

Слайд 1 Титульный лист

Слайд 2 Постановка задачи

В рамках данной работы мне была представлена задача выполнить на основе реальных данных: 1) предварительный стат. анализ; 2) вариограммный анализ; 3) исследование стат. свойств оценки вариограммы 4) прогнозирование методом кригинг

Слайд 3 Содержание. Структура данной работы имеет вид.

Слайд 4 Исходные данные

Исходные данные получены от учебно-научного центра «Нарочанская биологическая станция им. Г.Г.Винберга», состоящие из значений средней температуры воды в июле месяце каждый год в период с 1975 по 2012. Для анализа результатов прогнозирования были использованы наблюдения только с 1975 по 2006 год.

Слайд 5 Обзор ПО. Особенности

Для решения поставленной задачи на языке программирования R мной было реализовано клиент-серверное приложение, включающее в себя 3 модуля: модуль предварительного анализа данных, модуль анализа остатков и модуль вариограммного анализа. Приложение доступно по адресу на экране с любого устройства. Имеет простой, в случае необходимости легко расширяемый интерфейс.

Слайд 6 Обзор ПО. Модуль предварительного стат. анализа. Первичный анализ

Страница первичного анализа представляет возможности по подбору закона распределения проверкой как различными тестами, так и визуально: на гистограмме и графике квантилей. Также на данной странице отображается таблица с вычисленными описательными статистиками.

Слайд 7 Проверка на нормальность

По описательным статистикам, графикам гистограммы и квантилей и проверкой тестов показана близость выборочного распределения к нормальному.

Слайд 8 Обзор ПО. Модуль предварительного стат. анализа. Корреляционный анализ

Страница корреляционного анализа позволяет оценить зависимость исследуемых данных с помощью диаграммы рассеяния, вычисляет коэффициент корреляции и проверяет его значимость. Для исследуемых данных показана умеренная положительная зависимость температуры воды от времени и отсутствие выбросов.

Слайд 9 Обзор ПО. Модуль предварительного стат. анализа. Регрессионный анализ

Инструменты регрессионного анализа позволяет получить регрессионную модель по исследуемым данным, а также провести анализ модели: определить значимость вычисленных коэффициентов, адекватность модели и линейность.

Слайд 10 Регрессионный анализ

Найден вид исследуемого временного ряда и уравнение тренда. Доказана значимость коэффициентов регрессионной модели, показана адекватность, отклонение от линейности незначительно. Но при этом коэффициент детерминации оказался не высоким (< 0.7). //В таблице отображены известные значения, тренд и ряд остатков. //

Вывод Инструменты данного модуля позволяют быстро получить информацию по исследуемым данным и сделать первые выводы и наметить шаги по дальнейшему исследованию.

Слайд 11 Обзор ПО. Модуль анализа остатков. Автокорреляционная функция

Для исследования ряда остатков реализован соответствующий модуль включающий в себя первичный анализ (аналогичный рассмотренному) и анализ автокорреляционной функции. На слайде продемонстрирована страница анализа автокорреляционной функции, позволяющая как визуально, так и с помощью ряда тестов определить наличие автокорреляций и стационарности в исследуемом временном ряду.

Слайд 12 Анализ остатков

Показана близость распределения ряда остатков к нормальному. По графику и тестами Льюнга-Бокса и Дики-Фуллера сделано заключение об отсутствии значимых автокорреляций и стационарности в широком смысле.

Вывод В зависимости от результатов, полученных на рассмотренном этапе, можно закончить исследование в случае, если модель удовлетворительного качества, либо продолжить в модуле вариограммного анализа.

Слайд 13 Вариограммный анализ

Следует отметить, что в рамках модуля вариограммного анализа прогнозные значения вычисляются как сумма значений вычисленных по тренду и по кригингу. Для оценки качества используются коэффициент корреляции между известными значениями ошибки и интерполяционными, и среднеквадратическая ошибка.

Слайд 14 Обзор ПО. Модуль вариограммного анализа. Семивариограмма

Начальный шаг в вариограммном анализе состоит в подборе модели и её параметров к экспериментальной семивариограмме. В качестве оценки семивариограммы реализована возможность использовать оценку Матерона и робастную оценку Кресси-Хоккинса. А также реализовано два подхода по подбору: визуальными силами исследователя, и автоматическими методами.

Слайд 15 Обзор ПО. Модуль вариограммного анализа. Подбор параметров модели

Страница подбора параметров позволяет оценить как определённый параметр влияет на качество модели и конечный результат и как следствие выбрать оптимальное значение. В частности на данном слайде отображён

график зависимости MSE от значения параметра ранга. Выбрав значение ранга, соответствующее минимальному значению MSE, получаем оптимальное значение.

Слайд 16 Обзор ПО. Модуль вариограммного анализа. Прогнозирование кригинг

Страница кригинга является наглядной демонстрацией результатов подбора моделей. На ней изображается график с наблюдаемыми значениями и прогнозными значениями. График также сопровождается вспомогательными таблицами с произведёнными в процессе расчётами.

Слайд 17 Визуальный подход. Линейная

С помощью модуля вариограммного анализа мной были рассмотрены различные модели семивариограмм, из них линейная (на слайде) является простейшей. Представленная модель плохо вписывается в данные, что видно по характеристикам.

Слайд 18 Визуальный подход. Линейная с порогом

По аналогичной схеме рассмотрена линейная модель с порогом. Подобранные с помощью приложения (адаптивным методом) параметры позволили довольно точно предсказать неизвестные значения. Поэтому данная модель хороша для краткосрочных прогнозов. При этом КК не высок, а MSE наоборот. Что говорит о том, что данная модель не очень справляется с описанием всего ряда остатков.

Слайд 19 Визуальный подход. Сферическая

Следующая рассмотренная модель является широко используемой, результат её использования оказался хуже предыдущего.

Слайд 20 Визуальный подход. Периодическая

По графику оценки семивариограммы была замечена некоторая периодичность в значениях. Поэтому была использована следующая модель. В результате подбора, получился высокий КК, по сравнению со всеми подобранными. При этом по графику видно, что прогноз не точен, но модель лучше всего описывает всю выборку.

Слайд 21 Автоматический подход.

Как было сказано ранее также реализован функционал по автоматическому подбору моделей семивариограмм. В рамках данного подхода использовалась также оценка Кресси-Хоккинса.

Слайд 22 Автоматический подход. Периодическая модель

Как и в случае ручного подбора, данная модель лучше себя показывает для описания ошибки. И в целом дает неплохой результат по прогнозу. Но не точен.

Слайд 23 Автоматический подход. Волновая модель

По результатам построена такая модель с таким результатом. Плохо описывает ошибку в целом, но улавливает поведение неизвестных значений.

Вывод Визуальные методы точнее и надёжнее так как исследователь знает специфику данных, но требует определённых знаний у пользователя, как статистических так и опыт по вариограммному анализу. В свою очередь автоматический подбор может использоваться любыми пользователями, так как он сразу выдаёт результат. Ну и следует отметить так же, что температура воды является специфическим показателем (ОСОБЕННО В НАШЕМ СЛУЧАЕ), поскольку она может изменяться в широком диапазоне. В случае с данными, которые имеют более плавный характер изменений (например глубина), автоматический подбор будет показывать себя лучше.

Слайд 24 Оценка вариограммы

Дипломная работа включает в себя исследование статистических свойств оценки семивариограммы. Вводятся следующие понятия: вариограмма, оценка вариограммы.

Слайд 25 Первые два момента

Мной найдены первые два момента введённой оценки. В теореме доказана несмещённость оценки.

Слайд 26 Асимптотика

Проведено исследование асимптотического поведения оценки.

Слайд 27 Асимптотика

Как следствие из доказанных теорем, показано что рассмотренная оценка вариограммы является несмещённой и состоятельной в среднеквадратическом смысле оценкой неизвестной вариограммы.