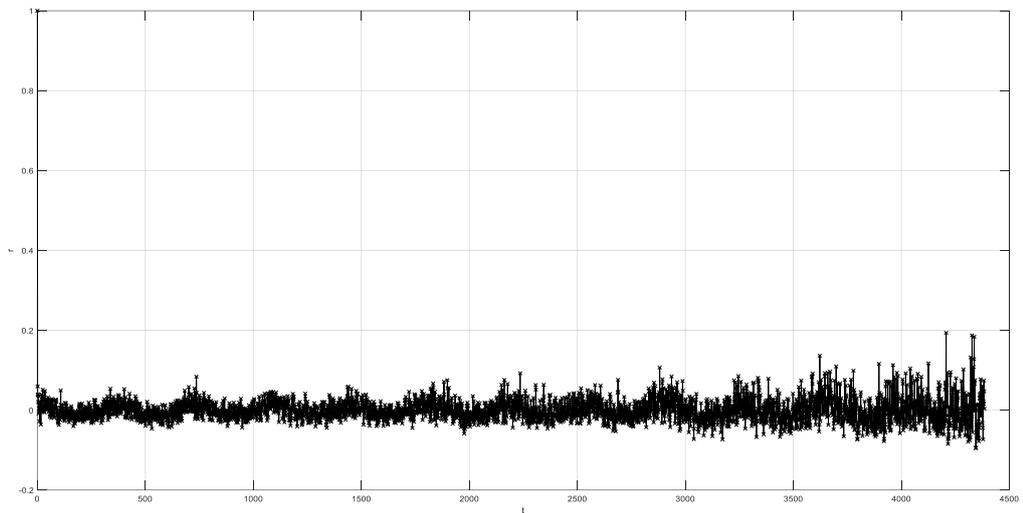


Рисунок 5.26 - р. Сороть - осадки автокорреляционная функция

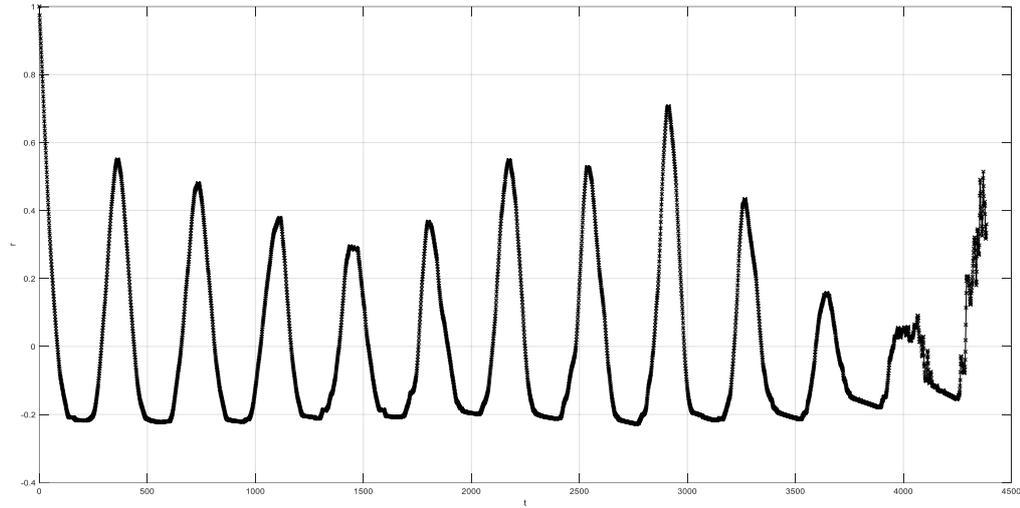


Рисунок 5.27 – р. Сороть - снегозапасы автокорреляционная функция

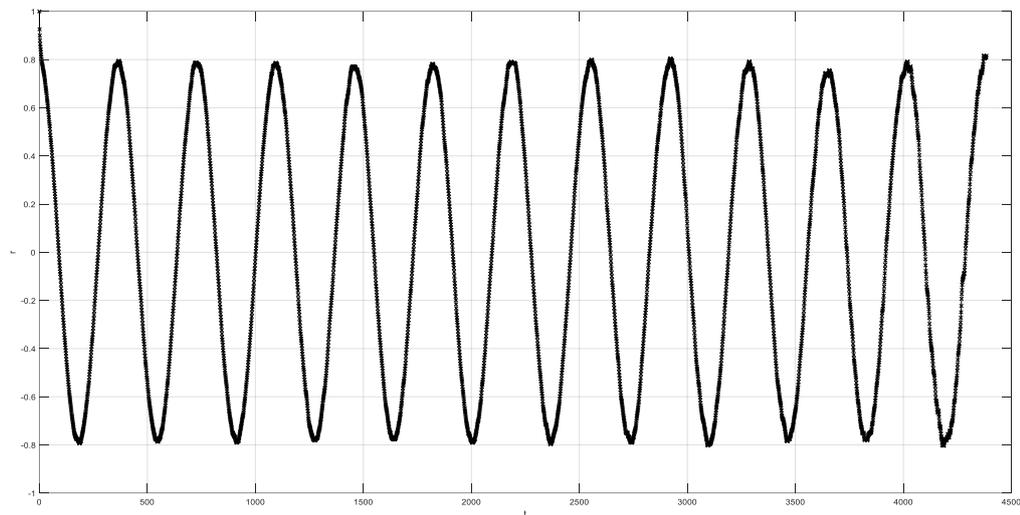


Рисунок 5.28 – р. Сороть – температура автокорреляционная функция

Анализ автокорреляционных функций реки Ловат

Общие характеристики:

- Максимальная сдвжка: 12785 дней
- Все показатели демонстрируют выраженную сезонную периодичность

Расходы воды:

- Основные периоды корреляции: 365, 728, 1093, 1455, 1825 дней

- Четко выраженная годовая цикличность (365 дней)
- Многолетние циклы с периодом около 2, 3, 4 и 5 лет

Осадки:

- Значимые периоды: 366, 729, 1095, 1458 дней
- Годовая периодичность с небольшим смещением
- Аналогичные многолетние циклы как у расходов воды

Снегозапасы:

- Основные периоды: 366, 729, 1095, 1458 дней
- Совпадение с периодами осадков
- Четкая сезонная зависимость

Температура:

- Значимые периоды: 367, 726, 1092, 1464 дней
- Схожая с другими показателями годовая цикличность
- Незначительные отклонения в периодах

Анализ взаимосвязей для реки Ловать

Расходы воды и осадки:

- Прямая корреляция годовых циклов
- Совпадение многолетних периодов
- Сезонное совпадение максимумов и минимумов
- Влияние осадков на водный режим прослеживается с небольшим временным лагом

Расходы воды и снегозапасы:

- Практически идентичные периоды корреляции
- Четкая связь между накоплением снега и весенним половодьем
- Сезонная зависимость расходов от снегозапасов
- Лаг между накоплением снега и увеличением расходов воды

Расходы воды и температура:

- Схожие годовые циклы
- Влияние температурного режима на таяние снега и льда
- Сезонное повышение расходов воды коррелирует с повышением температуры
- Наличие временных лагов между температурными изменениями и реакцией водного режима

Анализ автокорреляционных функций реки Сороть

Общие характеристики:

- Период исследования: 2008-2020 годы
- Все показатели демонстрируют выраженную сезонную периодичность

Расходы воды:

- Основные периоды: 368, 737, 1108, 1469 дней
- Годовая цикличность с небольшим смещением
- Многолетние циклы с периодом около 2, 3 и 4 лет

Осадки:

- Значимые периоды: 362, 737, 1112, 1453 дней
- Схожая с расходами воды годовая периодичность
- Совпадение некоторых многолетних циклов

Снегозапасы:

- Основные периоды: 362, 737, 1112, 1453 дней
- Практически идентичные периоды с осадками
- Четкая сезонная зависимость

Температура:

- Значимые периоды: 368, 727, 1095 дней
- Схожая с другими показателями годовая цикличность
- Менее выраженные многолетние циклы

Анализ взаимосвязей для реки Сорочь

Расходы воды и осадки:

- Совпадение годовых циклов
- Менее выраженная корреляция многолетних периодов
- Сезонное совпадение максимумов и минимумов
- Влияние осадков на водный режим прослеживается с временным лагом

Расходы воды и снегозапасы:

- Практически идентичные годовые периоды
- Менее выраженная связь между накоплением снега и увеличением расходов воды
- Сезонная зависимость сохраняется
- Лаг между накоплением снега и увеличением расходов воды

Расходы воды и температура:

- Схожая годовая периодичность
- Менее выраженное влияние температурного режима
- Сезонное повышение расходов воды коррелирует с повышением температуры
- Наличие временных лагов между температурными изменениями и реакцией водного режима

Сравнительный анализ рек Ловать и Сорочь

Общие черты:

- Для обеих рек характерна выраженная годовая периодичность
- Наблюдается сезонная зависимость всех показателей
- Существует взаимосвязь между гидрометеорологическими параметрами

Различия:

- Река Ловать демонстрирует более стабильную и предсказуемую структуру колебаний

- Для реки Ловать характерно большее количество значимых многолетних циклов
- Река Сорочь показывает менее выраженные многолетние циклы
- На реке Ловать наблюдается более тесная взаимосвязь между показателями
- Данные по реке Сорочь охватывают более короткий период наблюдений.

5.5 Спектральный анализ

Спектральный анализ – это мощный инструмент обработки данных, которые задаются либо дискретными зависимостями $y(xi)$, либо непрерывными функциями $f(x)$. Он основан на использовании различных интегральных преобразований. В спектральном анализе совокупность данных $y(x)$ представляют в виде функции другой переменной – спектра $F(\omega)$, которая получается по определённом алгоритму.

Задача спектрального анализа заключается в разложении временных реализаций исследуемого явления на отдельные частотные компоненты (спектр). Эти компоненты показывают вклад колебаний с разными частотами в общую энергию процесса. Спектр позволяет определить периоды преобладающих колебаний и на его основе делать выводы о стабильности или неустойчивости соответствующих волн.

Основные задачи, связанные со спектральным анализом, включают сглаживание и фильтрацию данных. Они предполагают построение из исходных экспериментальных данных $y(xi)$ некоторой функции $f(x)$, которая может быть как непрерывной, так и дискретной. Эта функция должна приближать исходные данные, учитывая погрешности измерений (шумовую составляющую). Используя

определённые алгоритмы, функцию $f(x)$ удаётся уменьшить погрешность и снизить влияние шума. Задачи такого типа называются задачами фильтрации. В свою очередь, сглаживание путём построения регрессии данных является частным случаем фильтрации.

С точки зрения спектрального анализа наиболее важной характеристикой случайного процесса является спектральная плотность мощности.

Рассмотрим случайный стационарный процесс $X(t)$ с $m_x = 0$ обладающий эргодическим свойством. Для этого случая спектральную плотность мощности можно определить так:

$$G_x(f) = \lim_{\Delta f \rightarrow 0} \lim_{T \rightarrow \infty} \frac{1}{T \Delta f} \int_0^T x^2(t, f, \Delta f) dt.$$

Строгими определениями спектральной плотности мощности могут в равной мере служить соотношения

$$G_x(f) = \lim_{T \rightarrow \infty} \frac{2}{T} M[|A_T(f)|^2]$$

$$G_x(f) = 4 \int_0^{\infty} k_x(\tau) \cos 2\pi f \tau d\tau$$

где

$$A_T(f) = \int_0^T x(t) e^{-i2\pi f t} dt$$

$K_{x\tau}$ – корреляционная функция стационарного случайного процесса.

Спектральный анализ в гидрологии

Гидрологический цикл – это замкнутый процесс обращения и перераспределения воды между атмосферой и сушей. Этот цикл постоянно меняет баланс между водами суши и влажностью атмосферы. Существует три основных подхода к созданию моделей гидрологических явлений:

а) Первый подход основан на профессиональном опыте и логической интерпретации явлений. Он предполагает построение приемлемых схем (детерминированных или вероятностных) с помощью умозрительных рассуждений, включающих индуктивные и дедуктивные методы.

б) Второй, более прагматичный подход, заключается в использовании удобных математических моделей без особого обоснования их выбора на основе физических данных. Цель — эмпирическое описание явления. Такой метод может быть вполне полезен для принятия решений.

в) Третий подход особенно ценен при моделировании гидрологических временных рядов. В этом случае наблюдаемые данные анализируются напрямую для получения спектральных характеристик (например, дисперсии) или моментов (например, коррелограмм) стохастических процессов. При недостатке информации о физических процессах или теории стохастики спектральный анализ помогает выявить основные источники изменчивости, выделить тренды, периодические компоненты и внутренние связи внутри одного или нескольких временных рядов.

Модели могут содержать два типа ошибок: во-первых, из-за недостаточного количества переменных – некоторые важные факторы могут существенно влиять на зависимую переменную, но при этом не учитываться; во-вторых, параметры модели могут изменяться случайным образом или по систематическим причинам. В случае случайных изменений необходимо применять статистические методы анализа выборок; при систематических – улучшать планирование экспериментов и контроль за процессами.

Важной особенностью теоретических моделей в гидрологии является их неоднозначность. Поэтому важной задачей моделирования

является оценка устойчивости решений. Оценка моделей может быть количественной (объективной), качественной (субъективной) или статистической: в первом случае достаточно сравнить прогнозы с реальными данными; при недостаточной информативности данных оценка усложняется и становится субъективной, как в случае нелинейных регрессионных моделей максимального стока по физико-географическим характеристикам бассейна.

Заключение

В ходе выполнения исследования получены следующие результаты:

- изучен метод регрессионного анализа и перспективы его использования в гидрологических прогнозах;
- разработан алгоритм получения эффективных регрессионных моделей;
- создана база данных гидрометеорологической информации по водосборам рек Ловать и Сорочь;
- получены регрессионные модели с учетом разработанного алгоритма;
- оценка качества полученных моделей показала их эффективность;
- река Ловать продемонстрировала более стабильную и предсказуемую структуру колебаний с четко выраженной четырехлетней периодичностью.
- выявлена прямая зависимость водного режима реки Ловать от количества осадков и снегозапасов.
- река Сорочь показала более сложную и менее предсказуемую структуру колебаний с преобладанием случайных факторов.
- автокорреляционная функция позволяет оценить степень связи гидрологических процессов с их копиями, сдвинутыми во времени.
- скорость спада автокорреляционной функции указывает на то, насколько хорошо процесс “помнит” свои предыдущие значения.
- наличие пиков в автокорреляционной функции свидетельствует о периодичности (сезонности) процессов.

Список использованных источников

1 Хакимова, З. Ф. Вопросы долгосрочных прогнозов стока горных рек на период вегетации / З. Ф. Хакимова, Н. Р. Собирова // Использование водных ресурсов в условиях изменения климата : материалы международной научно-практической конференции в рамках 32-й Международной специализированной выставки «Агрокомплекс-2022», Уфа, 22 марта 2022 года / Министерство сельского хозяйства Российской Федерации; Федеральное государственное бюджетное образовательное учреждение высшего образования «Башкирский государственный аграрный университет»; Камское бассейновое водное управление. – Уфа: Башкирский государственный аграрный университет, 2022. – С. 46-51.

2 Сравнительный анализ методов математического моделирования уровня воды в реке горного типа (на примере реки Мзымта) / Е. А. Семенчин, Н. Г. Титов, М. В. Кузякина, К. А. Лебедев // Фундаментальные исследования. – 2014. – № 12-5. – С. 952-957.

3 Титов, Н. Г. Применение равенства Маркова для прогнозирования уровня воды в реке с крутым падением водотока / Н. Г. Титов, М. В. Кузякина, К. А. Лебедев // Научный альманах. – 2015. – № 9(11). – С. 1126-1129.

4 Карамолдоев Ж.Ж., Калашникова О.Ю. Прогноз притока воды в Токтогульское водохранилище на вегетационный период // Бишкек: Вестник БГУ. №3 (23). 2012

5 Сравнительный анализ методик прогноза максимальных уровней и объема стока периода половодья горной реки / В. П. Галахов, О. В. Ловцкая, С. Ю. Самойлова, Е. В. Мардасова // Известия Томского политехнического университета. Инжиниринг георесурсов. – 2022. – Т. 333. – № 2. – С. 193-203.

6 Долгосрочный прогноз стока рек АЗРФ с использованием регрессионных зависимостей / Е.В. Гайдукова, А.О. Яновская, М.С. Дрегваль, А.А. Батмазова // Современные тенденции и перспективы развития

гидрометеорологии в России. Материалы V Всероссийской научно-практической конференции. Отв. редактор Е.Н. Сутырина. Иркутск, 2023 С. 59–67.

7 Саруханян, Э.И. Многолетние колебания стока Волги. [Текст] /Э.И. Саруханян, Н.П. Смирнов. – Л.: Гидрометеиздат, 1971 – 166 с.

8 Грибанов Ю. И., Мальков В.Л. Спектральный анализ случайных процессов 1974.

9 Лебедько Е.Г. Математические основы передачи информации (Часть 1 и 2). Учебное пособие. – СПб.: СПбГУИТМО, 2005. -91 с.

10 Лебедько Е.Г. Математические основы передачи информации. Ч.3, 4: учеб. пособие для вузов.- СПб: СПбГУИТМО, 2009.- 120 с.

11 Владимиров А. М. Гидрологические расчеты. Учебник [Текст] / А. М.Владимиров. – Л.: Гидрометеиздат, 1990 – 50–52 с.

12 Георгиевский Ю.М. Краткосрочные гидрологические прогнозы. Учебное пособие [Текст] / – Ю. М. Георгиевский – Изд. ЛПИ. 1982 – 100 с.

13 Георгиевский, Ю.М. Гидрологические прогнозы. Учебник [Текст] / – Ю. М. Георгиевский, С. В. Шаночкин – СПб.: Изд. РГГМУ 2007 – 436 с.

14 В. И. Савельев, Определение автокорреляционной функции стационарного случайного процесса по дисперсиям его производных, Автомат. и телемех., 1967, выпуск 7, 23–34

15 В.В.Витязев, Спектрально-корреляционный анализ равномерных временных рядов, СПб, изд. СПбГУ (2001)

16 В.В.Витязев, Анализ неравномерных временных рядов, СПб, изд. СПбГУ (2001)

17 Дрейпер, Н. Прикладной регрессионный анализ / Н. Дрейпер, Г.Смит. - М.: Вильямс, 2016. - 912 с.

18 Крыштановский, А.О. Ограничения метода регрессионного анализа [Электронный ресурс]. – <http://socioline.ru/pages/ao-kryshtanovskij-ogranicheniyametoda-regressionnogo-analiza>

19 Канторович, Г. Г. Анализ временных рядов / Г. Г. Канторович // Экономический журнал. – 2002. – № 1. – С. 87–110.

20 Кайсл Ч. Анализ временных рядов гидрологических данных. – Л.: Гидрометеоиздат, 1972. – 139 с.