

Санкт-Петербургский государственный университет

Селиверстова Ксения Александровна

Выпускная квалификационная работа

Статистический анализ и прогноз земных температур для оценки глобального потепления

Уровень образования:

Направление 01.04.02 «Прикладная математика и информатика»

Основная образовательная программа ВМ.5505.2021 «Математическое и информационное обеспечение экономической деятельности»

Научный руководитель:
кандидат физ.-мат. наук,
доцент кафедры управления
медико-биологическими системами
Корников Владимир Васильевич

Рецензент:
профессор, Жиглявский
Анатолий Александрович

Санкт-Петербург

2023

ОГЛАВЛЕНИЕ

Введение.....	3
Обзор литературы.....	5
Глава 1. Постановка задачи.....	7
1.1. Описание предметной области.....	7
1.1. Описание математической модели	12
Глава 2. Анализ временных рядов.....	16
2.1. Первичная обработка и анализ данных.....	16
2.2. Исследование временных рядов в R	22
Глава 3. Реализация метода SSA в R.....	35
3.1. Обзор функций и библиотек.....	37
3.2. Обсуждение полученных результатов	38
Глава 4. Реализация метода SSA в Python	45
4.1. Обзор функций	45
4.2. Обсуждение полученных результатов	46
5. Заключение	53
6. Список использованной литературы.....	55
7. Приложение № 1	59
8. Приложение № 2	68

Введение

Последние несколько лет стали самыми теплыми за весь период метеорологических наблюдений, на данный момент температуры превышает по разным данным на 1-1,5 градуса Цельсия температурные значения доиндустриальной эпохи [3]. Таяние ледников и сокращение ледникового покрова происходит значительно быстрее, чем когда-либо прежде, что приводит к повышению уровня мирового океана. А это значит, что множество городов мира, расположенных в районах, подверженных риску затопления могут уйти под воду уже в нашем веке. В связи с этим возникает потребность не только в более глубоком изучении процесса, но и в прогнозировании дальнейшего его поведения.

Целью данной работы является вероятностно-статистический анализ и прогнозирование глобального потепления с помощью метода *Singular Spectrum analysis (SSA)*.

Для достижения поставленной цели решены следующие задачи:

1. Изучение предметной области и построение информационно-логической модели для процесса глобального потепления;
2. Сбор и систематизация метеорологических показателей для анализа климатических условий выбранных регионов;
3. Проведение корреляционно-регрессионного анализа для выбранных параметров;
4. Изучение соответствующей литературы и способов программной реализации для метода (*SSA*);
5. Программная реализация метода *SSA* с помощью среды разработки *Rstudio* и языка программирования *Python*;
6. Сравнение прогнозных значений, полученных в ходе реализации в разных средах разработки с реальными данными;
7. Анализ полученных результатов и оценка адекватности модели

Данная исследовательская работа имеет следующую структуру. В 1 главе сформулирована постановка задачи, в которую входят описание предметной области и построение соответствующей информационно-логической модели, а также математическая модель. В главе номер 2 представлены результаты первичного вероятностно-статистического анализа и построенные ARIMA модели. Главы 3 и 4 включают в себя программную реализацию метода Singular spectrum analysis в среде *Rstudio* и *Python*, с описанием полученных результатов. Также итогом 2, 3 и 4 глав стал сравнительный анализ, критериями которого выступали преимущества и недостатки методов, а также способов реализации и точность прогнозных значений. В заключении формулируются основные выводы и полученные результаты с возможными направлениями для дальнейших исследований.

Обзор литературы

Научные публикации, использованные при написании исследовательской работы, можно разделить на 2 части:

- научные статьи и доклады, связанные с изменением климата, глобальным потеплением их причинами и последствиями;
- литература, в которой изложены теоретические основы методов многомерной статистики и, в частности, метода SSA, а также приведено множество примеров практического применения.

Актуальные проблемы и основные достижения в области метеорологии, загрязнений природной среды можно найти в таких научных журналах, как «Метеорология и гидрология» [14], а также «Фундаментальная и прикладная климатология» [18]. Более полные оценки рисков, к которым приводит глобальное потепление, а также меры борьбы и адаптаций к изменениям можно найти в оценочных докладах Межправительственной группы экспертов по изменению климата (МГЭИК) [13]. Конкретно об изменении климата и его последствиях, а также количестве выбросов парниковых газов в каждом регионе и Российской Федерации (РФ) в целом можно найти в ежеквартальных и ежегодных докладах Федеральной службы по гидрометеорологии и мониторингу окружающей среды (Росгидромет) [17]. Для более полного изучения вопросов климатологии, причинах и последствиях изменения климата можно воспользоваться курсом лекций, разработанных Всемирным фондом дикой природы World Wildlife Fund (WWF) России и WWF Германии [4]. Исследования изменений климата с помощью SSA в других странах, например, в республике Беларусь, приводятся в статье [12]. В ней представлен анализ влияния глобального потепления на экономическую и социальную сферы жизни в конкретных регионах.

Общая теоретическая основа метода SSA и его разновидностей представлены в пособии Д.Л. Данилова и А.А. Жиглявского [8]. Множество практических примеров, в которых используются реальные данные приведено

в книге Н.Э. Голяндиной [6]. Применение методов многомерной статистики для исследования авиационной эргономики описывается в статье М. М. Кислицина [11]. О модификации метода анализа сингулярного спектра и ее применении в современных глобальных навигационных спутниках для прогнозирования параметров вращения Земли подробно описано в статье [7].

Говоря о статьях зарубежных авторов о применении метода SSA хочется отметить труды Hossein Hassani и Dimitrios D. Thomakos [27] в которых описано применение метода для анализа экономических и финансовых временных рядов. Также в статье Pa Ousman Vojang, Tao-Chang, Yang Quoc, Bao Pham и Pao-Shan Yu [28] применяют коллаборацию метода анализа сингулярного спектра и методов машинного обучения для прогнозирования осадков. В статье R. Vautard, M. Ghil [30] рассматривается применяет SSA к 4 палеоклиматическим записям, при этом устанавливая статистическую или динамическую размерности. Также большое количество работ публикуется Калифорнийским университетом (The University of California), объединившим 10 публичных калифорнийских университетов. Так, например, Robyn Schelenz в своей статье пишет об опасности горячего воздуха и высокой влажности преобладающих в Калифорнии, а также рассматривает возможности охлаждения городов и граждан [26]. Из других работ иностранных авторов хочется отметить статью Fraedrich K. [21], в которой приводится оценка размерности наблюдаемых атмосферных аттракторов, также несколько статей Ghil и др. [23, 24] о выделении динамики атмосферных и температурных данных.

Глава 1. Постановка задачи

1.1. Описание предметной области

В данной работе исследуется одна из глобальных проблем человечества – изменение климата. Процесс перемены климата представляет собой колебания климата Земли в целом или отдельных ее регионов с течением времени. Колебания в свою очередь выражаются в статистически достоверных отклонениях параметров погоды от многолетних значений за определенный период времени (от десятилетий до миллионов лет).

Для проведения исследования были найдены данные о среднесуточных температурах, высоте снежного покрова и количестве осадков для восьми населенных пунктов, расположенных в северной части России. Рассматриваемый временной промежуток длиной в 83 года взят по причине доступности данных, так как измерения для взятых населенных пунктов начались с 1930-х годов. Требуется оценить степень изменений для наблюдаемых параметров в условиях глобального потепления, а также спрогнозировать дальнейшее поведение для каждого из параметров.

Парниковый эффект был обнаружен Жозефом Фурье в 1824 году и впервые был количественно исследован Сванте Аррениусом в 1896. *Выбросы парниковых газов* включают в себя газы с высокой прозрачностью в видимом диапазоне и с высоким поглощением в среднем и дальнем инфракрасном диапазонах [5], т.е. это водяные пары, метан, углекислый газ, озон, оксиды азота и тому подобные. Все эти отходы выделяются в процессах, в которых принимает участие человек, например, сжигание топлива, работа промышленных и энергетических комплексов, сельское хозяйство и другие.

Видимый солнечный свет, падающий на землю преобразуется в инфракрасное излучение, исходящее от Земли в космос. Парниковые газы затрудняют этот процесс, частично поглощая инфракрасное излучение и

удерживая уходящую в космос энергию в атмосфере, вызывая тем самым рост температуры у поверхности Земли.

Следующим источником влияния на глобальное потепление являются *вулканические выбросы*. В 60-х – 90-х годах наблюдалось глобальное затемнение – постепенное уменьшение количества солнечного света, достигающего поверхности Земли. Основной причиной данного явления являются частицы пыли, попадающие в атмосферу при вулканических выбросах и в результате деятельности производств. Затемнение дает охлаждающий эффект благодаря способности частиц отражать солнечный свет, поэтому углерод и аэрозоли, образующиеся после сжигания ископаемого топлива, уменьшали эффект потепления в данный период.

Выделяемые при извержении вулкана частицы сажи являются вторым по величине источником влияния на изменение температур. Характер влияния данного источника определяется положением частиц: в атмосфере или на поверхности Земли. При нахождении в атмосфере происходит поглощение солнечной радиации и нагрев воздуха, но поверхность, наоборот, охлаждается. Если же сажа выпадает на поверхность, в частности на ледники и снежный покров, то происходит, нагрев поверхности, который объясняется снижением ее альбедо.

Еще одной естественной причиной выступает *солнечная активность*. Под солнечной активностью подразумеваются совокупность доступных нашим наблюдениям изменений на Солнце, не считая самых мелких, не отражающихся на обычном состоянии светила. Еще во времена Галилея, в средние века, после первых наблюдений Солнца в телескоп, была высказана мысль о том, что солнечные пятна — это охладители и поэтому при увеличении количества пятен на Солнце температура на Земле должна падать. Однако тогда же было обнаружено, что если солнечные пятна и влияют на погоду и климат, то неодинаково на различных географических широтах. В одних районах при увеличении числа солнечных пятен становится теплее, в

других — холоднее [5]. В то же время при исследовании солнечной активности во времена ледникового периода один из соавторов исследования, доктор Реймунд Мушелер заявил: "Результатом снижения солнечной активности могут стать более холодные зимы в Северной Европе, потому что ультрафиолетовое излучение Солнца влияет на атмосферную циркуляцию. Интересно, что в результате тех же самых процессов зимы в Гренландии могут стать более теплыми, с большим количеством снегопадов и бурь. Так же исследование показывает, что различные солнечные процессы необходимо включать в климатические модели для того, чтобы более точно прогнозировать будущие изменения климата на глобальном и региональном уровне" [9].

Следующей причиной глобального потепления является орбитальное движение Земли вокруг Солнца.

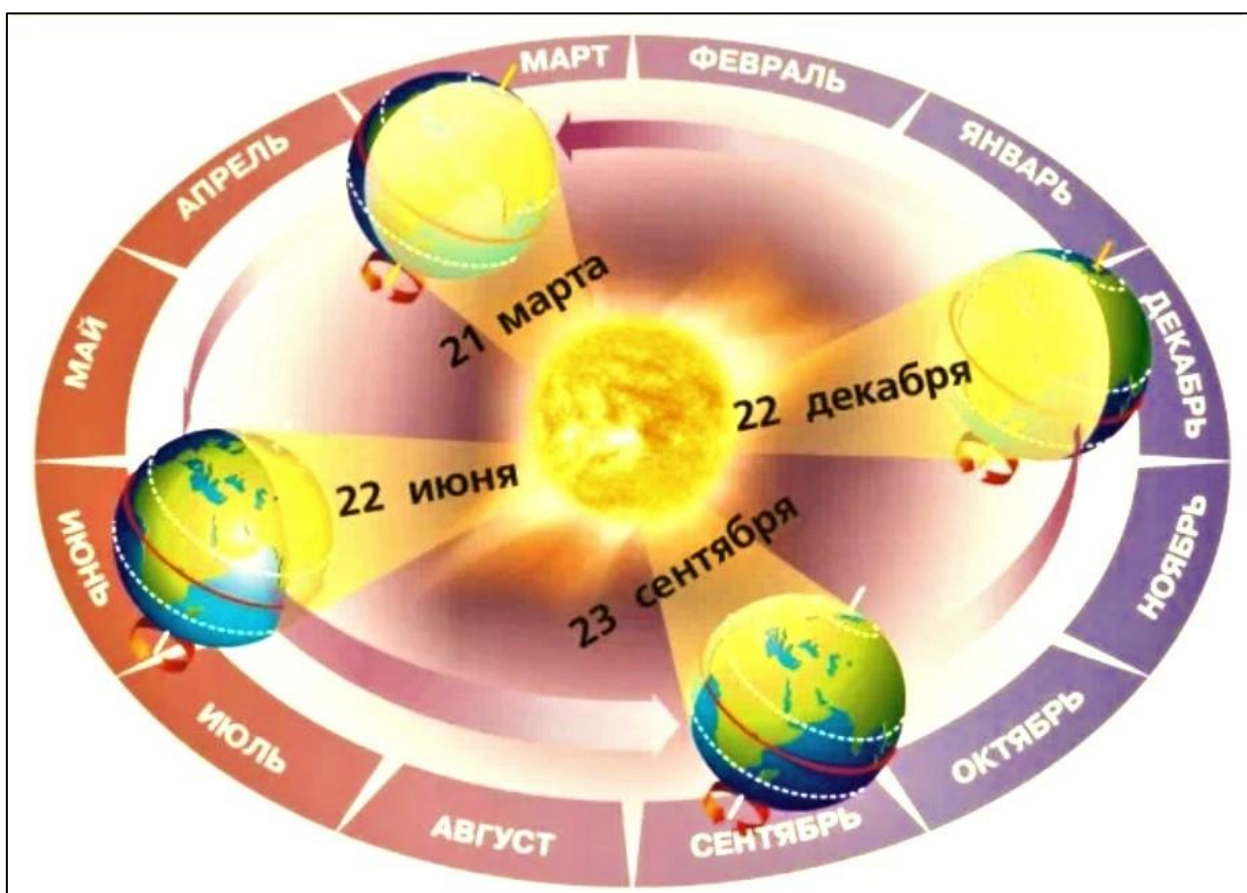


Рисунок 1. Цикл солнечной активности

Из-за наклона земной оси к плоскости орбиты при орбитальном движении Солнце освещает больше то Северное, то Южное полушарие планеты. Неравномерность освещения и нагрева земной поверхности вызывает смену сезонов года. Орбитальное движение определяет и изменение продолжительности светового дня — времени между восходом и заходом солнца. Сербский астроном Миланкович предположил, что циклические изменения климата во многом связаны с изменением орбиты вращения Земли вокруг Солнца, а также изменением угла наклона оси вращения Земли, по отношению к Солнцу. Подобные орбитальные изменения положения и движения планеты вызывают изменение радиационного баланса Земли, а значит и её климата. Миланкович, руководствуясь своей теорией, вполне точно рассчитал времена и протяжённость ледниковых периодов в прошлом нашей планеты. Но климатические изменения, вызванные изменением орбиты Земли, происходят обычно в течение десятков, а то и сотен тысяч лет. Наблюдаемое же в настоящий момент времени относительно быстрое изменение климата, по-видимому, происходит в результате действия ещё каких-то факторов.

Говоря о последствиях глобального потепления хочется отметить, что они все связаны друг с другом и каждое предыдущее способствует появлению следующего. Например, повышение температуры океана приводит к таянию арктического льда, которое в свою очередь вызывает увеличение уровня мирового океана, а, следовательно, и к частым возникновением чрезвычайных ситуаций.

Для сокращения темпов глобального потепления были приняты ряд соглашений, основными идеями которых были уменьшение количества выбросов, формирования эмиссионного бюджета углерода страны и переход к «зеленой» энергетике. Эмиссионный бюджет представляет собой допустимую суммарную антропогенную эмиссию углекислого газа за установленный период времени, определяемую исходя из целевого уровня глобального

потепления в конце периода [5]. Единственной страной, установившей национальный бюджет эмиссии CO₂, является Великобритания. Это делается для достижения законодательно установленной цели снижения национальной эмиссии к 2050 г. на 80 % по сравнению с 1990 г. Бюджет принимается на пятилетний период. Также был проведен ряд исследований, целью которых являлось определение бюджета CO₂, соответствующего потеплению 2°C. В каждом из исследований были различные временные рамки, статистические методы и набор учитываемых факторов. Подробнее ознакомиться с исследованиями можно в [5].

Возобновляемые источники энергии, как и не возобновляемые имеют свои преимущества и недостатки. В числе преимуществ, во-первых, неограниченность, так как восстановление происходит природным путем и за короткий период времени, во-вторых, количество выбросов в атмосферу при использовании стремится к минимуму и, в-третьих, доступность – добыча большинства из них производится легко. Среди недостатков можно выделить высокую стоимость, трудности транспортировки, а также ненадежность, так как некоторые источники зависят от погодных условий.

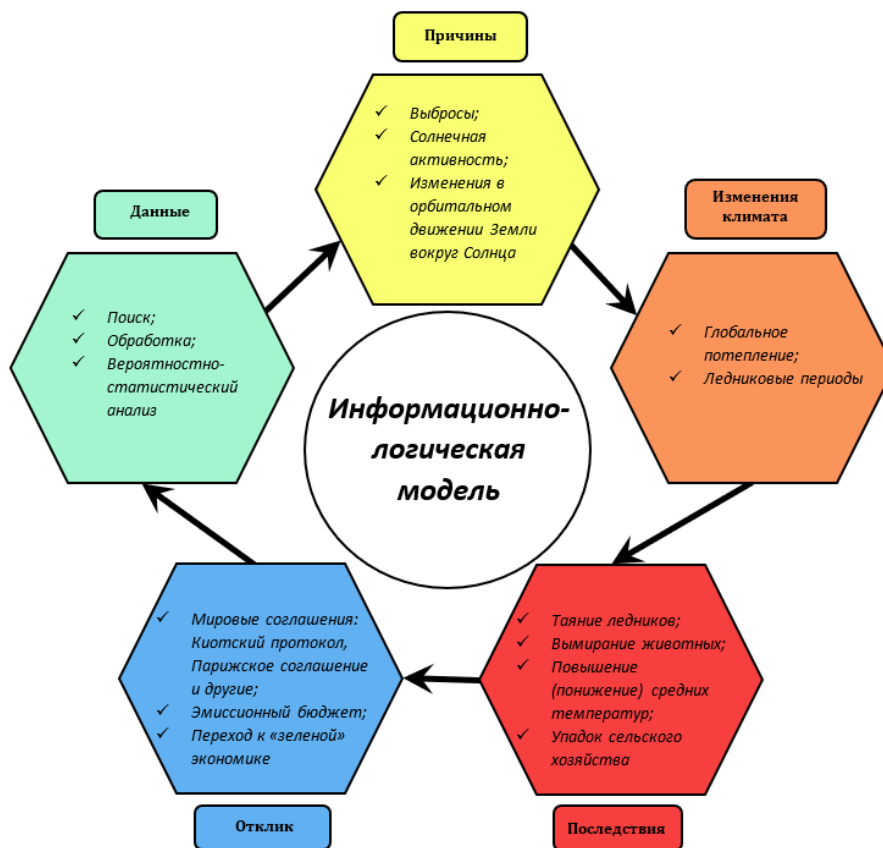


Рисунок 2. Информационно-логическая модель

1.1. Описание математической модели

Рассмотрим подробнее понятие временного ряда и методы его анализа. Временной ряд представляет собой совокупность упорядоченных во времени наблюдений некоторого параметра (параметров) в рамках исследуемого процесса. Анализ временного ряда, как правило, производится для получения декомпозиции ряда, то есть выделения тренда, сезонной и случайной составляющих для последующего детального изучения, а также для прогнозирования будущего поведения рассматриваемого процесса. Каждая из компонент временного ряда оказывает определенное влияние на его поведение. Так, например, сезонная компонента отвечает за циклические колебания ряда, а случайная отражает влияние стохастических факторов, объединение которых оказывает существенное влияние на поведение временного ряда. Для получения более точного результата прогнозирования необходимо построить правильную модель и провести ее оценку на

адекватность. Перед построением модели необходимо обработать данные путем замены отсутствующих значений на медианные, избавления от выбросов (шума) и проанализировать каждую из компонент декомпозиции временного ряда.

Далее, описывается математическая составляющая метода анализа сингулярного спектра. Рассматривается вещественнозначный временной ряд $F = (f_0, \dots, f_{N-1})$ длины N . Предполагается, что ряд F – ненулевой, т.е. существует, по крайней мере, одно i такое что $f_i \neq 0$. Обычно считается, что $f_i = f(i\Delta)$ для некоторой функции $f(t)$, где t – время, а Δ – некоторый временной интервал, однако это не будет играть особой роли в дальнейшем. Более того, числа $0, \dots, N - 1$ могут быть интерпретированы не только как дискретные моменты времени, но и как некоторые метки, имеющие линейно-упорядоченную структуру. Более подробное описание алгоритма можно найти в [6].

В данной работе используется базовый алгоритм метода, который состоит из двух дополняющих друг друга этапов: разложение и восстановление. Рассмотрим каждый этап подробнее.

1.1. Первый этап: разложение

Шаг 1: Вложение

В ходе выполнения процедуры вложения исходный временной ряд преобразуется в последовательность многомерных рядов.

Пусть L – некоторое целое число (длина окна), причем $1 < L < N$. Процедура вложения образует $K = N - L + 1$ векторов вложения

$$X_i = (f_{i-1}, \dots, f_{i+L-2})^T, \quad 1 \leq i \leq K,$$

имеющих размерность L .

L-траекторная матрица ряда F

$$X = [X_1: \dots: X_K]$$

состоит из векторов вложения в качестве столбцов. Другими словами, траекторная матрица – это матрица

$$X = (x_{ij})_{i,j=1}^{L,K} \begin{pmatrix} f_0 & f_1 & f_2 & \dots & f_{K-1} \\ f_1 & f_2 & f_3 & \dots & f_K \\ f_2 & f_3 & f_4 & \dots & f_{K+1} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ f_{L-1} & f_L & f_{L+1} & \dots & f_{N-1} \end{pmatrix} \quad (1)$$

Данная матрица является ганкелевой, так как X имеет одинаковые элементы на «диагоналях» $i + j = \text{const}$ и $x_{ij} = f_{i+j-2}$. Заметим, что существует взаимно-однозначное соответствие между ганкелевыми матрицами размерности $L \times K$ и рядами длины $N = L + K - 1$.

Шаг 2: Сингулярное разложение

Результатом данного шага является сингулярное разложение ($SVD = \text{Singular Value Decomposition}$) траекторной матрицы ряда.

Пусть $S = XX^T$. Пусть $\lambda_1, \dots, \lambda_L$ собственные числа матрицы S , взятые в неубывающем порядке ($\lambda_1 \geq \dots \geq \lambda_L \geq 0$) и U_1, \dots, U_L – ортонормированную систему собственных векторов матрицы S , соответствующих собственным числам. Пусть $d = \max\{i, \text{таких что } \lambda_i > 0\}$.

Если обозначить $V_i = \frac{X^T U_i}{\sqrt{\lambda_i}}, i = 1, \dots, d$, то сингулярное разложение матрицы X может быть записано как

$$X = X_1 + \dots + X_d \quad (2)$$

где $X_i = \sqrt{\lambda_i} U_i V_i^T$. Каждая из матриц X_i имеет ранг 1, поэтому их можно назвать элементарными матрицами. Набор $(\sqrt{\lambda_i} U_i V_i)$ – i -я собственная тройка сингулярного разложения (2).

1.2. Второй этап: восстановление

Шаг 3: Группировка

На основе разложения (2) процедура группировки делит все множество индексов $\{1, \dots, d\}$ на m непересекающихся подмножеств I_1, \dots, I_m .

Пусть $I = \{i_1, \dots, i_p\}$. Тогда результирующая матрица X_I соответствующая группе I , определяется как

$$X_I = X_{i_1} + \dots + X_{i_p}$$

Такие матрицы вычисляются для $I = I_1, \dots, I_m$, тем самым разложение (2) может быть записано в сгруппированном виде

$$X = X_{I_1} + \dots + X_{I_m} \quad (3)$$

Процедура выбора множеств I_1, \dots, I_m и называется группировкой собственных троек.

Шаг 4: Диагональное усреднение

На последнем шаге базового алгоритма каждая матрица сгруппированного разложения (3) переводится в новый ряд длины N .

Пусть $Y - L \times K$ матрица с элементами $y_{ij}, 1 \leq i \leq L, 1 \leq j \leq K$. Положим $L^* = \min(L, K), K^* = \max(L, K)$ и $N = L + K - 1$. Пусть $y_{ij}^* = y_{ij}$, если $L < K$, и $y_{ij}^* = y_{ji}$ иначе.

Диагональное усреднение переводит матрицу Y в ряд g_0, \dots, g_{N-1} по формуле

$$g_k = \begin{cases} \frac{1}{k+1} \sum_{m=1}^{k+1} y_{m, k-m+2}^* & \text{для } 0 \leq k < L^* - 1 \\ \frac{1}{L^*} \sum_{m=1}^{L^*} y_{m, k-m+2}^* & \text{для } L^* - 1 \leq k < K^*, \\ \frac{1}{N-k} \sum_{m=k-K^*+2}^{N-K^*+1} y_{m, k-m+2}^* & \text{для } K^* \leq k < N \end{cases} \quad (4)$$

Выражение (4) соответствует усреднению элементов матрицы вдоль «диагоналей» $i + j = k + 2$: выбор $k = 0$ дает $g_0 = y_{11}$, для $k = 1$ получаем $g_1 = \frac{(y_{12} + y_{21})}{2}$ и т.д. Заметим, что если матрица Y является траекторной матрицей некоторого ряда (h_0, \dots, h_{N-1}) (другими словами, если матрица Y является ганкелевой), то $g_i = h_i$ для всех i .

Применив диагональное усреднение (4) к результирующим матрицам X_{I_k} , мы получаем ряды $\tilde{F}^{(k)} = (\tilde{f}_0^{(k)}, \dots, \tilde{f}_{N-1}^{(k)})$ и, следовательно, исходный ряд (f_0, \dots, f_{N-1}) раскладывается в сумму m рядов:

$$f_n = \sum_{k=1}^m \tilde{f}_n^{(k)} \quad (5)$$

Глава 2. Анализ временных рядов

2.1. Первичная обработка и анализ данных

Данные для исследования были выгружены из специализированных массивов для климатических исследований от Всероссийского Научно-исследовательского института Гидрометеорологической Информации - Мировой центр Данных (ВНИИГИ-МЦД) [1]. Далее была произведена обработка данных, которая включала в себя:

1. Представление данных в удобном формате: значения всех метеорологических показателей были представлены в виде сводных таблиц в программном продукте MS Excel. Сводные таблицы позволяют быстро найти и представить данные в нужном разрезе.
2. Для некоторых городов часть значений отсутствовала и заполнение пропусков производилось за счет медианных значений для каждого месяца в отдельности. В Тикси было пропущено около 100 значений, поэтому данный населенный пункт был исключен из рассмотрения.

В силу большого объема данных для анализа были подсчитаны среднемесячные температуры. При первичном вероятностно-статистическом анализе были построены гистограммы с соответствующей линией тренда для случайно выбранных месяцев в каждом из сезонов. Результаты для каждого населенного пункта представлены ниже (см. рис. 3-8).

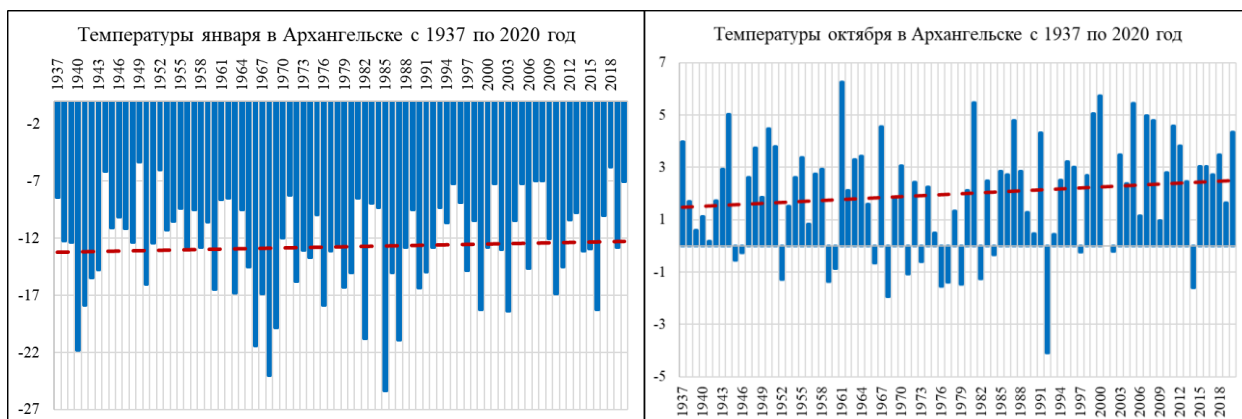


Рисунок 3. Гистограммы среднемесячных температур января и октября в Архангельске

На гистограммах пунктирной линией выделен тренд. Заметим, что на всех рисунках тренд является восходящим, но в период межсезонья среднее значение повышается значительно быстрее, то есть осенью и весной температура увеличилась 1,5-4 °С, когда как зимой и летом всего на 1 °С.

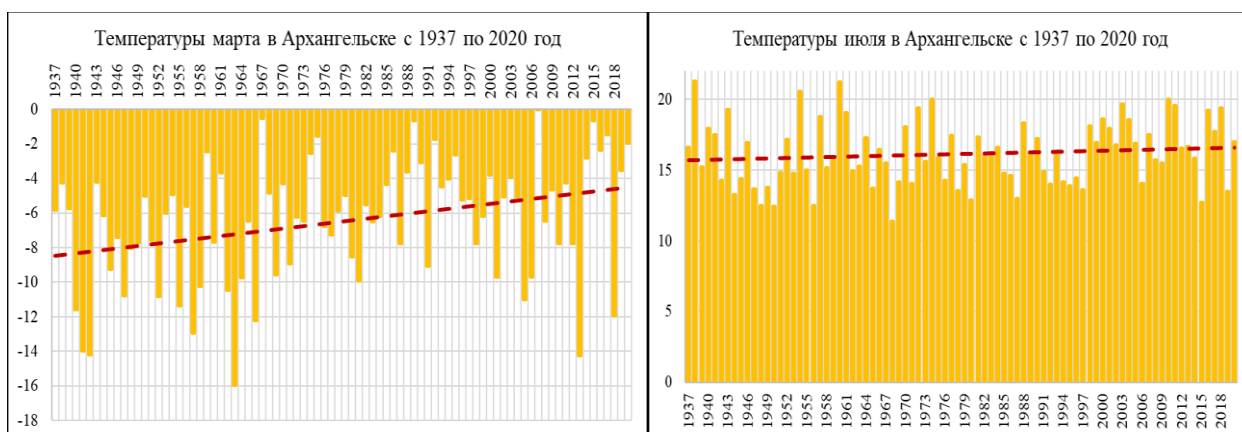


Рисунок 4. Гистограммы среднемесячных температур марта и июля в Архангельске

Для г. Верхоянска гистограммы представлены на рисунках 5 и 6.

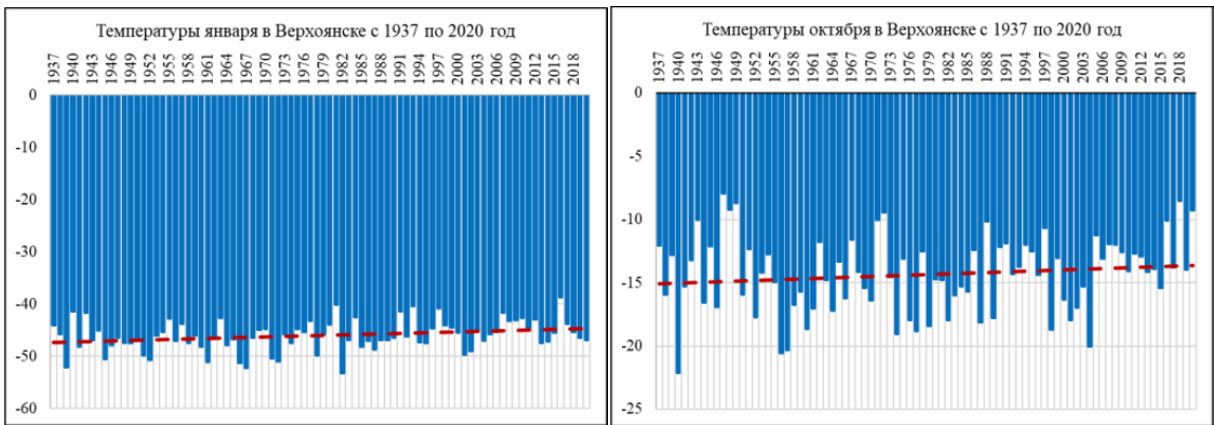


Рисунок 5. Гистограммы среднемесячных температур января октября в Верхоянске

Как зимой и летом, так и в межсезонье наблюдается восходящий тренд – незначительное повышение среднемесячных температур (1 °С). Также, заметим, что все чаще температурные значения превышают среднее, даже с учетом его возрастания.

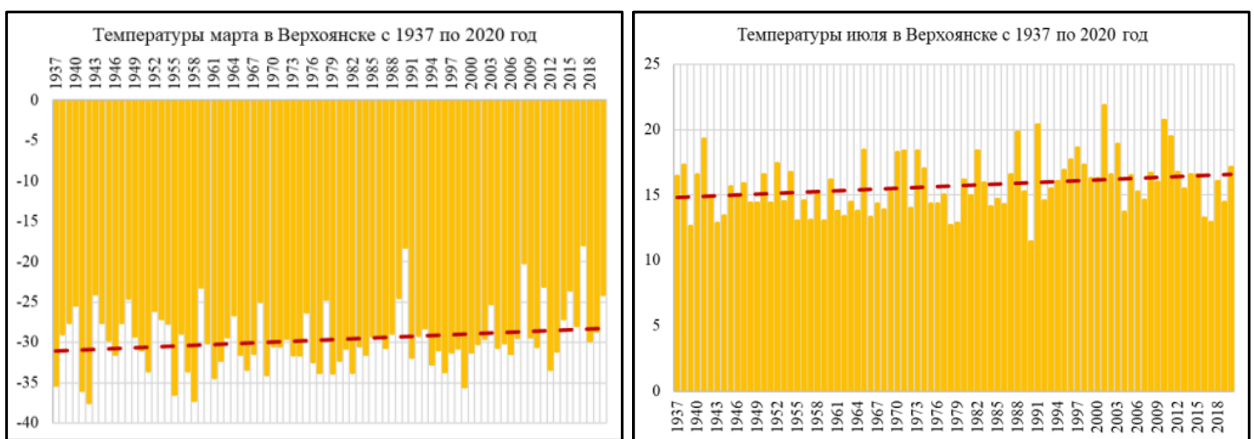


Рисунок 6. Гистограммы среднемесячных температур марта и июля в Верхоянске

В построенных гистограммах для Якутска (см. рис. 7-8) наблюдается значительное повышение средних температур зимой и весной на 7-8 °С.

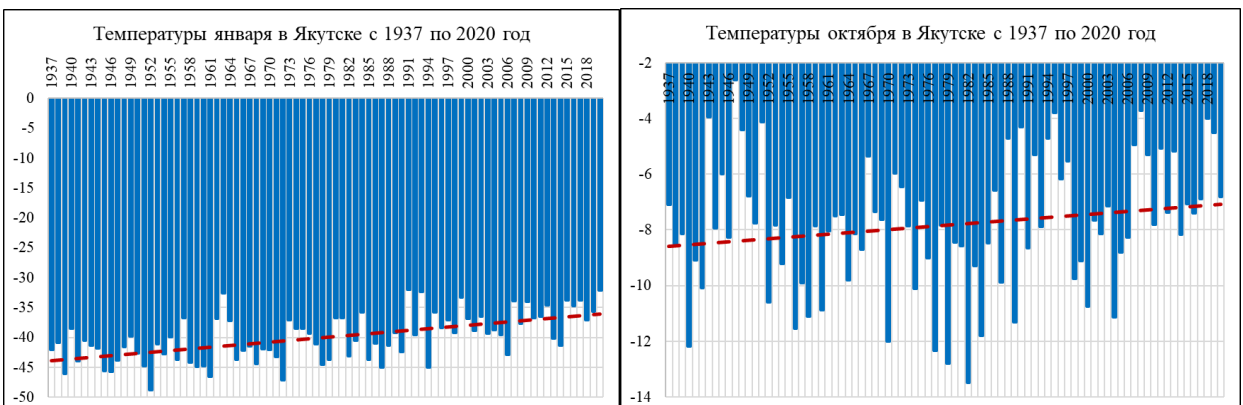


Рисунок 7. Гистограммы среднемесячных температур января и октября в Якутске

В то же время в октябре и июле температура повысилась всего на 0,9-1°С.

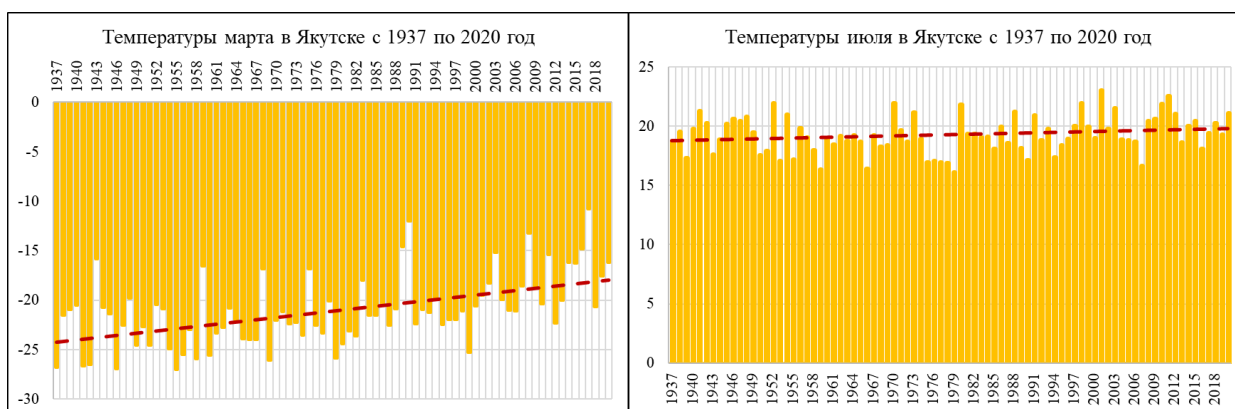


Рисунок 8. Гистограммы среднемесячных температур марта и июля в Якутске

При первичном вероятностно-статистическом анализе в разных городах изменение климата происходит по-разному, с разной интенсивностью, потому что климат зависит в большей степени от расположения города. С гистограммами для других населенных пунктов можно ознакомиться в Приложении № 1. Для более конкретных результатов необходимо воспользоваться основными статистическими методами анализа временных рядов.

Одним из этапов первичного анализа данных также является вычисление показателей динамики временного ряда. Для адекватности результатов было решено использовать среднегодовые температурные значения для каждого населенного пункта. Данные показатели позволяют оценить степень изменчивости временного ряда. В таблице № 1 приведены значения параметров динамики для г. Архангельска.

Таблица № 1. Показатели динамики временного ряда для г. Архангельска

Среднее значение температуры	1,380
Средний абсолютный прирост	0,009
Средний коэффициент роста	1,003
Средний темп роста	100,286
Средний темп прироста	0,286
Средняя величина абсолютного значения 1 % прироста	0,033

Среднее значение температуры за весь период наблюдений составляет 1,38°C. Значение абсолютного прироста, равно 0,009°C, то есть среднегодовые температуры различаются примерно на 0,009 градуса. Коэффициент роста характеризует во сколько раз различаются значения, наблюдаемого и базисного периодов. Заметим, что с каждым годом среднегодовая температура повышалась в среднем на 0,286%, о чем говорит средний темп прироста.

В таблице № 2 приведены значения показателей для г. Верхоянска.

Таблица № 2. Показатели динамики временного ряда для г. Верхоянска

Среднее значение температуры	-14,734
Средний абсолютный прирост	0,057
Средний коэффициент роста	0,995
Средний темп роста	99,494
Средний темп прироста	-0,506
Средняя величина абсолютного значения 1 % прироста	-0,113

Среднее значение температуры за весь период наблюдений составляет -14,734°C, что значительно ниже среднегодовой в Архангельске. Значение абсолютного прироста, составляет 0,057°C, что говорит о ежегодном повышении температуры на 0,057°C. Коэффициент роста характеризует во сколько раз различаются значения, наблюдаемого и базисного периодов. Также, отметим, что в среднем температура каждого года составляла 99,494%

от температуры предыдущего года. Средний темп прироста сигнализирует о ежегодном понижении температуры в среднем на 0,506%.

Значения показателей динамики для Якутска приведены в таблице № 3.

Таблица № 3. Показатели динамики временного ряда для г. Якутска

Среднее значение температуры	-9,184
Средний абсолютный прирост	0,090
Средний коэффициент роста	0,985
Средний темп роста	98,469
Средний темп прироста	-1,531
Средняя величина абсолютного значения 1 % прироста	-0,058

В Якутске температура на всем исследуемом промежутке составила - 9,184°C, что выше, чем в Верхоянске, но по-прежнему значительно ниже Архангельска. Температура каждого следующего года составляла 98,469% среднегодовой температуры предыдущего. Средний темп прироста говорит о ежегодном понижении температуры в среднем на 1,531%. Средняя величина абсолютного значения 1% прироста является отношением абсолютного прироста к темпу прироста, выраженному в процентах. Данный параметр показывает значимость каждого процента прироста за один и тот же период времени [16].

Значения показателей интенсивности изменений для города Мурманска представлены в таблице № 4.

Таблица № 4. Показатели динамики временного ряда для г. Мурманска

Среднее значение температуры	0,539
Средний абсолютный прирост	0,010
Средний коэффициент роста	1,005
Средний темп роста	100,502
Средний темп прироста	0,502
Средняя величина абсолютного значения 1 % прироста	0,021

Среднегодовая температура в Мурманске $0,539^{\circ}\text{C}$, что является самым высоким температурным значением, в сравнении с предыдущими. В среднем разница между среднегодовыми температурами составляет 0,01 градус, а относительная величина прироста составляет 0,502 %, то есть ежегодно значение температуры увеличивается в среднем на 0,502 %.

Для Котласа, Салехарда и Хатанги также подсчитаны значения параметров изменчивости, которые представлены в таблицах в Приложении № 1.

Выводы:

1. Самая высокая средняя температура за рассматриваемый промежуток времени в Котласе ($1,825^{\circ}\text{C}$), а самая низкая в городе Верхоянске ($-14,734^{\circ}\text{C}$);
2. Только в 3 из 7 оставшихся исследуемых населенных пунктов наблюдается положительный средний темп прироста, причем максимальный в Котласе - 0,702%. В 4 оставшихся наблюдается отрицательный прирост, что говорит о ежегодном понижении температур. Максимальное понижение наблюдается в Салехарде и составляет $-1,782\%$.

2.2. Исследование временных рядов в R

Среда разработки *Rstudio*, работа в которой осуществляется с помощью языка программирования *R*, является одним из удобных и простых

инструментов для анализа временных рядов. В ней присутствуют множество библиотек, которые позволяют получить декомпозицию ряда, применить различные вариации сглаживания, провести множество тестов для проверки статистических гипотез и многое другое.

После загрузки среднемесячных температурных значений для каждого города в среду, наборы данных были преобразованы во временные ряды, графическое представление которых практически идентично, поэтому на рисунке 9 представлен график для одного из населенных пунктов – Архангельска.

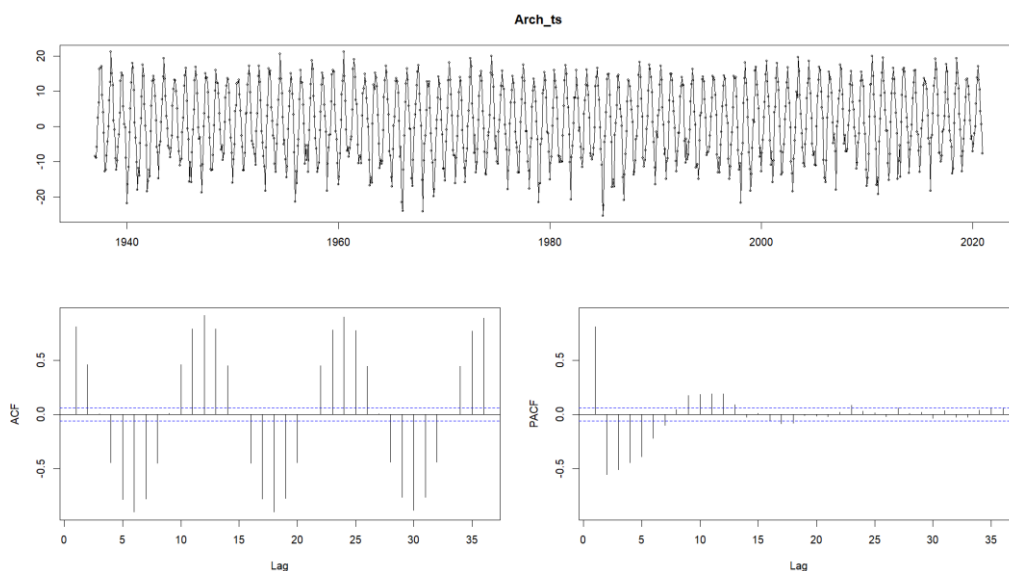


Рисунок 9. График среднемесячных температурных значений и автокорреляционных функций г. Архангельска

На графике по горизонтальной оси отмечены года, по вертикальной – соответствующее значение температур. Заметим, наличие явной сезонности и отсутствие тенденции к повышению. Также на рисунке построены графики автокорреляционных функций: черные столбцы являются значениями автокорреляционной функции, синие пунктирные линии – это ее доверительные интервалы. Данная функция показывает наличие зависимости между текущими наблюдениями и значениями ряда, которые находятся на расстоянии t интервалов (лагов). Говоря о значениях автокорреляционной функции, можно наблюдать как положительную, так и отрицательную

автокорреляцию, что является логичным. Максимумы АСФ функции приходятся на отметки в 12, 24 и 36 лагов, а минимумы на 6, 18 и 30 лагов, что является логичным, так как для исследования были взяты данные среднемесячных температур. Заметим, что температуры декабря, как и июля каждый год остаются примерно в одном и том же диапазоне значений, что и в предыдущем году. Данное замечание объясняется наличием сезонной компоненты временного ряда, которая далее будет рассмотрена подробнее.

Частная автокорреляционная функция показывает «чистую» зависимость между наблюдениями, то есть как изменится значение n элемента, если изменится значение $n - k$ (k – лаг) элемента, при этом значения остальных элементов останутся неизменными. По данной функции также можно увидеть влияние предыдущих значений показателя на текущие. Ознакомиться с графическим представлением для других временных рядов, в силу их идентичности, можно в Приложении № 1.

Для формулировки более конкретных выводов с помощью функции *decompose()* была получена декомпозиция ряда, представленная на рисунке 10. Разложение также имеет схожую визуализацию для каждого из рассматриваемых городов.

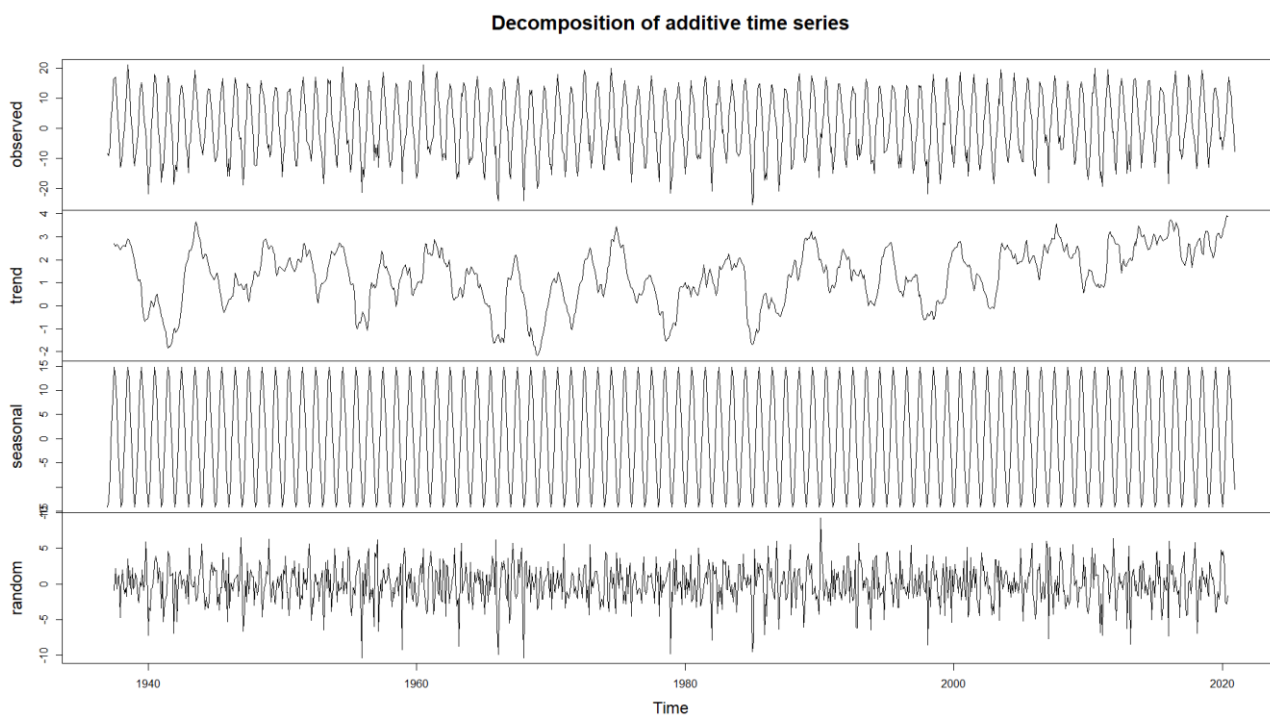


Рисунок 10. Декомпозиция временного ряда (Архангельск)

Декомпозиция представляет собой разложение временного ряда на компоненты. График расположенный сверху является графическим представлением исходного временного ряда. Расположенный под ним иллюстрирует наличие незначительного восходящего тренда, причем возрастание началось примерно с 1985-х годов, когда шел переход к постиндустриальной эпохе и наблюдались огромные скачки вперед во всех сферах общества. Данное замечание является одним из аргументов в пользу антропогенного влияния на изменение климата. Однако, говорить о наличии строгого повышения температур пока рано.

Третий график сверху показывает сезонность временного ряда. Наличие данной компоненты объясняется в основном зависимостью температурных значений от соответствующего времени года, а также расположением населенного пункта. Наконец, случайная составляющая (последний график на рисунке 10) представляет собой значения, получившиеся в результате воздействия случайных факторов на временной ряд.

Перед построением моделей, описывающих поведение ряда, необходимо провести расширенный тест Дики-Фуллера (ADF-тест) на стационарность. Его принцип заключается в следующем:

- Выдвигаются нулевая и альтернативная гипотезы

H_0 : Временной ряд является нестационарным.

H_1 : временной ряд является стационарным.

Далее выбирается уровень значимости (обычно $\alpha = 0,05$) и подсчитывается значения статистики *Dickey – Fuller* и *p – value*. Результаты ADF теста для температурных значений Архангельска приведены на рисунке 11.

```
Augmented Dickey-Fuller Test
data: Arch_ts
Dickey-Fuller = -8.031, Lag order = 10, p-value = 0.01
alternative hypothesis: stationary
```

Рисунок 11. Результаты ADF-теста (Архангельск)

Значение *p – value*, меньше выбранного уровня значимости 0,05, следовательно, необходимо отвергнуть гипотезу H_0 в пользу альтернативной H_1 .

Вывод: для всех рассматриваемых населенных пунктов значение *p – value* < 0,05, а, следовательно, временной ряд является стационарным, то есть значения среднего и дисперсии в любом из рассматриваемых периодов будут постоянными, а, следовательно, тренд отсутствует.

Так как в работе преследуется цель спрогнозировать поведение процесса глобального потепления в будущем, то необходимо выбрать наилучшую модель, объясняющую явление.

Одной из универсальных моделей является модель ARMA и ее улучшенные версии (ARIMA, SARIMA, SARIMAX), позволяющие учесть тренд, сезонную компоненту, а также случайную составляющую. Для наблюдаемого

временного ряда целесообразно использовать модель $SARIMAX(p, d, q)(P, D, Q)_m$

где p – параметр несезонной составляющей $AR(p)$;

q – параметр несезонной компоненты $MA(q)$;

d – порядок разности, такой как процесс $ARIMA$ становится $ARMA(p, q)$;

P – параметр сезонной составляющей $AR(p)$;

Q – параметр сезонной компоненты $MA(q)$;

D – порядок разности сезонной составляющей;

m – сезонная частота параметра [19].

После подбора параметров для данных по Архангельску была получена модель $ARIMA(1,0,0)(2,1,0)$ [12]. Для каждого города получились разные параметры модели, поэтому каждая в отдельности будет проверена на адекватность. В случае пригодности модели можно будет использовать ее в качестве прогнозной.

Первым шагом необходимо проверить остатки: их распределение и автокорреляцию. Гистограмма распределения остатков представлена на рисунке 12.

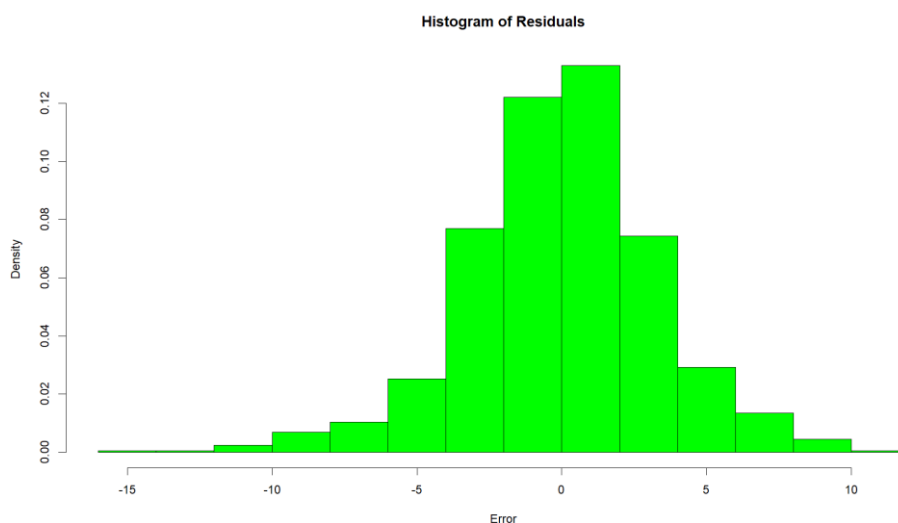


Рисунок 12. Гистограмма распределения остатков (Архангельск)

Как видно из рисунка распределение остатков построенной модели является близким к нормальному. Проверим нормальность остатков с помощью соответствующих тестов. Результаты представлены на рисунке 13.

	Name	Statistic	p-value
1	Lilliefors (Kolmogorov-Smirnov) normality test	0.0399555	6.637187e-04
2	Anderson-Darling normality test	3.0731113	1.030693e-07
3	Cramer-von Mises normality test	0.5059096	2.415967e-06
4	Shapiro-wilk normality test	0.9863080	4.194775e-08
5	Pearson chi-square normality test	66.1587302	9.978536e-05

Рисунок 13. Результаты тестов на нормальность остатков (Архангельск)

Значение p – *value* меньше 0,05 во всех тестах. Заметим, что проверка на нормальное распределение остатков проводилась среди всех остаточных значений модели в совокупности, но SARIMA использует разные параметры для каждого месяца, следовательно, целесообразно посмотреть распределение остатков каждого конкретного месяца (см. рис. 14-15)

Проверка на нормальность для Jan			
	Name	Statistic	p-value
1	Lilliefors (Kolmogorov-Smirnov) normality test	0.05360037	0.7985148
2	Anderson-Darling normality test	0.27378725	0.6566651
3	Cramer-von Mises normality test	0.04221245	0.6361927
4	Shapiro-wilk normality test	0.98818034	0.6447970
5	Pearson chi-square normality test	8.85714286	0.4505644
[1] -----			
Проверка на нормальность для Feb			
	Name	Statistic	p-value
1	Lilliefors (Kolmogorov-Smirnov) normality test	0.07539217	0.2817068
2	Anderson-Darling normality test	0.34256552	0.4835277
3	Cramer-von Mises normality test	0.05292458	0.4646135
4	Shapiro-wilk normality test	0.98434647	0.4028713
5	Pearson chi-square normality test	8.28571429	0.5056293
[1] -----			
Проверка на нормальность для Mar			
	Name	Statistic	p-value
1	Lilliefors (Kolmogorov-Smirnov) normality test	0.08067201	0.1948843
2	Anderson-Darling normality test	0.47722761	0.2314549
3	Cramer-von Mises normality test	0.08724792	0.1644317
4	Shapiro-wilk normality test	0.98592471	0.4950429
5	Pearson chi-square normality test	9.14285714	0.4241932
[1] -----			
Проверка на нормальность для Apr			
	Name	Statistic	p-value
1	Lilliefors (Kolmogorov-Smirnov) normality test	0.05950029	0.6552681
2	Anderson-Darling normality test	0.42574542	0.3086039
3	Cramer-von Mises normality test	0.06382421	0.3331326
4	Shapiro-wilk normality test	0.98088383	0.2466768
5	Pearson chi-square normality test	10.00000000	0.3504852

Рисунок 14. Результаты тестов на нормальность остатков для каждого месяца по отдельности (Архангельск)

Как видно из результатов выполнения программы значение $p - value$ значительно больше 0,05 для первых 4 месяцев.

Проверка на нормальность для May				
		Name	Statistic	p-value
1	Lilliefors (Kolmogorov-Smirnov)	normality test	0.1000180	0.03717352
2	Anderson-Darling	normality test	0.8775640	0.02356359
3	Cramer-von Mises	normality test	0.1599820	0.01712243
4	Shapiro-Wilk	normality test	0.9739487	0.08611725
5	Pearson chi-square	normality test	14.0000000	0.12232523
[1] -----				
Проверка на нормальность для Jun				
		Name	Statistic	p-value
1	Lilliefors (Kolmogorov-Smirnov)	normality test	0.05525945	0.7611380
2	Anderson-Darling	normality test	0.33894689	0.4928488
3	Cramer-von Mises	normality test	0.04702849	0.5508115
4	Shapiro-Wilk	normality test	0.98580757	0.4877805
5	Pearson chi-square	normality test	12.85714286	0.1691777
[1] -----				
Проверка на нормальность для Jul				
		Name	Statistic	p-value
1	Lilliefors (Kolmogorov-Smirnov)	normality test	0.05892264	0.6700729
2	Anderson-Darling	normality test	0.33769173	0.4961194
3	Cramer-von Mises	normality test	0.04865268	0.5253864
4	Shapiro-Wilk	normality test	0.98558553	0.4741898
5	Pearson chi-square	normality test	7.14285714	0.6222487
[1] -----				
Проверка на нормальность для Aug				
		Name	Statistic	p-value
1	Lilliefors (Kolmogorov-Smirnov)	normality test	0.1134080	0.009489557
2	Anderson-Darling	normality test	0.8982751	0.020927554
3	Cramer-von Mises	normality test	0.1446788	0.027302555
4	Shapiro-Wilk	normality test	0.9672000	0.030660019
5	Pearson chi-square	normality test	13.4285714	0.144153206

Рисунок 15. Результаты тестов на нормальность остатков для каждого месяца по отдельности (Архангельск)

Результаты тестов для месяцев с Мая по Август чуть хуже, в некоторых случаях значение $p - value$ меньше 0,05, но в большинстве месяцев оно превышает данное значение, что говорит о нормальном распределении остаточных значений, а значит согласуется с наблюдаемой выше гистограммой.

Проверка на автокорреляцию остатков осуществляется с помощью теста Дарбина-Уотсона. Сформулируем нулевую гипотезу об отсутствии автокорреляции и ей альтернативную

$$1) H_0: d = 0$$

$$H_1: d \neq 0$$

Далее было вычислено значение статистики Дарбина-Уотсона

$$DW = 2,01$$

На рисунке 16 представлены границы для принятия и отклонения нулевой гипотезы, а также определения характера автокорреляции при ее наличии.



Рисунок 16. Области принятия (отклонения) нулевой гипотезы

Выводы:

1. Значение статистики попадает в область принятия нулевой гипотезы, следовательно, автокорреляция остатков в модели отсутствует;
2. Распределение остатков является нормальным;
3. Модель является адекватной и ее можно использовать в качестве прогнозной.

Функция *forecast()* позволяет получить прогнозные значения, а также доверительные интервалы с разными уровнями значимости на заданный период времени. Так как период наблюдений до 2020 года, то прогноз осуществлялся на следующие 2 года. Результаты приведены на рисунке 17.

	Point Forecast	Lo 80	Hi 80	Lo 95	Hi 95
Jan 2021	-8.7367557	-13.050480	-4.4230314	-15.3340276	-2.1394837
Feb 2021	-8.5050252	-12.887377	-4.1226732	-15.2072541	-1.8027962
Mar 2021	-5.8718709	-10.256406	-1.4873357	-12.5774388	0.8336970
Apr 2021	0.7816170	-3.602988	5.1662222	-5.9240580	7.4872920
May 2021	7.9935542	3.608947	12.3781617	1.2878758	14.6992326
Jun 2021	13.0840043	8.699397	17.4686118	6.3783258	19.7896828
Jul 2021	16.8581058	12.473498	21.2427133	10.1524273	23.5637843
Aug 2021	13.4290597	9.044452	17.8136672	6.7233812	20.1347382
Sep 2021	10.1720142	5.787407	14.5566217	3.4663357	16.8776927
Oct 2021	3.2896279	-1.094980	7.6742354	-3.4160506	9.9953064
Nov 2021	-1.1327179	-5.517325	3.2518897	-7.8383964	5.5729607
Dec 2021	-6.5241792	-10.908787	-2.1395717	-13.2298577	0.1814993
Jan 2022	-9.6400384	-14.309318	-4.9707586	-16.7810855	-2.4989912
Feb 2022	-6.9971352	-11.675258	-2.3190123	-14.1517068	0.1574364
Mar 2022	-3.9569114	-8.635318	0.7214948	-11.1119163	3.1980935
Apr 2022	0.8632777	-3.815138	5.5416931	-6.2917411	8.0182965
May 2022	7.6994459	3.021030	12.3778616	0.5444267	14.8544652
Jun 2022	13.3748944	8.696479	18.0533101	6.2198752	20.5299137
Jul 2022	15.7680281	11.089612	20.4464437	8.6130088	22.9230473
Aug 2022	12.6847417	8.006326	17.3631574	5.5297225	19.8397610
Sep 2022	9.8690104	5.190595	14.5474260	2.7139911	17.0240297
Oct 2022	3.0330944	-1.645321	7.7115100	-4.1219249	10.1881137
Nov 2022	-1.4840961	-6.162512	3.1943196	-8.6391153	5.6709232
Dec 2022	-5.5559159	-10.234332	-0.8775003	-12.7109352	1.5991033

Рисунок 17. Прогнозные значения на 2021 и 2022 года (Архангельск)

Соответствующий график исходного временного ряда и прогнозных значений приведены на рисунке 18.

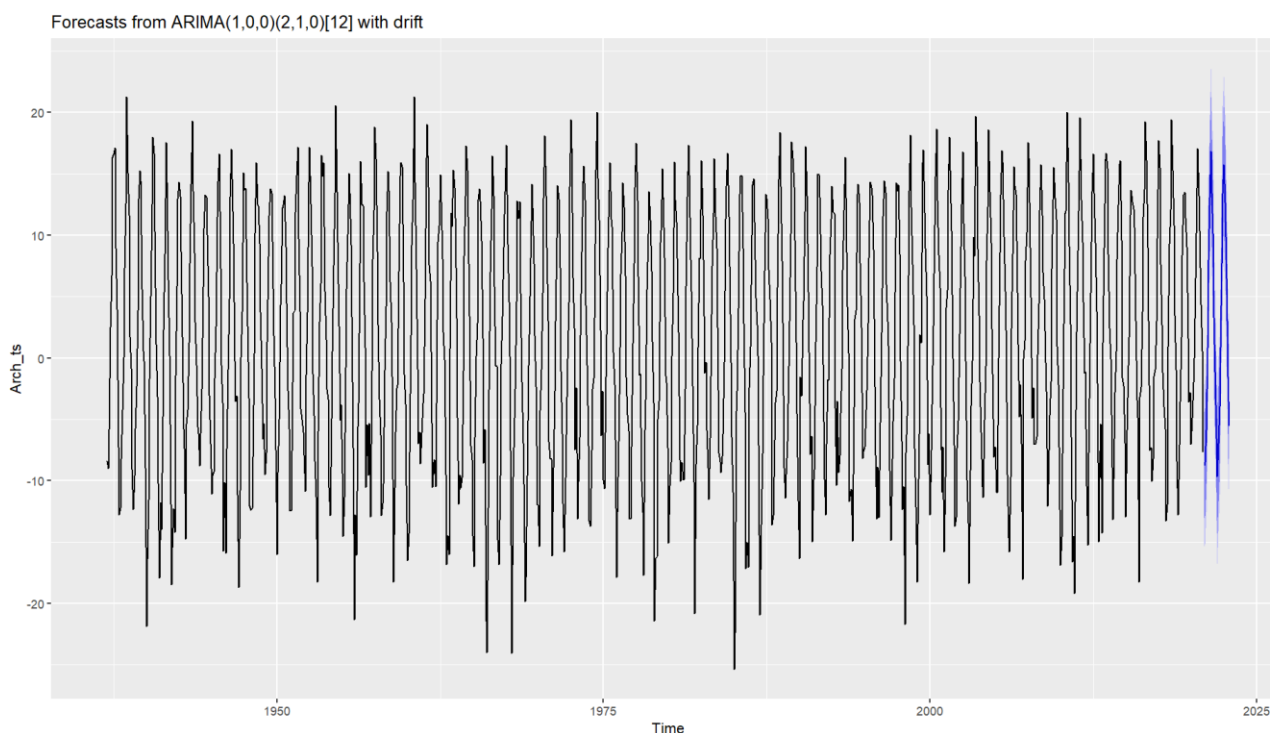


Рисунок 18. График полученных прогнозных значений (Архангельск)

Следующим шагом необходимо проверить точность модели и точность прогноза, используя функции из библиотеки *MLmetrics*, которая возвращает значения показателей MAPE, RMSE и других (см. табл. №5).

Точность модели также определялась и с помощью написанной функции, которая вычисляла разницу между фактическими и прогнозными значениями, а затем, высчитывала отношение модуля найденной разницы к длине массива прогнозных значений. Также для работы функции необходимо выбрать отклонение, которое считать допустимым. В данном случае отклонение в 1,5 градуса и меньше считалось несущественным. Точность модели составила 87,5%, что говорит о хорошем качестве прогноза.

Выводы:

1. Получена декомпозиция ряда и проинтерпретирована каждая его компонента (тренд, сезонность и шум);
2. Построена адекватная модель *ARIMA*, которую можно использовать в качестве прогнозной;
3. Получены прогнозные значения, для 2021 и 2022 годов, а точность модели 87,5 %.

Для каждого из рассматриваемых городов были получены декомпозиция временного ряда, построена модель *ARIMA* с различными значениями параметров. Для каждой из моделей была проведена проверка на адекватность, то есть соответствие остатков нормальному распределению, а также отсутствие их автокорреляции. В случае пригодности модели были рассчитаны прогнозные значения с доверительными интервалами в 80% и 95% уровнями значимости. После была произведена оценка качества модели и точности прогноза используя различные показатели: *RMSE*, *MAPE* и другие. Все результаты представлены в виде таблицы 5.

Таблица №5. Результаты анализа временных рядов в *R*

Название населенного пункта	Стац-ость	Параметры <i>ARIMA</i>	Адекватность модели		Качество модели	Точность прогноза
			Норм-ость	Авток-ция		

<i>Архангельск</i>	+	<i>ARIMA</i> (1,0,0) (2,1,0)[12]	+	-	<i>RMSE</i> = 1,024 <i>MAPE</i> = 0,16	83,3%
<i>Верхоянск</i>	+	<i>ARIMA</i> (2,0,2) (2,1,0)[12]	+	-	<i>RMSE</i> = 1,35 <i>MAPE</i> = 0,29	70,8%
<i>Мурманск</i>	+	<i>ARIMA</i> (0,0,1) (1,1,0)[12]	+	-	<i>RMSE</i> = 0,95 <i>MAPE</i> = 0,08	91,6%
<i>Котлас</i>	+	<i>ARIMA</i> (1,0,1) (2,1,0)[12]	+	-	<i>RMSE</i> = 1,413 <i>MAPE</i> = 0,177	83,3%
<i>Салехард</i>	+	<i>ARIMA</i> (0,0,2) (1,1,0)[12]	+	-	<i>RMSE</i> = 2,24 <i>MAPE</i> = 0,58	41,6%
<i>Хатанга</i>	+	<i>ARIMA</i> (1,0,0) (2,1,0)[12]	+	-	<i>RMSE</i> = 3,71 <i>MAPE</i> = 0,54	45,8%
<i>Якутск</i>	+	<i>ARIMA</i> (1,0,2) (2,1,0)[12]	+	-	<i>RMSE</i> = 1,66 <i>MAPE</i> = 0,29	70,8%

В силу идентичности графиков и рисунков, визуализации исходного временного ряда и его декомпозиции представлены в Приложении № 1.

Каждый из рассматриваемых рядов является стационарным, следовательно, тенденция к повышению отсутствует. После проверки рядов на стационарность были построены модели *ARIMA* с подобранными параметрами. Заметим, что вторая тройка параметров модели, отвечающая за сезонную компоненту для 5 из 7 городов, является одинаковой, исключение составляют Мурманск и Салехард. В то же время первые тройки, описывающие трендовую составляющую практически у всех населенных пунктов разные.

Для прогнозирования дальнейшего поведения температурных значений необходимо проверить условия адекватности модели, а именно:

- остатки модели подчиняются нормальному закону распределения;
- отсутствует автокорреляция остатков.

Соответствие остатков нормальному распределению проверялось с помощью тестов Колмогорова-Смирнова, Шапиро-Уилка, Пирсона и других популярных тестов для каждого конкретного месяца в отдельности, в силу

причин, описанных ранее. Автокорреляция остатков была проверена с помощью теста Дарбина-Уотсона, который имеет различные области принятия решений для каждого диапазона значений.

Так как все модели прошли проверку на адекватность (см. табл. №5), то следующим шагом были получены прогнозные значения, с которыми можно ознакомиться в таблице 6.

Таблица №6. Прогнозные значения для каждого города на 2021 и 2022

Название города/месяца	Архангельск		Верхоянск		Мурманск		Котлас		Салехард		Хатанга		Якутск	
	2021	2022	2021	2022	2021	2022	2021	2022	2021	2022	2021	2022	2021	2022
Январь	-10,89	-10,9	-44,46	-45,94	-10,58	9,66	-8,09	-8,34	-20,89	-19,13	-25,08	-27,88	-35,36	-34,12
Февраль	-9,53	-9,54	-40,05	-40,25	-7,56	-6,7	-8,33	-6,9	-17,91	-15,16	-23,73	-22,93	-30,44	-29,42
Март	-4,61	-4,62	-27,1	-26,46	-4,46	-4,16	-4,71	-2,7	-8,83	-8,18	-23,23	-20,82	-18,07	-17,09
Апрель	2,03	2,02	-6,22	-6,27	0,39	-0,29	2,25	2,34	-4,45	-3,64	-12,33	-11,5	-1,69	-1,49
Май	8,55	8,54	4,57	4,78	4,82	4,91	10,55	10,74	3,21	4,27	-3,64	-2,97	9,35	9,26
Июнь	14,88	14,88	17,75	18,36	9,68	10,18	14,17	14,55	9,06	9,48	12,54	12,01	18,64	19,07
Июль	17,98	17,97	15,96	15,94	13,29	14,44	17,95	17,22	16,2	15,89	11,41	12,12	20,18	20,33
Август	15,72	15,71	11,85	10,99	11,23	11,47	14,02	13,19	12,59	13,05	13,76	14,32	15,69	15,14
Сентябрь	10,14	10,13	3,55	4	8,47	8,54	9,83	9,51	7,5	8,06	4,9	5,47	7,44	8,04
Октябрь	3,81	3,8	-10,64	-11,17	1,14	2,04	3,69	3,72	-1,61	-1,53	-6,86	-7,85	-4,93	-5,45
Ноябрь	-2,33	-2,34	-31,91	-32,16	-3,02	-1,48	-2,88	-3,02	-11,47	-9,83	-20,33	-18,61	-22,32	-21,67
Декабрь	-7,26	-7,27	-42,42	-41,7	-4,95	-5,17	-7,43	-6,17	-17,45	-17,72	-20,9	-19,78	-36,86	-38,02

Говоря о точности модели можно выделить два основных параметра: *RMSE* и *MAPE*. *RMSE* показывает значение среднеквадратической ошибки, то есть разницу между фактическими и прогнозируемыми значениями. Чем меньше значение данного показателя, тем выше точность модели. Максимальное отклонение для построенных моделей ARIMA составляет 3,7 градусов, а минимальное – 0,95°C. *MAPE* – средняя абсолютная ошибка, выраженная в процентах. В целом, значения, данного показателя являются довольно хорошими.

Каждая модель построила прогнозные значения (см. табл. № 6) и одним из показателей качества модели является точность прогноза, которая высчитывается как отношение модуля разности прогнозных и фактических значений к длине массива прогнозных значений. Также необходимо было выбрать какое значение отклонения в градусах считать несущественным. В нашем случае длина массива составляет 24 месяца, а разница в 1,5 градуса и меньше считалась несущественной. Как видно из таблицы 5 не у всех моделей точность прогноза является высокой. Это обусловлено отсутствием значений

у нескольких городов в исходных данных, которые при первичной обработке были заменены медианными значениями.

Использование классических инструментов для анализа временных рядов в *Rstudio* является одним из универсальных способов исследования данных, но у них имеется ряд недостатков:

- *ARIMA* предполагает, что данные временных рядов стационарны, что не всегда встречается в реальных процессах;
- Для эффективной работы *ARIMA* требуется значительный объем данных;
- Данный вид модели является одномерной моделью прогнозирования, поэтому она может плохо работать с данными из нескольких временных рядов;
- *ARIMA* не отражает нелинейные тенденции в данных.

Альтернативными моделями, описывающими поведение временных рядов, являются экспоненциальное сглаживание, сезонная декомпозиция временных рядов, вейвлет-анализ и другие. Далее будет рассмотрен один из часто используемых методов в исследованиях временных рядов – метод анализа сингулярного спектра (*SSA*), позволяющий проанализировать поведение каждой компоненты в отдельности. А затем, отобрав наиболее информативные из них, спрогнозировать поведение явления в будущем.

Глава 3. Реализация метода SSA в R

Метод анализа сингулярного спектра был разработан Голяндиной Ниной Эдуардовной, Некруткиным Владимиром Викторовичем и Жиглявским Анатолием Александровичем в 1990-е годы и сейчас является одним из часто используемых в исследованиях.

Singular spectrum analysis предоставляет мощный инструмент для анализа данных временных рядов, особенно в случаях, когда данные зашумлены, нестационарны и сложны. Его способность извлекать значимую информацию

из различных временных рядов сделала его популярным выбором в различных областях. В науке о климате *SSA* используется для анализа метеорологических показателей, а также для обнаружения и прогнозирования таких климатических явлений, как Эль-Ниньо и Ла-Нинья. В финансовом сегменте его применяют для анализа поведения цен на акции и прогнозирования рынка в целом. В биомедицинской обработке сигналов *SSA* используется для анализа данных электрокардиограммы (ЭКГ) и электроэнцефалограммы (ЭЭГ), а также для обнаружения отклонений в сигналах.

Говоря о преимуществах метода *SSA* отметим следующие:

- Является непараметрическим методом, то есть он не предполагает использование конкретной модели или распределения временных рядов, что приводит к более точным результатам;
- Обработка нелинейных и нестационарных временных рядов, что делает его универсальным инструментом для различных приложений;
- Сжатие данных: *SSA* может уменьшить размерность временного ряда, извлекая его наиболее важные компоненты. Это позволяет более эффективно хранить и анализировать большие наборы данных временных рядов;
- Шумоподавление: *SSA* может эффективно разделять компоненты сигнала и шума во временном ряду. Разбивая временной ряд на составляющие, *SSA* устраняет влияние шума и усиливает компоненты сигнала;
- Прогнозирование: *SSA* можно использовать для прогнозирования будущих значений временного ряда. Экстраполируя компоненты сигнала, *SSA* может обеспечить точные прогнозы будущих тенденций и закономерностей;
- Другие

Таким образом, *SSA* является универсальным и эффективным методом анализа временных рядов с рядом преимуществ, включая уменьшение шума, сжатие данных, извлечение тенденций и сезонных колебаний, обнаружение аномалий и прогнозирование.

Данный метод можно реализовать в различных средах разработки, но в настоящей работе будут использованы *Rstudio* и язык программирования *Python*. В *Rstudio* разработан специальный пакет - *Rssa*, который позволяет применить *SSA* к любому временному ряду.

3.1. Обзор функций и библиотек

Первым шагом реализации был импорт необходимых для реализации библиотек:

- *readxl* – библиотека предназначена для чтения файлов формата *.xls* и *.xlsx*;
- *stringr* – пакет, предоставляющий большой набор функций для упрощения работы со строковыми элементами;
- *dplyr* – пакет, помогающий осуществить качественную обработку данных;
- *scales* – пакет, отвечающий за масштабирование, является частью инфраструктуры *ggplot2* [22];
- *Cairo* – графическая библиотека, которая используется для обеспечения высококачественного вывода изображений в различных форматах, включая растровые (*PNG, JPEG, TIFF*), векторные (*PDF, PostScript, SVG*) и экранные (*X11, Windows*) [20].
- *ggplot2* – это система декларативного создания графики, основанная на *The Grammar of Graphics* [22].
- *Rssa* – библиотека, обеспечивающая реализацию *SSA*, имеющая множество инструментов для анализа временных рядов и визуализации результатов.

Функция, представленная на листинге 1, позволяет найти и подгрузить файлы с заданным именем, находящиеся в той же папке что и файл с программным кодом.

Листинг № 1

```
wd <- getwd()
Arch_path <- str_c(wd, "/Arch.xlsx")
Arch <- read_xlsx(Arch_path)
```

Листинги с оставшимся программным кодом представлены в Приложении № 2.

3.2. Обсуждение полученных результатов

Для каждого из исследуемых городов была проведена программная реализация метода SSA. Далее будут представлены результаты для одного из населенных пунктов, так как результаты практически идентичны. Ознакомится с результатами для других городов можно в приложении № 1.

Ко всем загруженным данным был применен метод анализа сингулярного спектра с помощью функции `ssa()`. Аргументами функции в данном случае являются температурные значения и параметр L , представляющий собой длину окна «гусеницы». Подробнее об аргументах этой и других возможных функций входящих в пакет *Rssa* можно познакомиться в книге *Singular Spectrum Analysis with R* под авторством Нины Эдуардовны Голяндиной, Антона Ивановича Коробейникова и Анатолия Александровича Жиглявского. Книга служит универсальным учебником по пакету *Rssa*, в котором описана актуальная методология и основные алгоритмы SSA, а также области его приложения с соответствующими примерами [25].

С помощью функции `reconstruct()` была получена декомпозиция временного ряда для Салехарда.

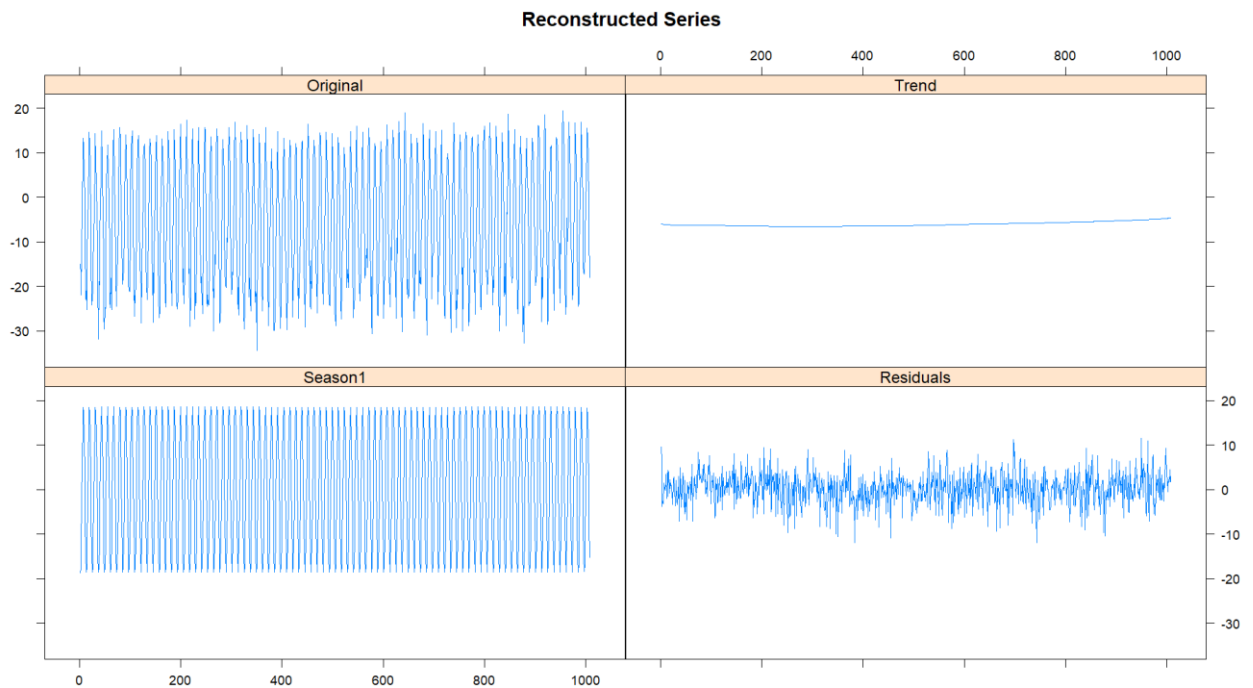


Рисунок 19. Декомпозиция, полученная с помощью *SSA* в *RStudio* (Салехард)

Данный результат согласуется с разложением, полученным при анализе временных рядов с помощью простых методов. Также наблюдается сезонность, обусловленная сменами времен года и незначительно восходящий тренд.

Далее были получены графические представления собственных векторов, а также 2D-диаграммы рассеяния для пар собственных векторов (см. рис. 20-21) с характерной долей описания временного ряда.

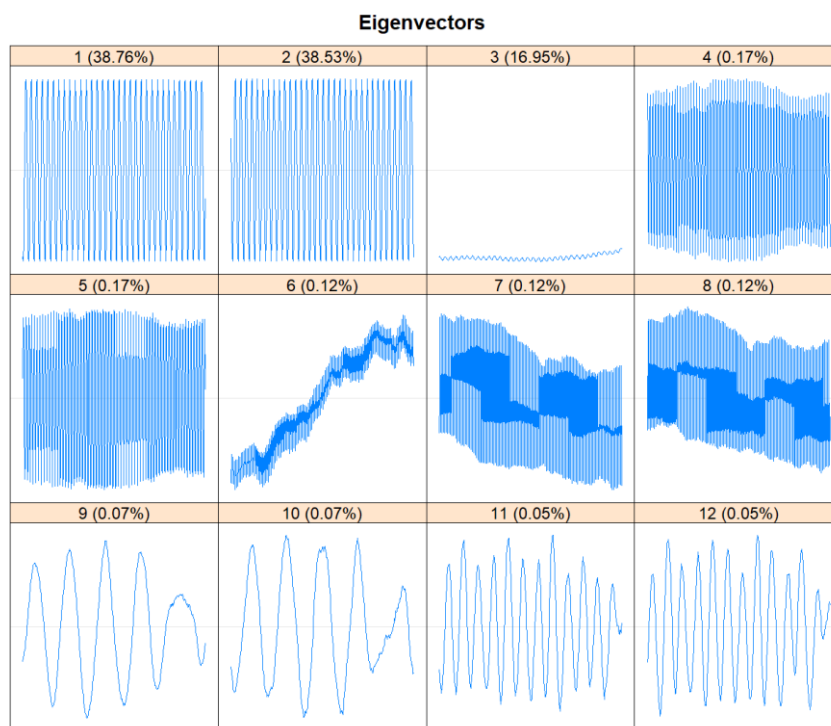


Рисунок 20. Первые 12 собственных векторов (Салехард)

Интересными для изучения представляются первые 3 собственных вектора: 1 и 2 описывают сезонную составляющую, а 3 медленно возрастает и его можно отнести к трендовой компоненте.

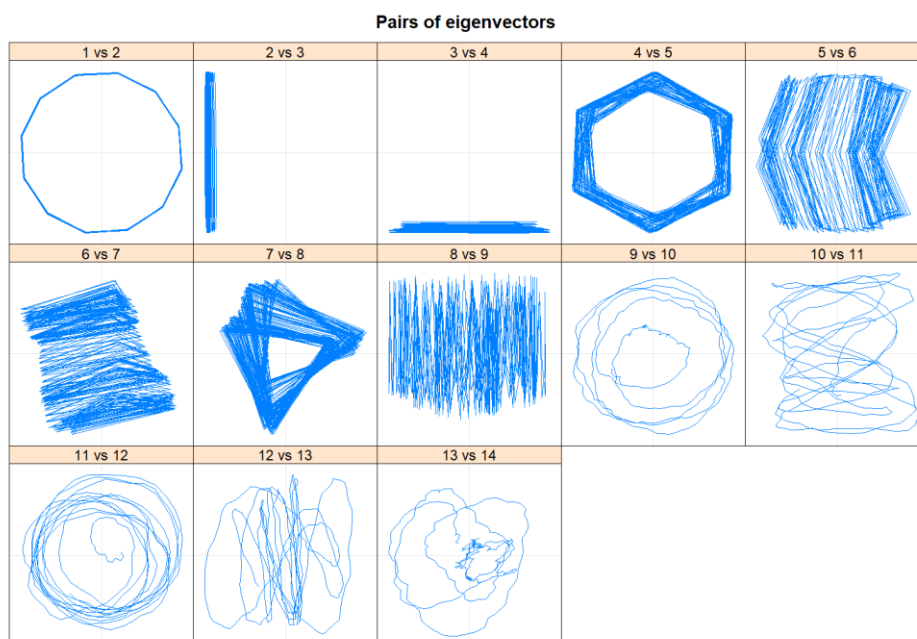


Рисунок 21. 2-D диаграммы рассеяния (Салехард)

На рис. 21 показано, что пары 1–2, 4–5 и 7–8 создаются модулированными синусоидами, поскольку соответствующие 2D-диаграммы рассеяния

собственных векторов напоминают правильные многоугольники. Этот способ идентификации основан на следующих свойствах: синусоидальная волна имеет ранг 2 и порождает две собственные тройки, которые представляют собой синусоидальные волны с одинаковой частотой и имеют фазовый сдвиг точно или приблизительно равный $\pi/2$ из-за ортогональности собственных векторов [25].

Отметим, что для каждого из исследуемых городов разные пары собственных векторов образуют правильные многоугольники на 2D-диаграммах рассеяния. Так, например, для села Хатанга (см.рис.22) правильные многоугольники образованы парами собственных векторов: 2-3, 4-5 и 6-7.

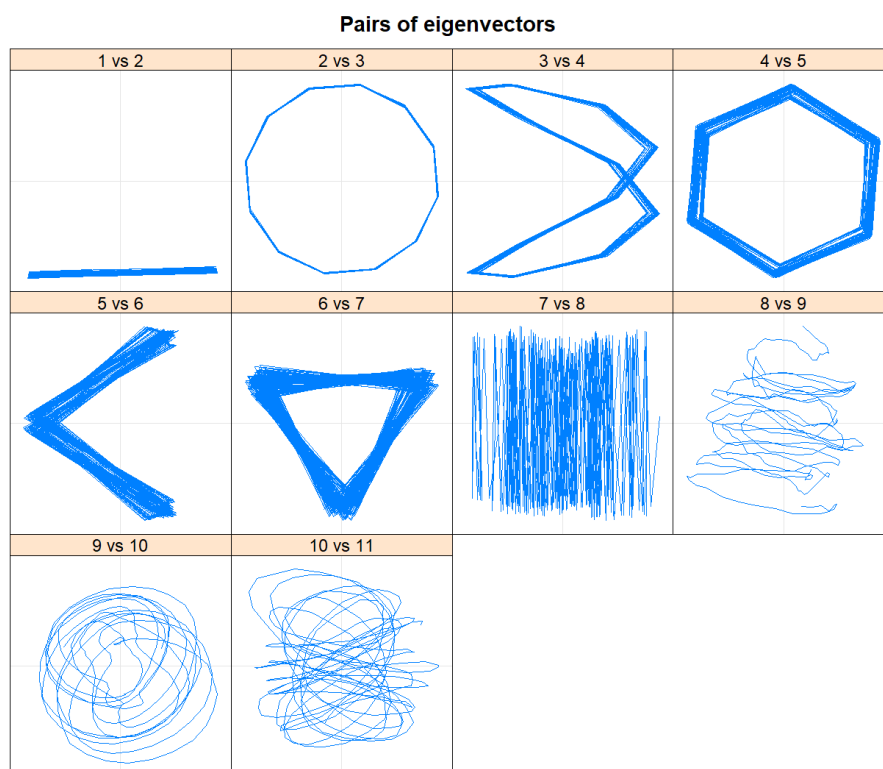


Рисунок 22. 2-D диаграммы рассеяния (Хатанга)

Следующим шагом была построена корреляционная матрица абсолютных значений в оттенках серого: белый цвет свидетельствует о нулевых абсолютных значениях, а черный об единичных. Участки, сочетающие в себе

все три цвета, говорят о смещении компонент и, скорее всего, образованы случайной составляющей (шумом).

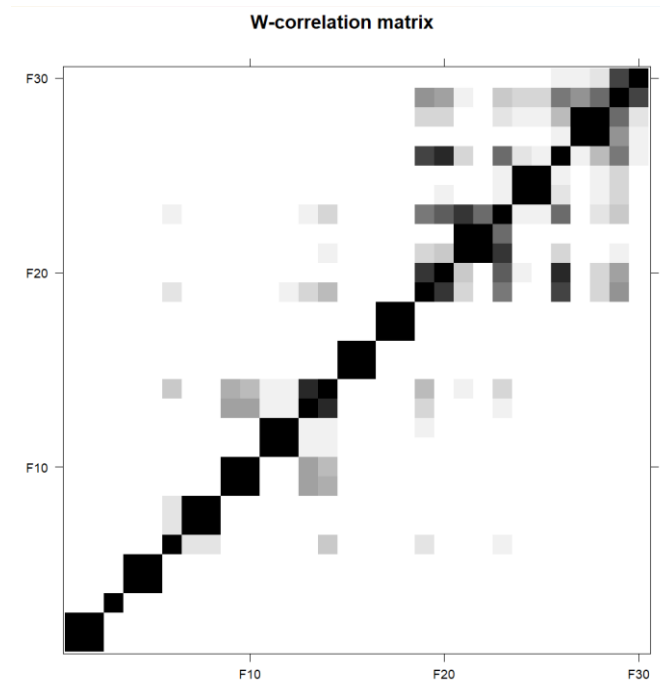


Рисунок 23. Пары собственных векторов (Салехард)

Если посчитать количество вершин в многоугольниках для пар собственных векторов, то можно определить период синусоид. Их также можно определить и с помощью функции *parestimate()*, ниже представлен ее вывод.

\$F1						
period	rate		Mod	Arg		Re
12.003	0.000000		1.00000	0.52		0.86608
\$F2						
period	rate		Mod	Arg		Re
6.009	0.000000		1.00000	1.05		0.50129
\$F3						
period	rate		Mod	Arg		Re
3.002	0.000000		1.00000	2.09		-0.49881
						Im
						0.49990
						0.86528
						0.86671

Рисунок 24. Периоды синусоид для выбранных пар собственных векторов (Салехард)

Как видно из рис. 24 результаты автоматического подсчета периодов и количество вершин в многоугольниках для выбранных пар собственных векторов согласуются.

На рис. 25 представлены 3 реконструированных синусоидальных сигнала.

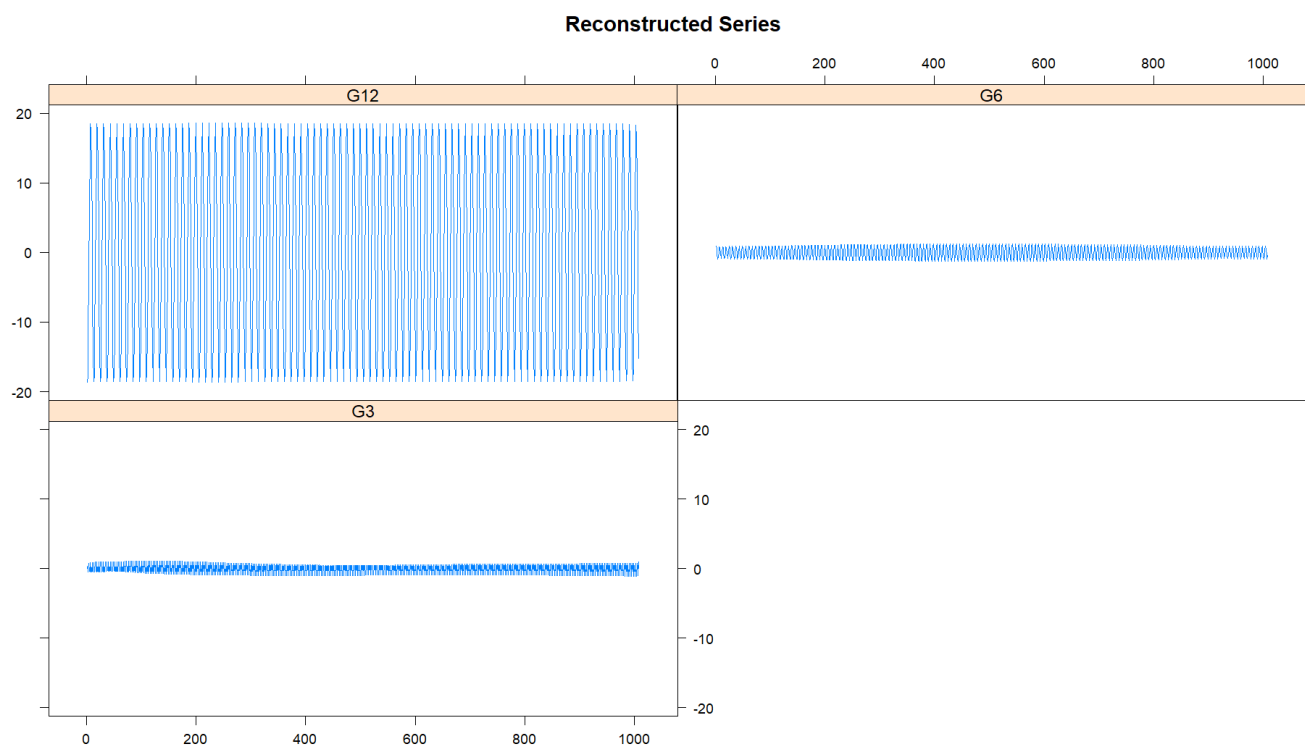


Рисунок 25. Реконструированные синусоидальные сигналы (Салехард)

Видно, что амплитуда колебаний то уменьшается, то возрастает. В силу большого количества значений данные изменения слабо выражены.

Заметим, что исследуемый набор температурных значений находится в промежутке с 1937 до 2020 года включительно, следовательно, оценить точность прогноза на 2021 и 2022 годы можно с помощью сравнения фактических значений и вычисленных прогнозных. Будущие значения получаются путем применения функции *forecast()* к полученным, в ходе применения метода анализа сингулярного спектра, данным. Прогнозные значения представлены в таблице № 7.

Таблица № 7. Прогнозные значения на 2021 и 2022 годы (Салехард)

Название города/месяца	Салехард	
	2021	2022
Январь	-22,28	-22,18
Февраль	-21,27	-21,15
Март	-13,77	-13,63
Апрель	-6,36	-6,23
Май	1,33	1,47
Июнь	10,88	11,01
Июль	14,55	14,65
Август	12,31	12,41
Сентябрь	7,57	7,68
Октябрь	-2,97	-2,87
Ноябрь	-14,11	-13,99
Декабрь	-19,24	-19,12

Точность прогноза вычислялась с помощью следующего программного кода (см. листинг 2).

Листинг № 2

```
sucesss_S=0
for(i in res_S){
  if(abs(i)<1.5){
    sucesss_S=sucesss_S+1
  }
}
print(sucesss_S/length(res_S))
```

Здесь

- *res_S* – массив прогнозных значений;
- *sucesss_S* – значение успеха, изначально равное 0. В ходе выполнения цикла оно будет увеличиваться при условии, что разница между фактическим и прогнозным значениями будет меньше 1,5 °С.

В конце выводится значение точности прогноза выраженное, как отношение числа успехов к общему числу будущих значений, то есть знаменатель будет 24. Точность прогноза составила 91,67%, в то время как при прогнозировании на основе простых методов анализа временных рядов точность составила всего 41,6%.

Ниже представлена сравнительная таблица точности прогнозных значений для простых методов анализа (ПМА) и метода SSA.

Таблица № 8. Сравнительная таблица точности прогнозных значений

Название города/месяца	Точность прогноза	
	ПМА	SSA
Архангельск	87,50%	91,60%
Верхоянск	70,80%	66,70%
Мурманск	87,50%	91,60%
Котлас	83,30%	95,80%
Салехард	37,50%	91,60%
Хатанга	45,80%	54,17%
Якутск	66,60%	83,30%

Как видно из сравнительной таблицы точность прогноза повысилась практически для всех исследуемых городов, за исключением Верхоянска. Данный факт может быть обусловлен значительным влиянием внешних факторов. В целом, можно сделать вывод, что метод SSA является более точным для исследования и прогнозирования временных рядов.

Глава 4. Реализация метода SSA в Python

4.1. Обзор функций

Первым шагом реализации был импорт необходимых для реализации библиотек:

- *numpy* - работа с многомерными массивами и матрицами
- *matplotlib.pyplot* - визуализация данных
- *pandas* - обработка и анализ структурированных данных
- *scipy.linalg* - решение СЛАУ, оптимизация, интеграция, интерполяция и обработка сигналов
- Ряд самостоятельно написанных функций для построения графиков и проведения различных тестов для проверки адекватности модели и точности прогноза.

4.2. Обсуждение полученных результатов

После задания всех необходимых параметров и заполнения матрицы температурными значениями была построена траекторная (Ганкелевая) матрица, представленная на рисунке 26 и найдены собственные числа с соответствующими им векторами.

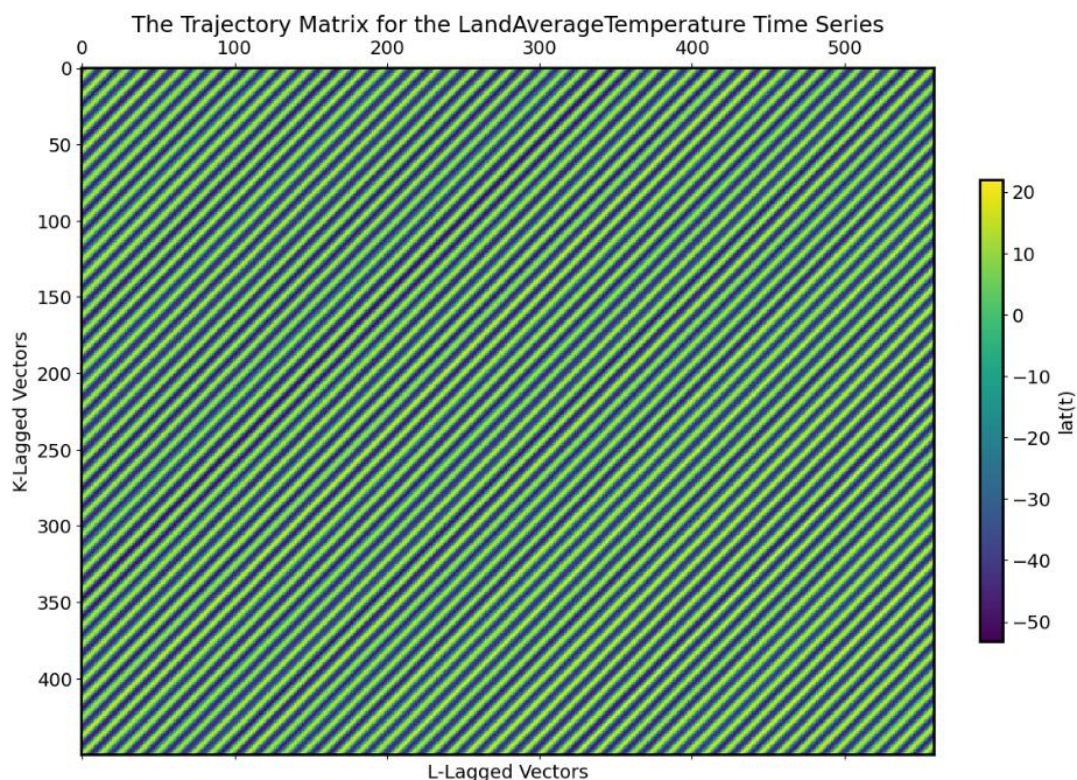


Рисунок 26. Траекторная матрица для временного ряда (Верхоянск)

В силу идентичности рисунков траекторных матриц для оставшихся населенных пунктов они представлены в приложении № 1. На рисунке 27 представлены 15 первых собственных векторов. Наиболее информативными являются 2, 9, 10, 11 и 14. Каждый из этих собственных векторов отвечает за определенные компоненты. Так, например 2 отвечает за трендовую составляющую, а 9 и 10 за сезонность.

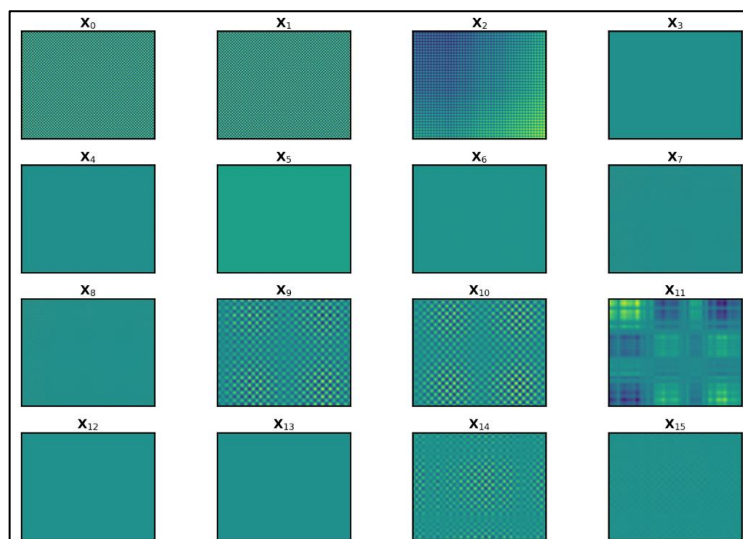


Рисунок 27. Визуализация собственных векторов (Верхоянск)

Для того, чтобы проверить правильность предположений высказанных выше построим собственные векторы на графике (см. рис. 28).

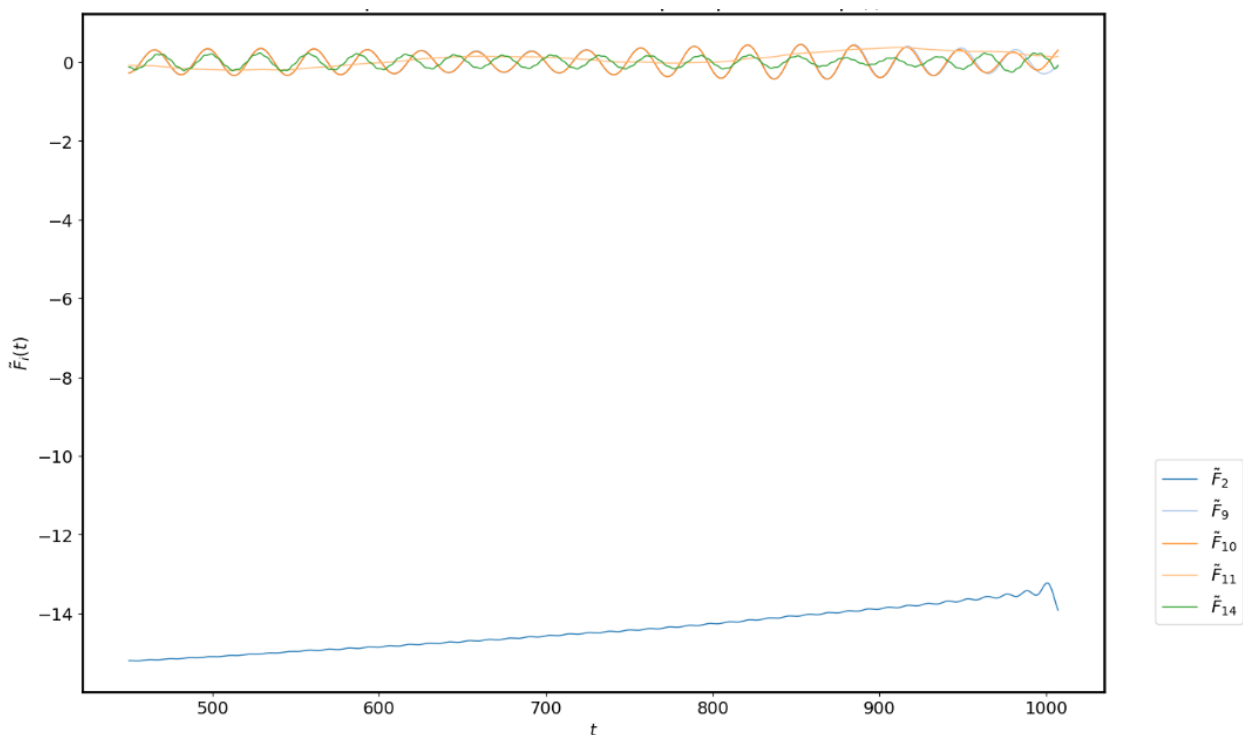


Рисунок 28. Графическое представление собственных векторов (Верхоянск)

Из рисунка 28 видно, что 2 собственный вектор и 11, вероятно, отвечают за тренд. Также наблюдается наличие 2 периодик: 9 и 10 векторы отвечают за первый вид периодической составляющей, а 14 за второй – менее выраженной. Остальные вектора можно определить как шум, случайную компоненту

временного ряда. Теперь, когда четко выделены составляющие ряда, построим их вместе с исходными данными (см. рис. 29). Здесь

LAT – исходный временной ряд;

trend1 и *trend2* – трендовые составляющие;

periodic1 и *periodic2* – сезонные компоненты ряда;

noise – шум или случайная составляющая временного ряда

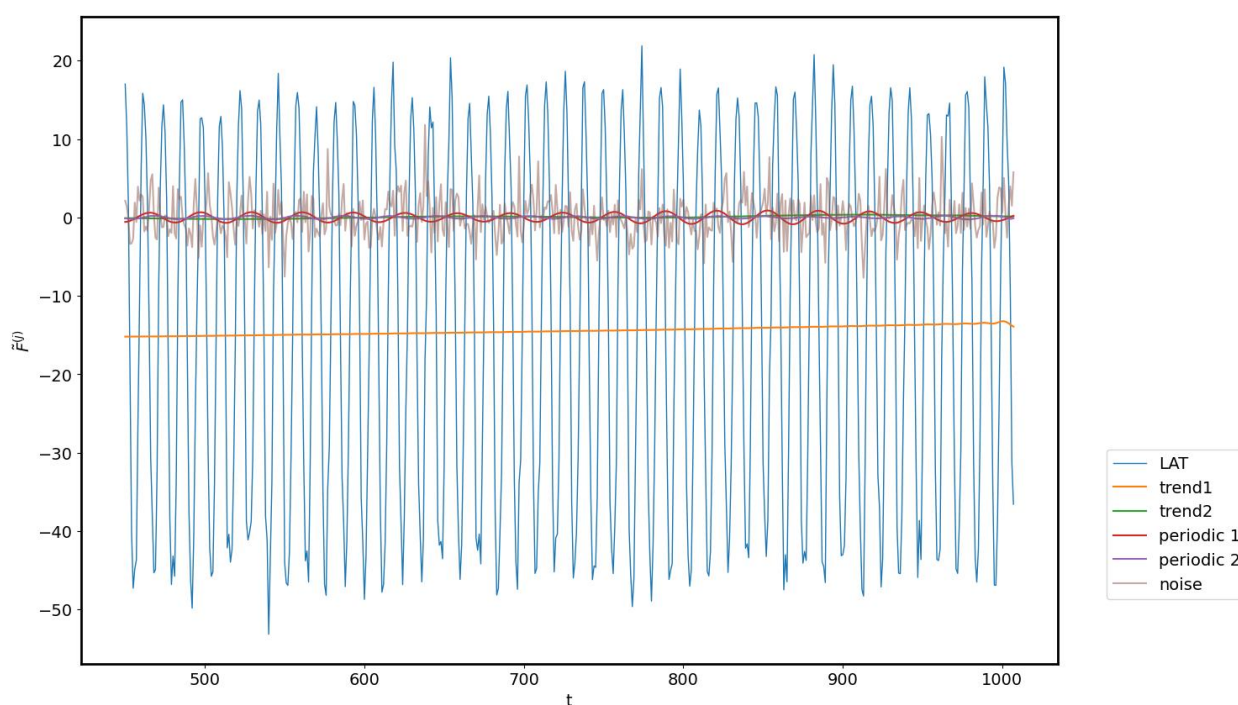


Рисунок 29. Графическое представление пар собственных векторов (Верхоянск)

Из рисунка 29 видно незначительное повышение температур и явную сезонность в данных, что согласуется с результатами, полученными при применении простых методов анализа временных рядов. Построим каждую компоненту по отдельности (см. рис. 30)

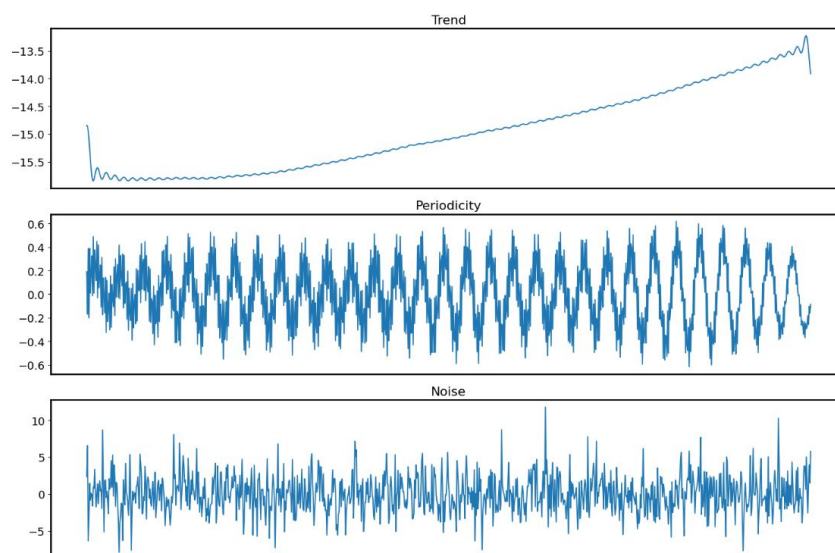


Рисунок 30. Декомпозиция временного ряда (Верхоянск)

У всех населенных пунктов присутствует периодичность в 12, в 6 и в 3 месяца, но при графическом представлении каждой из составляющих временного ряда сезонная компонента выглядит по-разному. Так, например, у Котласа (см. рис. 31) наблюдается уменьшение амплитуды колебаний с течением времени.

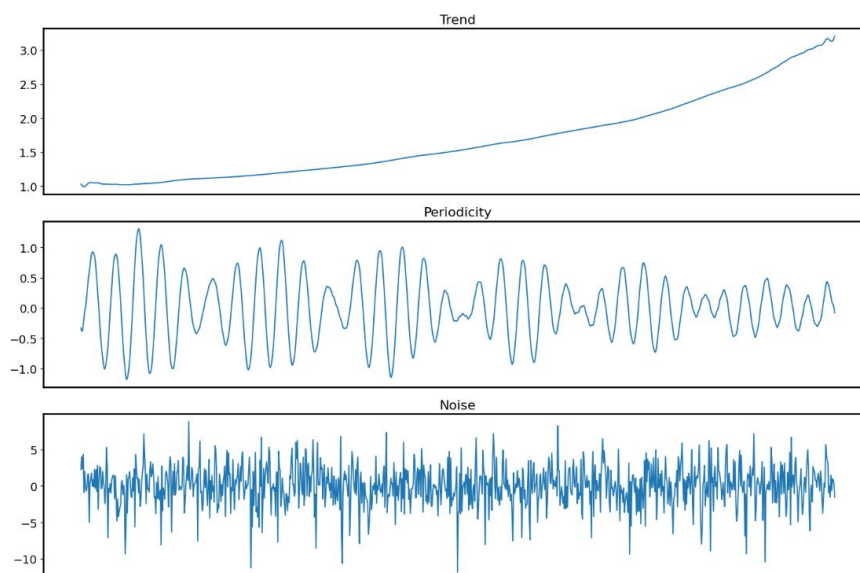


Рисунок 31. Декомпозиция временного ряда (Котлас)

В то время как у Салехарда (см. рис. 32) практически не наблюдается значительных изменений.

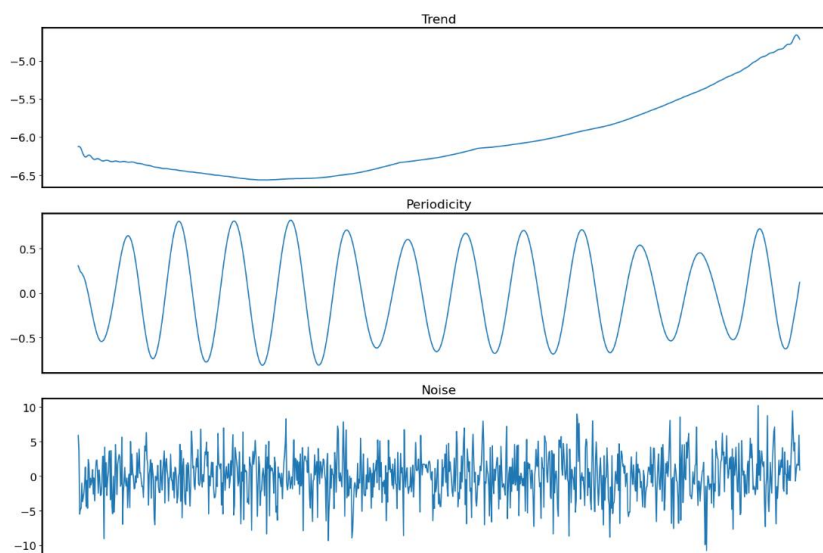


Рисунок 32. Декомпозиция временного ряда (Салехард)

Далее была построена корреляционная матрица, которая позволяет проверить насколько качественно разделены компоненты временного ряда, а цветовая шкала – определить соответствующие абсолютные значения.

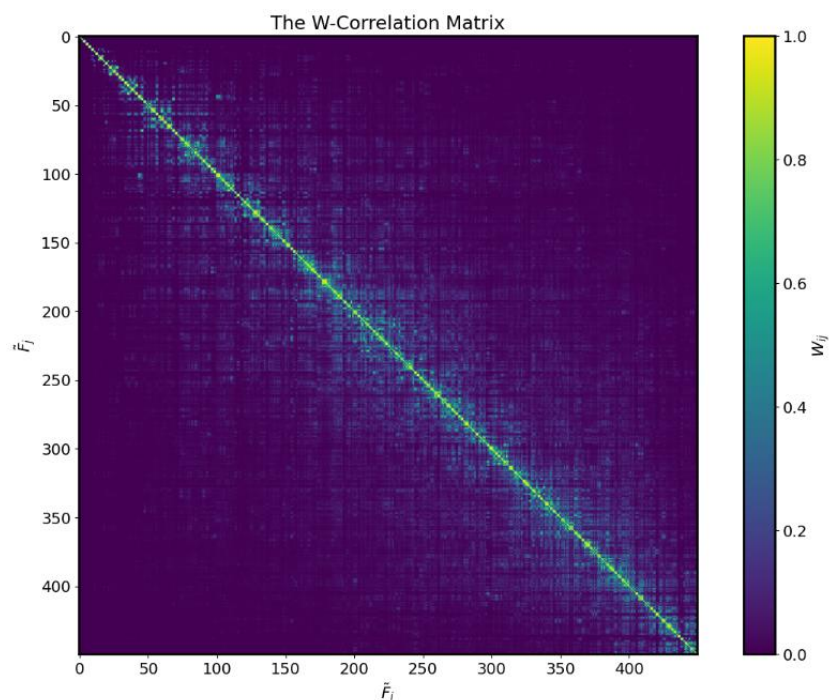


Рисунок 33. Корреляционная матрица временного ряда (Верхоянск)

Следующий график (см. рис. 34) иллюстрирует насколько отобранные компоненты описывают исходный временной ряд.

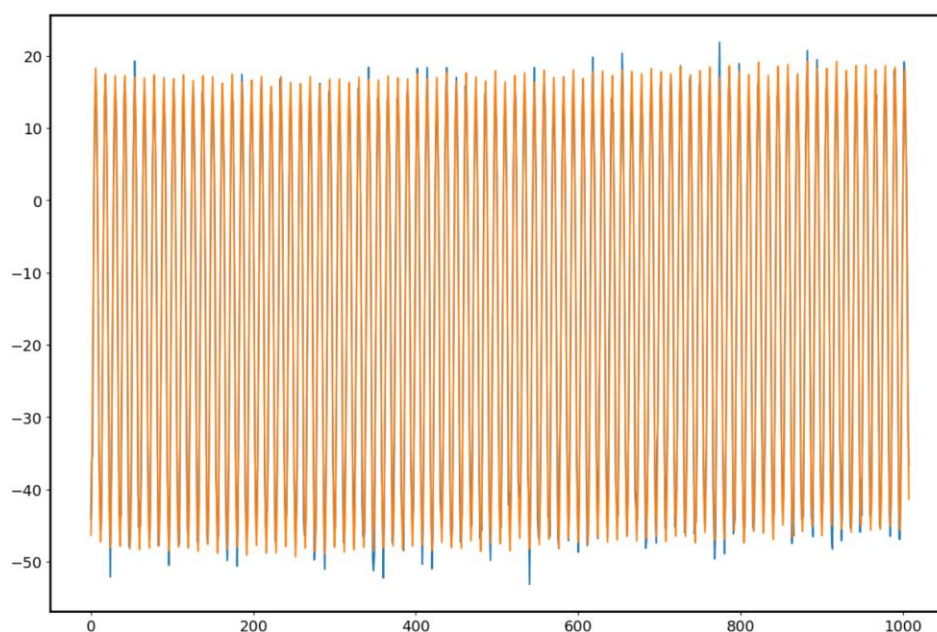


Рисунок 34. Исходный и восстановленный временные ряды (Верхоянск)

Заметим, что восстановленный ряд не включает всплески температурных значений и на протяжении всего времени остается примерно в тех же диапазонах, что также согласуется с результатами полученными ранее.

Построим прогноз на следующие 2 года на основе полученной модели (см. табл. 9).

Таблица № 9. Прогнозные значения, полученные в Python на 2021 и 2022 года

Название города/месяца	Архангельск		Верхоянск		Мурманск		Котлас		Салехард		Хатанга		Якутск	
	2021	2022	2021	2022	2021	2022	2021	2022	2021	2022	2021	2022	2021	2022
Январь	-11,47	-11,2	-45,19	-45	-9,97	-10,11	-11,36	-11,3	-22,49	-22,2	-29,19	-29,23	-37,45	-37,33
Февраль	-9,54	-9,72	-40,79	-40,68	-9,04	-9,24	-8,87	-8,77	-21,1	-20,94	-29,06	-29	-30,65	-30,54
Март	-4,07	-3,91	-26,7	-26,71	-5,13	-5,14	-3,22	-3,12	-14,26	-14,02	-21,96	-21,92	-17,76	-17,67
Апрель	2,52	2,52	-9,78	-9,87	-0,16	1	3,58	3,66	6,08	-5,76	-12,98	-12,9	-3,32	-3,25
Май	8,85	8,63	3,64	3,45	4,48	4,73	10,47	10,61	2,45	2,83	-2,74	-2,75	8,95	9
Июнь	13,83	14	14,8	14,5	8,73	8,75	16,15	16,34	10,95	11,29	9,58	9,64	17	17,04
Июль	15,93	16,05	18,7	18,42	11,54	11,37	18,23	18,45	14,32	14,67	14,95	15,11	19,77	19,8
Август	14,41	14,67	12,78	12,38	10,99	10,77	15,73	15,93	12,39	12,53	11,51	11,63	16,36	16,39
Сентябрь	9,95	9,72	2,18	1,73	6,86	6,72	10,15	10,33	6,94	7,29	4,08	4,24	6,46	6,49
Октябрь	3,54	3,63	-13,27	-13,73	1,31	1,29	3,36	3,57	-2,73	-2,39	-8,39	-8,41	-8,37	-8,32
Ноябрь	-3,68	-3,81	-31,28	-31,75	-3,65	-3,47	-3,56	-3,37	-12,79	-12,49	-21,01	-20,98	-24	-23,92
Декабрь	-9,36	-9,5	-42,2	-42,67	-7,73	-7,31	-9,21	-9,12	-18,91	-18,64	-26,57	-26,41	-34,99	-34,9

Точность прогноза составляет 79,2% что совпадает с точностью, полученной при реализации SSA в Rstudio и превышает тот же показатель на 8,4%, но полученный при прогнозировании с помощью ARIMA. В сравнительной таблице 10 представлена точность для рассмотренных методов анализа и способов их реализации.

Таблица № 10. Точность прогнозных значений

Название города/месяца	Точность прогноза		
	ПМА	SSA в R	SSA в Python
Архангельск	83,30%	95,80%	95,80%
Верхоянск	70,80%	79,20%	79,20%
Мурманск	91,60%	91,60%	91,60%
Котлас	83,30%	95,80%	95,80%
Салехард	41,60%	91,60%	91,60%
Хатанга	45,80%	54,17%	66,70%
Якутск	70,80%	83,30%	83,30%

Отметим, что для Хатанги при реализации метода анализа сингулярного спектра в *Python* удалось достичь большей точности прогнозных значений, чем в ходе использования *SSA* в *Rstudio*. В таблице № 11 представлены прогнозные значения на 2023 год для каждого из исследуемых населенных пунктов.

Таблица № 11. Прогнозные значения на 2023 год

Название города/месяца	Архангельск	Верхоянск	Мурманск	Котлас	Салехард	Хатанга	Якутск
	2023	2023	2023	2023	2023	2023	2023
Январь	-10,99	-43,69	-8,42	-10,9	-22,08	-29,27	-35,42
Февраль	-9,43	-41,08	-7,08	-9,51	-21,03	-29,1	-30,67
Март	-4,4	-28,77	-3,61	-3,95	-13,5	-21,93	-17,47
Апрель	2,4	-10,43	1,24	3,87	-6,11	-13,07	-1,69
Май	9,2	4,54	6,57	11,18	1,59	-2,85	9,22
Июнь	14,29	14,63	11,18	15,66	11,12	9,6	17,89
Июль	15,99	17,48	13,67	16,56	14,72	15,04	20,99
Август	14,15	12,46	12,99	14,55	12,48	11,5	15,7
Сентябрь	9,91	2,7	9,09	10,36	7,76	4,35	7,37
Октябрь	3,97	-14,1	3,19	3,92	-2,8	-7,84	-6,45
Ноябрь	-3,17	-33,21	-2,72	-3,95	-13,92	-20,33	-24,77
Декабрь	-9,22	-42,43	-6,82	-10,14	-19,05	-25,79	-34,78

Так как в библиотеке *pySSA* отсутствуют некоторые функции, то было принято решение реализовать метод самостоятельно. Ниже представлены некоторые различия при реализации в различных средах разработки:

1. У метода *SSA* существует множество вариаций, предполагающих такие операции как центрирование и нормирование, а в ходе реализации может потребоваться устранение сезонности и тренда. Библиотека *Rssa* в *Rstudio* позволяет автоматически выполнить

любую из данных операций, в то время как в *Python* необходимо их реализовать самостоятельно, что может приводить к несогласованности в результатах.

2. Также в библиотеке *Rssa* в *Rstudio* доступны дополнительные функции, позволяющие выбрать наилучшую длину окна и число компонент. В библиотеке *pySSA* для *Python* такие возможности отсутствуют.
3. Для выполнения программного кода *SSA* в *Python* потребовалось меньше времени, чем в *Rstudio*, следовательно, при большом объеме данных для сокращения времени лучше использовать язык *Python*;
4. Для многих исследователей, студентов *Python* более удобен в использовании, чем язык программирования *R*.

Вывод: оба языка программирования имеют свои плюсы и минусы при реализации метода анализа сингулярного спектра, но выбор между ними зависит от конкретных требований.

5. Заключение

В ходе работы было проведено исследование предметной области процесса глобального потепления, которое включало в себя изучение причин, последствий и соответствующих мер борьбы и поддержки, оказываемых для сдерживания темпов глобального потепления.

Проведен первичный вероятностно-статистический анализ по итогам которого сформулированы следующие выводы:

1. На гистограммах среднемесячных температур наблюдаются разные темпы повышения температурных значений в разные времена года. В исследуемых населенных пунктах наибольшее повышение температур наблюдалось в различных месяцах;
2. Подсчитаны показатели динамики для каждого временного ряда:

- самая высокая средняя температура за рассматриваемый промежуток времени в Котласе ($1,825^{\circ}\text{C}$), а самая низкая в городе Верхоянске ($-14,734^{\circ}\text{C}$);
- только в 3 из 7 оставшихся исследуемых населенных пунктов наблюдается положительный средний темп прироста, причем максимальный в Котласе - $0,702\%$. В 4 оставшихся наблюдается отрицательный прирост, что говорит о ежегодном понижении температур. Максимальное понижение наблюдается в Салехарде и составляет $-1,782\%$.

На основе первичного вероятностно-статистического анализа можно сделать вывод о потеплении, но не глобальном, а скорее локальном, так как в нескольких из исследуемых городов наблюдается понижение средних температур.

С помощью языка программирования *R* и инструментов среды разработки *Rstudio* были проведены исследования временных рядов на стационарность, выполнена их декомпозиция и построена соответствующая модель *ARIMA* учитывающая тренд и сезонную компоненты. Для каждой модели были проведены оценки адекватности, включающие в себя тесты на нормальное распределение остатков и их автокорреляцию. После были получены прогнозные значения на 2021 и 2022 год, а также подсчитаны показатели качества модели и точность прогнозных значений.

Помимо простых методов анализа временных рядов был реализован один из универсальных и более точных методов - метод *SSA*. Для его использования были выбраны две среды разработки: *Python* и *Rstudio*. И в той и в другой среде удалось отобрать наиболее информативные компоненты временного ряда, выполнить его декомпозицию, построить корреляционную матрицу и получить прогноз на 2021, 2022 и 2023 года. Определить точность прогнозных значений удалось только для 2021 и 2022 годов, так как для них уже

присутствуют фактические значения. Практически для всех моделей качество прогноза составило 80 и более процентов.

Перед применением различных методов анализа временных рядов были проведены тесты на стационарность, которые не выявили изменение средних значений во времени, что говорит об отсутствии тренда. Данное замечание также подтверждает результаты полученные при первичном вероятностно-статистическом анализе.

Говоря о направлениях для дальнейших исследований хочется провести анализ климатических показателей экваториальной зоны, так как именно в ней наблюдаются высокие значения температур в сочетании с повышенной влажностью воздуха. Также интересным представляется представить метеорологические величины как зависимые от расположения города: близость к морю, рельеф местности, уровень солнечной радиации и т.д.

6. Список использованной литературы

1. Автоматизированная Информационная Система Обработки Режимной Информации (АИСОРИ) // Удаленный доступ к ЯОД архивам. [Электронный ресурс]. – Режим доступа: <http://aisori-m.meteo.ru/waisori/index0.shtml> (дата обращения: 23.11.2021)
2. Башмаков И. Низко углеродная Россия: перспективы после кризиса. Вопросы экономики [Электронный ресурс]. 2009 -. Режим доступа: <https://doi.org/10.32609/0042-8736-2009-10-107-120>. (Дата обращения: 10.10.2021)
3. Борьба с изменением климата — гонка, из которой мы все еще можем выйти победителями [Электронный ресурс]. - Режим доступа: <https://www.un.org/ru/un75/climate-crisis-race-we-can-win> (дата обращения: 03.05.2022)
4. Всемирный фонд дикой природы World Wildlife Fund (WWF) России и WWF Германии [Электронный ресурс]. – Режим доступа: <https://wwf.ru/> (дата обращения: 23.11.2021)

5. Глобальное потепление [Электронный ресурс]. – Режим доступа: <https://ru.wikipedia.org/wiki> (дата обращения: 03.05.2022).
6. Голяндина Н. Э. Метод «Гусеница»-SSA: анализ временных рядов: Учеб. пособие. – СПб. : Изд-во СПбГУ, 2004. – 76с.
7. Гречкосеев А.К. [и др.] Модификация базового метода сингулярного спектрального анализа для повышения точности прогнозирования неравномерности вращения Земли // Вычислительные технологии. - 2020. – Т. 25, № 3. - С. 54-65. [Электронный ресурс]. – режим доступа: https://elibrary.ru/download/elibrary_43066617_78456857.pdf (дата обращения 20.03.2023)
8. Данилов Д.Л., Жиглявский А.А. (ред.) Главные компоненты временных рядов: метод «Гусеница». – СПб. : Изд-во СПбГУ, 1997. – 307 с.
9. Жиглявский А.А, Корников В.В., Пепельшев А.Н. Анализ и прогноз земных температур и ледяного покрытия. – СПб. : Изд-во СПбГУ, 2016.
10. Качурин Л. Г. Физические основы воздействия на атмосферные процессы. – Л.: Гидрометеоиздат, 1989. [Электронный ресурс]. – Режим доступа: <http://www.twirpx.com>. (дата обращения 25.05.2022).
11. Кисцилин, М. М. Многомерная статистика временных рядов наблюдений в авиационной эргономике // Вопросы кибернетики. - М., 1978.
12. Лысенко С.А., Буюков И.В. Особенности современного изменения климата в республике Беларусь // Фундаментальная и прикладная климатология - 2020. - С. 22-41. – Режим доступа: https://elibrary.ru/download/elibrary_44336345_12077462.pdf
13. Межправительственная группа экспертов по изменению климата (МГЭИК). [Электронный ресурс] – Режим доступа: [IPCC — Intergovernmental Panel on Climate Change](https://www.ipcc.ch/) (дата обращения: 20.12.2022).
14. Научно-технический журнал «Метеорология и гидрология». [Электронный ресурс]. – Режим доступа: <http://www.mig-journal.ru/> (дата обращения: 15.01.2023)

15. Пичугин Ю. А. Учет сезонных эффектов в задачах прогноза и контроля данных о приземной температуре воздуха // *Метеорология и гидрология* - 1996. Т. № 4. - С. 53 – 64
16. Показатели динамики: темп роста и темп прироста. [Электронный ресурс] – Режим доступа: <https://univer-nn.ru/statistika/pokazateli-dinamiki-temp-rosta-i-prirosta/> (дата обращения: 15.03.2023)
17. Федеральная служба по гидрометеорологии и мониторингу окружающей среды (Росгидромет). [Электронный ресурс] - Режим доступа: <https://www.meteorf.gov.ru/> (дата обращения: 23.12.2022).
18. Научный журнал «Фундаментальная и прикладная климатология». [Электронный ресурс] – Режим доступа: <https://fa-climatology.ru/index.php/fpk-ru> (дата обращения: 17.01.2023).
19. A Complete Introduction To Time Series Analysis (with R): SARIMA models. [Электронный ресурс] – Режим доступа: <https://medium.com/analytics-vidhya/a-complete-introduction-to-time-series-analysis-with-r-sarima-models-ff86d526d1d7> (дата обращения: 8.04.2023)
20. Cairo [Электронный ресурс] – Режим доступа: <https://github.com/s-u/Cairo> (дата обращения: 7.04.2023)
21. Fraedrich, K., (1986), Estimating the dimension of weather and climate attractors, *J. Atmos. Sci.*, 43, 419 – 432.
22. ggplot2 [Электронный ресурс] – Режим доступа: <https://ggplot2.tidyverse.org/> (дата обращения: 7.04.2023)
23. Ghill, M., Vautard, R., (1991), Interdecadal oscillations and the warming trend in global temperature time series, *Nature*, 350, 324 – 327.
24. Ghill, M., Yiou, P., (1996), Spectral methods: What they can and cannot do for climatic time series, in *Decadal Climate Variability: Dynamics and Predictability*, Eds. D. Anderson and J. Willebrand, Elsevier, Amsterdam.

25. Golyandina, N., Korobeynikov, A., Zhigljavsky, A. Singular Spectrum Analysis with R. – 2018 [Электронная версия книги] - <https://link.springer.com/book/10.1007/978-3-662-57380-8>
26. Heat waves aren't going away. Here's how we can prepare // [Электронный ресурс]: <https://www.universityofcalifornia.edu/news/heat-waves-arent-going-away-heres-how-we-can-prepare> (дата обращения: 20.02.2023).
27. Hossein H, Dimitrios D., Thomakos. A review on Singular Spectrum Analysis for economic and financial time series. – Режим доступа: https://www.researchgate.net/publication/266685235_A_review_on_Singular_Spectrum_Analysis_for_economic_and_financial_time_series
28. Introducing SSA for Time Series Decomposition. [Электронный ресурс] – Режим доступа: <https://www.kaggle.com/code/jdarcy/introducing-ssa-for-time-series-decomposition> (дата обращения: 25.02.2023)
29. Pa Ousman Bojang, Tao-Chang, Yang, Quoc, Bao Pham, Pao-Shan Yu Linking Singular Spectrum Analysis and Machine Learning for Monthly Rainfall Forecasting // National Cheng-Kung University – 2020. [Электронный ресурс] – Режим доступа: <https://www.mdpi.com/2076-3417/10/9/3224> (дата обращения: 24.04.2023)
30. R. Vautard, M. Ghil Singular spectrum analysis in nonlinear dynamics, with applications to paleoclimatic time series // University of California – 2002. [Электронный ресурс] – Режим доступа: https://www.researchgate.net/publication/222436693_Singular_spectrum_analysis_in_nonlinear_dynamics_with_applications_to_paleoclimatic_time_series (дата обращения: 15.04.2023)

7. Приложение № 1

Проверка на нормальность для May			
	Name	Statistic	p-value
1	Lilliefors (Kolmogorov-Smirnov) normality test	0.04050717	0.9818379
2	Anderson-Darling normality test	0.14070349	0.9722699
3	Cramer-von Mises normality test	0.02147934	0.9527546
4	Shapiro-wilk normality test	0.99596842	0.9968763
5	Pearson chi-square normality test	0.85714286	0.9997022
[1] -----			
Проверка на нормальность для Jun			
	Name	Statistic	p-value
1	Lilliefors (Kolmogorov-Smirnov) normality test	0.08078119	0.1933382
2	Anderson-Darling normality test	0.48224815	0.2249622
3	Cramer-von Mises normality test	0.06316474	0.3398803
4	Shapiro-wilk normality test	0.97798165	0.1597406
5	Pearson chi-square normality test	10.85714286	0.2856284
[1] -----			
Проверка на нормальность для Jul			
	Name	Statistic	p-value
1	Lilliefors (Kolmogorov-Smirnov) normality test	0.06062913	0.6261820
2	Anderson-Darling normality test	0.19855014	0.8834063
3	Cramer-von Mises normality test	0.03417324	0.7816180
4	Shapiro-wilk normality test	0.99477201	0.9847507
5	Pearson chi-square normality test	8.85714286	0.4505644
[1] -----			
Проверка на нормальность для Aug			
	Name	Statistic	p-value
1	Lilliefors (Kolmogorov-Smirnov) normality test	0.08521270	0.1387465
2	Anderson-Darling normality test	0.45175783	0.2671037
3	Cramer-von Mises normality test	0.07633555	0.2281297
4	Shapiro-wilk normality test	0.97863910	0.1764647
5	Pearson chi-square normality test	12.00000000	0.2133093

Рисунок 1. Результаты тестов на нормальность остатков для каждого месяца по отдельности (Верхоянск)

Таблица № 1 и 2. Показатели динамики временного ряда для г. Котласа (слева) и г. Салехарда

Среднее значение температуры	1,825	Среднее значение температуры	-5,716
Средний абсолютный прирост	0,024	Средний абсолютный прирост	0,047
Средний коэффициент роста	1,007	Средний коэффициент роста	0,982
Средний темп роста	100,702	Средний темп роста	98,218
Средний темп прироста	0,702	Средний темп прироста	-1,782
Средняя величина абсолютного значения 1 % прироста	0,034	Средняя величина абсолютного значения 1 % прироста	-0,026

Таблица 3. Показатели динамики временного ряда для с. Хатанга

Среднее значение температуры	-12,414
Средний абсолютный прирост	0,061
Средний коэффициент роста	0,993
Средний темп роста	99,330
Средний темп прироста	-0,670
Средняя величина абсолютного значения 1 % прироста	-0,091

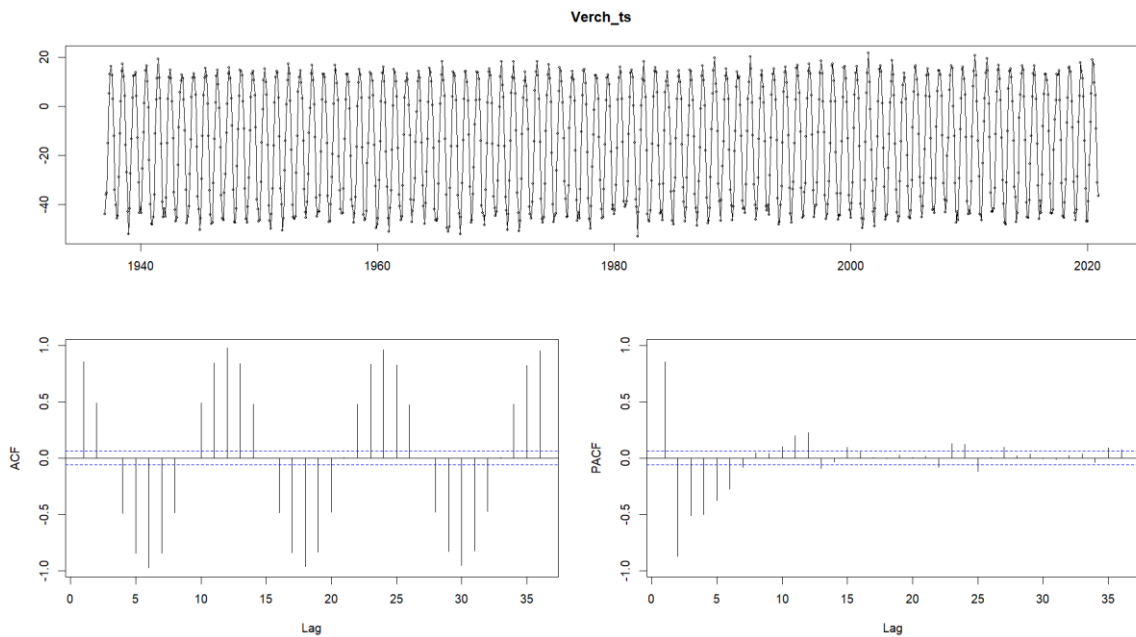


Рисунок 7. Исходный временной ряд и его автокорреляционные функции (Верхоянск)

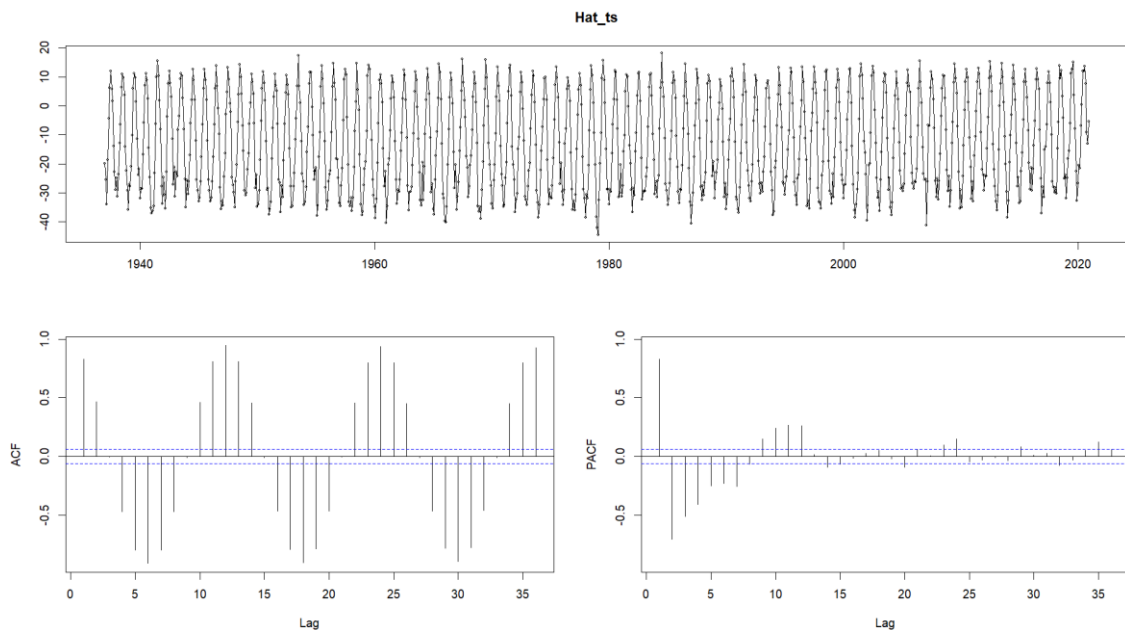


Рисунок 8. Исходный временной ряд и его автокорреляционные функции (Хатанга)

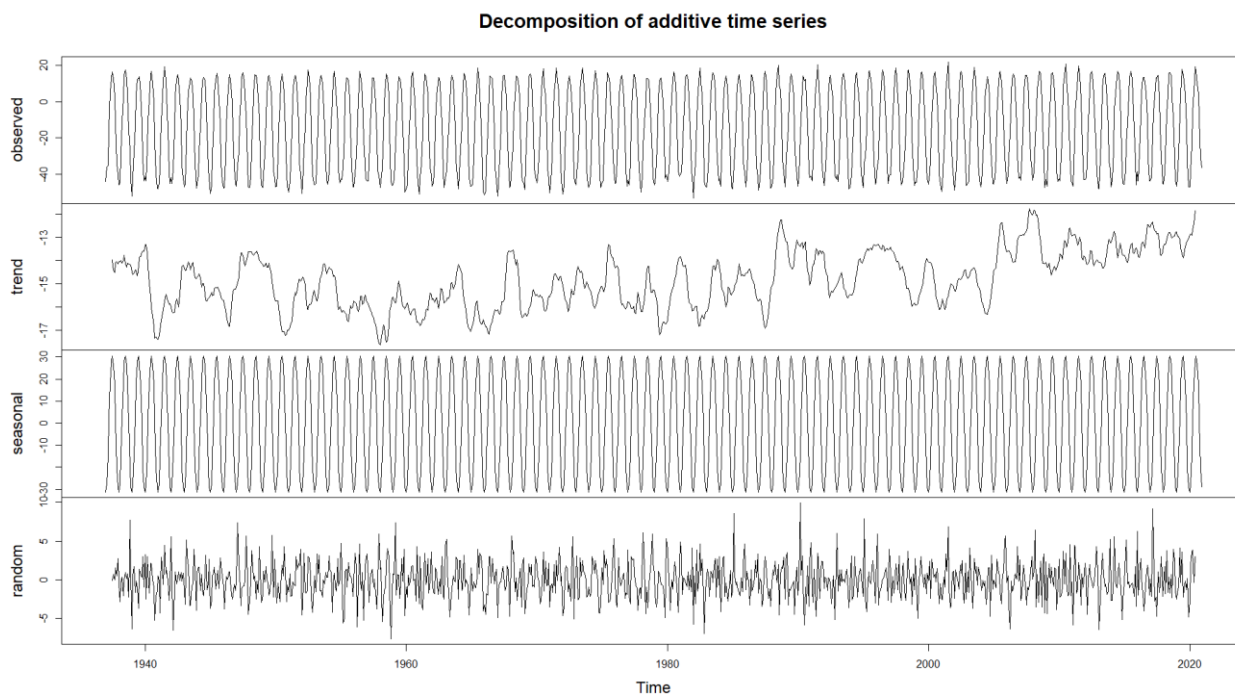


Рисунок 9. Декомпозиция временного ряда (Верхоянск)

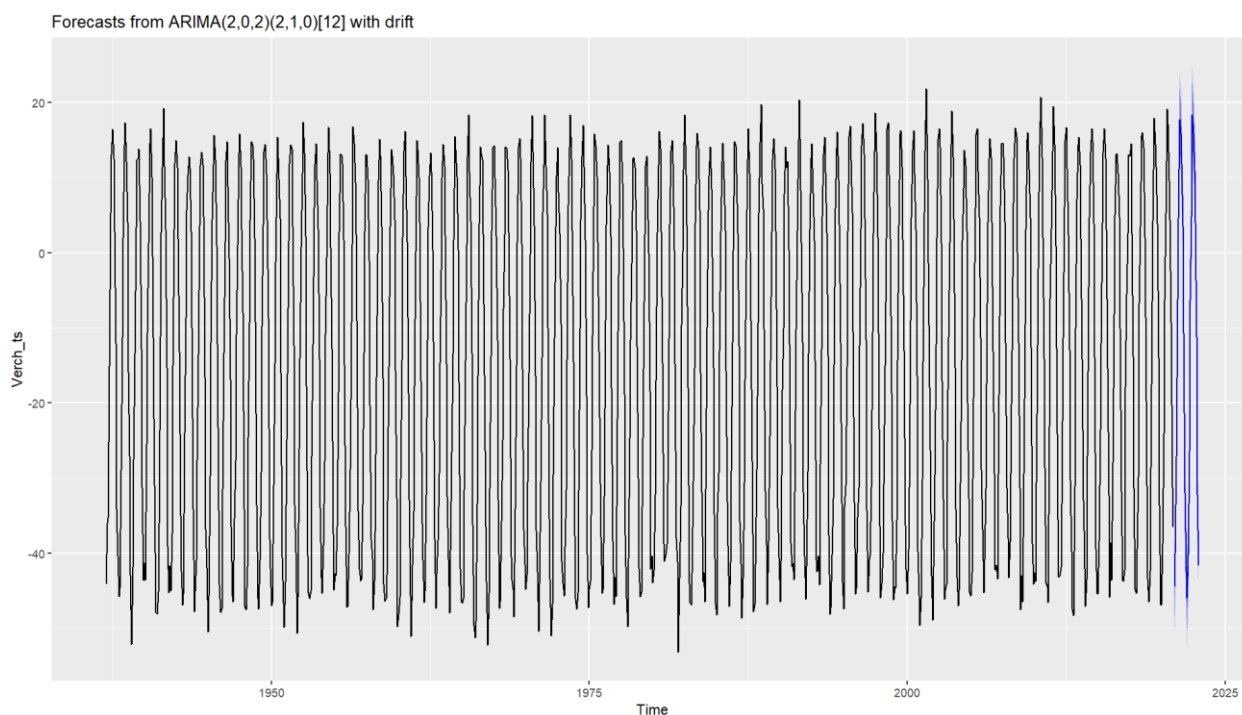


Рисунок 10. График полученных прогнозных значений (Верхоянск)

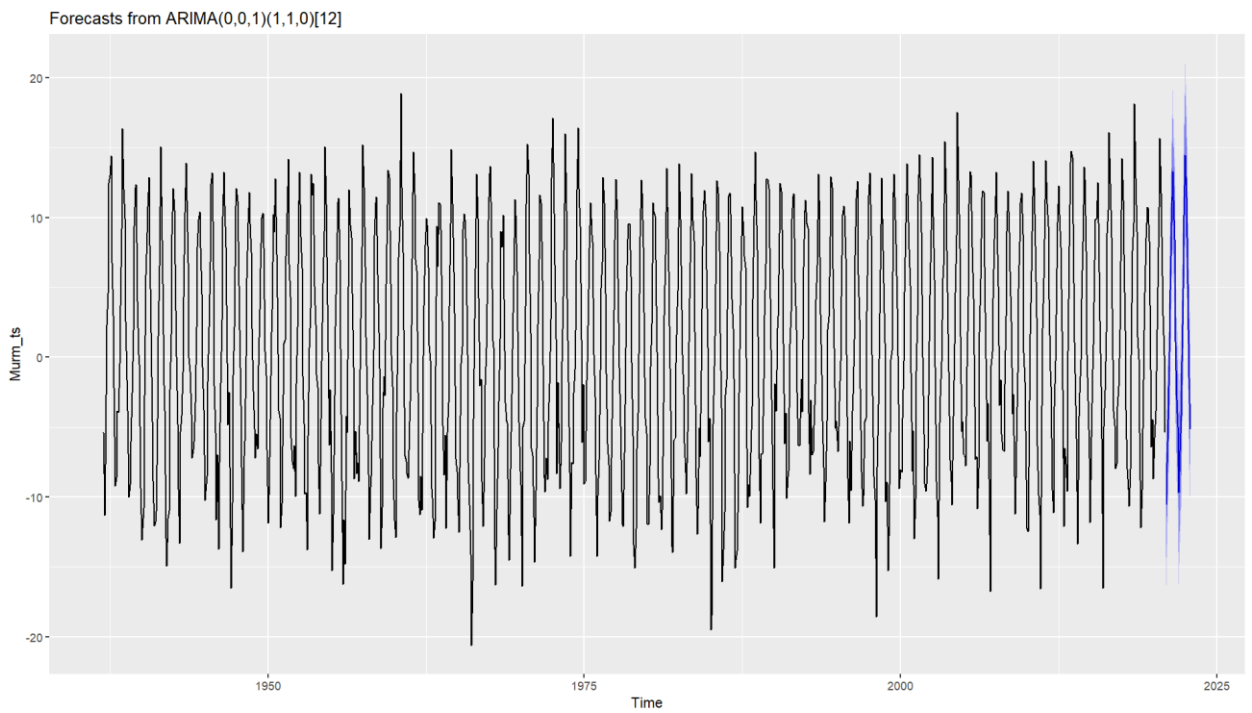


Рисунок 11. График исходного временного ряда и полученных прогнозных значений (Мурманск)

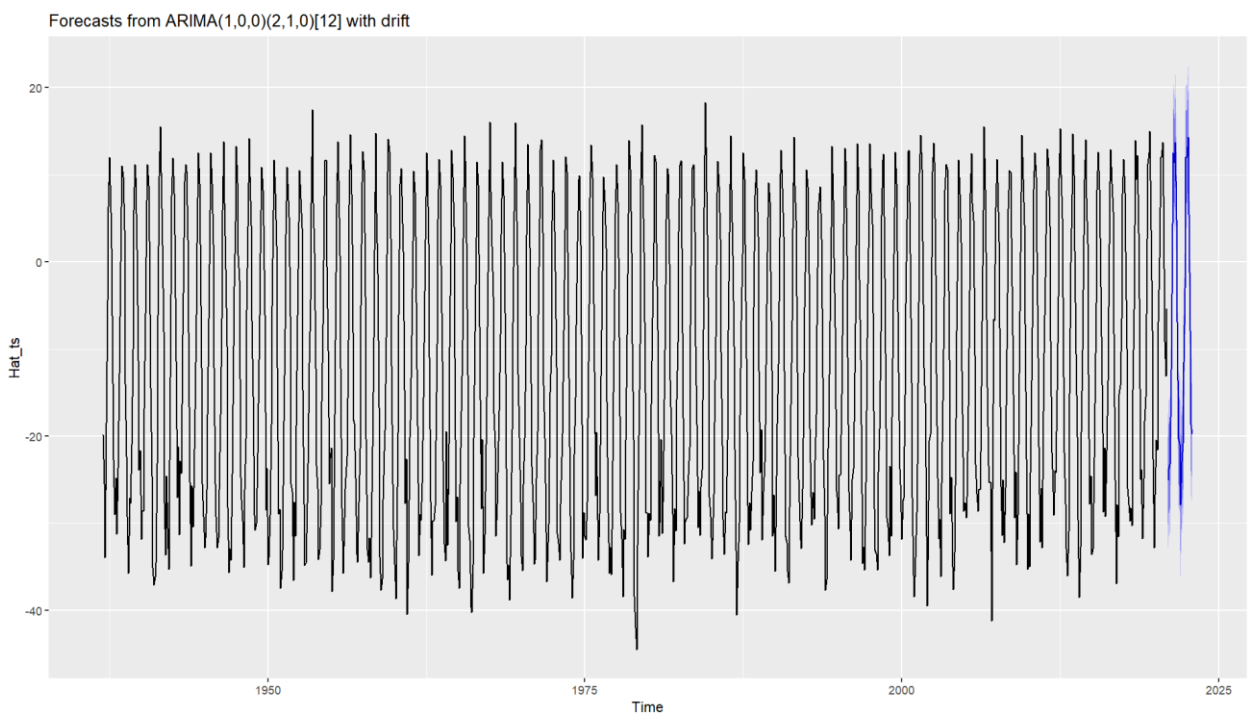


Рисунок 12. График исходного временного ряда и полученных прогнозных значений (Хатанга)

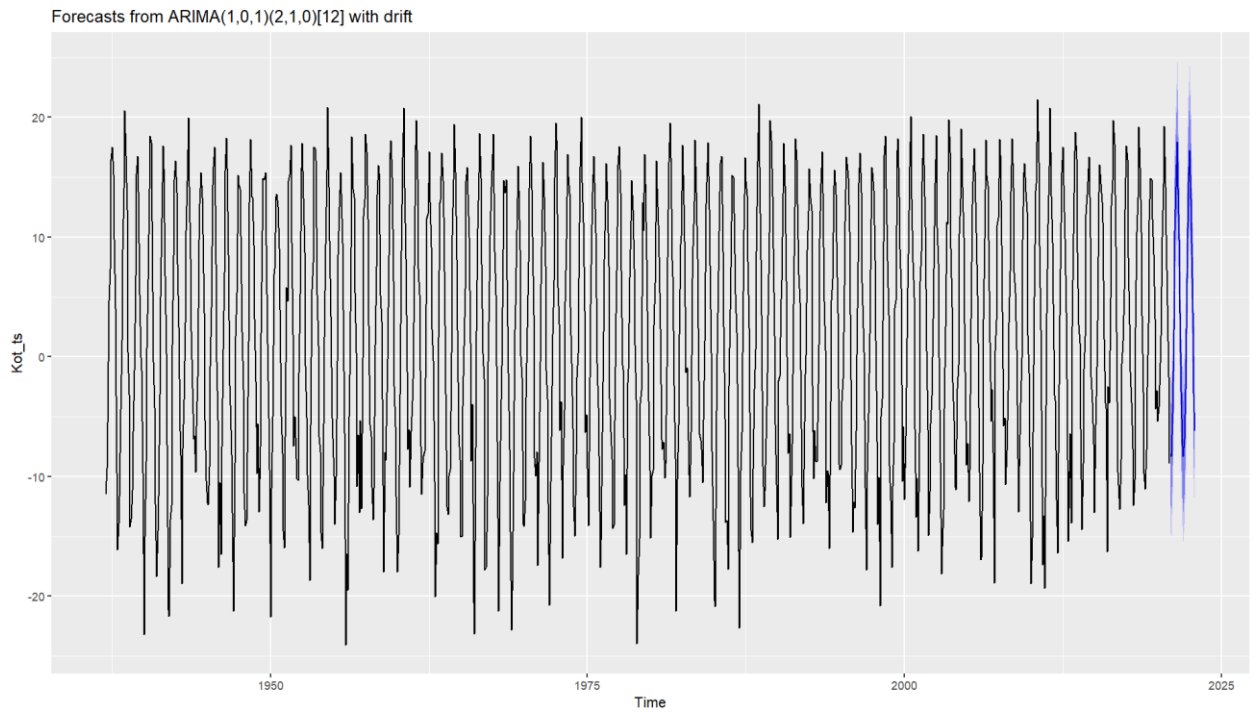


Рисунок 13. График исходного временного ряда и полученных прогнозных значений (Котлас)

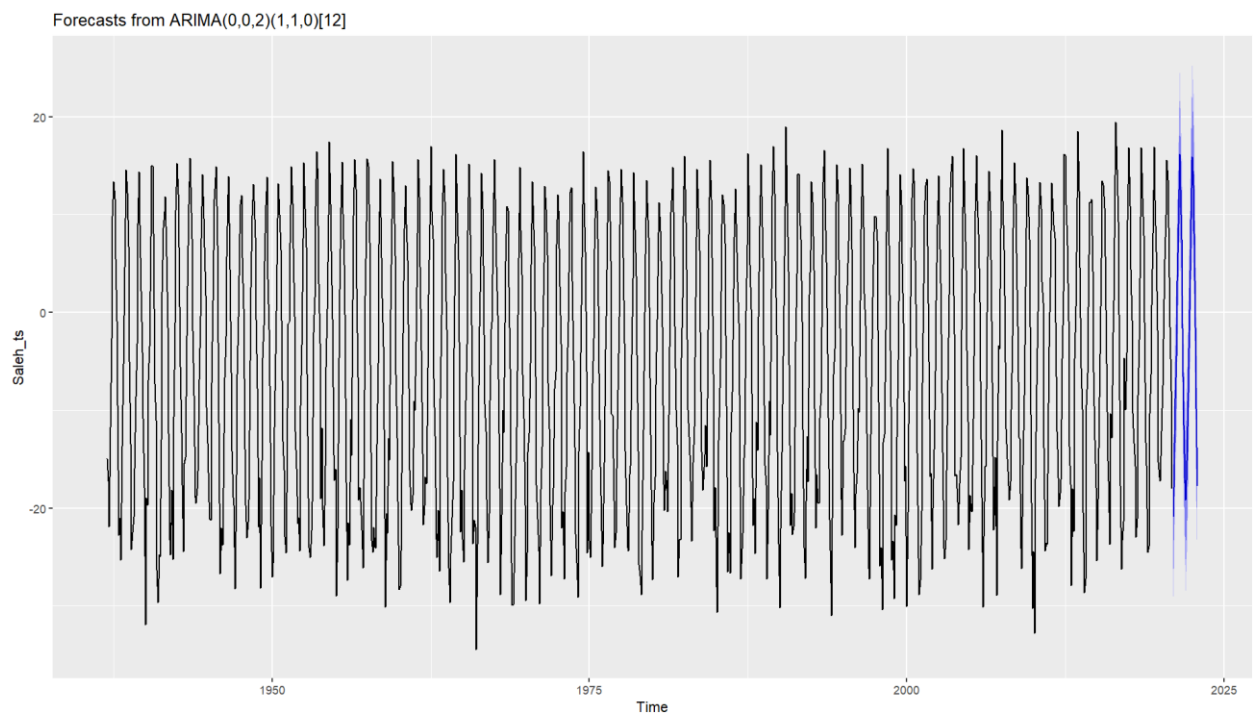


Рисунок 14. График исходного временного ряда и полученных прогнозных значений (Салехард)

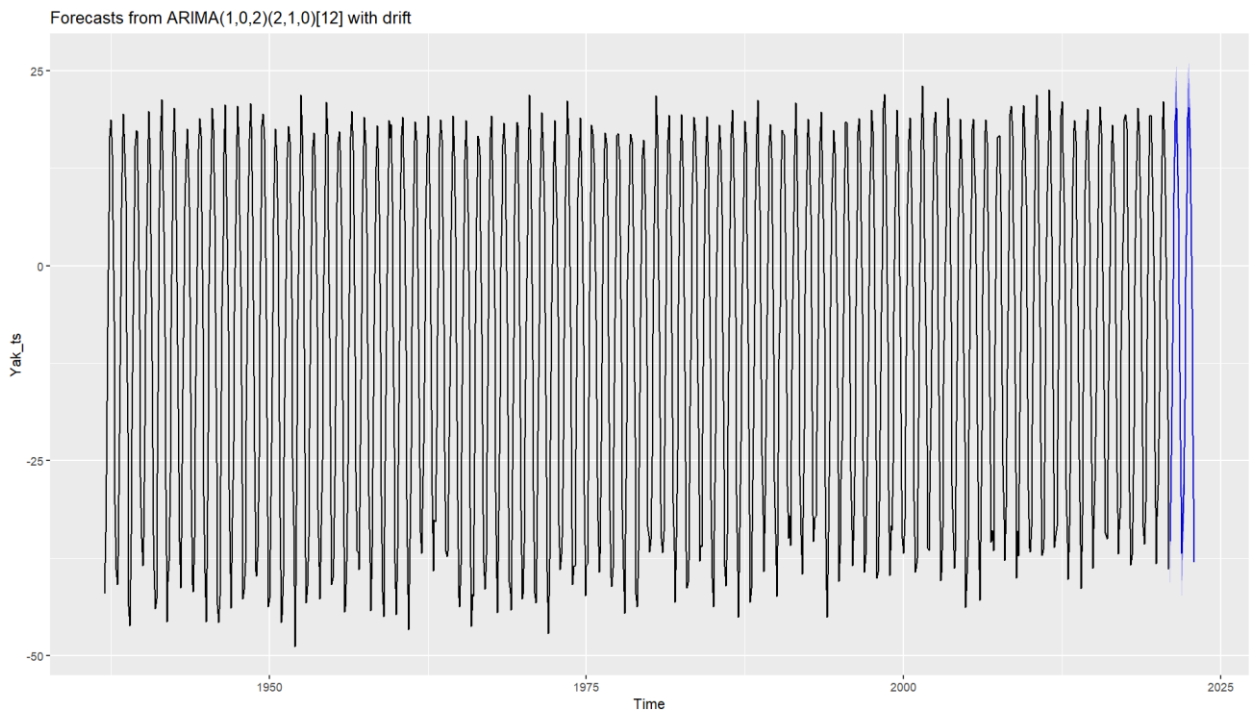


Рисунок 15. График исходного временного ряда и полученных прогнозных значений (Якутск)

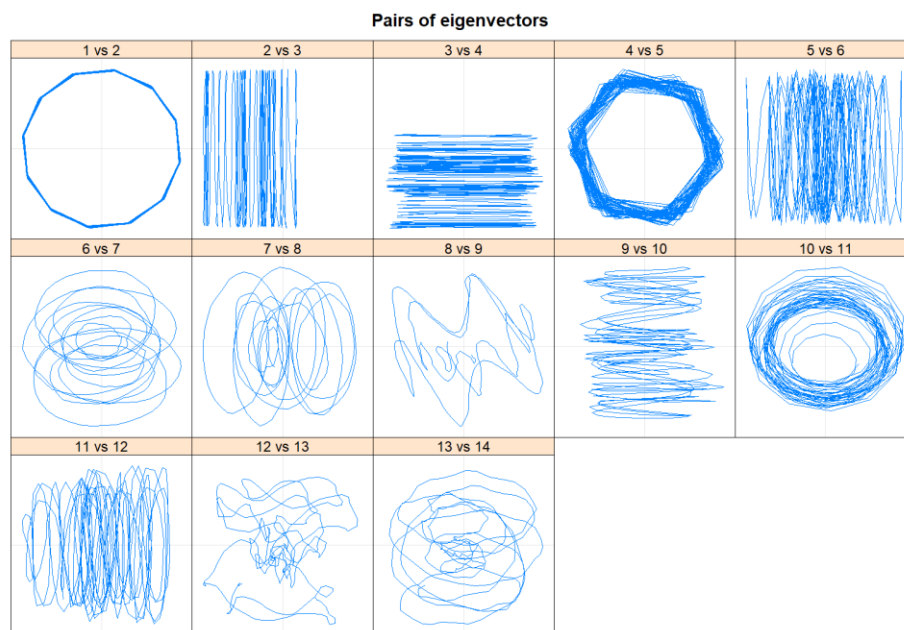


Рисунок 16. 2-D диаграммы рассеяния (Мурманск)

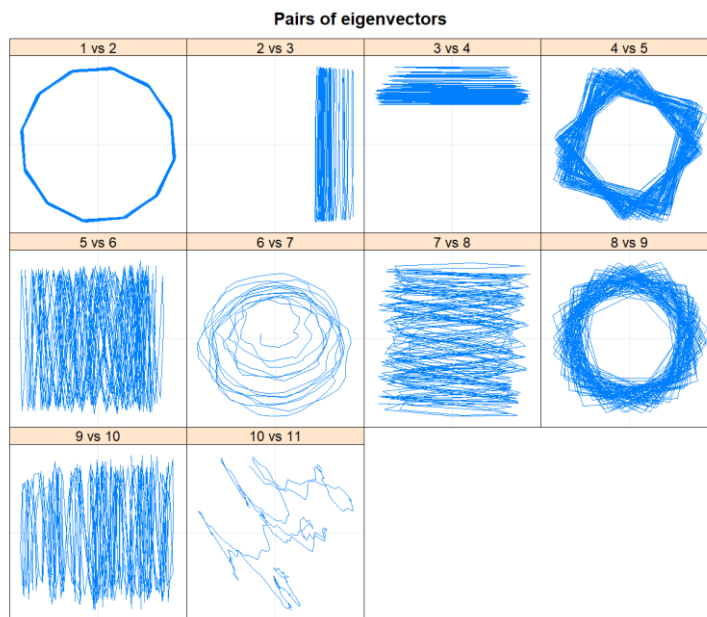


Рисунок 17. 2-D диаграммы рассеяния (Архангельск)

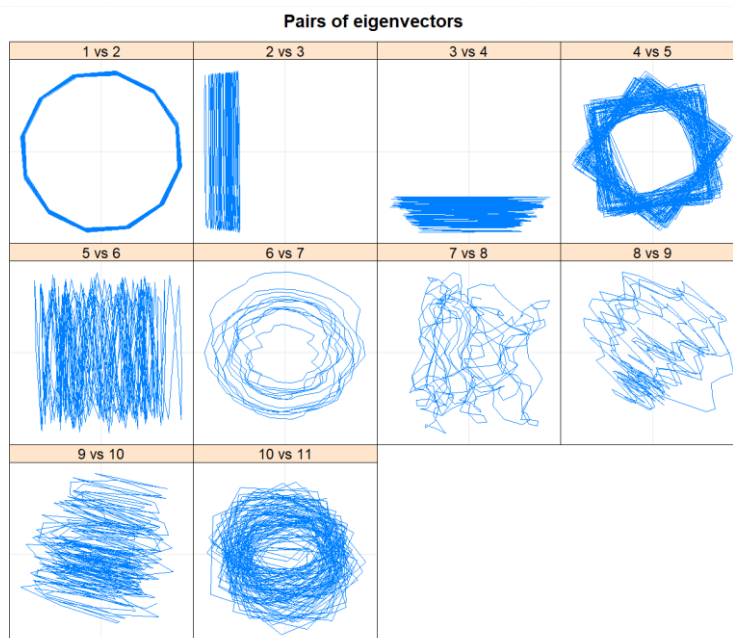


Рисунок 18. 2-D диаграммы рассеяния (Котлас)

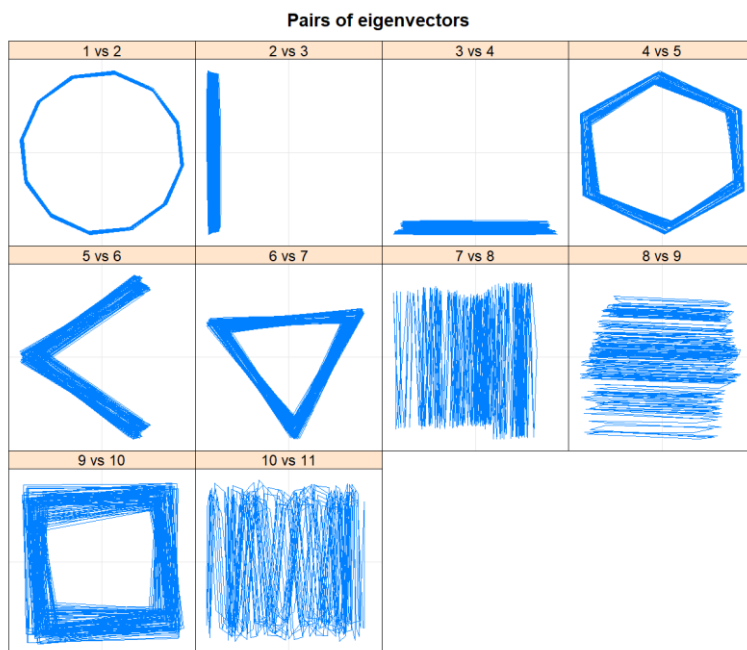


Рисунок 19. 2-D диаграммы рассеяния (Якутск)

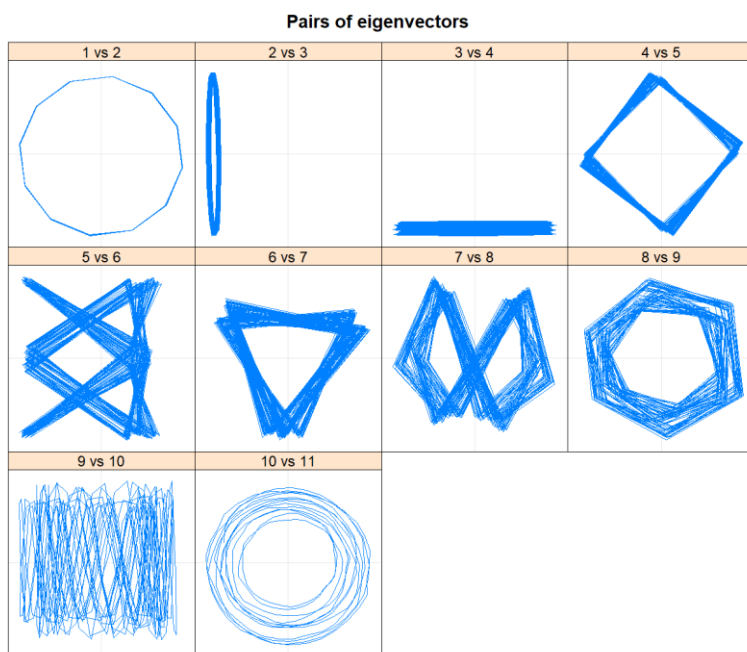


Рисунок 20. 2-D диаграммы рассеяния (Верхоянск)

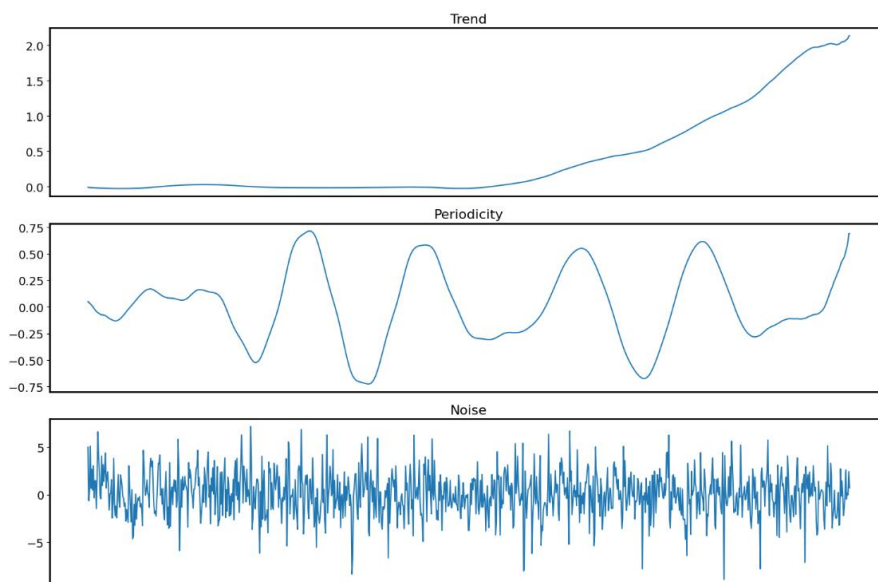


Рисунок 21. Декомпозиция полученная в Python (Мурманск)

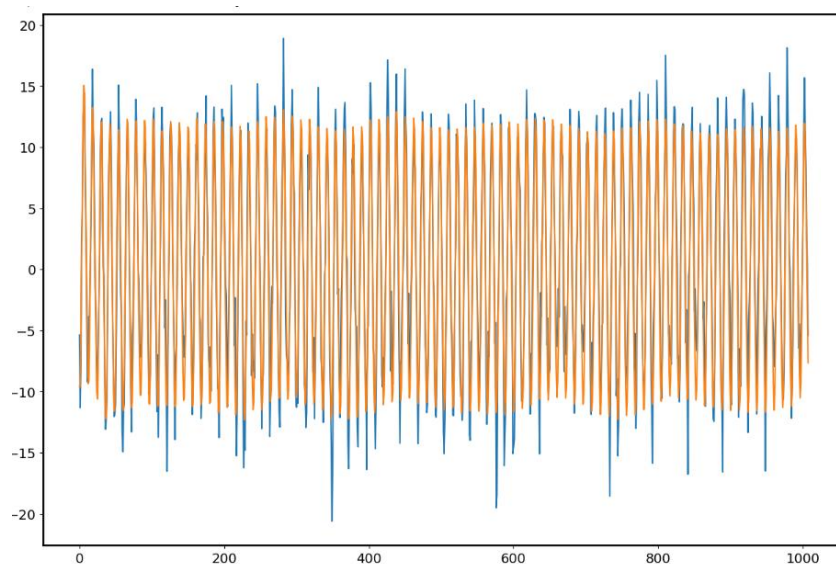


Рисунок 22. Исходный и восстановленный ряд (Мурманск)

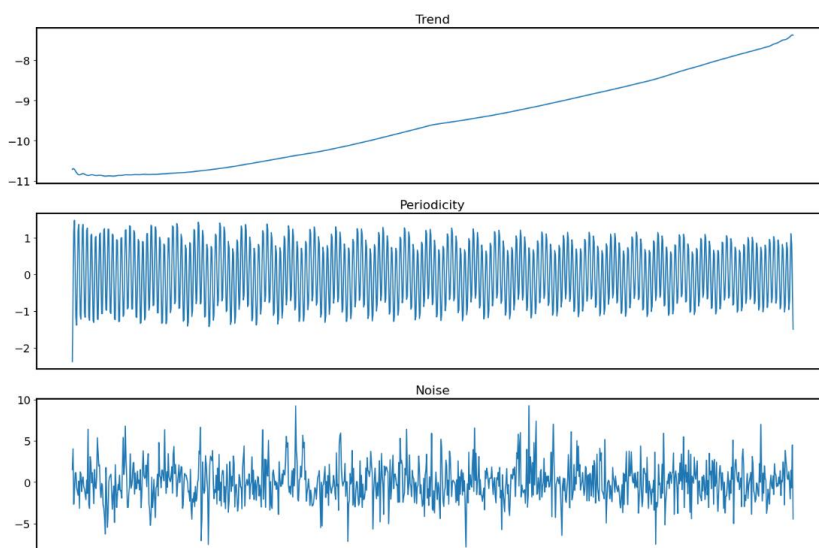


Рисунок 23. Декомпозиция полученная в Python (Якутск)

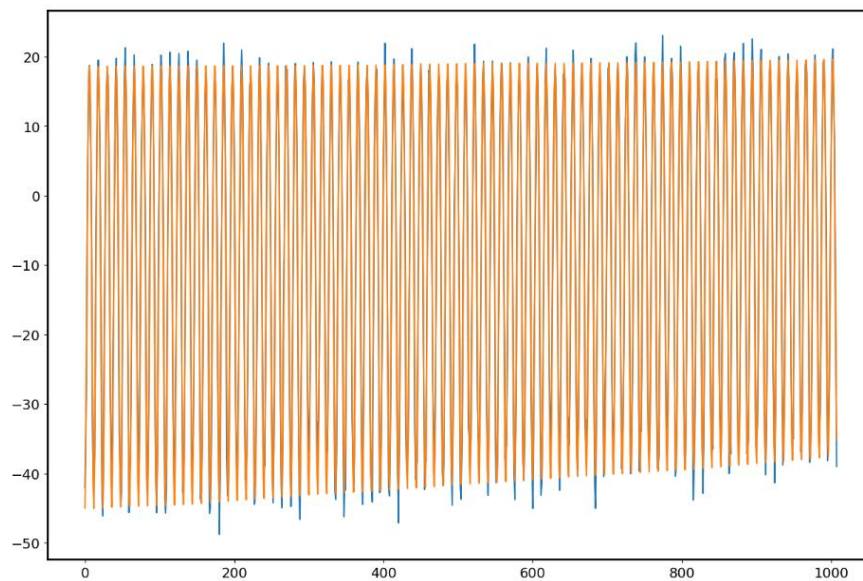


Рисунок 24. Исходный и восстановленный ряд (Якутск)

8. Приложение № 2

Листинг № 1. Фрагмент кода из *Rstudio*

```

library(ggplot2)
library(readxl)
library(Cairo)
library(scales)
library(dplyr)
library(stringr)
library(Rssa)
wd <- getwd()
Hat_path <- str_c(wd, "/Hatanga.xlsx")
Hat <- read_xlsx(Hat_path)
Hat_ssa <- ssa(Hat$Temp, L = 400, kind = "1d-ssa")
# 2. Reconstruct
Hat_rec <- reconstruct(Hat_ssa, groups = list(Trend=c(3), Season1 = c(1,2)))
plot(Hat_rec, add.residuals = TRUE, add.original = TRUE, plot.method = "xyplot",
     superpose=TRUE, anti.key=list(colSums = 2))

plot(Hat_ssa, type = "vectors", idx = 1:12)
plot(Hat_ssa, type = "paired", idx = 2:13, plot.contrib = FALSE)

print(parestimate(Hat_ssa, groups = list(2:3, 4:5, 6:7), method = "pairs"))
plot(wcor(Hat_ssa, groups=1:30), scales=list(at=c(10,20,30)))
plot(Hat_ssa, type = "paired", idx = 1:10, plot.contrib = FALSE)
plot(reconstruct(Hat_ssa, groups = list(G12 = 2:3, G6 = 4:5, G3 = 6:7)),
     plot.method = "xyplot", layout = c(2, 2), add.residuals = FALSE,
     add.original = FALSE)

```

Листинг № 2. Фрагмент кода из *Python* [29]

```

import numpy as np
from numpy import pi
import matplotlib.pyplot as plt
import pandas as pd
import numpy.linalg as npla

plt.rcParams['figure.figsize'] = (15,10)
plt.rcParams['font.size'] = 14
# plt.rcParams['image.cmap'] = 'plasma'
plt.rcParams['axes.linewidth'] = 2
from cycler import cycler
cols = plt.get_cmap('tab10').colors
plt.rcParams['axes.prop_cycle'] = cycler(color=cols)

def plot_2d(m, title=""):
    plt.imshow(m)
    plt.xticks([])
    plt.yticks([])
    plt.title(title)
data1 = pd.read_excel('/content/drive/MyDrive/Гусеница/Натанга.xlsx')
datas_1=data1['Temp']
N = len(datas_1)
L = 502 # длина окна
K = N - L + 1 # Количество столбцов в траекторной матрице

d = npla.matrix_rank(X) # Размерность траекторной матрицы
U, S, VT = npla.svd(X)
# S = np.sqrt(S)
V = np.transpose(VT)

```