



МИНИСТЕРСТВО НАУКИ И ВЫСШЕГО ОБРАЗОВАНИЯ РОССИЙСКОЙ
ФЕДЕРАЦИИ

федеральное государственное бюджетное образовательное учреждение
высшего образования

«РОССИЙСКИЙ ГОСУДАРСТВЕННЫЙ
ГИДРОМЕТЕОРОЛОГИЧЕСКИЙ УНИВЕРСИТЕТ»

Кафедра инженерной гидрологии

ВЫПУСКНАЯ КВАЛИФИКАЦИОННАЯ РАБОТА
(бакалаврская работа)

На тему **Использование регрессионных моделей**
в прогностических целях

Исполнитель Филатов Дмитрий Михайлович
(фамилия, имя, отчество)

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

Гайдукова Екатерина Владимировна
(фамилия, имя, отчество)

Консультант _____
(ученая степень, ученое звание)

(фамилия, имя, отчество)

**«К защите допускаю»
Заведующий кафедрой**

(подпись)

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

Хаустов Виталий Александрович
(фамилия, имя, отчество)

«05» июня 2023 г.

Санкт-Петербург
2023

Содержание

Введение	3
1 Регрессионный анализ данных	5
1.1 Алгоритм регрессионного анализа	5
1.2 Прогностические возможности регрессионных моделей	14
1.3 Алгоритм увеличения эффективности прогнозных регрессионных моделей	20
2 Формирование базы данных для исследования	21
2.1 Выбор речных водосборов	21
2.2 Анализ исходных гидрометеорологических данных	21
3 Применение разработанного эффективного алгоритма для прогнозов по регрессионным моделям	28
4 Анализ полученных результатов и рекомендации для использования разработанного алгоритма	44
Заключение	46

ВВЕДЕНИЕ

Эффективное применение регрессионных зависимостей для прогноза гидрологических характеристик горных рек отмечают некоторые исследователи. Так, например, для некоторых горных рек Узбекистана рассчитаны многофакторные зависимости между стоком рек вегетационного периода и атмосферными осадками [1]. Для прогноза уровня воды реки Мзымта (Краснодарский край) предложены методики, основанные на регрессионном анализе и на использовании нейросетевых технологий, дающие примерно равные результаты [2] и на основе теории Марковских процессов с дискретным временем [3]. 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.

1 Регрессионный анализ данных

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}	
.....	
$x_{(n-L)1}$	$x_{(n-L)2}$	$X_{(n-L)j}$	$x_{(n-L)k}$	
$x_{(n-L)1}$	$x_{(n-L)2}$	$X_{(n-L)j}$	$x_{(n-L)k}$	$y_{(n-L)(k+1)}$
.....
$x_{(n-L)1}$	$x_{(n-L)2}$	$X_{(n-L)j}$	$x_{(n-L)k}$	$y_{(n-L)(k+1)}$
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.

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

или

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

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

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

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

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 x_y 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 мультилинейные 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.

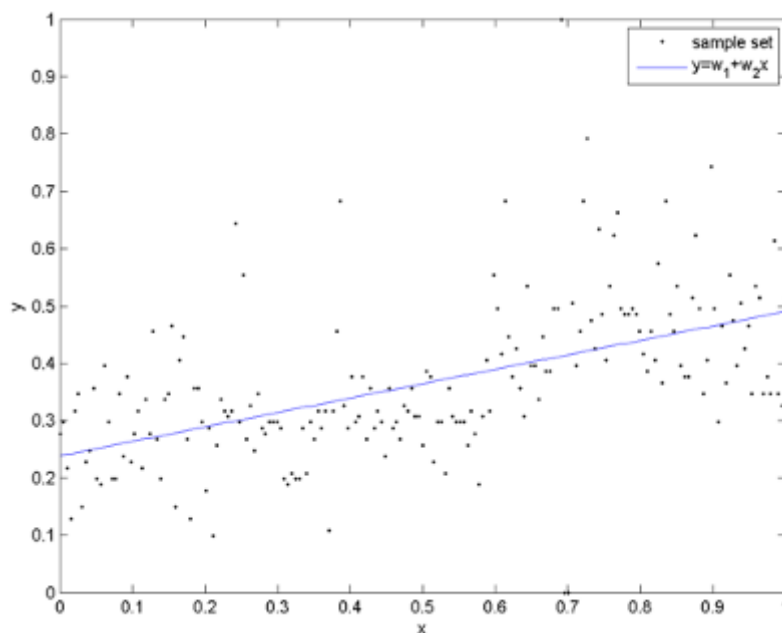


Рисунок 1.1 – Линия линейной регрессии

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

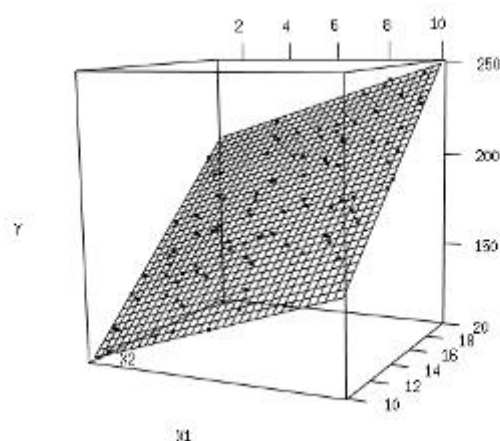


Рисунок 1.2 – Множественная регрессия

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

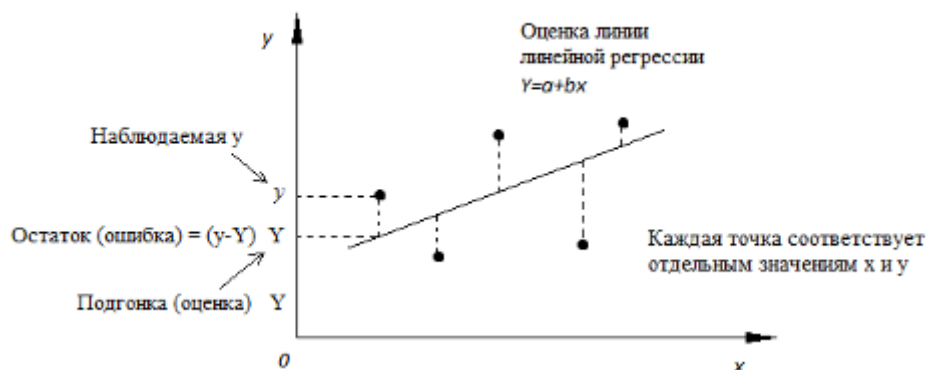


Рисунок 1.3 – Линия линейной регрессии. Пунктиром изображено расстояние $y - Y$ для каждой точки.

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 с 1955 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 находиться 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 $Q=308 \text{ м}^3/\text{с}^3$.

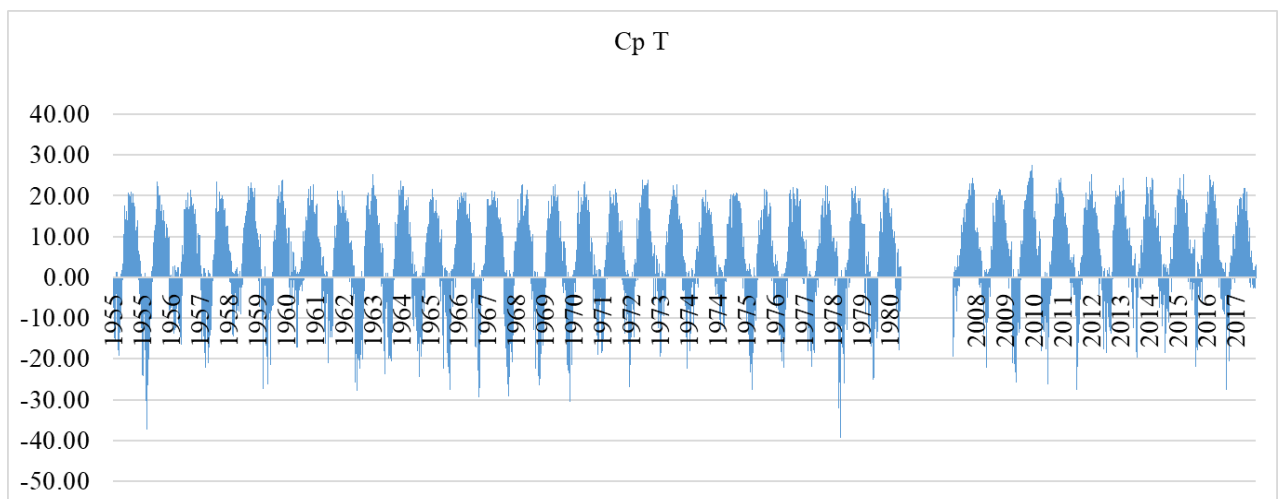


Рисунок 2.2 – Средняя температура по годам метеостанция Великие Луки

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

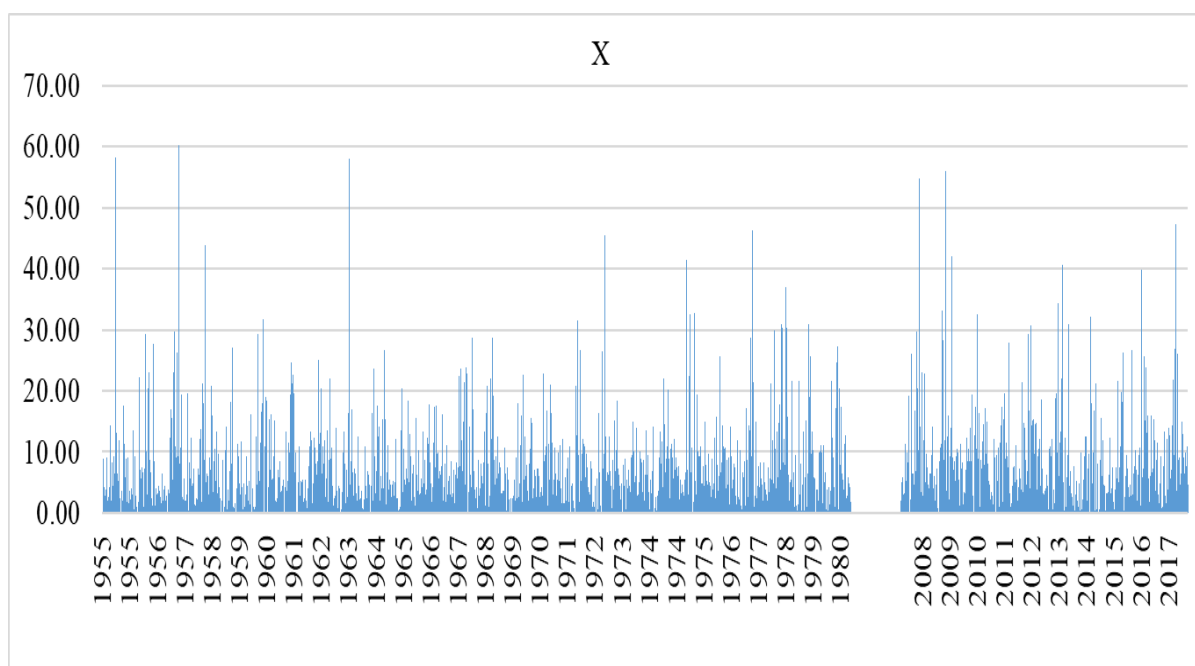


Рисунок 2.3 – Осадки метеостанция Великие Луки

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.

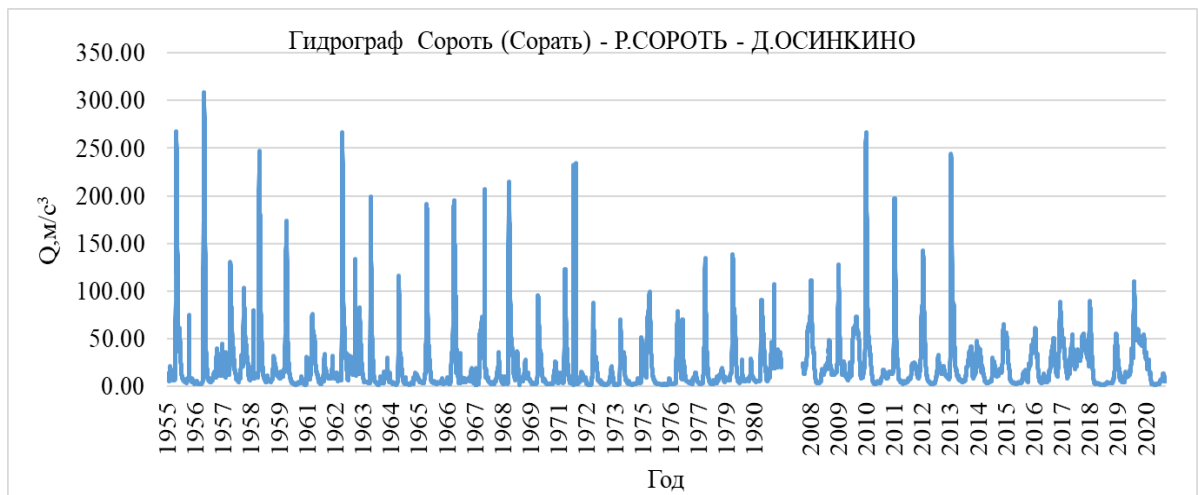


Рисунок 2.5 – Гидрограф Сороть (Сорать) - Р.СОРОТЬ - Д.ОСИНКИНО

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.

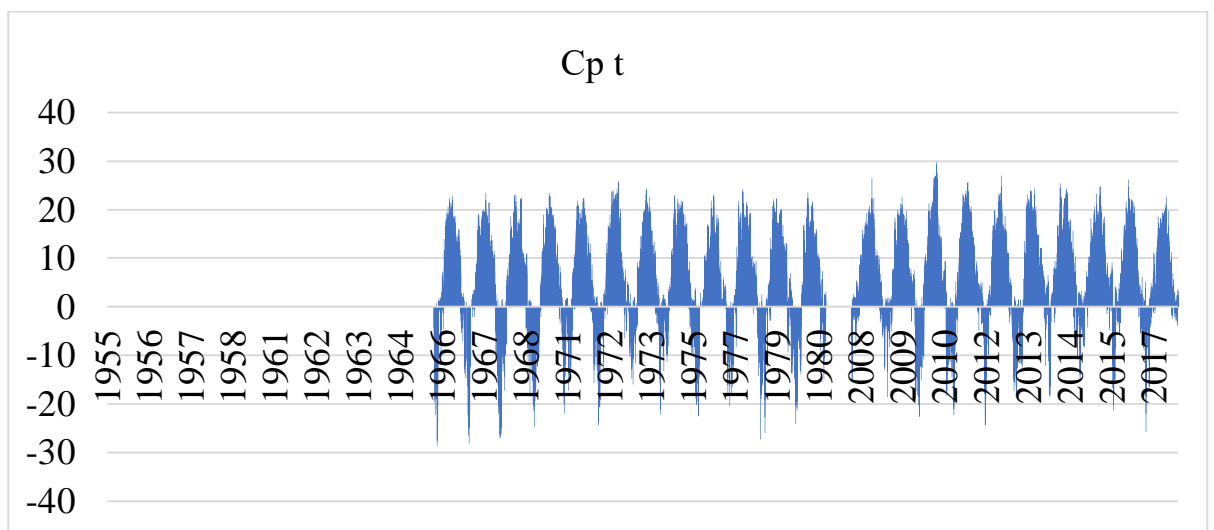


Рисунок 2.6 – Средняя температура по годам метеостанция Пушкинские горы

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

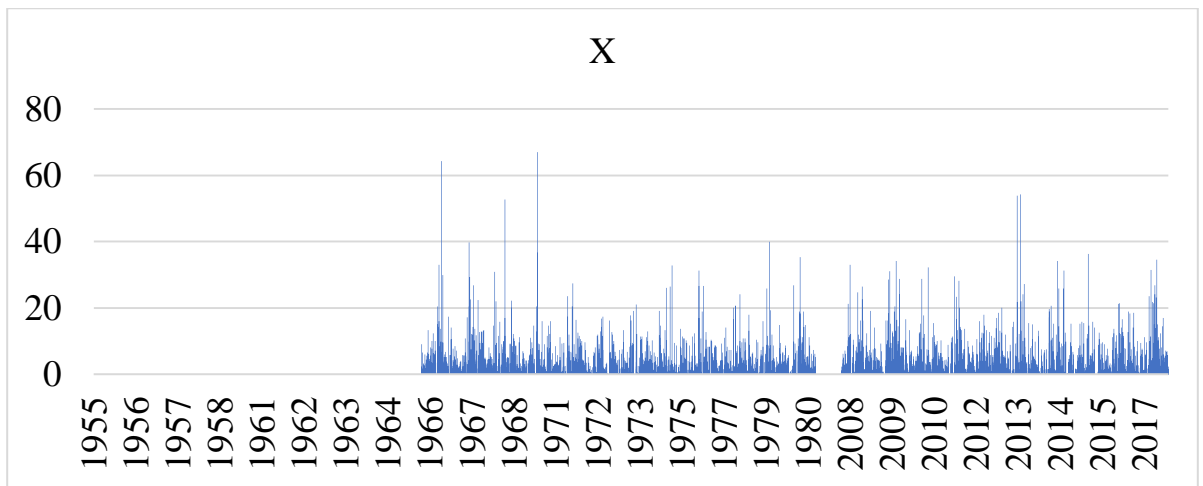


Рисунок 2.7 – Осадки по годам метеостанция Пушкинские горы

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

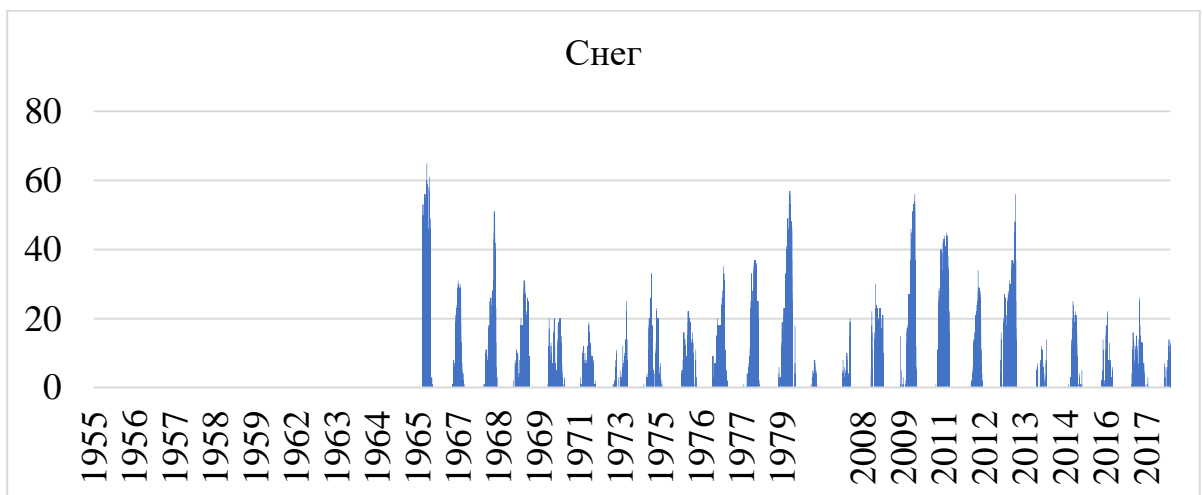


Рисунок 2.8 – Снегозапасы метеостанция Пушкинские горы

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 коррелируют с 2018,2019,2020 годами. Сформировать общую таблицу, проанализировать её.

Таблица 2.3 – Подбор года аналога

год	Ловать				год	Сороть			
	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 Сороть (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.

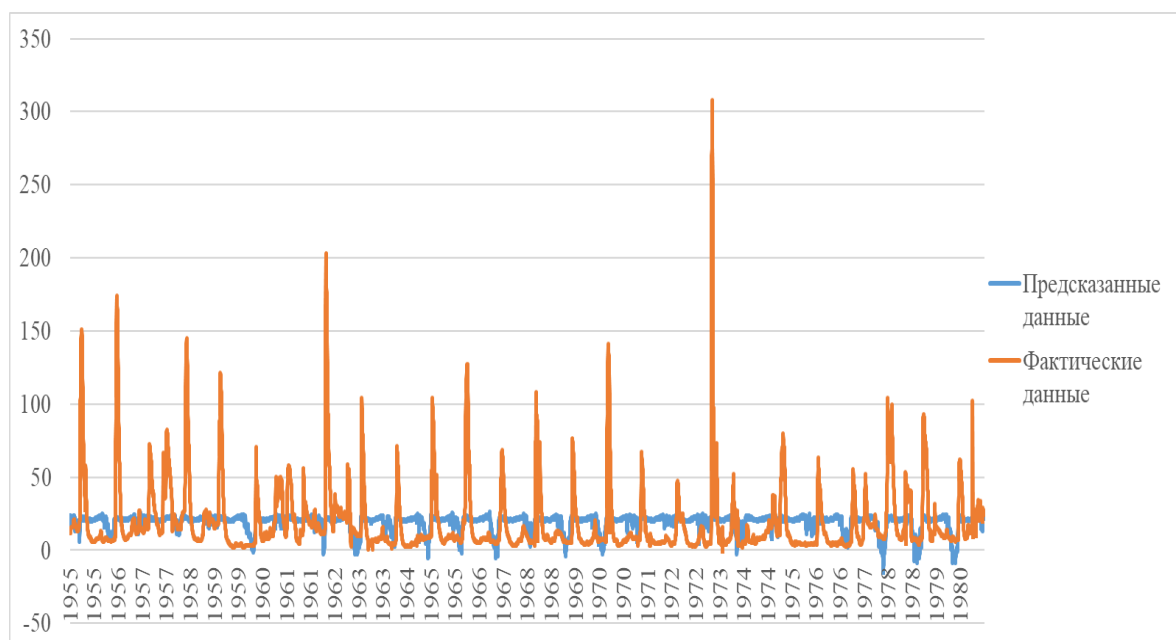


Рисунок 3.1 – Предсказанные и фактические данные с 1955 по 1980 г.

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.

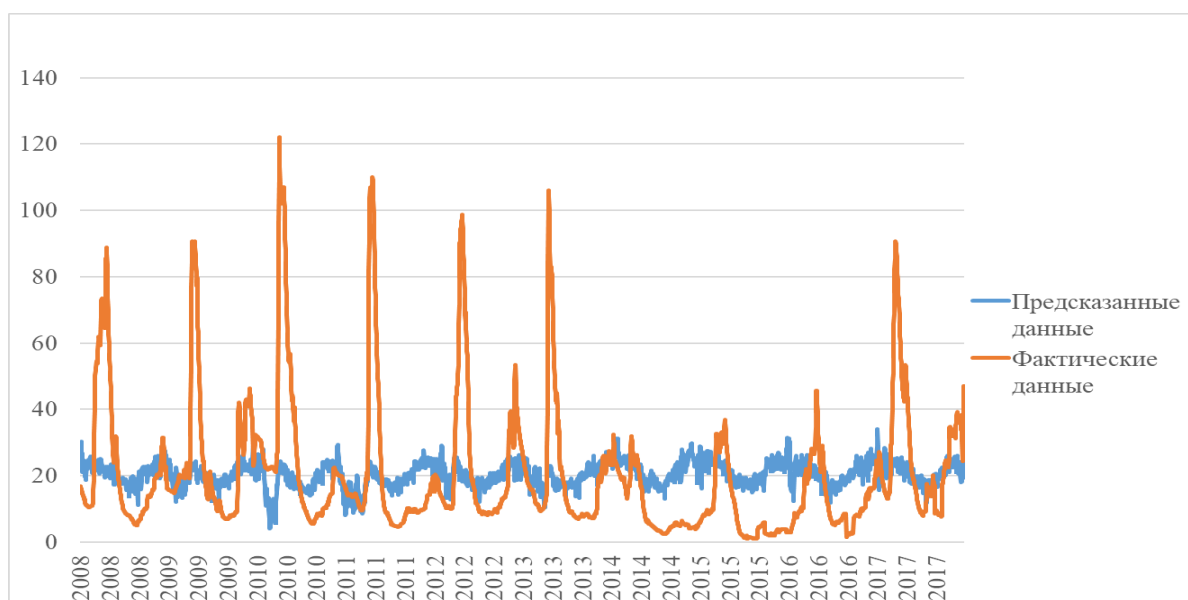


Рисунок 3.2 – Предсказанные и фактические данные с 2008 по 2017 г.

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.

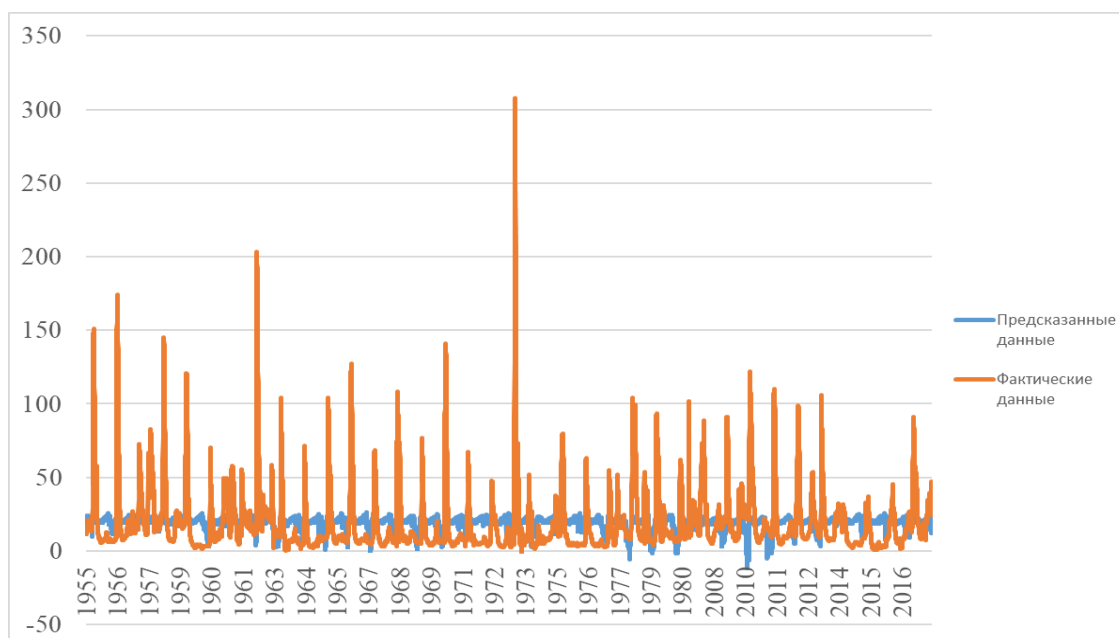


Рисунок 3.2 – Предсказанные и фактические данные с 1955 по 1980 и с 2008 по 2017 г.

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 стандартная ошибка статистик</i>								
	<i>Значени</i>	<i>Нижние 95%</i>		<i>верхние 95%</i>		<i>нижние 95.0</i>		<i>верхние 95.0</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

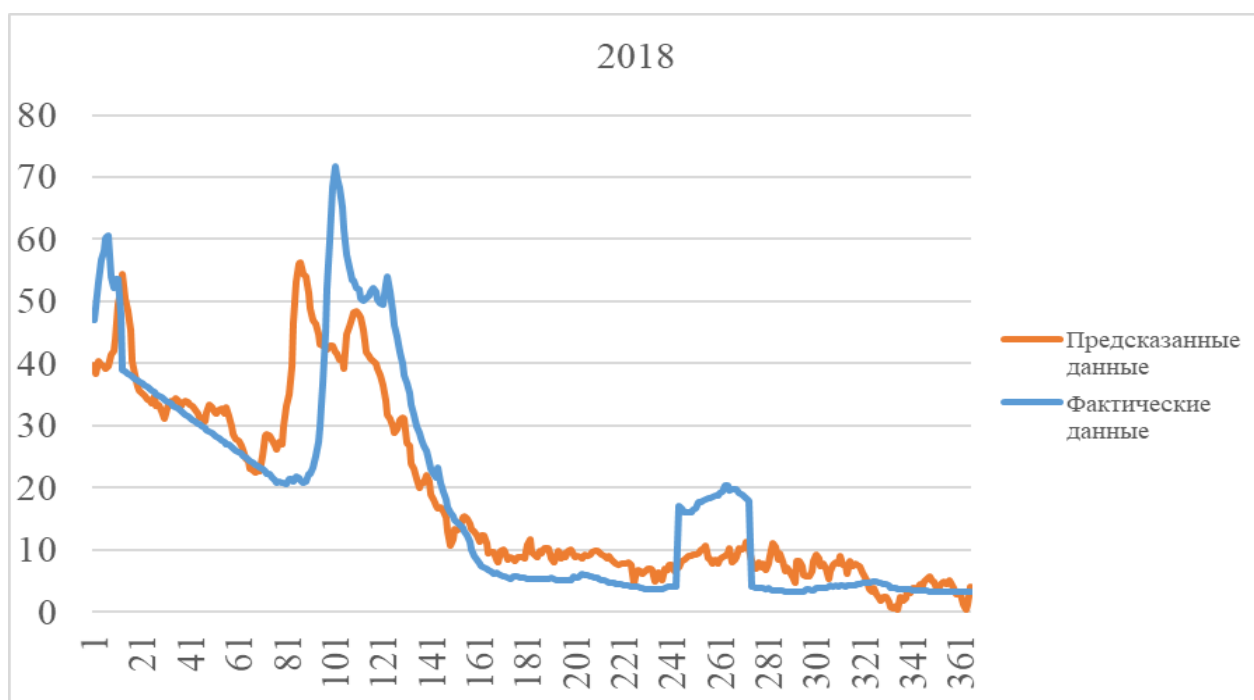


Рисунок 3.3 – Сопоставление графиков фактических значений и модельных, р. Ловать, 2018 год.

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	

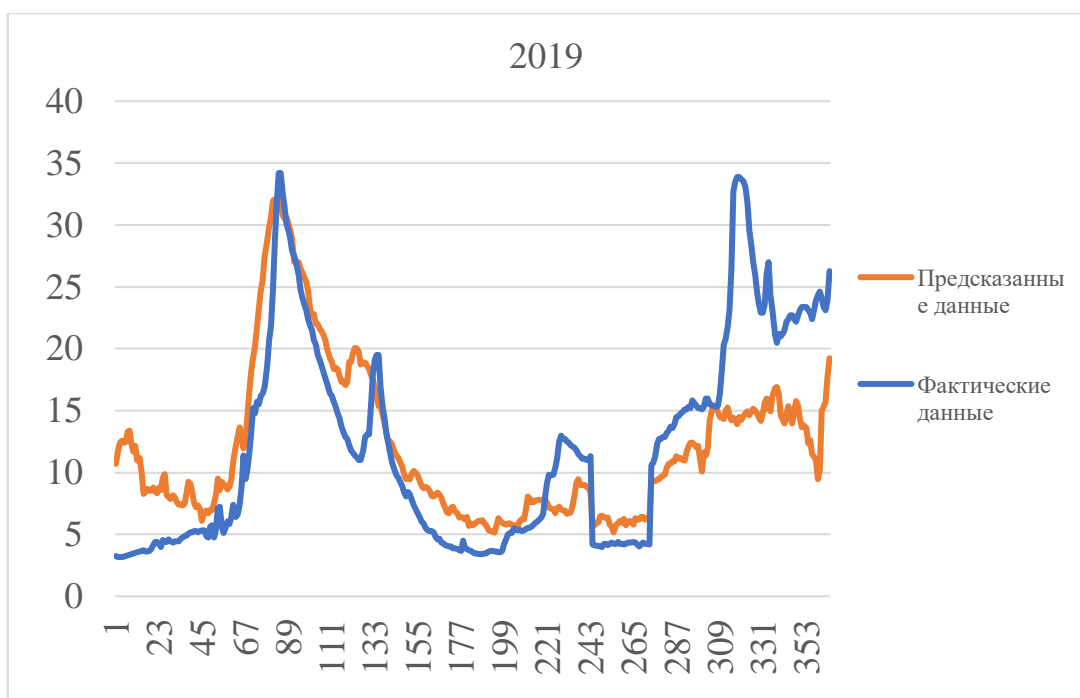


Рисунок 3.4 – Сопоставление графиков фактических значений и модельных, р. Ловать, 2019 год.

Таблица 3.3 – Регрессионный анализ р. Ловать, 2020 год

ВЫВОД ИТОГОВ								
2020								
<i>Регрессионная статистика</i>								
Множестве	0.8509851							
R-квадрат	0.7241757							
Нормирова	0.721111							
Стандартна	6.355916							
Наблюдени	365							
<i>Дисперсионный анализ</i>								
	<i>df</i>	<i>SS</i>	<i>MS</i>	<i>F</i>	<i>Значимость F</i>			
Регрессия	4	38183.009	9545.7524	236.29	2.7058E-99			
Остаток	360	14543.16	40.397668					
Итого	364	52726.17						
<i>Коэффициент стандартная ошибка статистика Значение Нижние 95% Верхние 95% Нижние 95.0 Верхние 95.0</i>								
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

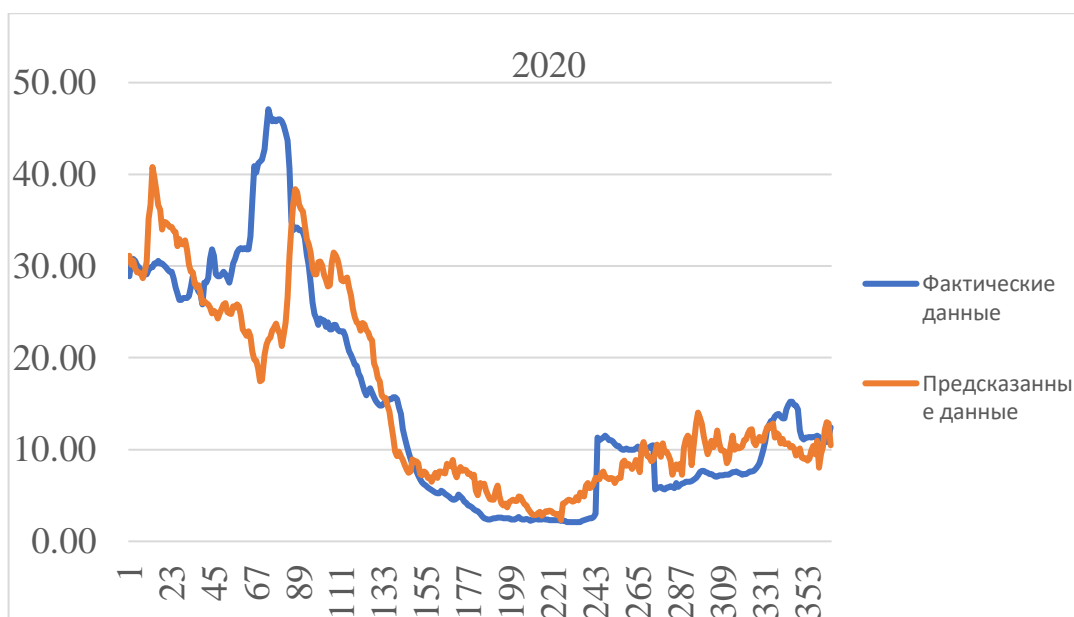


Рисунок 3.4 – Сопоставление графиков фактических значений и модельных, р. Ловать, 2020 год.

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. 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	

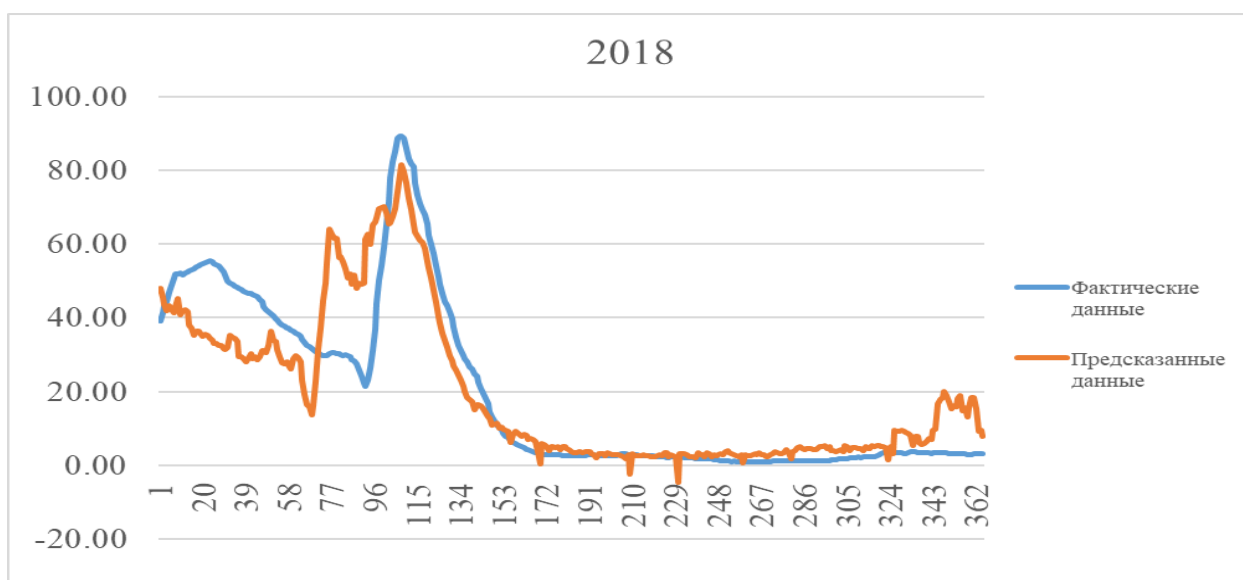


Рисунок 3.5 – Сопоставление графиков фактических значений и модельных, р. Сорочь, 2018 год.

Таблица 3.4 – Регрессионный анализ р. Сорочь, 2019 год

ВЫВОД ИТОГОВ								
2019								
<i>Регрессионная статистика</i>								
Множестве	0.6529318							
R-квадрат	0.4263199							
Нормирова	0.4199457							
Стандартна	17.610685							
Наблюдени	365							
<i>Дисперсионный анализ</i>								
	<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						
<i>Коэффициент</i>								
<i>Статистик</i>	<i>Значени</i>	<i>Нижние 95%</i>	<i>Верхние 95%</i>	<i>Нижние 95.0</i>	<i>Верхние 95.0</i>			
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

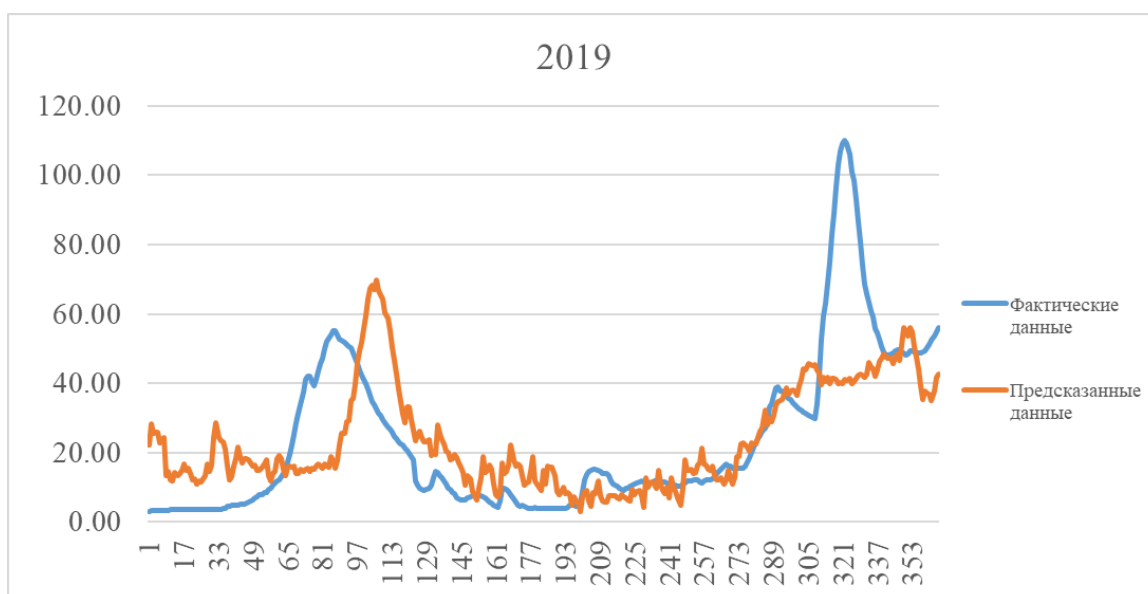


Рисунок 3.6 – Сопоставление графиков фактических значений и модельных, р. Сорочь, 2019 год.

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</i>									
<i>Y-пересече</i>	<i>Стандартная ошибка</i>	<i>t-статистика</i>	<i>P-Значение</i>	<i>Нижние 95%</i>	<i>Верхние 95%</i>	<i>Нижние 95.0%</i>	<i>Верхние 95.0%</i>		
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	

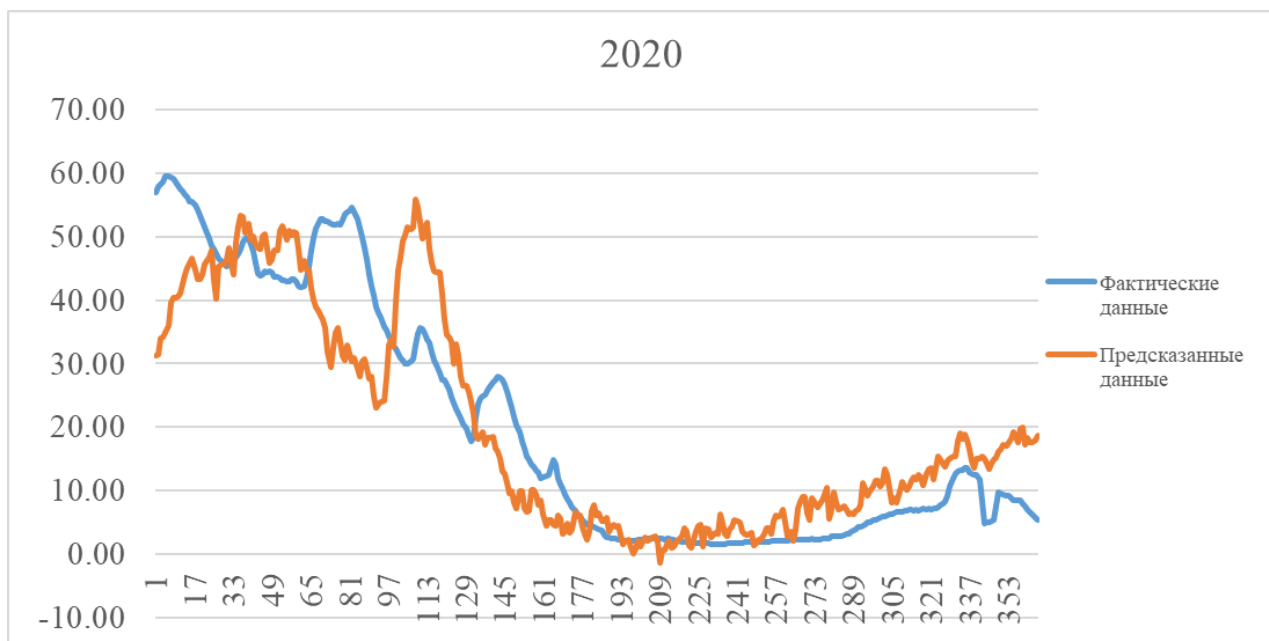


Рисунок 3.7 – Сопоставление графиков фактических значений и модельных, р. Сорочь, 2020 год.

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</i>	<i>Нижние 95%</i>	<i>Верхние 95%</i>	<i>нижние 95.0%</i>	<i>верхние 95.0%</i>	
У-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	

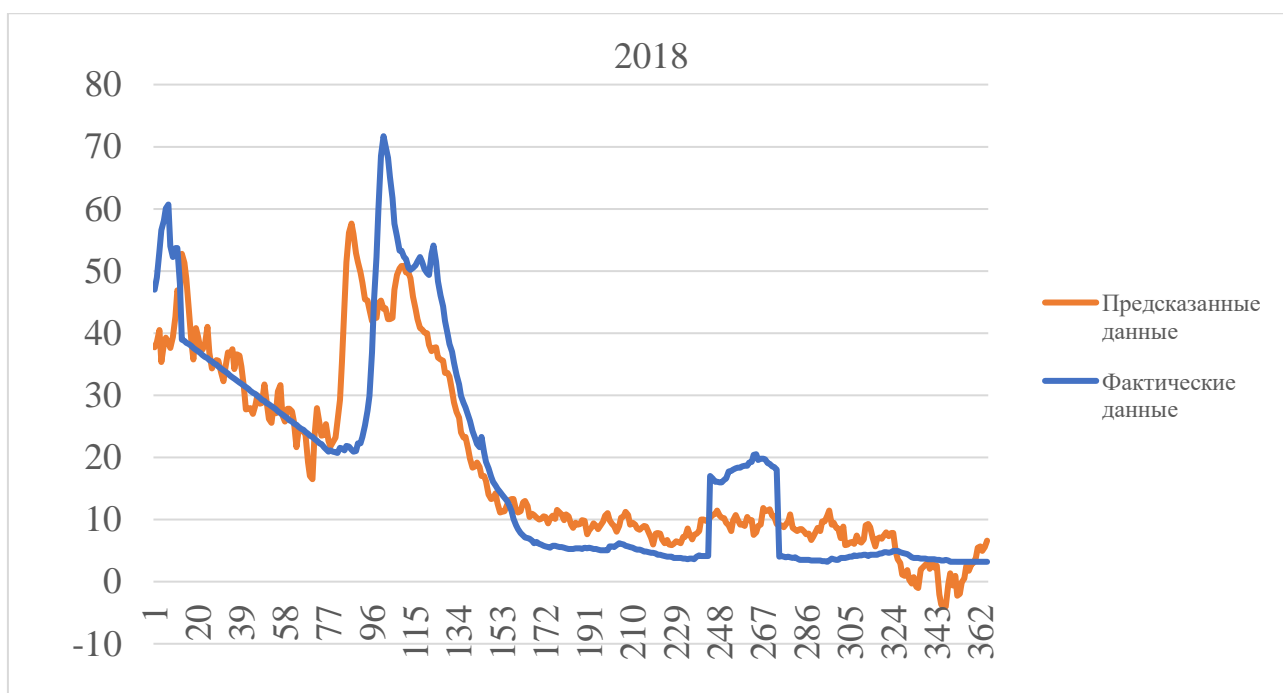


Рисунок 3.8 – Сопоставление графиков фактических значений и модельных, р. Ловать, 2018 год.

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 стандартная ошибка t-Значение p-Значение нижние 95% верхние 95% нижние 95% верхние 95%								
Y-пересеч	9.035469	0.738917	12.22799	5.39E-29	7.582333	10.48861	7.582333	10.48861
Переменная 1	0.253062	0.015485	16.34279	2.65E-45	0.22261	0.283513	0.22261	0.283513
Переменная 2	-0.09475	0.060591	-1.56382	0.118737	-0.21391	0.024403	-0.21391	0.024403
Переменная 3	-0.41554	0.038721	-10.7318	1.69E-23	-0.49169	-0.3394	-0.49169	-0.3394
Переменная 4	-0.26168	0.031653	-8.26708	2.68E-15	-0.32393	-0.19943	-0.32393	-0.19943

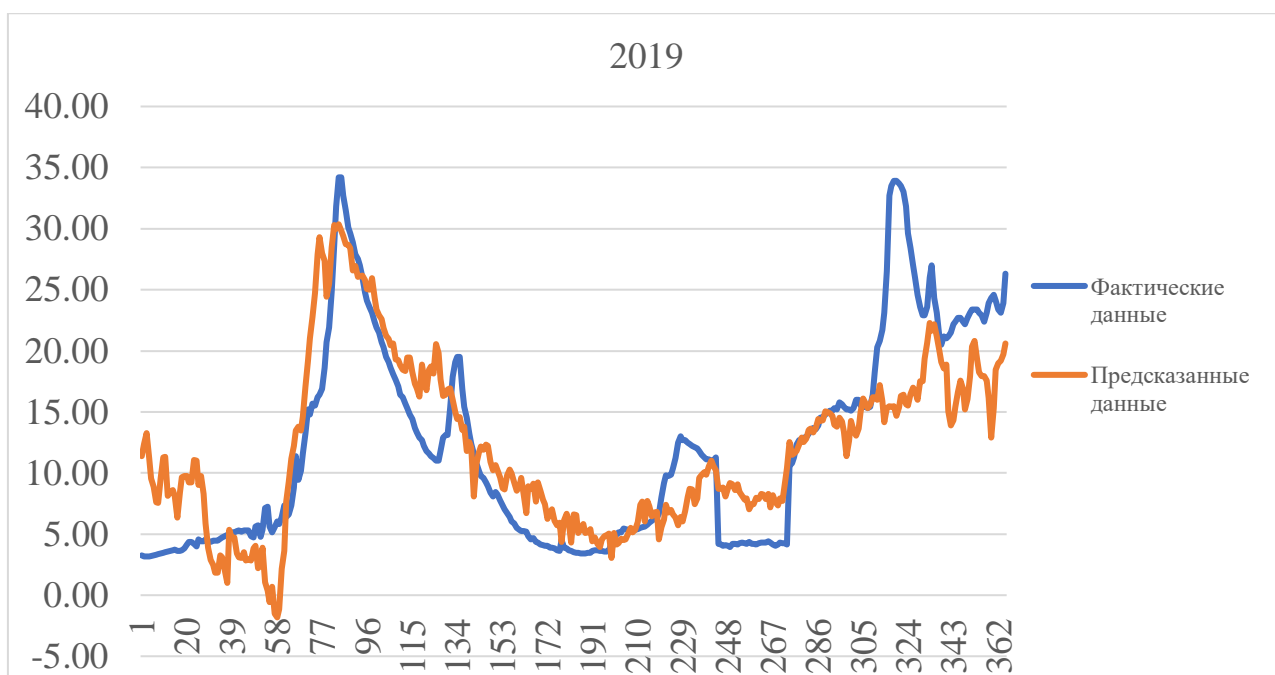


Рисунок 3.9 – Сопоставление графиков фактических значений и модельных, р. Ловать, 2019 год.

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

ВЫВОД ИТОГОВ							
2020							
Регрессионная статистика							
Множественный коэффициент корреляции R	0.866063						
Коэффициент детерминации R-квадрат	0.750066						
Нормированный коэффициент детерминации	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					
Кoeffициент							
	Статистика t	Значимость	Верхние 95%	Нижние 95%	Верхние 95.0%	Нижние 95.0%	
Y-пересеч	7.036113	0.783726	8.977773	1.56E-17	5.494857	8.577369	8.577369
Переменная 1	0.985124	0.042586	23.13267	3.39E-73	0.901376	1.068872	1.068872
Переменная 2	0.114091	0.116692	0.977709	0.328875	-0.11539	0.343575	-0.11539
Переменная 3	-0.35533	0.154235	-2.30383	0.021801	-0.65865	-0.05202	-0.65865
Переменная 4	-0.40401	0.047215	-8.55678	3.4E-16	-0.49686	-0.31116	-0.49686

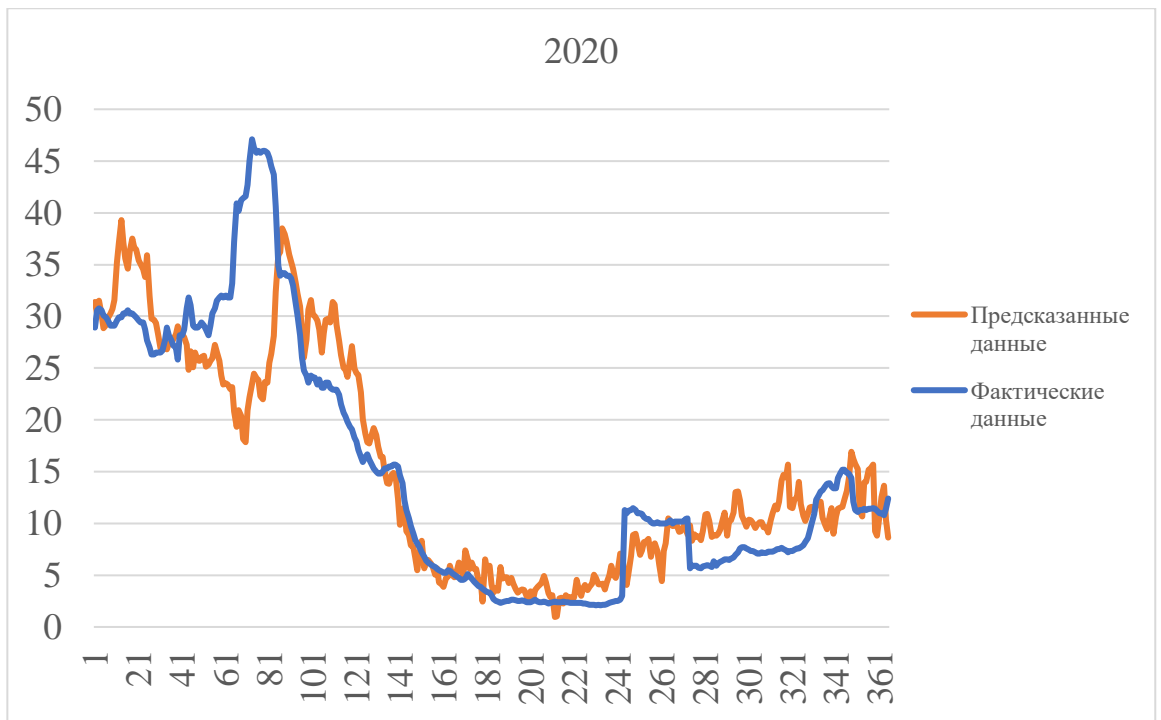


Рисунок 3.10 – Сопоставление графиков фактических значений и модельных, р. Ловать, 2020 год.

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	0.914765							
R-квадрат	0.836795							
Нормированный коэффициент	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 стандартная ошибка t-Значение нижние 95% верхние 95%								
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

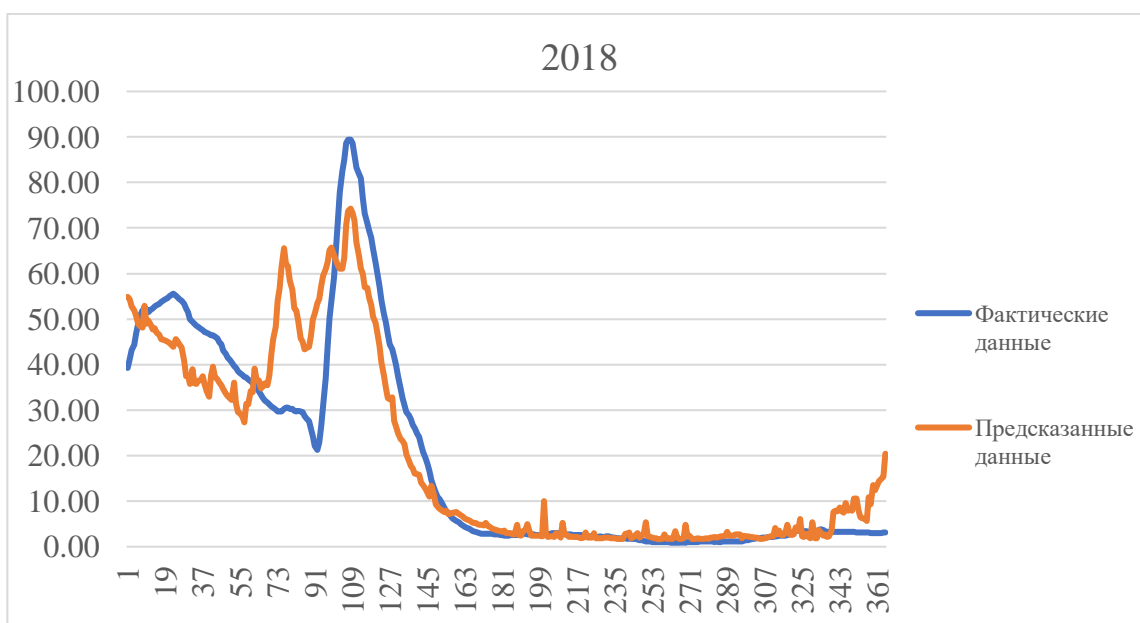


Рисунок 3.11 – Сопоставление графиков фактических значений и модельных, р. Сороть, 2018 год.

Таблица 3.10 – Регрессионный анализ, р. Сорочь, 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						
Коэффициент стандартной ошибки								
	Значение	Верхние 95%	Нижние 95%	Верхние 95.0%	Нижние 95.0%	Верхние 95.0%	Нижние 95.0%	Верхние 95.0%
У-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

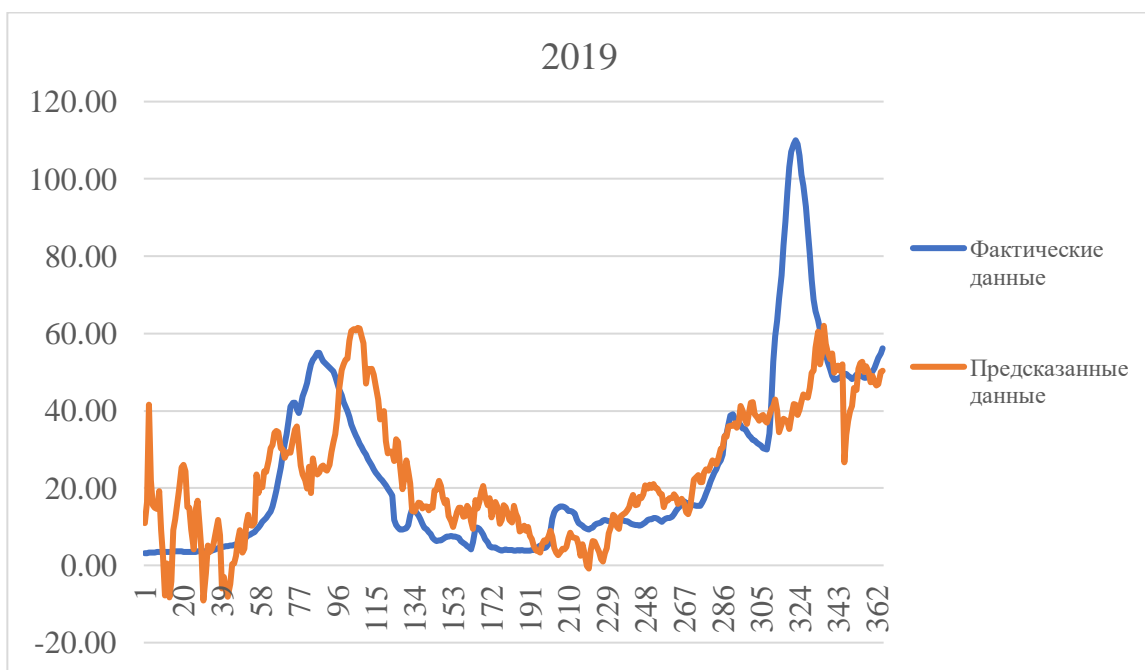


Рисунок 3.12 – Сопоставление графиков фактических значений и модельных, р. Сорочь, 2019 год.

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

ВЫВОД ИТОГОВ								
2020								
Регрессионная статистика								
Множественный коэффициент	0.842166							
R-квадрат	0.709244							
Нормированный коэффициент	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						
Коэффициент стандартная ошибка значимость нижние 95% верхние 95% нижние 95.0% верхние 95.0%								
У-пересеч	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

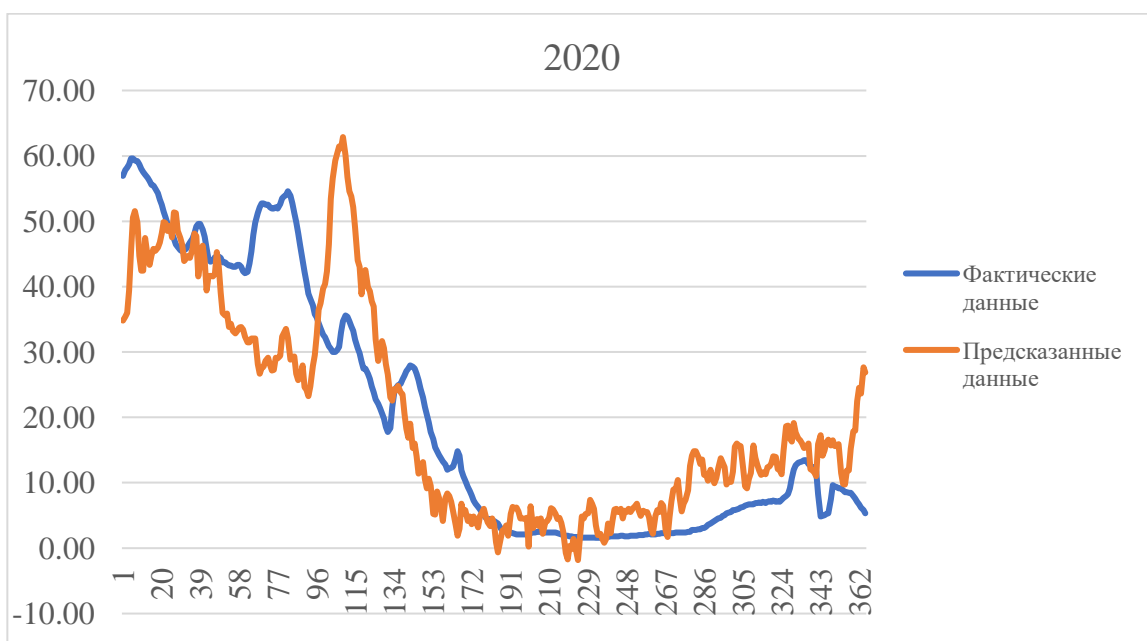


Рисунок 3.13 – Сопоставление графиков фактических значений и модельных, р. Сорочь, 2020 год.

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. Сороть

1 методика по гидропосту Сороть (Сорать) - р.Сороть - 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.

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.

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 Свод Правил 33–101–2003 «Определение основных расчетных гидрологических характеристик» – М: Госстрой России, ФГУП ЦПП, 2004.

7 Методические рекомендации по определению расчетных гидрологических характеристик при недостаточности данных гидрометрических наблюдений. // СПб.: — ГГИ. 2004 г. стр. 13-18.