



МИНИСТЕРСТВО ОБРАЗОВАНИЯ И НАУКИ РОССИЙСКОЙ ФЕДЕРАЦИИ

федеральное государственное бюджетное образовательное учреждение
высшего образования

«РОССИЙСКИЙ ГОСУДАРСТВЕННЫЙ
ГИДРОМЕТЕОРОЛОГИЧЕСКИЙ УНИВЕРСИТЕТ»

Кафедра морские информационные системы

ВЫПУСКНАЯ КВАЛИФИКАЦИОННАЯ РАБОТА

На тему Методика непараметрического анализа статистической связи

геоэкологических явлений

Исполнитель Зарукина Мария Арсеньевна

(фамилия, имя, отчество)

Руководитель доктор технических наук, доцент

(ученая степень, ученое звание)

Завгородний Владимир Николаевич

(фамилия, имя, отчество)

«Б защите допускаю»

Заведующий кафедрой

(подпись)

д.т.н., доцент

(ученая степень, ученое звание)

Завгородний Владимир Николаевич

(фамилия, имя, отчество)

« » 2018 г.

Санкт-Петербург 2018



МИНИСТЕРСТВО ОБРАЗОВАНИЯ И НАУКИ РОССИЙСКОЙ ФЕДЕРАЦИИ

федеральное государственное бюджетное образовательное учреждение
высшего образования

«РОССИЙСКИЙ ГОСУДАРСТВЕННЫЙ
ГИДРОМЕТЕОРОЛОГИЧЕСКИЙ УНИВЕРСИТЕТ»

Кафедра морские информационные системы

ВЫПУСКНАЯ КВАЛИФИКАЦИОННАЯ РАБОТА

На тему Методика непараметрического анализа статистической
связи

геоэкологических явлений

Исполнитель Зарукина Мария
Арсеньевна

(фамилия, имя, отчество)

Руководитель доктор технических наук,
доцент

(ученая степень, ученое звание)

Завгородний Владимир Николаевич

(фамилия, имя, отчество)

«К защите допускаю»

Заведующий кафедрой

(подпись)

(ученая степень, ученое звание)

(фамилия, имя, отчество)

ОГЛАВЛЕНИЕ

ВВЕДЕНИЕ.....	3
ГЛАВА 1. АНАЛИЗ ВОЗМОЖНОСТИ ПРИМЕНЕНИЯ МЕТОДОВ НЕПАРАМЕТРИЧЕСКОЙ И ПАРАМЕТРИЧЕСКОЙ СТАТИСТИКИ ДЛЯ ИССЛЕДОВАНИЯ ОДНОРОДНОСТИ ЭКОЛОГИЧЕСКИХ ЯВЛЕНИЙ НА ПРИМЕРЕ РЕКИ ОХТА.....	4
1.1 Характеристика системы экологического мониторинга реки Охта	4
1.2 Сравнительный анализ методов непараметрической и параметрической статистики.....	7
Выводы по главе:.....	16
ГЛАВА 2. ХАРАКТЕРИСТИКА ИНСТРУМЕНТАЛЬНЫХ СРЕДСТВ СТАТИСТИЧЕСКОГО АНАЛИЗА ПРОГРАММНОГО ПРОДУКТА МАТНСАД.....	18
2.1 Особенности пакета MathCad.....	18
2.2 Ввод выборочных значений многомерных временных рядов.....	21
2.3 Показатели, характеризующие тенденцию динамики изменения признаков	25
2.4 Анализ однородности и динамики изменений признаков.....	29
Выводы по главе:.....	35
ГЛАВА 3. МЕТОДИКА НЕПАРАМЕТРИЧЕСКОГО АНАЛИЗА СТАТИСТИЧЕСКОЙ ОДНОРОДНОСТИ ГЕОЭКОЛОГИЧЕСКИХ ЯВЛЕНИЙ НА ПРИМЕРЕ ЭКОЛОГИЧЕСКИХ ПРОБ РЕКИ ОХТА С ИСПОЛЬЗОВАНИЕМ СРЕДСТВ МАТНСАД.....	36
3.1 Проверка однородности данных параметрическими методами.....	36
3.2 Проверка статистической однородности и связи между признаками непараметрическими методами	39
3.3 Основные положения методики непараметрического анализа статистической связи геоэкологических явлений	45
Выводы по главе:.....	50
ЗАКЛЮЧЕНИЕ	51
СПИСОК ИСПОЛЬЗУЕМЫХ ИСТОЧНИКОВ:.....	54

ВВЕДЕНИЕ

Актуальность темы. В настоящее время проблема загрязнения водных объектов становится все более актуальной и представляет большой интерес и для специалистов в области охраны окружающей среды. Река Охта – одна из самых загрязненных рек Санкт-Петербурга. В данной работе рассмотрена методика анализа данных имеющихся об экологическом состоянии реки Охта. В створах г. Санкт-Петербург в 2014 г. вода характеризовалась как грязная.

В данной работе проведен анализ данных, полученных студентами экологического факультета на летней практике. Анализ проведен параметрическими и непараметрическими методами, проверена однородность данных.

Целью настоящей выпускной квалификационной работы является разработка методики непараметрического анализа для оценки статистической связи геоэкологических явлений и ее применение к экологическим данным о состоянии реки Охта.

Задачи:

- 1) Анализ параметрических и непараметрических методов исследования;
- 2) Выбор и обоснование статистических критериев для определения однородности и связи временных рядов, полученных при экологических наблюдениях реки Охта
- 3) Выполнение расчетов по выбранным критериям с использованием пакетов Excel и MathCad.

Объект данной работы: гидрохимический режим реки Охта.

Предмет: статистический анализ однородности и связи гидрохимических показателей геоэкологических явлений реки Охта.

Научная новизна данной работы данной работы состоит в использовании критериев непараметрического анализа применительно к экологическим данным реки Охта.

ГЛАВА 1. АНАЛИЗ ВОЗМОЖНОСТИ ПРИМЕНЕНИЯ МЕТОДОВ НЕПАРАМЕТРИЧЕСКОЙ И ПАРАМЕТРИЧЕСКОЙ СТАТИСТИКИ ДЛЯ ИССЛЕДОВАНИЯ ОДНОРОДНОСТИ ЭКОЛОГИЧЕСКИХ ЯВЛЕНИЙ НА ПРИМЕРЕ РЕКИ ОХТА

1.1 Характеристика системы экологического мониторинга реки Охта

Основная цель экологического мониторинга заключается в создании информационной системы, позволяющей получать достоверные сведения о состоянии окружающей среды и ее изменениях в биотических и абиотических компонентах под действием естественных и антропогенных факторов [1].

Системы экологического мониторинга призваны решить следующие задачи:

Сбор первичной информации, ее накопление, систематизация, анализ и формирование банков данных;

Обработка и представление данных в виде различных таблиц, графиков, карт и др.;

Усовершенствование и разработка методов получения исходной информации, оценка текущего состояния окружающей среды;

Анализ причин, наблюдаемых и вероятных изменения состояния;

Оперативное обеспечение необходимой информацией всех заинтересованных лиц. При этом оценка текущего состояния среды является основной для принятия тактических (оперативных) решений в области природопользования, а прогноз – для принятия решений стратегических (долговременных) [1].

Целью мониторинга вод реки Охта состоит в получении объективной информации о состоянии водных объектов суши, направленной на разработку методик по улучшению экологической обстановки. Объект исследования очень обширный.

Бассейн реки Охты относится к системе водотоков и водоемов Балтийского бассейнового округа. Исток реки расположен в районе Лемболовских высот во Всеволожском районе Ленинградской области. Река течет с севера на юг и впадает в Неву в 12,5 км от ее устья [2].

Река Охта – один из самых крупных правых приток реки Невы в черте города Санкт-Петербурга. Она берет начало в районе Лемболовских высот на карельском перешейке. Длина реки 90 км, причем 9 км река течет в городе, водосборная площадь 768 км². Сток реки на территории города зарегулирован плотиной Охтинского водохранилища, поэтому режим уровня и расхода воды в ней зависит от работы гидротехнических сооружений. Кроме того, на уровень воды р.Охты в нижнем течении влияет изменение уровня воды в р.Неве [3].

На гидрохимический режим оказывают влияние загрязняющие вещества сточных вод и площадных стоков. На ряде участков реки имеются набережные, реку пересекает более десяти автомобильных, железнодорожных и пешеходных мостов [4].

Охта является судоходной рекой на протяжении 8 км от устья [5]. На территории бассейна Охты осуществляется интенсивное использование водных ресурсов, поскольку данный регион имеет высокоразвитую хозяйственно-экономическую деятельность [6]. На загрязнение реки влияют техногенные, антропогенные и климатические факторы.

С 1997 года Российский Государственный Гидрометеорологический университет ежегодно проводит летние наблюдения за состоянием реки Охта. Наблюдения проводятся на 13 станциях по течению реки, а также у устья притоков реки Лубья и Оккервиль.

В качестве исходных данных использовались данные с 2006 по 2015 гг. полученные со станций, находящихся на р.Охта. Достоверность оценок и результатов обеспечивается использованием в качестве информационной базы материалов государственной системы наблюдений за состоянием поверхностных вод, применением стандартных методов математической обработки данных наблюдений [7].

Статистические методы – это методы анализа данных, причем обычно достаточно большого количества данных. Статистические данные могут иметь различную природу. Исторически самыми ранними были два вида данных – сведения о числе объектов, удовлетворяющих тем или иным условиям, и числовые результаты измерений [8].

При гидрологических обобщениях используются различные методы статистического анализа. Выбор этих методов и способов их использования обычно производится на основе предварительного (разведочного) анализа исходных данных, который, в свою очередь, основывается на простейших формах описания и представления ряда случайной величины.

Гидрологические наблюдения ограничены во времени. Поэтому их ряды представляют собой выборки, а рассчитанные по ним числовые характеристики являются выборочными [9].

Одной из важных задач гидрологического анализа является оценка однородности исходных данных.

Под однородностью принято понимать принадлежность рассматриваемых рядов всех частей одного ряда к одной генеральной совокупности, подчинение одному закону распределения.

В математической статистике разработаны приемы анализа однородности на основе гипотез того или иного вида, которые должны быть доказаны или опровергнуты в ходе проверки однородности.

В гидрологических исследованиях проверка однородности исходного ряда или исходных рядов производится чаще всего по следующим

параметрам: математическому ожиданию или дисперсии (параметрические гипотезы) и по закону распределения (непараметрические гипотезы) [9].

Для оценки однородности рядов наблюдений применялся параметрический t-критерий Стьюдента и непараметрические критерии Смирнова, Спирмена и Кендалла. Общий вывод для ряда наблюдения делался на основании сопоставления результатов по всем критериям. Однородностью рядов принято считать подчинение всех признаков одному общему закону распределения.

Естественное состояние рек без влияния человеческой деятельности характеризуется чистотой вод. Формирование химического состава таких вод определяется грунтовыми водами, а его изменение за счет грунтовых вод может варьироваться в узких пределах в течение долгого времени. При этом любые отклонения химического состава воды за определенные пределы могут негативно сказываться на состоянии речной биоты. Именно такую картину можно наблюдать на экологическом состоянии реки Охты, где воды загрязнены вредными веществами, и структура водных экосистем нарушена [10].

1.2 Сравнительный анализ методов непараметрической и параметрической статистики

Приступая к статистической обработке какого-либо массива данных, необходимо определить наиболее подходящие методы по особенностям выборок – параметрические или непараметрические. Для этого, сравним эти методы.

Если критерий основан на том или другом конкретном типе распределения генеральной совокупности или использует параметры этих совокупностей, то такой критерий называется параметрическим. Например, параметрическим критерием является широко используемый на практике критерий, основанный на t-распределении Стьюдента, поскольку этот критерий исходит из предположения о нормальном законе распределения

случайных величин в генеральных совокупностях. Если критерий не опирается на предположение о конкретном типе распределения генеральных совокупностей и не использует параметры этих совокупностей, то он называется непараметрическим критерием (иначе, его называют «критерий, свободный от распределения» - также используемое в литературе название).

Возможности практического применения параметрических и непараметрических критериев существенно зависят от характера тех измерений, которые может осуществить исследователь. Параметрические критерии требуют для своего применения количественных измерений, т.е. измерений по шкале интервалов или шкале отношений. В то же время, большинство непараметрических критериев может применяться и в тех случаях, когда результаты выборок измерены с помощью шкал наименования и порядка [11].

При параметрическом анализе находят математическое ожидание и дисперсию случайного процесса.

Заключение о значимости различий между выборочными совокупностями при использовании параметрических критериев осуществляется на основании сравнения параметров распределений, т.е. сводных числовых характеристик. В задачах проверки параметрических гипотез используются параметрические критерии, в расчете которых используются параметры распределений, например, средние значения, дисперсии и т.д. Параметрические критерии позволяют принять оценки основных параметров генеральных совокупностей, допустимые погрешности в средних значениях и дисперсиях. Критерии способны выявить тенденции изменения признака при различных условиях, оценить воздействие двух и более совместно действующих факторов на изменение признака [12].

Параметрические методы обладают для исследователя многими преимуществами, но нельзя забывать о том, что применение их правомерно

только тогда, когда обрабатываемые данные показывают распределение, лишь несущественно отличающееся от гауссовского.

Наиболее распространенными среди параметрических критериев в статистическом анализе экологических явлений являются критерии Стьюдента и Фишера. Критерий Стьюдента служит для оценки гипотез об однородности исходной информации по математическому ожиданию [9].

Критерий, основанный на сравнении средних значений выборок – критерий Стьюдента, является наиболее известным. С одной стороны, анализ средних значений сравнительно прост для вычислений. С другой стороны, средние величины наиболее понятны. Существенным ограничением применения критерия Стьюдента является требование нормального распределения признака и в той, и в другой выборке, когда данные двух серий измерений, соответствующих той и другой выборке, коррелируют.

При невозможности применить параметрические надлежит обратиться к непараметрическим методам. Эти методы успешно разрабатывались в последние 3-4 десятилетия, и их разработка была вызвана прежде всего потребностями ряда наук. Они показали свою высокую эффективность [13].

Несмотря на широкое использование параметрических методов при решении геоэкологических задач, иногда нет возможности для их применения - например, в случае, когда данные не подчиняются закону нормальному распределения. В экологических исследованиях часто неизвестны и не могут быть установлены с помощью качественного анализа типы вероятностных законов распределения случайных процессов. Тогда можно обратиться к непараметрическим методам.

Для непараметрических методов математический аппарат, более разработан. Непараметрические методы не нуждаются в предварительных предположениях касательно вида исходного распределения. Они допускают использование данных экологических экспериментов, полученных в результате не только количественных, но и качественных измерений с

помощью дискретных шкал наименований и порядка. Достоинством непараметрических методов является простота расчётов. Однако, мощность статистических критериев, построенных на их основе, уступает решающим аналогичные задачи параметрическим критериям (например, критериям Стьюдента, Фишера) [12].

При наличии ограниченного количества показателей, именно непараметрические критерии служат единственным средством для проверки гипотез. Кроме того, важным достоинством непараметрических критериев является относительная простота вычислительных процедур, связанных с практическим применением этих критериев [11].

Непараметрические критерии проверки гипотез основаны на операциях с частотами, рангами и т.п. Оценка гипотез непараметрическим методом грубее, нежели параметрическим, однако не требует от выборки знания закона и параметров распределения.

Часто данные, получаемые в экологических экспериментах, имеют распределение, отличное от нормального, поэтому применение некоторых параметрических критериев при анализе результатов исследований может привести к ошибкам в статистических выводах. В таком случае непараметрические критерии становятся более подходящими.

Непараметрические методы наиболее приемлемы, когда объем выборок мал [12]. Если данных много (например, при объёме выборки $n > 100$), то нет необходимости использовать непараметрические критерии. Дело в том, что когда выборки становятся очень большими, то выборочные средние подчиняются нормальному закону в силу центральной предельной теоремы, даже если исходная переменная не является нормальной или измерена с погрешностью.

Для анализа временных рядов наблюдений за изменением экологической ситуации могут найти применение такие непараметрические методы как:

- выборочный коэффициент ранговой корреляции Спирмена, Кендалла, используемые для оценки степени связи признаков;
- критерий Вилкоксона, служащий для проверки двух независимых выборок. Достоинством этого критерия является то, что он применим к случайным величинам, распределение которых неизвестно, требуется лишь, чтобы величины были непрерывными[15];
- критерий Смирнова используемый для проверки гипотезы о принадлежности одному закону распределения всех независимых выборок.

Критерий Колмогорова-Смирнова

Критерий предназначен для выявления различия двух совокупностей по состоянию некоторого свойства. Критерий чувствителен в улавливании любого различия функций распределения этого свойства в рассматриваемых совокупностях (средних значений, дисперсий, эксцессов и др.) [11].

Непараметрические критерии Колмогорова-Смирнова являются основой непараметрической математической статистики. Они продолжают быть образцом при разработке новых методов и широко применяются на практике.

Критерий предназначен для сопоставления двух распределений:

- а) эмпирического с теоретическим, например, равномерным или нормальным;
- б) одного эмпирического распределения с другим эмпирическим распределением.

Критерий позволяет найти точку, в которой сумма накопленных расхождений между двумя распределениями является наибольшей, и оценить достоверность этого расхождения.

С помощью критерия Смирнова можно сравнивать эмпирическое распределение с теоретическим или два эмпирических распределения друг с другом. В данном критерии сравниваются накопленные (кумулятивные) частоты по каждому разряду (альтернативе). При этом если разность

накопленных частот в двух распределениях оказывается большой, то различия между двумя распределениями являются существенными.

Для применения критерия Колмогорова–Смирнова необходимо соблюдать следующие условия:

1. Измерение может быть проведено шкале интервалов и отношений.
2. Выборки должны быть случайными и независимыми.
3. Желательно, чтобы суммарный объем двух выборок ≥ 50 . С увеличением объема выборки точность критерия повышается.
4. Эмпирические данные должны допускать возможность упорядочения по возрастанию или убыванию какого-либо признака и обязательно отражать какое-то его однонаправленное изменение. В том случае, если трудно соблюсти принцип упорядоченности признака, лучше использовать критерий χ^2 -квадрат.

Коэффициент корреляции рангов Спирмена

Коэффициент корреляции рангов, предложенный К. Спирменом, относится к непараметрическим показателям связи между переменными, измеренными в ранговой шкале. При расчете этого коэффициента не требуется никаких предположений о характере распределений признаков в генеральной совокупности. Этот коэффициент определяет степень тесноты связи порядковых признаков, которые в этом случае представляют собой ранги сравниваемых величин.

Величина коэффициента корреляции Спирмена лежит в интервале +1 и -1. Он, может быть положительным отрицательным, характеризуя направленность связи между двумя признаками, измеренными в ранговой шкале.

Для применения коэффициента корреляции Спирмена, необходимо соблюдать следующие условия:

- Сравнимые переменные должны быть получены в порядковой (ранговой) шкале, но могут быть измерены также в шкале интервалов и отношений.

- Характер распределения коррелируемых величин не имеет значения.

- Число варьирующих признаков в сравниваемых переменных X и Y должно быть одинаковым.

Если обе переменные, между которыми изучается связь, представлены в порядковой шкале, или одна из них - в порядковой, а другая - в метрической, то применяются ранговые коэффициенты корреляции: Спирмена или τ -Кендалла. И тот, и другой коэффициент требует для своего применения предварительного ранжирования обеих переменных.

Коэффициент ранговой корреляции Спирмена - это непараметрический метод, который используется с целью статистического изучения связи между явлениями. В этом случае определяется фактическая степень параллелизма между двумя количественными рядами изучаемых признаков и дается оценка тесноты установленной связи с помощью количественно выраженного коэффициента.

Мощность коэффициента ранговой корреляции Спирмена несколько уступает мощности параметрического коэффициента корреляции. Коэффициент ранговой корреляции целесообразно применять при наличии небольшого количества наблюдений. Данный метод может быть использован не только для количественно выраженных данных, но также и в случаях, когда регистрируемые значения определяются описательными признаками различной интенсивности.

Коэффициент ранговой корреляции Спирмена при большом количестве одинаковых рангов по одной или обоим сопоставляемым переменным дает огрубленные значения. В идеале оба коррелируемых ряда должны представлять собой две последовательности несовпадающих значений.

Коэффициент Кендалла

Коэффициент корреляции Кендалла относится к числу непараметрических, т.е. при вычислении этого коэффициента не играет роли характер распределения сравниваемых переменных. Иногда, этот коэффициент можно использовать вместо коэффициента корреляции Спирмена, поскольку способ его вычисления более прост и основан на вычислении суммы инверсий и совпадений. Он предназначен для определения взаимосвязи между двумя ранговыми переменными.

Для одних и тех же значений переменных значения коэффициента корреляции Спирмена будет всегда немного больше, чем значения коэффициента ранговой корреляции Кендалла, тогда как уровень значимости будет одинаков или же у коэффициента корреляции Кендалла будет немного больше.

Коэффициент корреляции Кендалла используется в случае, когда переменные представлены двумя порядковыми шкалами при условии, что связанные ранги отсутствуют.

Можно выделить 2 ограничения в использовании коэффициента Кендалла:

- невозможность рассчитать согласованность мнений экспертов по каждой переменной в отдельности.
- коэффициент измеряет согласованность мнений в смысле их коррелированности, но не совпадения.

t-критерий Стьюдента

t-критерий Стьюдента используется для определения статистической значимости различий средних величин. Может применяться как в случаях сравнения независимых выборок, так и при сравнении связанных совокупностей. Критерий для определения статистической значимости различий двух средних значений [14. С. 372].

Параметрический критерий, позволяющий проверить гипотезу о том, что среднее значение изучаемого признака отличается от некоторого

известного значения [14. 164]. 2. Критерий, предназначенных для сравнения среднего значения распределения переменной с некоторой эталонной величиной [14 с.373].

Критерий t Стьюдента для независимых выборок (t ; критерий Стьюдента). Параметрический критерий, позволяющий проверить гипотезу о том, что средние значения двух совокупностей, из которых извлечены сравниваемые независимые выборки, отличаются друг от друга [14 С. 165].

Несмотря на то, что открытие Стьюдента в свое время совершило переворот в статистике, t -критерий все же довольно сильно ограничен в возможностях применения, т.к. сам по себе происходит из предположения о нормальном распределении исходных данных. Если данные не являются нормальными (что обычно и бывает), то и t -критерий уже не будет иметь распределения Стьюдента. Однако в силу действия центральной предельной теоремы средняя даже у ненормальных данных быстро приобретает колоколообразную форму распределения. В случае применения двухвыборочного критерия для независимых выборок также необходимо соблюдение условия равенства (гомоскедастичности) дисперсий [16].

Распределение Стьюдента при малой выборке дает не вполне определенные результаты (широкий доверительный интервал), однако, это вовсе не говорит о слабости метода Стьюдента, а объясняется тем, что малая выборка содержит малую информацию об интересующем нас признаке [15].

Альтернативным критерию Стьюдента, является непараметрический критерий Вилкоксона, если распределение, хотя бы для одной выборки существенно отличается от нормального.

Критерий Вилкоксона, служит для проверки двух независимых выборок. Достоинством этого критерия является то, что он применим к случайным величинам, распределение которых неизвестно, требуется лишь, чтобы величины были непрерывными [15].

Критерий может применяться к результатам, полученным на основе однократного измерения некоторого свойства в признаках проб воды из реки Охта.

Условия применения T–критерия Вилкоксона:

Измерение может быть проведено во всех шкалах, кроме шкалы номинальной.

Выборка должна быть связной.

Число элементов в сравниваемых выборках должно быть равным.

Нулевые сдвиги из рассмотрения исключаются, и количество наблюдений n уменьшается на количество этих нулевых сдвигов. Можно обойти это ограничение, сформулировав гипотезы, включающие отсутствие изменений.

Выводы по главе:

Изучение сложных закономерностей экологических процессов в водной среде требует проведения объективного количественного анализа. Безусловно, выбор методов анализа данных зависит от условий задачи [25]. Указанные выше особенности экологических исследований не всегда позволяют применять методы параметрической статистики. Непараметрические методы исследования актуальны при анализе концентрации вредных веществ в воде и ее изменении со временем, при анализе изменения количества донных отложений или при выявлении связи между сокращением численности животных и количеством бытовых отходов в среде их обитания. В проблеме проверки статистических гипотез, возникающих в экологических исследованиях, непараметрические методы анализа данных о загрязнении морей могут иметь преимущества перед параметрическими в силу их разнообразия, простоты расчётов и возможности применения для небольших объемов выборок.

Река Охта является обширным объектом для исследований поэтому для статистического анализа имеющихся данных о пробах воды, необходима

методика. Необходимо установить, являются ли данные полученные студентами РГГМУ однородными, проверить статистическую значимость признаков и установить связь между различными признаками. С этой целью будут применены параметрические t-критерий Стьюдента и непараметрические критерии Вилкоксона, Спирмена и Кендалла.

ГЛАВА 2. ХАРАКТЕРИСТИКА ИНСТРУМЕНТАЛЬНЫХ СРЕДСТВ СТАТИСТИЧЕСКОГО АНАЛИЗА ПРОГРАММНОГО ПРОДУКТА MATHCAD

2.1 Особенности пакета MathCad

В последнее время для проведения различного рода расчетов на компьютере, все большее применение находят не традиционные языки программирования, а специальные математические пакеты, такие как например MathCad. MathCad – один из самых популярных, на сегодняшний день, пакетов, позволяющий специалистам в конкретной предметной области, не погружаясь в тонкости программирования, реализовывать математические модели, производить статистические, математические и вероятностные расчеты.

Несмотря на то, что эта программа, в основном, ориентирована на пользователей, не являющихся программистами, MathCad также используется в сложных проектах, чтобы визуализировать результаты математического моделирования путём использования распределённых вычислений и традиционных языков программирования. Также MathCad часто используется в крупных инженерных проектах, где большое значение имеет соответствие стандартам.

Преимущества пакета MathCad:

В среде MathCad математические выражения записываются в общепринятом виде. Текстовый процессор пакета позволяет оформлять тексты, не прибегая к текстовым редакторам (таким как Microsoft Office Word). Кроме того, MathCad – полноценное Windows-приложение, поэтому буфер обмена позволяет перенести фрагменты из документа MathCad в Word-документ и при необходимости продолжить работу уже в текстовом редакторе; Процесс создания программы в среде MathCad происходит одновременно с отладкой;

В пакет MathCad внедрен мощный математический аппарат, который позволяет производить решение математических задач без вызова внешних процедур. Перечень некоторых вычислительных инструментов, доступных в среде MathCad:

- Решение алгебраических уравнений и систем (линейных и нелинейных);
- Решение систем обыкновенных дифференциальных уравнений (задача Коши и Краева задача);
- Решение дифференциальных уравнений в частных производных;
- Работа с векторами и матрицами (линейная алгебра и др.);
- Поиск максимумов и минимумов функциональных зависимостей;
- Статистическая обработка данных

Также, пакет MathCad содержит в себе справочник по основным математическим и физико-химическим формулам и константам, которые можно автоматически переносить в документ;

Система MathCad содержит средства символьной математики, что позволяет решать математические задачи не только численно, но и аналитически;

Пакет MathCad оборудован средствами анимации, что позволяет реализовывать созданные модели не только в статичном виде (числа, таблицы), но и в динамике (анимационные представления);

Возможность создания средствами Mathcad высококачественных технических отчетов с таблицами, графиками и текстом;

возможность интеграции Mathcad с множеством других программных систем (Excel, MatLab, Lotus, Power Point и др.);

Mathcad чрезвычайно прост в использовании и легок в обучении. Большинство действий, необходимых для управления программой, являются интуитивно понятными, и на освоение основных ее возможностей человеку, работавшему ранее в среде Windows, требуется два - три часа [17].

Из приведенных выше характеристик можно заключить, что пакет MathCad обладает огромными возможностями для решения разнообразных задач [18].

Все эти особенности Mathcad не только облегчают процесс достижения конечного результата, осуществление более глубокого анализа проблем, но и позволяют использовать альтернативные пути решения тех же задач [19].

Mathcad был создан не для профессиональных математиков – для них существуют разнообразные системы, действующие в областях символьной математики, а скорее, как мощный микрокалькулятор для работы расчетчиков-инженеров, позволяющий справляться с рутинными задачами инженерной практики, ежедневно встречающимися в работе:

Решение дифференциальных или алгебраических уравнений с постоянными или переменными параметрами;

Численное и аналитическое дифференцирование и интегрирование;

Анализ функций, поиск их экстремумов;

Вывод таблиц и графиков при анализе найденных решений [20].

Проблема создания специализированного программного обеспечения резко упрощается с привлечением универсальных программных систем компьютерной математики (СКМ), поскольку при этом исключается необходимость разработки программ на языках высокого уровня. Применение СКМ в процессе проведения практических занятий позволяет уделять должное внимание многовариантному решению широкого спектра алгоритмически сложных задач по теории вероятностей и математической статистике [18].

Для статистического анализа существует большое множество программ. Тем не менее, в большей их части алгоритмы подсчета оказываются скрытыми. В некоторых случаях это бывает удобно, так как позволяет не отвлекаться от основной цели, но иной раз требуется и более тонкая настройка. В таких случаях, MathCad очень удобен, так как позволяет

пользователю не только работать с уже известными алгоритмами статистического анализа, но и реализовывать свои собственные.

Возможности инструментальных средств MathCad для статистического анализа весьма многообразно:

Применение Mathcad для оценки постоянных величин и параметров математических моделей переменных величин, зависящих от одного или нескольких аргументов.

Построение математической модели по экспериментальным данным. Решение задачи интерполяции. Использование метода наименьших квадратов для построения аппроксимирующей функции. Проверка адекватности модели.

Проведение дисперсионного и регрессионного анализа. Линейная регрессия — самый простой и наиболее часто используемый вид регрессии. Использование метода наименьших квадратов для вычисления коэффициентов уравнения регрессии.

Полиномиальная регрессия. В Mathcad реализована регрессия одним полиномом, отрезками нескольких полиномов, а также двумерная регрессия массива данных.

Применение статистического комплекса для оценки качества изделий, характеризующихся совокупностью разнородных величин.

Планирование эксперимента. Использование системы Mathcad при планировании эксперимента. Оптимизация планов экспериментов. Полный факторный план и дробный факторный план эксперимента. Минимизации дисперсии определения коэффициентов уравнения регрессии: D — оптимальные планы.

2.2 Ввод выборочных значений многомерных временных рядов

Студентами экологического факультета РГГМУ были собраны данные о реке Охта и представлены в виде таблицы содержащей год проведения забора проб воды, номера станций с их координатами и большое количество

признаков таких как: время, температура воды, БПК, концентрация кислорода, рН воды, щелочность, концентрация хлоридов, кальция, магния, железа, нитритов, фосфатов, жесткость воды, окисление перманганата, и т.д.

Для статистического анализа экологического состояния реки были отобраны следующие признаки: БПК, концентрации кислорода, железа, нитритов и фосфатов, так как в данной задаче они представляют наибольший интерес.

БПК (биохимическая потребность в кислороде) – косвенный показатель содержания нестойких, легко окисляющихся органических веществ в воде водоемов. Главным образом, к таким веществам относятся органические вещества бытовых и сточных вод, масса отмирающего зоо- и фитопланктона. Величина БПК выражается в миллиграммах кислорода из расчета на один литр воды. Процессы биохимического окисления веществ происходят под влиянием различных видов микрофлоры, в том числе обычно присутствующей в водоемах, сточных водах. БПК является одним из важнейших и наиболее часто употребляемых показателей санитарного состояния водоемов и характера сточных вод.

Таблица 2 - БПК в речных водах

Степень загрязнения (классы водоемов)	БПК₅
Очень чистые	0.5 - 1.0
Чистые	1.1 - 1.9
Умеренно загрязненные	2.0 - 2.9
Загрязненные	3.0 - 3.9
Грязные	4.0 - 10.0
Очень грязные	> 10.0

Концентрация кислорода является основным из показателей качества воды. Кислород находится в природной воде в виде молекул O₂. Процессы, происходящие в воде делятся на две группы: те, что повышают

концентрацию кислорода и те, что наоборот, уменьшают ее. К первой группе процессов, обогащающих воду кислородом, следует отнести: процесс абсорбции кислорода из атмосферы; выделение кислорода водной растительностью в процессе фотосинтеза; поступление в водоемы с дождевыми и снеговыми водами, которые обычно пересыщены кислородом. К группе процессов, уменьшающих содержание кислорода в воде, относятся реакции потребления его на окисление органических веществ: биологическое, биохимическое и химическое.

В поверхностных водах содержание растворенного кислорода варьирует в широких пределах - от 0 до 14 мг/л - и подвержено сезонным и суточным колебаниям. Суточные колебания зависят от интенсивности процессов его продуцирования и потребления и могут достигать 2,5 мг/л растворенного кислорода. В зимний и летний периоды распределение кислорода носит характер стратификации. Дефицит кислорода чаще наблюдается в водных объектах с высокими концентрациями загрязняющих органических веществ и в вырождающихся (перерождающихся) водоемах, содержащих большое количество биогенных и гумусовых веществ.

Концентрация железа. Главными источниками соединений железа в поверхностных водах являются процессы химического выветривания горных пород, сопровождающиеся их механическим разрушением и растворением. В процессе взаимодействия с содержащимися в природных водах минеральными и органическими веществами образуется сложный комплекс соединений железа, находящихся в воде в растворенном, коллоидном и взвешенном состоянии. Значительные количества железа поступают с подземным стоком и со сточными водами предприятий металлургической, металлообрабатывающей, текстильной, лакокрасочной промышленности и с сельскохозяйственными стоками [21] ПДК железа в речных водах – 0,1 мг/дм³.

Концентрация нитритов. Нитриты представляют собой промежуточную ступень в цепи бактериальных процессов окисления аммония до нитратов (нитрификация - только в аэробных условиях) и, напротив, восстановления нитратов до азота и аммиака (денитрофикация - при недостатке кислорода). Подобные окислительно-восстановительные реакции характерны для станций аэрации, систем водоснабжения и собственно природных вод. Кроме того, нитриты используются в качестве ингибиторов коррозии в процессах водоподготовки технологической воды и поэтому могут попасть и в системы хозяйственно-питьевого водоснабжения. Широко известно также применение нитритов для консервирования пищевых продуктов. В поверхностных водах нитриты находятся в растворенном виде. В кислых водах могут присутствовать небольшие концентрации азотистой кислоты (HNO_2) (не диссоциированной на ионы). Повышенное содержание нитритов указывает на усиление процессов разложения органических веществ в условиях более медленного окисления NO_2^- в NO_3^- , что указывает на загрязнение водного объекта, т.е. является важным санитарным показателем. Концентрация нитритов в поверхностных водах составляет сотые (иногда даже тысячные) доли миллиграмма в 1 дм³ [22].

Концентрация фосфатов. Соединения минерального фосфора поступают в природные воды в результате выветривания и растворения пород, содержащих ортофосфаты (апатиты и фосфориты) и поступления с поверхности водосбора в виде орто-, ме-та-, пиро- и полифосфат-ионов (удобрения, синтетические моющие средства, добавки, предупреждающие образование накипи в котлах, и т. п.), а также образуются при биологической переработке остатков животных и растительных организмов. Избыточное содержание фосфатов в воде, особенно в грунтовой, может быть отражением присутствия в водном объекте примесей удобрений, компонентов хозяйственно-бытовых сточных вод, разлагающейся биомассы.

Концентрация фосфатов в природных водах обычно очень мала - сотые, редко десятые доли миллиграммов фосфора в 1 дм³, в загрязненных водах она может достигать нескольких миллиграммов в 1 дм. [21]

Рассматриваемые признаки были измерены в период с 2006 по 2015 годы на пятнадцати станциях. В данной работе рассмотрены первая и тринадцатая станции, так как первая станция находится в устье реки Охты, тринадцатая у охтинского водохранилища.

В пакет прикладных программ MathCad также были введены данные для последующего их анализа. Были отдельно введены данные по всем выбранным признакам в виде векторов.

2.3 Показатели, характеризующие тенденцию динамики изменения признаков

Для представления общей картины о состоянии реки Охты, необходимо выявить тенденции динамики в концентрации тех или иных веществ.

Одно из основных положений научной методологии — необходимость изучать все явления в развитии, во времени. Это относится и к статистике: она должна дать характеристику изменений статистических показателей во времени.

Таблица 1 - Содержание кислорода (мл/л) в пробах реки Охта 2006 - 2015 гг.

2006	2007	2008	2009	2010	2011	2012	2013	2014	2015
1,13	3,00	0,54	1,36	2,21	0,88	4,03	2,35	4,36	5,17

Ряд динамики состоит из двух строк или столбцов: промежутков или моментов времени, к которым относятся уровни, и самих уровней признака (показателя). Ряд, в котором время задано в виде промежутков — лет, месяцев, суток, называется интервальным динамическим рядом. Ряд, в котором время задано в виде конкретных дат (моментов времени), называется моментным динамическим рядом.

Сравнивая уровни разных лет в таблице 1, мы замечаем, что в целом концентрация кислорода возрастает. Однако нередко концентрация

кислорода следующего года оказывается ниже уровня предыдущего. Иногда рост по сравнению с предыдущим годом велик, как в 2012 г., а иногда мал. Следовательно, рост уровня концентрации кислорода наблюдается лишь в среднем, как тенденция. В отдельные же годы уровни испытывают колебания, отклоняясь от основной тенденции.

Тенденция динамики изменения показателей связана с действием долговременно существующих причин и условий развития, хотя, после какого-то периода эти причины и условия тоже могут измениться и породить уже другую тенденцию развития изучаемого объекта. Колебания же, напротив, связаны с действием краткосрочных, или циклических, факторов, влияющих на отдельные уровни динамического ряда и отклоняющих уровни от тенденции то в одном, то в другом направлении.



Рисунок 1 - Динамика изменения концентрации кислорода в реке Охта

Тенденцию и колебания наглядно показывает график (рисунок 1). По оси абсцисс отложены годы, в которые производились заборы воды, по оси ординат – уровни концентрации кислорода в воде.

На рис.1 хорошо заметно, что рост концентрации кислорода 2006 - 2015 гг., характеризовался линейной тенденцией, а колеблемость была хаотической, без явной цикличности.

Колеблемость – отклонение значений от уровней тренда в одну и другую сторону. Типы колебаний статистических показателей весьма разнообразны, но все же можно выделить три основных: пилообразную, или маятниковую, колеблемость, циклическую долгопериодическую и случайно распределенную по времени колеблемость.

Пилообразная или маятниковая колеблемость состоит в попеременных отклонениях уровней от тренда в одну и в другую сторону. Таковы автоколебания маятника. Такие автоколебания можно наблюдать в динамике урожайности при невысоком уровне агротехники: высокий урожай при благоприятных условиях погоды выносит из почвы больше питательных веществ, чем их образуется естественным путем за год; почва обедняется, что вызывает снижение следующего урожая ниже тренда, он выносит меньше питательных веществ, чем образуется за год, плодородие возрастает и т.д.

Циклическая долгопериодическая колеблемость свойственна, например, солнечной активности (10-11-летние циклы), а значит, и связанным с ней на Земле процессам - полярным сияниям, грозовой деятельностью, урожайности отдельных культур в ряде районов, некоторым заболеваниям людей, растений. Для этого типа характерны редкая смена знаков отклонений от тренда и кумулятивный (накапливающийся) эффект отклонений одного знака, который может тяжело отражаться на экономике. Зато колебания хорошо прогнозируются.

Случайно распределенная во времени колеблемость - нерегулярная, хаотическая. Она может возникать при наложении (интерференции) множества колебаний с разными по длительности циклами. Но может возникать в результате столь же хаотической колеблемости главной причины существования колебаний.

При аналогичном анализе других признаков (таких как концентрация нитритов, фосфатов, железа в воде), можно увидеть аналогичную картину – в большинстве случаев тенденция возрастает. Это значит, что загрязнение Охты продолжается и нужно предпринимать действия, направленные на устранение и профилактику дальнейшего загрязнения вод.

Также, благодаря линиям тренда проведенным на графиках, можно судить о том, что на поверхности концентрация неблагоприятных веществ выше чем на дне. (Пример – рисунок 2)

Абсолютное изменение уровней — в данном случае его можно назвать абсолютным приростом — это разность между сравниваемым уровнем и уровнем более раннего периода, принятым за базу сравнения.

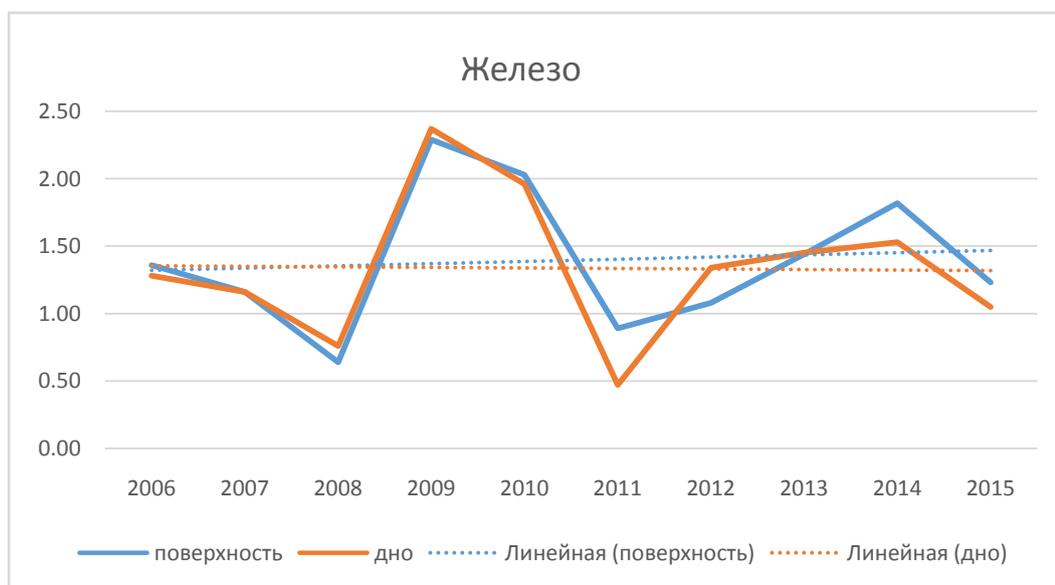


Рисунок 2 - Концентрация железа на дне и поверхности на первой станции

Если при изучении и измерении тенденции динамики колебания уровней играли лишь роль помех, «информационного шума», от которого следовало по возможности абстрагироваться, то в дальнейшем сама колеблемость становится предметом статистического исследования.

На рисунке 1 можно увидеть случайно распределенную по времени колеблемость. Для определения типа колебаний применяется метод Кендалла [23].

2.4 Анализ однородности и динамики изменений признаков

Ниже представлены графики изменения признаков с течением времени за период 2006 – 2015 гг. в устье и водохранилище реки.

Биохимическая потребность в кислороде

Таблица 2 - БПК данные по станциям

	2006	2007	2008	2009	2010	2011	2012	2013	2014	2015
1 пов.	3,46	5,84	5,11	4,72	3,48	5,50	3,12	1,91	3,142	10,93
13 пов.	4,79	5,73	4,88	5,62	6,20	2,06	6,78	5,02	2,44	5,42
1 дно	3,19	1,93	5,44	5,62	5,53	3,98	3,56	2,73	4,77	7,03
13 дно	5,01	5,38	4,14	7,02	4,90	5,77	5,54	5,26	2,76	6,25

На первой станции представленной в верхнем графике на рис.3 линии тренда проходит между значениями в 5мл/л и 6мл/л, что говорит о том, что воду можно считать загрязненной. В 2015 году, уровень БПК на поверхности пересекает отметку в 10мл/л, что говорит о переходе реки из класса грязной в очень загрязненный.

На тринадцатой станции происходит изменение в пределах 4.5мл/л и 5.5мл/л, что также говорит о загрязненности рек, но в 2011 году имеется сильный перепад на поверхности до 2мл/л, что считается пределом умеренной загрязненности. К тому-же, в целом тенденция идет на спад, что является хорошим показателем.

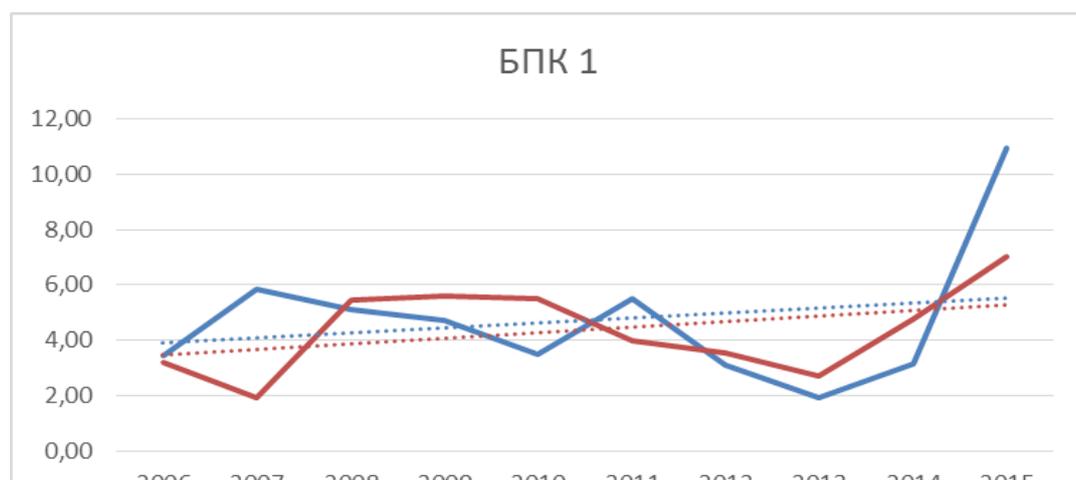




Рисунок 3б - Тенденция изменения БПК на станции №13

Концентрация кислорода

На рисунке 4 четко видно, что на обеих станциях концентрация кислорода возрастает. На поверхности ситуация немногим лучше, чем на дне.

Дефицит кислорода чаще наблюдается в водных объектах с высокими концентрациями загрязняющих органических веществ. На графиках можно увидеть, что концентрация кислорода в реке Охта очень низкая.

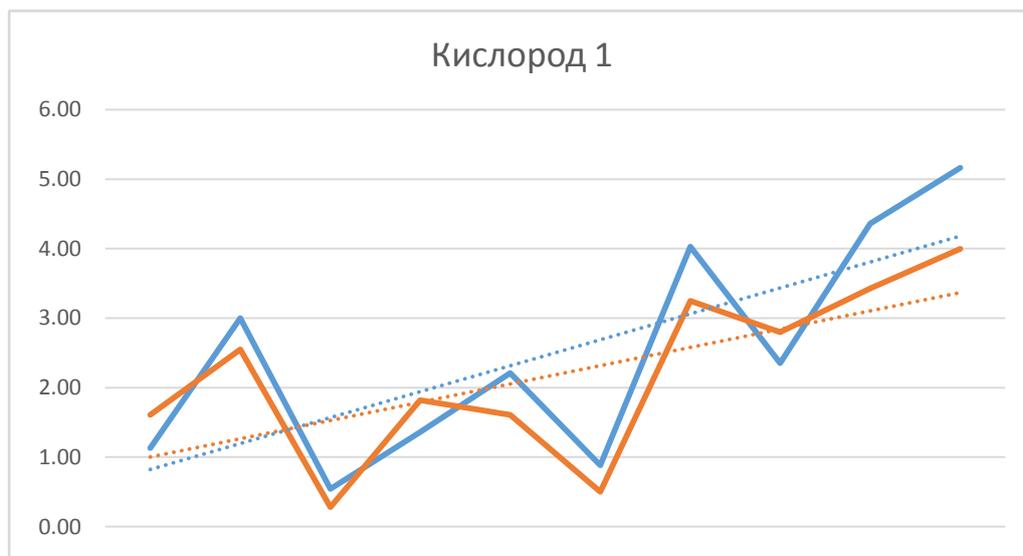


Рисунок 4а - Тенденция изменения концентрации кислорода на станции №1



Рисунок 4б - Тенденция изменения концентрации кислорода на станции №13

Таблица 3. Концентрация кислорода данные по станциям

	2006	2007	2008	2009	2010	2011	2012	2013	2014	2015
1 пов.	1,13	3,00	0,54	1,36	2,21	0,88	4,03	2,35	4,36	5,17
13 пов.	4,89	5,83	3,91	4,88	4,90	4,87	7,66	5,21	4,76	5,83
1 дно	1,61	2,55	0,28	1,82	1,61	0,50	3,25	2,80	3,43	4,00
13	5,28	6,36	3,51	5,01	2,14	4,61	7,4	5,32	5,11	5,50

дно										
-----	--	--	--	--	--	--	--	--	--	--

Концентрация железа

На рис.5 представлены графики изменения концентрации железа в водах реки Охта. Причем, концентрация имеет тенденцию увеличиваться. Предельно-допустимая концентрация железа в речных водах равняется примерно 0,1 мг/л. В нашем случае, концентрация превышает 1,5 мг/л что является достаточно сильным превышением. Сильный скачок наблюдается в 2009 году.

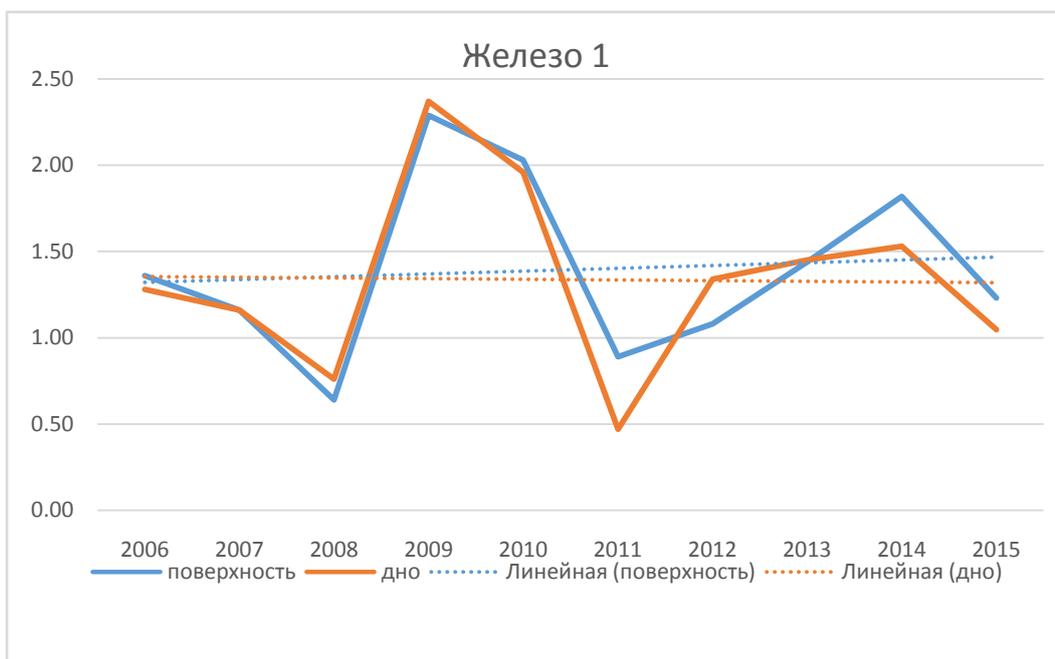


Рисунок 5б - Тенденция изменения концентрации железа на станции №13



Рисунок 5б - Тенденция изменения концентрации железа на станции №13

Таблица 4 - Концентрация железа данные по станциям

	2006	2007	2008	2009	2010	2011	2012	2013	2014	2015
1 пов.	1,36	1,16	0,64	2,29	2,03	0,89	1,08	1,44	1,82	1,23
13 пов.	0,80	2,00	1,32	2,51	2,18	0,96	1,31	1,67	2,34	1,66
1 дно	1,28	1,16	0,76	2,37	1,96	0,47	1,34	1,45	1,53	1,05
13 дно	1,20	1,60	1,44	2,44	2,47	1,18	0,52	1,67	2,53	1,71

Концентрация нитритов

Таблица 5. Концентрация нитритов данные по станциям

	2006	2007	2008	2009	2010	2011	2012	2013	2014	2015
1 пов.	59,1	95,1	4,4	55,10	69,10	198,00	63,3	45,3	59,38	72,08
13 пов.	47,1	78,1	68,1	69,10	105,10	102,00	53,1	67,9	135,1	80,62
1 дно	69,1	79,1	3,1	57,10	33,10	198,00	60,7	40,7	66,58	67,49
13 дно	43,1	99,1	38,1	70,10	103,10	126,00	53,1	58,1	137,1	80,86



Рисунок 6а - Тенденция изменения концентрации нитритов на 1 станции

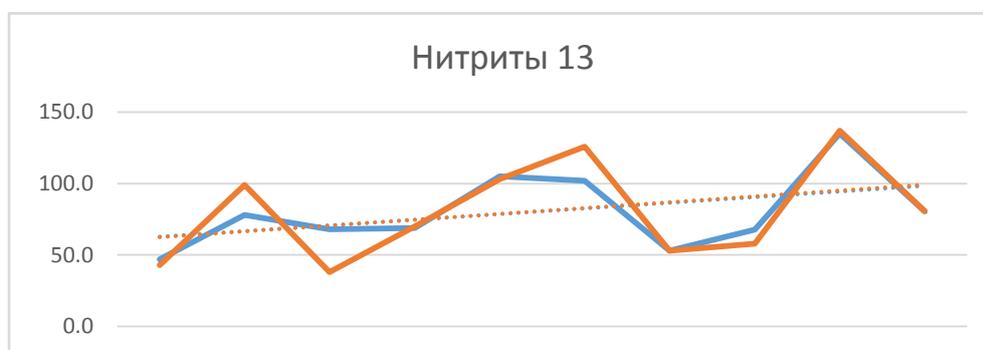


Рисунок 6б - Тенденция изменения концентрации нитритов на станции №1

На рисунке 6 данные представлены в размерности мкг/л. ПДК нитритов в воде составляет 3,3 мг/дм³, то есть 3300 мкг/л. Это говорит о том, что в реке Охте концентрация Нитритов в нормальном состоянии, но смотря на тенденцию, можно сделать вывод, что медленно концентрация возрастает.

Концентрация фосфатов

Как и в данных о концентрации нитритов. Фосфаты представлены с размерностью мкг/л. Но рис.7 четко видно, что концентрация фосфатов в оде сильно уменьшилась с годами

Таблица 6 - концентрация фосфатов данные по станциям

	2006	2007	2008	2009	2010	2011	2012	2013	2014	2015
1 пов.	118,30	128	106,50	212,96	106,20	131,60	93,38	65,76	358,67	146,81
13 пов.	74,62	94,4	152,00	151,46	106,20	13,50	132,53	59,35	29,37	90,38
1 дно	161,98	126	99,40	205,28	95,60	125,96	97,59	57,75	102,8	119,15
13 дно	63,70	84,2	152,00	159,15	113,30	76,95	92,77	54,54	38,37	92,77

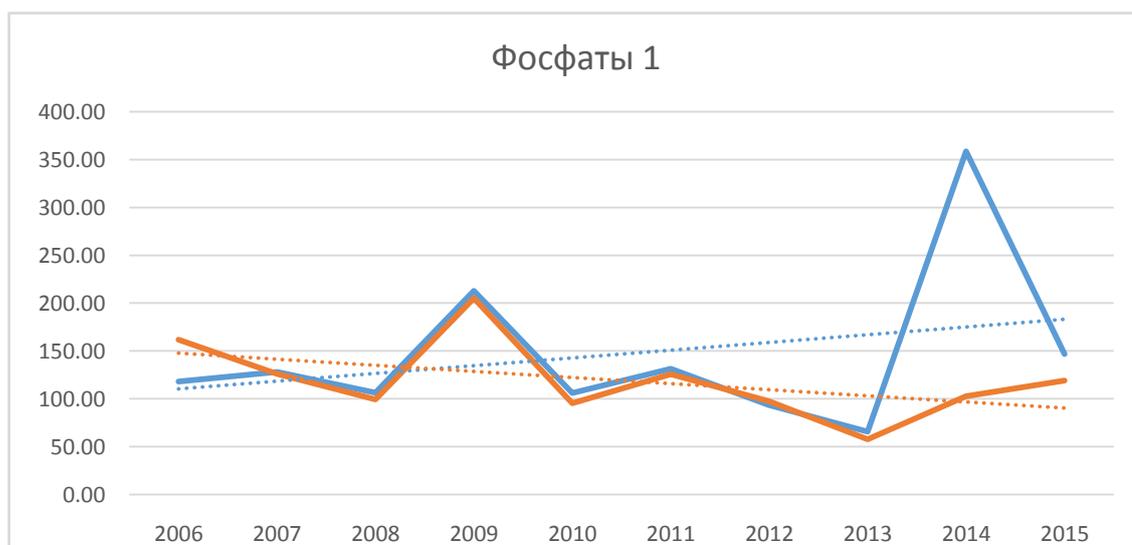


Рисунок 7а – Тенденция изменения концентрации фосфатов на станции №1

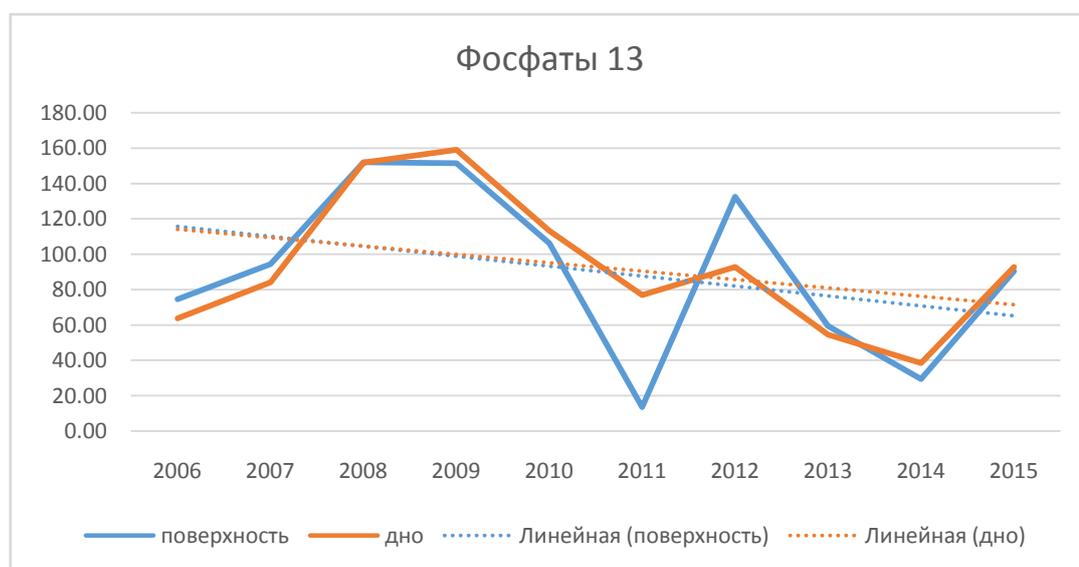


Рисунок 7б - Тенденция изменения концентрации фосфатов

Выводы по главе:

Состояние реки Охты ухудшается, что хорошо видно по тенденциям концентраций загрязняющих веществ в воде, особенно железа. Для более детального анализа состояния реки необходима некая методика. Предлагается выполнить статистический анализ с применением непараметрических критериев в пакете прикладных программ MathCad.

Выбор программного продукта обусловлен рядом его достоинств, таких как легкость и удобность работы, возможность выполнять большое количество расчетов, наличие справочника, а также получение наглядных результатов.

Исследование будет проводиться по пяти признакам: БПК, концентрации кислорода, нитритов, фосфатов и железа в воде, так как в данной задаче, эти признаки играют важную роль в формировании понимания об общем состоянии воды в реке Охта.

ГЛАВА 3. МЕТОДИКА НЕПАРАМЕТРИЧЕСКОГО АНАЛИЗА СТАТИСТИЧЕСКОЙ ОДНОРОДНОСТИ ГЕОЭКОЛОГИЧЕСКИХ ЯВЛЕНИЙ НА ПРИМЕРЕ ЭКОЛОГИЧЕСКИХ ПРОБ РЕКИ ОХТА С ИСПОЛЬЗОВАНИЕМ СРЕДСТВ MATHCAD

3.1 Проверка однородности данных параметрическими методами

Важно понимать являются ли имеющиеся данные однородными. Под однородными данными принято понимать некоторый уровень их рассеивания, при котором рассчитываемые статистические показатели давали бы надежную и качественную характеристику анализируемой совокупности. Граница, разделяющая неоднородные данные о однородных – не четкая.

Для начала проверим данные за один год. Поведем расчеты коэффициента вариации в 2007 году на дне и поверхности с первой по четырнадцатую станции. Выбор года обусловлен полнотой имеющихся данных.

Для расчеты среднеквадратического отклонения сначала нужно вычислить математическое ожидание. Затем, производится расчет отклонений показателей от средней величины, и рассчитывается дисперсия. Из дисперсии берется квадратный корень и таким образом вычисляется среднеквадратическое отклонение. Далее, делим среднеквадратическое отклонение на среднее значение, получаем:

Таблица 7 – результаты расчетов на дне реки, за 2007 г.

	БПК (мл/л)	Кислород	Железо	Нитриты	Фосфаты
--	------------	----------	--------	---------	---------

		(мл/л)	(мг/л)	(мкг/л)	(мкг/л)
МО	3,84	3,21	1,73	136,9	120
Коэф. вариации	0,32	0,57	0,43	0,21	0,43

Из полученных результатов, приведенных в таблице 7, можно сделать выводы, что на дне реки Охты однородными являются только БПК и концентрация нитритов, остальные признаки менее однородны.

Также, определим однородность признаков в 2007 году на поверхности реки

Таблица 8 – результаты расчетов на поверхности реки, за 2007 г.

	БПК (мл/л)	Кислород (мл/л)	Железо (мг/л)	Нитриты (мкг/л)	Фосфаты (мкг/л)
МО	4,50	3,60	1,71	135,9	129,94
Коэф. вариации	0,32	0,42	0,33	0,18	0,33

Из результатов полученных в ходе расчетов и приведенных в таблице 8, можно сделать выводы, что на поверхности реки Охты однородными являются все признаки кроме концентрации кислорода в воде.

Теперь, поверим на однородность данные по всем станциям, всем годам в выбранных признаках. Для этого проведем расчет МО. Полученные значения приведены в таблице 9.

Таблица 9 МО показателей в разные годы на дне и поверхности

Дно	2006	2007	2008	2009	2010	2011	2012	2013	2014	2015
БПК	4,32	4,50	4,80	5,46	4,60	4,60	6,16	4,26	4,57	4,91
Кислород	3,46	3,60	1,70	2,96	2,76	2,40	2,63	2,93	5,16	4,53
Железо	1,16	1,71	0,96	2,85	3,33	1,34	1,02	1,56	2,08	1,60
Нитриты	55,0	135,9	83,2	74,9	93,5	137,6	69,6	69,0	109	76,4
Фосфаты	94,6	129,9	203,4	192,6	144,2	113,3	138,5	92,7	79,3	118,6
Поверхн.	2006	2007	2008	2009	2010	2011	2012	2013	2014	2015

БПК	4,13	3,89	4,81	5,55	4,44	5,39	6,11	3,65	4,67	3,85
Кислород	3,11	3,21	1,25	2,98	2,32	2,18	2,53	2,60	5,35	3,86
Железо	1,25	1,73	0,97	3,10	3,19	0,95	1,04	1,56	2,11	1,58
Нитриты	59,4	136,9	93,0	79,5	88,2	150,1	74,3	67,0	108	74,2
Фосфаты	139	120,0	194,9	208,8	144,2	118,8	137,4	89,6	62,4	105,9

По данным таблицы 9, можно судить об однородности имеющихся данных: БПК, концентрации кислорода и железа на поверхности и на дне, являются достаточно однородными, их разброс варьируется в пределах 30%. Концентрации фосфатов и нитритов менее однородны, особенно на поверхности.

Применение t-критерия Стьюдента для проверки однородности

Одним из параметрических методов проверки однородности выборок, является t-критерий Стьюдента. Для его расчета необходимо вычислить математическое ожидание, среднеквадратическое отклонение, затем вычислить эмпирическое значение и определить по таблице квантилей значения уровня значимости.

Уровень статистической значимости (или статистически значимый результат) показывает, какова вероятность случайного возникновения исследуемых показателей. Общая статистическая значимость явления выражается коэффициентом p-уровень. В любом эксперименте или наблюдении существует вероятность, что полученные данные возникли из-за ошибок выборки. То есть статистически значимой является величина, чья вероятность случайного возникновения крайне мала либо стремится к крайности. Крайностью в этом контексте считают степень отклонения статистики от нулевой гипотезы (гипотезы, которую проверяют на согласованность с полученными выборочными данными) [24].

Для реки Охта, t-критерием Стьюдента, были проверены на однородность данные по всем станциям за 2007 и 2014 гг.

При расчете t-критерия Стьюдента получились приведенные в таблице 10.

Таблица 10 – Результаты расчета t-критерия Стьюдента

	БПК		Кислород		Железо		Нитриты		Фосфаты	
	t	a	t	a	t	a	t	a	t	a
Поверхность	0,14	0,8	3,43	0,002	1,76	0,1	2,52	0,02	1,89	0,1
Дно	1,74	0,1	3,54	0,002	1,65	0,1	2,51	0,02	3,23	0,002

По полученным результатам можно заключить: практически все признаки однородны, за исключением биохимической потребности в кислороде на поверхности. В экологическом анализе считается нормальным уровень значимости в 10%. В наших результатах вероятность ошибки присутствует более маленькая, такая как 0,2% для концентрации кислорода на поверхности и на дне, и для концентрации фосфатов на дне реки.

3.2 Проверка статистической однородности и связи между признаками непараметрическими методами

Критерий Вилкоксона

Проверка непараметрическим критерием Вилкоксона была произведена для выявления однородности между значениями за 2007 и 2014 гг., годы были выбраны благодаря полноте имеющихся данных.

Для расчета критерия Вилкоксона, была принята выборка данных за 2014 год, за типичный сдвиг, будем принимать сдвиг в отрицательную сторону (отрицательное значение разности значения 2014, 2007 года), а сдвиг в положительную сторону, примем за нетипичный. Далее, все значения выписываются по модулю и модули ранжируются. Затем, суммируются ранги типичного сдвига, таким образом получается t-эмпирическое. Значение эмпирическое находится по таблице и сверяется с критическим. Если критическое значение больше, значит выборка однородна, в противном случае, если t-критическое сильно крупнее t-эмпирического – можно

утверждать, что выборка является неоднородной. В нашем случае, при $n=14$ и $p<0,05$, t -критическое равно 25.

Таблица 11 – результаты расчета критерия Вилкоксона

	БПК	Железо	Кислород	Нитриты	Фосфаты
Поверхность	50	6	28	93	90
Дно	25	10	25	94	102

По результатам, приведенным в таблице 11, можно сделать вывод, что однородными являются: значение БПК у дна реки Охта, концентрации железа и кислорода.

Критерии Спирмена и Кендалла

Проверим непараметрическими методами Спирмена и Кендалла – коррелируют ли данные по крайним станциям за все года по разным признакам.

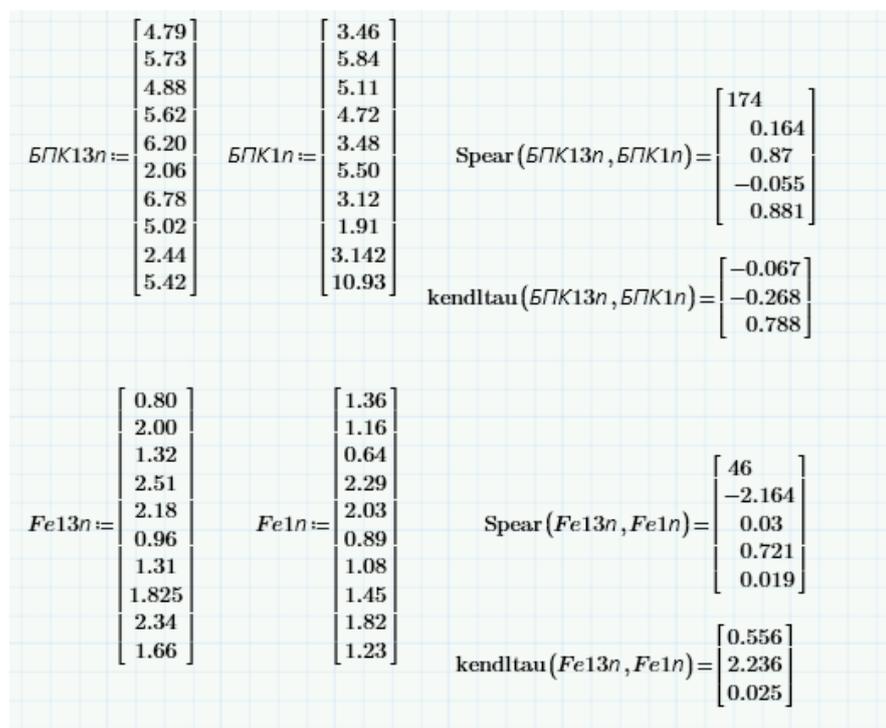


Рисунок 8 – пример расчёта критериев в MathCad

Для этого составим матрицы данных, присвоив им идентификаторы такие как: БПК13п (БПК на тринадцатой станции, на поверхности), Fe1д (концентрация железа на первой станции, у дна) и т.д. Произведем

необходимые расчеты по критериям Спирмена и Кендалла в пакете прикладных программ MathCad. На рисунке 8 представлен пример расчета критериев для поверхности реки Охты по первой и тринадцатой станциям, таких признаков как биохимическая потребность в кислороде и концентрация железа в воде. Выбор станций обусловлен тем, что они находятся на приличном расстоянии друг о друга. Одна является водохранилищем, другая устьем.

Таблица 11 – результаты расчетов по критериям Спирмена и Кендалла

	БПК		Железо		Кислород		Нитриты		Фосфаты	
	Пов.	Дно	Пов.	Дно	Пов.	Дно	Пов.	Дно	Пов.	Дно
Спирмен	-0,05	0,212	0,721	0,6	0,608	0,705	0,503	0,394	-0,23	0,103
Кендалл	-0,07	0,156	0,556	0,378	0,584	0,539	0,289	0,289	-0,16	0,022

Из приведенных результатов можно сделать вывод, что связь между значениями признаков на разных станциях есть, но не сильная. Например, концентрации железа, кислорода и нитритов имеют связь на разных станциях более сильную, чем биохимическая потребность в кислороде и концентрация фосфатов во воде.

Также, хорошо видно, что для одних и тех-же значений переменных, значение коэффициента Спирмена немного больше чем значение коэффициента Кендалла, тогда как уровень значимости одинаков или же, в некоторых случаях, у коэффициента корреляции Кендалла будет немного больше [26].

Проверка связи между признаками, непараметрическими методами

Как известно – метод ранговой корреляции Спирмена позволяет определить тесноту и направление корреляционной связи между двумя признаками.

Проверим тесноту связи данных попарно между имеющимися признаками (БПК, концентрация кислорода, железа, нитритов и фосфатов) на дне тринадцатой станции.

При использовании коэффициента ранговой корреляции условно оценивают тесноту связи между признаками, считая значения коэффициента равные 0,3 и менее – показателями слабой тесноты связи; значения более 0,4, но менее 0,7 – показателями умеренной тесноты связи, а значения 0,7 и более – показателями высокой тесноты связи.

Таблица 12а – коэффициент корреляции Спирмена признаков на поверхности реки

	БПК	Кислород	Железо	Нитриты	Фосфаты
БПК		0,723	0,297	-0,176	0,648
Кислород	0,723			-0,292	0,085
Железо	0,297	-0,091		0,527	0,2
Нитриты	-0,176	-0,292	0,527		-0,321
Фосфаты	0,648	0,085	0,2	-0,321	

Таблица 12б – коэффициент корреляции Спирмена признаков у дна реки

	БПК	Кислород	Железо	Нитриты	Фосфаты
БПК		0,345	-0,224	-0,006	0,333
Кислород	0,345		-0,297	-0,176	-0,297
Железо	-0,224	-0,297		0,539	0,018
Нитриты	-0,006	-0,176	0,539		-0,333
Фосфаты	0,333	-0,297	0,018	-0,333	

Таким образом, было вычислено, что **на дне:**

Связь практически отсутствует между БПК и концентрацией нитритов в воде и концентрациями железа и фосфатов;

Имеется очень слабая связь между БПК и концентрацией железа, концентрациями: железа и кислорода, нитритов и кислорода, кислорода и фосфатов;

Умеренная теснота связи присутствует между концентрациями нитритов и фосфатов, БПК и концентрацией фосфатов, БПК и концентрацией кислорода, концентрациями железа и нитритов.

На поверхности:

Связь практически отсутствует между БПК и концентрацией нитритов в воде, концентрациями железа и кислорода, железа и фосфатов, кислорода и фосфатов;

Имеется очень слабая связь между БПК и концентрацией железа и концентрациями: нитритов и кислорода, нитритов и фосфатов;

Умеренная теснота связи присутствует между БПК и концентрацией фосфатов, концентрациями железа и нитритов;

Высокая теснота связи между БПК и концентрацией кислорода.

Альтернативным коэффициенту корреляции Спирмена, является коэффициент корреляции Кендалла – предназначен для определения взаимосвязи между двумя ранговыми переменными.

Расчеты коэффициента ранговой корреляции Кендалла дают похожие цифры, что говорит в целом о непрочной связи между признаками в целом.

Таблица 13а - коэффициент корреляции Кендалла признаков на поверхности реки

	БПК	Кислород	Железо	Нитриты	Фосфаты
БПК		0,539	0,244	-0,268	0,556
Кислород	0,539		-0,135	-0,225	0,09
Железо	0,244	-0,135		-0,225	0,156
Нитриты	-0,268	-0,225	-0,225		-0,244

Фосфаты	0,556	0,09	0,156	-0,244	
---------	-------	------	-------	--------	--

Таблица 13б - коэффициент корреляции Кендалла признаков у дна реки

	БПК	Кислород	Железо	Нитриты	Фосфаты
БПК		0,289	-0,111	0,022	0,244
Кислород	0,289		-0,2	-0,067	-0,111
Железо	-0,111	-0,2		0,422	0,022
Нитриты	0,022	-0,967	0,422		-0,289
Фосфаты	0,244	0,422	0,022	-0,289	

$$\text{Spear}(БПК13n, Hum13n) = \begin{bmatrix} 194 \\ 0.527 \\ 0.598 \\ -0.176 \\ 0.627 \end{bmatrix} \quad \text{kendltau}(БПК13n, Hum13n) = \begin{bmatrix} -0.067 \\ -0.268 \\ 0.788 \end{bmatrix}$$

$$\text{Spear}(БПК13n, O13n) = \begin{bmatrix} 45.5 \\ -2.17 \\ 0.03 \\ 0.723 \\ 0.018 \end{bmatrix} \quad \text{kendltau}(БПК13n, O13n) = \begin{bmatrix} 0.539 \\ 2.171 \\ 0.03 \end{bmatrix}$$

$$\text{Spear}(БПК13n, \Phi13n) = \begin{bmatrix} 58 \\ -1.945 \\ 0.052 \\ 0.648 \\ 0.043 \end{bmatrix} \quad \text{kendltau}(БПК13n, \Phi13n) = \begin{bmatrix} 0.556 \\ 2.236 \\ 0.025 \end{bmatrix}$$

Рисунок 9 - Пример расчета коэффициентов Спирмена и Кендалла в MathCad

Благодаря пакету прикладных программ MathCad, расчёты практически не требуют затрат времени. Полученных результатов достаточно для того, чтобы судить о присутствии взаимосвязи между признаками на станции. В нашем случае, проверены временные ряды признаков.

Временной ряд - совокупность значений какого-либо показателя за несколько последовательных моментов или периодов времени. Каждый уровень временного ряда формируется под воздействием большого числа факторов [27]. В нашем случае, одним из главных факторов загрязнений реки, являются точки сброса сточных и возвратных вод [28].

3.3 Основные положения методики непараметрического анализа статистической связи геоэкологических явлений

Параметрический t-критерий Стьюдента в данной методике служит для того, чтобы получить общее представление о ситуации, для возможности дальнейшего сравнения результатов расчетов непараметрическими методами. Для грамотного применения t-критерия Стьюдента, необходимо, прежде всего, проверить выборку на подчинение закону нормального распределения. В экологических исследованиях часто неизвестны и не могут быть установлены с помощью качественного анализа типы вероятностных законов распределения случайных процессов. В этом случае и находят применение непараметрические методы анализа.

В данной методике предлагается применить непараметрические критерии Вилкоксона, Спирмена и Кендалла. Выбор данных критериев обусловлен сравнительной простотой расчетов и рядом особенностей критериев. Также, важную роль играет то, что для непараметрических методов, математический аппарат более разработан. Они допускают использование данных экологических экспериментов, полученных в результате не только количественных, но и качественных измерений с помощью дискретных шкал наименований и порядка. Достоинством непараметрических методов является простота расчётов.

В первую очередь необходимо установить однородность выборок, с этой целью применяется t-критерий Вилкоксона. Критерий предназначен для сопоставления показателей, в нашей задаче для выборок данных за 2007 и 2014 год. Необходимо установить – являются ли данные однородными на разных станциях за выбранные годы. В дальнейшем, можно провести такой анализ для всех лет, сравнивая их попарно. Проще говоря, мы определяем – является ли сдвиг показателей в одном направлении более интенсивным, чем в другом.

Критерий применим в тех случаях, когда признаки измерены, по крайней мере, в порядковой шкале. Целесообразно применять данный критерий, когда величина самих сдвигов варьирует в некотором диапазоне (10—15 % от их величины) [29].

Критерий Вилкоксона относится к маломощным критериям. Под мощностью критерий принято понимать его способность выявлять различия между признаками, если они есть. Чем мощнее критерий, тем лучше он отвергает нулевую гипотезу и подтверждает альтернативную. Маломощные критерии обладают рядом достоинств: простота; широкий диапазон, по отношению к самым разным данным; применимость к неравным по объему выборкам; большая информативность результатов.

Следующим шагом в исследовании, является проверка однородности данных распределенных по времени на двух станциях. В данной работе рассмотрены две крайние станции – первая и тринадцатая. Данная проверка проведена уже с помощью критериев Спирмена и Кендалла. Критерии Спирмена и Кендалла в проверке неоднородности устанавливают – имеется ли связь между изменениями признаков с течением времени. В дальнейшем, можно предположить, что в случае наличия связи, имеются факторы, влияющие на изменение значений признаков.

Преимуществами коэффициента Спирмена являются: высокая работоспособность (т.е. устойчивость к нетипичным наблюдениям) и широкая область применения. Коэффициент корреляции Спирмена – мера линейной связи между случайными величинами. Он определяет степень тесноты связи порядковых признаков, которые в этом случае представляют собой ранги сравниваемых величин [30].

Аналогом критерия Спирмена, является критерий Кендалла. Достоинством этого критерия является то, что он применим к случайным величинам, распределение которых неизвестно, требуется лишь, чтобы величины были непрерывными. Способ вычисления критерия более прост, но

результаты его расчетов дают более грубую картину происходящих процессов. Данный метод может быть использован не только для количественно выраженных данных, но также и в случаях, когда регистрируемые значения определяются описательными признаками различной интенсивности. Предлагается применять критерий Кендалла в случаях, когда ситуацию надо увидеть в целом, при необходимости более детального исследования, применять критерий Спирмена [9].

В исследовании однородности t-критерием Вилкоксона, был использован уровень значимости $p < 0,05$. Уровень значимости, это допустимая ошибка в нашем утверждении (например, что выборки являются однородными).

Уровень значимости меньше 0,05 – первый уровень значимости. До 5% составляет вероятность того, что мы ошибочно сделали вывод о том, что различия достоверны, в то время как они недостоверны на самом деле. Можно утверждать обратное: мы лишь на 95% уверены в том, что различия действительно достоверны. В данном случае можно написать и так: $P > 0,95$. Общий смысл критерия останется тем же.

Также, существует второй уровень значимости. При нем $p < 0,01$. Вероятность ошибочного вывода о том, что различия достоверны, составляет не более 1%. Иначе говоря, мы на 99% уверены, что различия действительно достоверны

Третий уровень значимости при $p < 0,001$. Всего 0,1% составляет вероятность того, что мы сделали ошибочный вывод о том, что различия достоверны. Это — самый надёжный вариант вывода о достоверности различий [30].

Вероятность в критерии Вилкоксона смотрится в таблице.

После получения представления об однородности данных, производится проверка их на связь (устанавливаем закономерности изменения признаков) Для этого использованы коэффициенты ранговой

корреляции Спирмена и Кендалла. Признаки проверяются попарно. При исследовании 5 признаков, неповторяемых пар выходит 10. Проверка на корреляцию в данной работе была проведена на одной станции – тринадцатой. В дальнейшей работе целесообразно провести подобную проверку на всех станциях.

После установления статистической связи между выборками, проверяем вероятность выявления ошибочной корреляции. Ниже приведены вероятности ошибок в том, что статистически большее значение может появиться для некоррелируемых образцов.

Таблица 15а – вероятность, что большее значение может появиться для некоррелируемых выборок (критерий Спирмена, поверхность)

	БПК	Кислород	Железо	Нитриты	Фосфаты
БПК		0,018	0,405	0,627	0,043
Кислород	0,018		0,802	0,413	0,815
Железо	0,405	0,802		0,117	0,58
Нитриты	0,627	0,413	0,117		0,365
Фосфаты	0,043	0,815	0,58	0,365	

Таблица 15б – вероятность, что большее значение может появиться для некоррелируемых выборок (критерий Спирмена, дно)

	БПК	Кислород	Железо	Нитриты	Фосфаты
БПК		0,328	0,533	0,987	0,347
Кислород	0,328		0,405	0,627	0,405
Железо	0,533	0,405		0,108	0,96
Нитриты	0,987	0,627	0,108		0,347
Фосфаты	0,347	0,405	0,96	0,347	

Таблица 16а – вероятность, что большее значение может появиться для некоррелируемых выборок (критерий Кендалла, поверхность)

	БПК	Кислород	Железо	Нитриты	Фосфаты
БПК		0,03	0,325	0,788	0,025

Кислород	0,03		0,587	0,366	0,717
Железо	0,325	0,587		0,089	0,531
Нитриты	0,788	0,366	0,089		0,325
Фосфаты	0,025	0,717	0,531	0,325	

Таблица 16б – вероятность, что большее значение может появиться для некоррелируемых выборок (критерий Кендалла, дно)

	БПК	Кислород	Железо	Нитриты	Фосфаты
БПК		0,245	0,655	0,929	0,325
Кислород	0,245		0,421	0,788	0,655
Железо	0,655	0,421		0,089	0,929
Нитриты	0,929	0,788	0,089		0,245
Фосфаты	0,325	0,655	0,929	0,245	

Из приведенных выше значений можно заключить, что *на поверхности*:

Высокая вероятность ошибки в суждении о связи данных присутствует между БПК и концентрацией нитритов, концентрациями: железа и кислорода, фосфатов и кислорода, железа и фосфатов.

На дне:

Высокая вероятность ошибки в суждении о связи данных присутствует между БПК и концентрацией железа, БПК и концентрацией нитритов, концентрациями: нитритов и кислорода, кислорода и фосфатов, фосфатов и железа.

Сверяем с полученными значениями тау Кендалла и Спирмена в разделе 3.2. Вполне возможно, что несмотря на то, что в ходе установления

наличия связи между признаками, был сделан вывод, об отсутствии таковой (во всех случаях), на самом деле, связь есть [31].

Исследование однородности признаков и связи между признаками и выявление вероятности ошибок, были проведены в пакете прикладных программ MathCad, что позволило избежать расчетов и ввода формул вручную.

Выводы по главе:

Из приведенных выше расчетов можно заключить, что в целом, имеющиеся значения по пяти исследуемым признакам являются однородными. Если свести все расчеты в один, то признаками с наибольшей однородностью являются: концентрация кислорода, концентрация железа и БПК. Однородность позволяет получить наиболее точные результаты в ходе статистического анализа. У параметрических и непараметрических методов оценки однородности есть свои преимущества и недостатки, однако анализ непараметрическими критериями позволяет производить наиболее точную оценку. ППП MathCad дает возможность произвести исследование однородности просто и не затрачивая большого количества времени.

ЗАКЛЮЧЕНИЕ

Целью работы – разработка методики непараметрического анализа статистической связи геоэкологических явлений выполнена в данной выпускной квалификационной работе.

В процессе исследования получены следующие основные результаты:

1. Проведен анализ параметрических и непараметрических методов.

Изучение сложных закономерностей экологических процессов в водной среде требует проведения объективного количественного анализа. Безусловно, выбор методов анализа данных зависит от условий задачи [25]. Указанные выше особенности экологических исследований не всегда позволяют применять методы параметрической статистики. Непараметрические методы исследования актуальны при анализе концентрации вредных веществ в воде и ее изменении со временем, при анализе изменения количества донных отложений или при выявлении связи между сокращением численности животных и количеством бытовых отходов в среде их обитания. В проблеме проверки статистических гипотез, возникающих в экологических исследованиях, непараметрические методы анализа данных о загрязнении морей могут иметь преимущества перед параметрическими в силу их разнообразия, простоты расчетов и возможности применения для небольших объемов выборок.

Река Охта является обширным объектом для исследований поэтому для статистического анализа имеющихся данных о пробах воды, необходима методика. Необходимо установить, являются ли данные полученные студентами РГГМУ однородными, проверить статистическую значимость признаков и установить связь между различными признаками. С этой целью будут применены параметрические t-критерий Стьюдента и непараметрические критерии Вилкоксона, Спирмена и Кендалла. Изучение сложных закономерностей экологических процессов в водной среде требует проведения объективного количественного анализа. Безусловно, выбор методов анализа данных зависит от условий задачи. Особенности экологических исследований не всегда позволяют применять методы параметрической статистики. Непараметрические методы исследования актуальны при анализе концентрации вредных веществ в воде и ее изменении со временем, при анализе изменения количества донных отложений или при выявлении связи между сокращением численности животных и количеством

бытовых отходов в среде их обитания. В проблеме проверки статистических гипотез, возникающих в экологических исследованиях, непараметрические методы анализа данных о загрязнении морей могут иметь преимущества перед параметрическими в силу их разнообразия, простоты расчётов и возможности применения для небольших объемов выборок.

2. Состояние реки Охты ухудшается, что хорошо видно по тенденциям концентраций загрязняющих веществ в воде, особенно железа. Для более детальной оценки состояния реки необходима методика. Предлагается выполнить статистический анализ с применением непараметрических критериев в пакете прикладных программ MathCad.

Выбор программного продукта обусловлен рядом его достоинств, таких как легкость и удобность работы, возможность выполнять большое количество расчетов, наличие справочника, а также получение наглядных результатов.

3. Из расчетов, выполненных в ходе данной выпускной квалификационной работы можно заключить, что в целом, имеющиеся значения являются однородными. Признаками с наибольшей однородностью являются: концентрация кислорода, концентрация железа и БПК. Однородность позволяет получить наиболее точные результаты в ходе статистического анализа. У параметрических и непараметрических методов оценки однородности есть свои преимущества и недостатки, однако анализ непараметрическими критериями позволяет производить наиболее точную оценку, ППП MathCad, позволяет произвести исследование однородности просто и не затрачивая большого количества времени.

Общий план предлагаемой методики непараметрического анализа статистической связи геоэкологических явлений:

К данным применяется t-критерий Стьюдента для представления общей картины об экологическом состоянии;

- Производится проверка однородности данных по всем станциям за два года (в данной работе 2007 и 2014 гг.) t-критерием Вилкоксона;

- Проверка наличия связи между признаками коэффициентами ранговой корреляции Спирмена и Кендалла;

- Результаты полученные в ходе установления связи между признаками коэффициентами Спирмена и Кендалла подвергаются проверке на вероятность ошибки;

Данная методика позволяет учесть такие факторы как вероятность ошибки, установить наличие связи между признаками и их однородность используя хорошо разработанный математический аппарат и относительную простоту использования непараметрических методов.

СПИСОК ИСПОЛЬЗУЕМЫХ ИСТОЧНИКОВ:

1. Алексеев Д. К. Экологический мониторинг: современное состояние, подходы и методы: учебное пособие / Д. К. Алексеев, В. В. Гальцова, В. В. Дмитриев; М-во образования и науки Российской Федерации,

Гос. образовательное учреждение высш. проф. образования Российский гос. гидрометеорологический ун-т. - Санкт-Петербург : РГГМУ, 2011. -302с.

2. Станскова А.Г. Оценка экологического состояния р. Охта в Санкт-Петербурге и Ленинградской области / А.Г. Станскова; - Санкт-Петербург: РГГМУ, 2016. -73с.

3. Куличенко А.Ю. Оценка состояния Р. Охта (г. Санкт-Петербург) на основе биотических методов / А.Ю. Куличенко // ФГБОУ ВО «Вятский государственный университет», ФГБУН Институт биологии Коми научного центра Уральского отделения Российской академии наук. Материалы XV Всероссийской научно-практической конференции с международным участием. Киров: Вятский государственный университет. -2017. - С. 110-107.

4. Зуева Н.В. Использование макрофитов в оценке экологического состояния малой реки (на примере реки Охты, Санкт-Петербург)/ Н.В.Зуева, А.А.Бобров// Биология внутренних вод. – 2018. - №1. - С.45-54.

5. Охта. Энциклопедия Санкт-Петербурга // Электронная энциклопедия и словарь [Электронный ресурс] URL: http://encdic.com/enc_spb/Ohta-971.html. (дата обращения 25.04.2018)

6. Охта //Wikipedia [Электронный ресурс] URL: [https://ru.wikipedia.org/wiki/Охта_\(Приток_Невы\)](https://ru.wikipedia.org/wiki/Охта_(Приток_Невы)) (дата обращения: 11.04.2018).

7. Урусова Е.С. Оценка загрязненности реки Охта в пределах Санкт-Петербурга на основе применения интегральных кривых / Е.С. Урусова // ОБЩЕСТВО. СРЕДА. РАЗВИТИЕ. – 2015. - №4 (37).- С.171-175.

8. Орлов А.И. Нечисловая статистика / А.И. Орлов; – М.: МЗ-Пресс, 2004. – 513 с.

9. Шелутко В.А. Методы обработки и анализа гидрологической информации. Учебно-методическое пособие / В.А. Шелутко; — Санкт-Петербург: Издательство СПбГУ, 2007. — 192 с.

10. Схема комплексного использования и охраны водных объектов (СКИОВО) бассейна реки Нева. В 6-х книгах/ Книга 2. Оценка экологического состояния и ключевые проблемы речного бассейна. // ФЕДЕРАЛЬНОЕ АГЕНТСТВО ВОДНЫХ РЕСУРСОВ Невско-Ладожское Бассейновое Водное Управление [Электронный ресурс] URL: <http://www.nord-west-water.ru> (дата обращения: 14.04.2018).

11. Грабарь М.И. Применение математической статистики в педагогических исследованиях. Непараметрические методы / М.И. Грабарь, К.А. Краснянская; - М.:Педагогика, 1977 - 136с

12. Кобзарь А.И. Прикладная математическая статистика. Для инженеров и научных работников/ А.И Кобзарь; – М.: ФИЗМАТЛИТ, 2006 – 816 с.

13. [Статистика в обработке материалов психологических исследований // allbest \[Электронный ресурс\] URL: https://knowledge.allbest.ru/psychology/3c0b65625a3ad68a5c53b88421306d27_0.html](https://knowledge.allbest.ru/psychology/3c0b65625a3ad68a5c53b88421306d27_0.html) (дата обращения: 14.04.2018).

14. Кузнецов В.М. Основы научных исследований в животноводстве // В.М. Кузнецов; – Киров: Зональный НИИСХ Северо-Востока, 2006. – 202 с.

15. Гмурман В.Е. Теория вероятностей и математическая статистика // В.Е. Гмурман; – М: Высшая школа, 2002. –343 с.

16. _Наследов, А.Д. Математические методы психологического исследования. Анализ и интерпретация данных / А. Д. Наследов; – СПб.: Речь, 2004. – 392 с.

17. Панферов А.И. Применение Mathcad в инженерных расчетах. Учебное пособие/ А.И.Панферов, А.В.Лопарев, В.К.Пономарев; - СПб: СПбГУАП, 2014. -83с.

18. Радченко Т.А. Статистический анализ данных в пакете Mathcad: Учебное пособие/ Т.А. Радченко, А.В. Дылевский ; - Воронеж: Изд-во ВГУ, 2004. - 23 с.

19. Ивановский Р. И. Теория вероятностей и математическая статистика. Основы, прикладные аспекты с примерами и задачами в среде Mathcad / Р.И. Ивановский; — СПб.: БХВ- Петербург, 2008. — 528 с.
20. Макаров Е.Г. Mathcad Учебный курс / Е.Г. Макаров; - СПб: Питер, 2009. -381с.
21. Конык О.А. Контроль качества воды, атмосферного воздуха и почвы. Учебное пособие / О.А.Конык, Т.В.Шахова; - СЫКТЫВКАР: СЛИ, 2013.- 145с.
22. Зенин А.А. Гидрохимический словарь / А.А. Зенин, Н.В. Белоусова; - Л.: Гидрометеиздат,1988. - 240с.
23. Елисеева И.И. Е51 Общая теория статистики: Учебник/ И.И. Елисеева, М.М Юзбашев / Под ред. чл.-корр. РАН И.И. Елисеевой. - М.: Финансы и статистика, 2001.- 480 с.
24. URL: <http://away.vk.com/away.php> (дата обращения: 29.05.2018)
25. Кривда А.А. Экологические проблемы мирового океана. Реферат / А.А. Кривда; – Киев, 2000. – с.3.
26. Коэффициент ранговой корреляции Кендалла // Stapsy [Электронный ресурс] URL: <http://statpsy.ru/kendall/correlation-kendall/> (дата обращения: 27.05.2018)
27. Временные ряды // studfiles [Электронный ресурс] URL: <https://studfiles.net/preview/2558995/> (дата обращения: 11.05.2018)
28. Гольцова В.В. «ОЦЕНКА экологического состояния водотоков Санкт-Петербурга на основе гидрохимических и гидробиологических методов» / В.В. Гольцова, Л.В. Кулангиева, Н.Ю. Ваганова // ЭКОЛОГИЧЕСКИЕ И ГИДРОМЕТЕОРОЛОГИЧЕСКИЕ ПРОБЛЕМЫ БОЛЬШИХ ГОРОДОВ И ПРОМЫШЛЕННЫХ ЗОН Материалы конференции; - СПб.: РГГМУ, 2004. - С. 122.
29. Достоинства малоомощных критериев [Электронный ресурс] URL: <https://megalektsii.ru/s35704t7.html> (дата обращения 20.04.2018)

30. Назовите преимущества и недостатки коэффициента корреляции Спирмена по сравнению с коэффициентом корреляции Пирсона [Электронный ресурс] URL: <https://lektsii.org/7-47322.html> (дата обращения 30.04.2018)

31. Лекция 11. Методы измерения тесноты парной корреляционной связи. [Электронный ресурс] URL: http://edu.tltsu.ru/sites/sites_content/site216/html/media96435/lec_11-1.pdf (дата обращения 3.05.2018)

32. Руководство к ППП MathCad