

*Ильичев Владимир Юрьевич, к.т.н., доцент кафедр «Системы обработки информации», «Тепловые двигатели и гидромашины»,*

*Калужский филиал ФГОУ ВО МГТУ им. Н.Э. Баумана (национальный исследовательский университет), г. Калуга, Россия*

*Перевозчикова Ирина Сергеевна, заведующий лабораторией кафедры «Защита информации»,*

*Калужский филиал ФГОУ ВО «Московский государственный технический университет имени Н.Э. Баумана (национальный исследовательский университет), г. Калуга, Россия, соавтор*

## **АВТОМАТИЗАЦИЯ ОБРАБОТКИ СТАТИСТИЧЕСКИХ МЕТЕОРОЛОГИЧЕСКИХ ДАННЫХ С ИСПОЛЬЗОВАНИЕМ СРЕДСТВ ЯЗЫКА PYTHON**

**Аннотация:** Авторы описывают исследования, посвященные созданию методики обработки статистических данных с целью обеспечения наглядности выводимых результатов для упрощения их анализа. В качестве инструмента исследования используется наиболее популярный на данный момент язык программирования Python и подключаемые к нему библиотеки математических функций, а также модуль визуализации. Приведено пошаговое описание применения разработанного алгоритма обработки информации. В качестве примера показано использование созданного программного комплекса для сортировки статистических метеорологических данных, в частности массивов данных по скоростям ветра в разных городах, для рассмотрения целесообразности использования альтернативных источников энергии - ветрогенераторов. По результатам работы сделаны выводы, а также приведены возможные пути дальнейшего использования разработанной авторами методики в других целях, кроме описанной в примере.

**Ключевые слова:** статистика, алгоритм сортировки, гистограмма, медианное значение, ветрогенератор, модули языка Python, визуализация результатов.

**Annotation:** The authors describe studies on the creation of a methodology for processing statistical data in order to ensure the visibility of the output results in order to simplify their analysis. The most popular Python programming language and libraries of mathematical functions connected to it, as well as a rendering module, are used as a research tool. A step-by-step description of the application of the developed information processing algorithm is given. As an example, the use of the created software system for sorting statistical meteorological data, in particular, arrays of data on wind speeds in different cities, to consider the feasibility of using alternative energy sources - wind generators, is shown. Based on the results of the work, conclusions were made, as well as possible ways of further using the method developed by the authors for other purposes than those described in the example.

**Keywords:** statistics, sorting algorithm, histogram, median, wind generator, Python modules, visualization of results.

## **Введение**

Для анализа механизмов протекания процессов различной природы, а также для решения прочих задач в науке и технике в настоящее время наиболее часто применяются методы, связанные с обработкой массивов статистической информации [1]. Данная информация собирается целенаправленно с использованием специальных приборов и экспериментальных установок, а её обработка осуществляется по особым алгоритмам, являющихся предметом высшей математики.

Решение задач обработки больших объёмов собираемой экспериментальным путём информации на данном этапе развития технологий невозможно представить без использования вычислительной техники (ЭВМ) и современных языков программирования.

Описываемая в статье работа посвящена разработке алгоритмов и программного обеспечения, позволяющего не только обрабатывать массивы информации, но и представлять результаты в наглядном графическом виде с целью более эффективного их анализа. В качестве примера рассматривается обработка баз метеорологических данных, в частности, данных по изменяющейся в течение года скорости ветра в различных регионах России. Данные исследования необходимы, например, при рассмотрении возможности применения на разных территориях альтернативных энергетических установок – ветрогенераторов [2].

Авторами осуществлена разработка методики статистического анализа с использованием современных средств автоматизации путём создания программного продукта на самом популярном на данный момент языке программирования Python. Особый упор сделан на использовании многих специальных библиотек математических функций и методов сортировки данных, а также модулей визуализации.

### **Материал и методы исследования**

Рассмотрим пошагово процесс получения материалов и применения выбранных методов исследования.

В качестве исходных статистических данных были использованы свободно распространяемые метеорологические базы, размещённые на сайте [3], на котором имеются сведения по 9 метеорологическим параметрам за каждые сутки – минимальной, максимальной и средней температуре воздуха, общем количестве осадков, глубине снежного покрова, направлении и скорости ветра (средней и пиковой), давлении воздуха и продолжительности солнечного сияния.

Структура указанного сайта позволяет выбрать промежуток времени для формирования базы, а также огромное количество метеорологических станций, после чего сформированную базу можно скачать в виде файла Excel в формате .xls.

Для конкретизации решаемой задачи и сравнения результатов выбраны

данные только по средней в течение каждых суток скорости ветра в двух городах нашей страны – в Москве и в Мурманске, и только за один год – 2022-й (для которого уже известен полный набор сведений по всем дням). Следует отметить, что это никоим образом не ограничивает применение разработанного метода исследования при обработке баз данных, существенно превышающих по объёму указанные.

Далее рассмотрим вычислительные методы и программные приёмы, использованные для обработки полученных массивов данных. Проще и нагляднее всего это можно сделать в ходе описания алгоритма созданного авторами программного продукта на языке Python, состоящего из следующих блоков:

1. Импорт используемых библиотек функций: Pandas [4] – для преобразования и обработки данных, NumPy [5] – для работы с массивами, Statistics – для проведения статистического анализа и Matplotlib.pyplot [6] – для вывода результатов расчётов и анализа в виде наглядных графиков.

2. Считывание данных из соответствующего изучаемому городу файлу Excel и их конвертация в базу данных Pandas.

3. Выбор одного кадра данных из полученной базы, соответствующего столбцу таблицы Excel, в котором содержатся данные по средней скорости ветра по дням года.

4. Транспонирование столбца в строку [7] и преобразование её в список с использованием функций модуля NumPy.

5. Вывод полученного списка на экран компьютера (для контроля правильности осуществлённого набора действий), а также формирование графической зависимости скорости ветра (км/ч) от номера дня года. Вывод графика в файл формата .png и на экран. Таким образом, на данном этапе осуществляется визуализация неотсортированного массива (списка) данных.

6. На втором этапе средствами модуля Matplotlib.pyplot реализован вывод того же массива данных, только в виде гистограммы [8], на которой по оси абсцисс отложены диапазоны скорости ветра, а по оси ординат – количество

дней, в которые скорости ветра попадают в каждый диапазон. При этом имеется возможность менять общий диапазон рассматриваемых скоростей и количество столбцов (баров) равной ширины на гистограмме.

7. В виде пользовательской функции осуществлена реализация сортировки массива скоростей по их возрастанию с использованием наиболее производительного математического рекурсивного метода «быстрой сортировки» [9], относящегося к классу методов «разделяй и властвуй» [10].

8. Вывод полученного массива (списка) на экран компьютера (также для контроля правильности выполнения алгоритма), с последующим построением графической зависимости скорости ветра (км/ч) от номера элемента списка. Таким образом, на третьем этапе осуществляется визуализация отсортированного массива данных.

9. С использованием модуля Statistics [11] производится вычисление медианного значения массива (медианной скорости ветра за год, км/ч) с выводом этого значения на график, построенный в п. 8. Далее осуществляется запись составного графика в файл формата .png и вывод его на экран.

### **Пример расчёта**

Произведём теперь обработку баз данных по скоростям ветра с помощью созданного программного продукта для городов Москва и Мурманск за 2022 г., с визуализацией и анализом полученных результатов.

Графические зависимости для неотсортированного массива исходных данных по скоростям ветра, полученным из файлов Excel, показаны на рис. 1 (*a* – для г. Москва, *б* – для г. Мурманск).

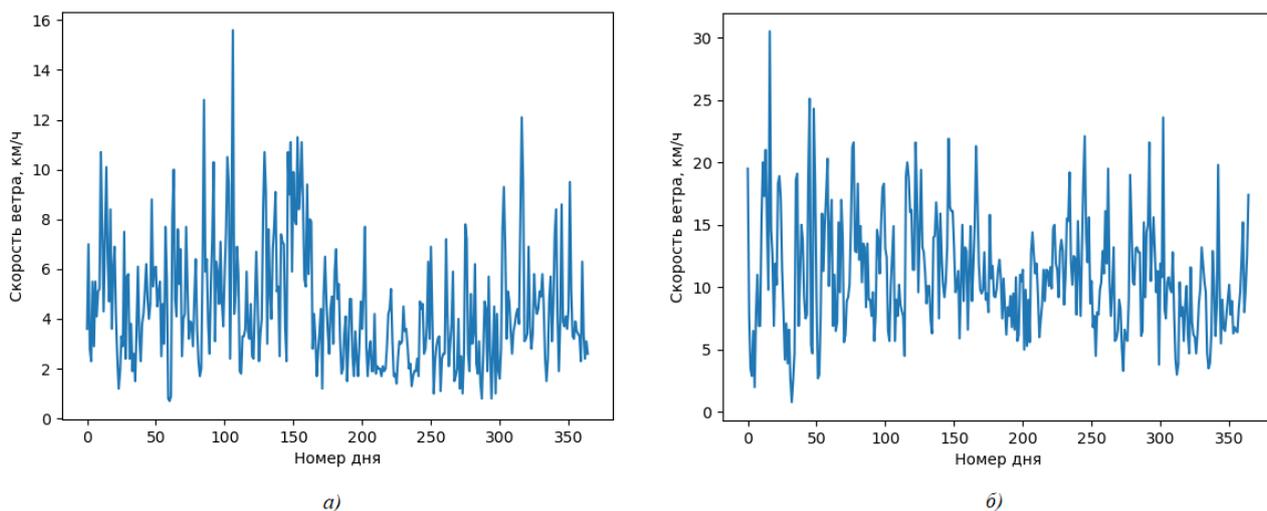


Рис. 1. Неотсортированные данные по скоростям ветра в г. Москва (график *а*) и в г. Мурманск (график *б*).

По графикам рис. 1 сложно сделать какие-либо выводы, кроме, пожалуй, определения значения максимальной за год скорости ветра, которое, однако, не является полезным для рассмотрения перспективности применения ветряных генераторов в выбранных городах.

На рис. 2 приведены гистограммы скоростей ветра (отсортированные по количеству дней, когда средняя скорость ветра попадала в каждый из показанных диапазонов значений). Для наглядности верхняя граница рассматриваемых значений скоростей для двух городов выбрана одинаковой – равной максимальной средней за день скорости ветра для г. Мурманск (30 км/ч).

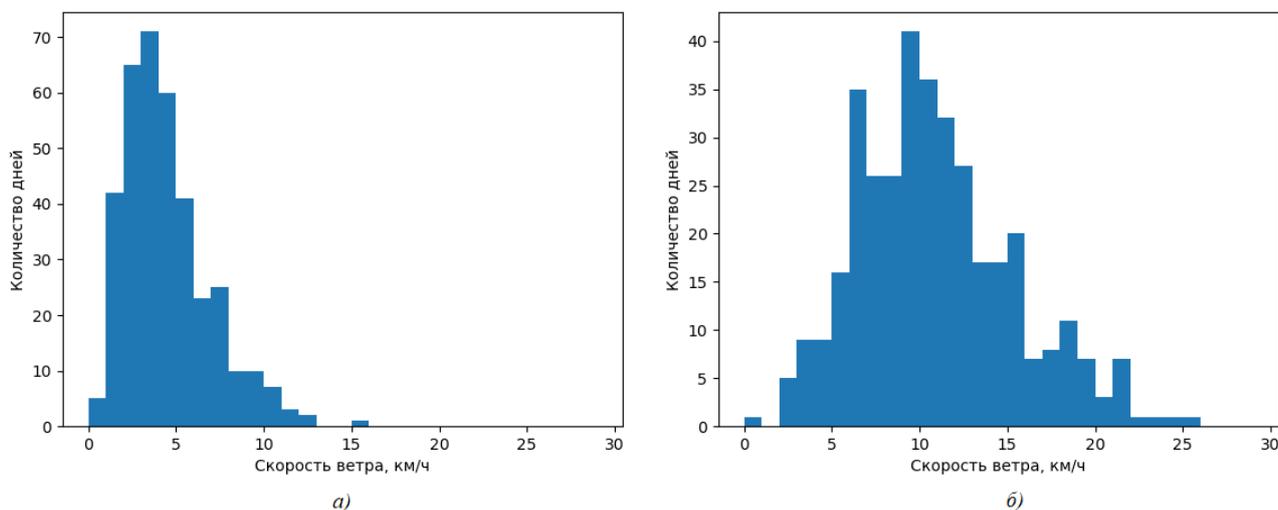


Рис. 2. Гистограммы данных по скоростям ветра в г. Москва (график *а*) и в г. Мурманск (график *б*), распределённых по диапазонам скоростей.

По гистограммам можно получить уже более полезную информацию для рассмотрения проектов установки в городах ветрогенераторов. В частности, наглядно видно, что преобладающая средняя по дням скорость ветра в г. Москва попадает в диапазон 3-4 км/ч (71 день), а в г. Мурманск – в диапазон 10-11 км/ч (41 день). Т.е., средняя скорость ветра в Мурманске примерно в 3 раза больше, чем в Москве.

Альтернативным методом, позволяющим достичь наглядности результатов с целью формулирования окончательных выводов, является анализ отсортированных в порядке возрастания данных по скоростям ветра, представленных на рис. 3 (синие кривые). Там же указаны медианные значения среднесуточных скоростей (оранжевые линии).

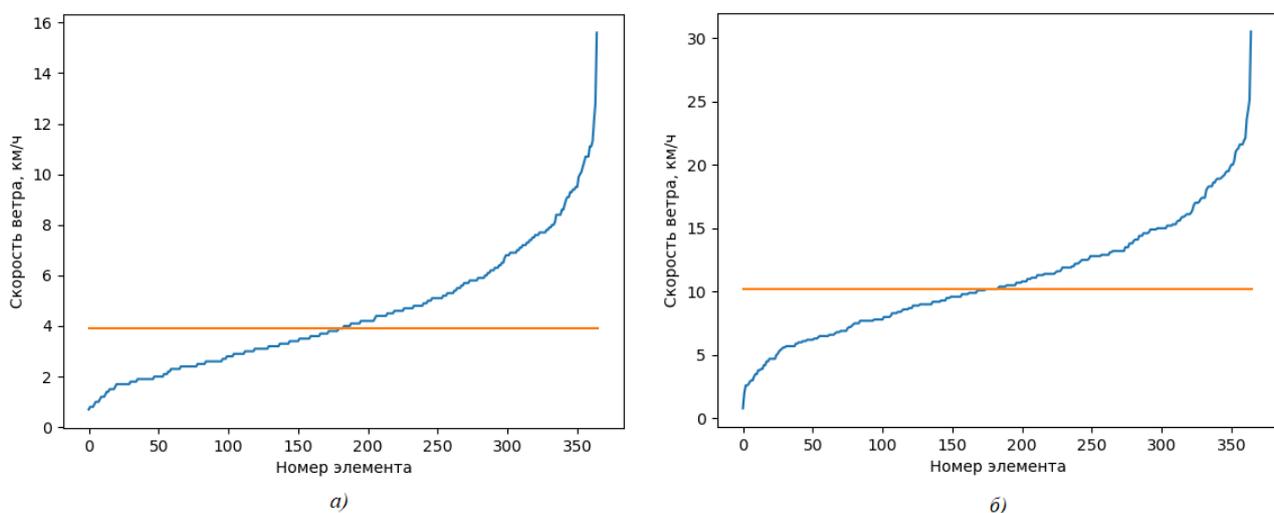


Рис. 3. Гистограммы данных по скоростям ветра в г. Москва (график *а*) и в г. Мурманск (график *б*), отсортированных в порядке возрастания с указанием медианных значений.

Графики рис. 3 подтверждают выводы, сделанные при анализе графиков рис. 2. Более пологие участки кривых соответствуют преобладающим скоростям ветра; через них же проходят графики медианных значений – для г. Москва это значение чуть менее 4 км/ч, а для г. Мурманск – чуть более

10 км/ч.

Рекомендуемое значение медианной скорости ветра для сооружения ветрогенераторов составляет от 4 м/с [12] или 14.4 км/ч, что в несколько раз превышает вычисленную скорость ветра в г. Москве и немного больше скорости ветра в г. Мурманск. Однако, в связи с тем, что в исходных базах данных приведены приземные скорости, рекомендуемого значения в г. Мурманск можно достичь, размещая ветрогенераторы на высоте (это справедливо, т.к. в атмосфере скорость ветра является степенной зависимостью от высоты, где степень больше единицы и зависит от конструкции ветряного колеса). В г. Москва рекомендуемой для ветряных установок скорости ветра не удастся достичь даже при значительных высотах размещения ветряка (достижимых в настоящее время).

### **Заключение**

Таким образом, цель исследования, заключающаяся в создании методики применения технологий сортировки и визуализации массива значений, полученных из баз данных, полностью достигнута. В результате работы программы создаётся ряд графиков, наглядно отображающих зависимости, необходимые для анализа данных в различных отраслях знаний.

Описанное исследование не только наглядно демонстрирует принципы обработки данных путём их сортировки, но и их конкретное применение для анализа скоростей ветра в разных регионах с целью рассмотрения целесообразности использования ветрогенераторов. Рекомендуется применять реализованные алгоритмы обработки массивов статистических данных в любых отраслях науки и техники, для исследования процессов, информация о которых собирается экспериментальным путём.

Читателю, изучающему среду программирования Python, безусловно будет интересно узнать описанные в статье возможности данного языка в совокупности с подключаемыми программными библиотеками при их применении к обработке баз данных.

В целом, можно сделать вывод, что описанные в статье программные

эксперименты увенчались успехом, что наглядно продемонстрировано их результатами.

### **Библиографический список:**

1. Афанасова Д.К., Афанасова К.А., Жукова Е.М. Применение аппарата математической статистики при обработке экспериментальных данных. // Молодой ученый. 2016. № 9-1 (113). С. 9-11.

2. Ильичев В.Ю., Шевелев Д.В. Расчёт характеристик мощности ветряных турбогенераторов с применением программного модуля Windpowerlib. // Известия МГТУ МАМИ. 2021. Т. 15. № 1. С. 23-31.

3. Хранитель рекордов погоды. [Электронный ресурс]. URL: <https://meteostat.net/> (Дата обращения 15.09.2023).

4. Ильичев В.Ю., Юрик Е.А. Анализ массивов данных с использованием библиотеки Pandas для Python. // Научное обозрение. Технические науки. 2020. № 4. С. 41-45.

5. Ильичев В.Ю., Юрик Е.А. Обработка статистических данных методом глубокого обучения с использованием модуля Keras. // Научное обозрение. Технические науки. 2020. № 5. С. 16-20.

6. Ильичев В.Ю., Юрик Е.А. Создание отчётов по доходам организаций с помощью языка Python. // Вектор экономики. 2020. № 6 (48). С. 10.

7. Токмазов Г.В., Панькина С.И., Лютикова М.Н. Решение простейших математических задач с использованием прикладных программ. // Учебное пособие. М.: Ай Пи Ар Медиа, 2023. 208 с.

8. Абдрахманов А.Н. Создание программы для построения гистограммы на основе графематического анализа данных из открытого интернет-источника. // В сборнике: Системы управления, информационные технологии и математическое моделирование. Материалы II Всероссийской научно-практической конференции с международным участием. В 2-х томах. Отв. редактор В.Н. Задорожный. 2020. С. 55-59.

9. Розанов А.В., Ермолаев Б.О. Алгоритмы быстрой сортировки больших

данных. // В сборнике: Экономико-математические методы анализа деятельности предприятий АПК. Материалы VII Международной научно-практической конференции, посвященной 110-летию Вавиловского университета. Под редакцией С.И. Ткачева. Саратов, 2023. С. 262-268.

10. Баушев А.Н., Семёнова О.Л. Об оценке временной сложности алгоритмов, построенных методом «разделяй и властвуй». // Интеллектуальные технологии на транспорте. 2019. № 1 (17). С. 45-47.

11. Кузнецова А.А. Применение инструментов Python в курсе статистики. // Актуальные проблемы преподавания математики в техническом вузе. 2023. № 10. С. 64-68.

12. Ветрогенераторы: вопросы и ответы. [Электронный ресурс]. URL: <https://www.eprussia.ru/epr/101/7689.htm> (Дата обращения 15.09.2023).