

Министерство науки и высшего образования Российской Федерации
Федеральное государственное бюджетное образовательное учреждение
высшего образования
АМУРСКИЙ ГОСУДАРСТВЕННЫЙ УНИВЕРСИТЕТ
(ФГБОУ ВО «АмГУ»)

Институт компьютерных и инженерных наук
Кафедра математического анализа и моделирования
Направление подготовки – 01.03.02 Прикладная математика и информатика
Направленность (профиль) образовательной программы «Прикладная математика и информатика»

ДОПУСТИТЬ К ЗАЩИТЕ

И.о. зав. кафедрой

_____ Н.Н. Максимова
«_____» _____ 2025 г.

БАКАЛАВРСКАЯ РАБОТА

на тему: Прогнозирование численности популяции лосося с использованием временных рядов

Исполнитель
студент группы 1101об

(подпись, дата)

С.А. Бызов

Руководитель
доцент, канд. физ.-мат. наук

(подпись, дата)

Е.М. Веселова

Нормоконтроль
ассистент

(подпись, дата)

В.О. Салмиянов

Благовещенск 2025

Министерство науки и высшего образования Российской Федерации
Федеральное государственное бюджетное образовательное учреждение
высшего образования
АМУРСКИЙ ГОСУДАРСТВЕННЫЙ УНИВЕРСИТЕТ
(ФГБОУ ВО «АмГУ»)

Институт компьютерных и инженерных наук
Кафедра математического анализа и моделирования

УТВЕРЖДАЮ
И.о. зав. кафедрой
_____ Н.Н. Максимова
« _____ » _____ 2025 г.

З А Д А Н И Е

К бакалаврской работе студента Бызова Степана Андреевича

1. Тема бакалаврской работы: Прогнозирование численности популяции лосося с использованием временных рядов (утверждена приказом от 03.04.2025 № 879-уч).
 2. Срок сдачи студентом законченной работы: 25.06.2025 г.
 3. Исходные данные к бакалаврской работе: научные статьи, официальное руководство Python, учебные и учебно-методические работы, отчет по производственной практике (научно-исследовательской работе), отчет по преддипломной практике, среда разработки – Google Colab.
 4. Содержание бакалаврской работы (перечень подлежащих разработке вопросов): теоретический анализ методов прогнозирования временных рядов, реализация моделей временных рядов для прогнозирования биологических популяций; визуализация результатов прогнозирования.
 5. Перечень материалов приложения: листинги вычислительных программ.
 6. Консультанты по бакалаврской работе: нормоконтроль – Салмиянов В.О., ассистент.
 7. Дата выдачи задания: 14.04.2025 г.
- Руководитель бакалаврской работы: Веселова Елена Михайловна, доцент, канд. физ.-мат. наук.

Задание принял к исполнению (14.04.2025): _____ Бызов С.А.

РЕФЕРАТ

Бакалаврская работа содержит 81 с., 7 рисунков, 4 таблицы, 8 приложений, 20 источников.

ПОПУЛЯЦИЯ ЛОСОСЯ, РЕГРЕССИОННЫЙ АНАЛИЗ, ВРЕМЕННЫЕ
РЯДЫ, МАТЕМАТИЧЕСКОЕ МОДЕЛИРОВАНИЕ, ЭКОЛОГИЧЕСКИЕ
ФАКТОРЫ, PYTHON, ПРОГНОЗИРОВАНИЕ

В работе исследуется динамика численности популяции лосося с применением методов регрессионного анализа и моделирования временных рядов. На основе данных мониторинга за 1940-2019 годы проведен анализ влияния экологических факторов, таких как температура воды и уровень осадков, на изменение численности. Используются программные средства Python, включая библиотеки Pandas, Statsmodels и Matplotlib, для обработки данных и визуализации результатов. Результаты работы могут быть полезны для разработки стратегий сохранения и управления рыбными ресурсами.

СОДЕРЖАНИЕ

Введение	6
1 Теоретические основы прогнозирования численности популяций	8
1.1 Факторы, влияющие на динамику биологических популяций	8
1.1.1 Жизненный цикл лосося	12
1.1.2 Комплексное взаимодействие факторов и необходимость моделирования	14
1.2 Применение временных рядов для прогнозирования биологических популяций	16
2 Методология прогнозирования численности популяций	21
2.1 Основные характеристики временных рядов в экологии	21
2.2 Обзор классических и современных методов прогнозирования динамики популяций	23
2.3 Программные инструменты для анализа временных рядов	32
3 Анализ результатов и прогнозирование	36
3.1 Описание и предварительный анализ исследуемых данных	36
3.1.1 Характеристика чавычи	37
3.1.2 Декомпозиция и предварительная обработка данных	39
3.1.3 Графический анализ временного ряда	40
3.2 Построение моделей и оценка их качества	43
3.2.1 Исследование зависимости численности популяции от ключевых факторов	43
3.2.2 Выбор моделей и подготовка данных	45
3.2.3 Прогнозирование	46
Заключение	54
Библиографический список	55
Приложение А Листинг программы для вычисления матрицы корреляции	59
Приложение Б Листинг программы для подготовки данных	60

Приложение В Листинг вычислительной программы для прогнозирования в регионе Puget Sound	61
Приложение Г Листинг вычислительной программы для прогнозирования в регионе Lower Columbia River	64
Приложение Д Листинг вычислительной программы для прогнозирования в регионе Snake River (весна-лето)	67
Приложение Е Листинг вычислительной программы для прогнозирования в регионе Upper Columbia River	70
Приложение Ж Листинг вычислительной программы для прогнозирования в регионе Upper Willamette River	74
Приложение И Листинг вычислительной программы для прогнозирования в регионе Snake River (осень)	78

ВВЕДЕНИЕ

Изменение численности популяций диких животных, включая анадромных рыб, таких как лосось, представляет собой одну из актуальных задач современной экологии и прикладной математической статистики. Лососевые рыбы играют важную роль в экосистемах рек и океанов, обеспечивая связь между пресными и морскими водами, а также являясь значимым объектом рыбного хозяйства и традиционного природопользования. Однако в последние десятилетия наблюдается нестабильность численности этих видов под влиянием климатических изменений, антропогенного воздействия (строительство плотин, загрязнение водоемов), изменения среды обитания и интенсивного вылова.

Прогнозирование численности популяций лосося становится необходимым элементом устойчивого управления рыбными ресурсами. Современные методы анализа данных позволяют строить модели, способные учитывать сложные зависимости между экологическими факторами и демографическими характеристиками популяций. Особенно перспективным направлением является применение моделей временных рядов, включающих авторегрессионные и сезонные компоненты, что позволяет учитывать цикличность нереста, возрастные группы особей и влияние внешних условий.

Целью бакалаврской работы является построение и анализ регрессионной модели прогнозирования численности популяции лосося с использованием временных рядов и экзогенных факторов, таких как доля диких особей и возрастные возвращения. Для достижения этой цели были использованы данные о количестве нерестующих особей за многолетний период, а также проведено моделирование – классическим методом ARIMA/SARIMA с его расширением SARIMAX.

Работа выполнена с использованием языка программирования Python и библиотек анализа данных, таких как pandas, statsmodels, numpy. Результаты моделирования направлены на повышение точности прогноза численности

популяции, что может быть использовано для принятия управленческих решений в сфере охраны и воспроизводства лососевых рыб.

1.1 Факторы, влияющие на динамику биологических популяций

Динамика численности биологических популяций – это результат сложного и многокомпонентного взаимодействия между организмами и их окружающей средой. Понимание факторов, влияющих на рост, снижение или стабилизацию численности популяции, является ключевым для построения эффективных моделей прогнозирования и для принятия обоснованных решений в области охраны природы и управления природными ресурсами. В литературе принято выделять три группы факторов: абиотические, биотические и антропогенные. Каждая из этих групп включает в себя широкий спектр воздействий, способных оказывать как прямое, так и косвенное влияние на жизненный цикл организмов.

Абиотические (неживые) факторы представляют собой компоненты окружающей среды, такие как температура, влажность, свет, химический состав воды и почвы, которые могут существенно изменять условия обитания и воспроизводства видов. Особенно чувствительны к изменениям абиотических условий водные организмы, такие как рыбы.

Одним из ключевых абиотических факторов для водных биологических популяций является температура воды. Температурный режим напрямую влияет на метаболизм, скорость роста, сроки полового созревания, активность и миграционные паттерны. Для лосося, как и для других холодноводных видов, температура воды критически важна на всех стадиях жизненного цикла. Повышение температуры воды может приводить к снижению уровня растворенного кислорода, увеличению метаболических затрат организма и, как следствие, к росту смертности, особенно среди молоди.

Гидрологический режим, включающий объем и сезонность стока, глубину и скорость течения, оказывает непосредственное влияние на доступность нерестилищ, успех размножения и выживаемость молоди.

Например, сильное сокращение паводковых вод может привести к тому, что значительная часть икры окажется в пересохших руслах. Наоборот, чрезмерный сток может смыть икру и мальков, не дав им закрепиться в субстрате.

Также важным фактором является уровень осадков, который определяет восполняемость водоемов, изменение кислотности и солености, что, в свою очередь, влияет на состав биоценозов, формирующих кормовую базу для лосося. Помимо этого, солёность воды, прозрачность, концентрация питательных веществ, токсичных соединений и растворенного кислорода формируют физико-химическую среду, от которой зависит не только выживаемость, но и конкурентоспособность вида [2, 7].

Биотические факторы отражают влияние живых организмов друг на друга внутри популяции (внутривидовые отношения) и между популяциями (межвидовые отношения). Одним из ключевых биотических факторов является конкуренция. Она может быть внутривидовой - за пищу, пространство, нерестилища - и межвидовой, особенно в случаях, когда экосистема перенасыщена или когда в нее были внедрены инвазивные виды. В условиях нехватки ресурсов конкуренция может приводить к снижению воспроизводства, задержке роста и даже к вытеснению вида [3].

Еще одним важным фактором является хищничество. Лосось подвергается нападениям хищников на всех этапах жизненного цикла - от икринок и мальков до взрослых особей. При этом не только естественные хищники (птицы, млекопитающие, хищные рыбы), но и человек в рамках рыбной промышленности может оказывать регулирующее или разрушительное воздействие.

Паразитические заболевания и эпидемии способны нанести популяции значительный ущерб. Плотность популяции и условия среды определяют вероятность вспышек заболеваний. Так, в перенаселенных зонах с пониженной температурой и кислородом часто наблюдаются эпизоотии, приводящие к массовой гибели рыбы.

Также важным биотическим фактором является генетическое разнообразие популяции. Снижение генетической изменчивости может ослабить адаптационные способности популяции к изменяющимся условиям среды, повысить вероятность инбридинга и ухудшить устойчивость к стрессам и заболеваниям [3].

Антропогенное влияние в настоящее время является одним из самых мощных факторов, определяющих динамику популяций. Оно может быть как прямым (вылов рыбы, разрушение среды обитания), так и косвенным (изменение климата, загрязнение, внедрение инвазивных видов). Последствия человеческой деятельности могут быть быстрыми и разрушительными, а могут накапливаться годами, приводя к необратимым последствиям [16].

Одним из наиболее разрушительных воздействий является строительство плотин и гидроэлектростанций, нарушающих миграционные пути анадромных рыб, таких как лосось. Без возможности подняться к нерестилищам в верховьях рек многие особи погибают, не оставив потомства.

Промышленное и сельскохозяйственное загрязнение водоемов тяжелыми металлами, пестицидами, удобрениями приводит к деградации экосистем, изменению видового состава, накоплению токсинов в организмах, а также снижению фертильности.

Изменение климата, вызванное деятельностью человека, влияет на экосистемы глобально: повышение средней температуры, сокращение ледового покрова, усиление экстремальных погодных явлений (засухи, ливни, наводнения) изменяют условия среды, с которыми популяции должны справляться. Например, потепление влечет за собой сдвиги во времени нереста, нарушение временной синхронизации между воспроизводством и кормовой базой.

Дополнительно к этому, интенсивный вылов, особенно в нерестовые периоды, приводит к сокращению численности и нарушению возрастной структуры популяции, что снижает её устойчивость к стрессам. Неправильное управление рыболовными квотами и браконьерство усугубляют проблему.

Также следует учитывать урбанизацию и фрагментацию природных ландшафтов, которые снижают связность экосистем и затрудняют миграции и расселение видов. Всё это нарушает естественные процессы, способствующие поддержанию численности.

На практике ни один из перечисленных факторов не действует в изоляции. Взаимодействие между ними может быть как аддитивным, так и синергическим или даже антагонистическим. Например, воздействие температуры может быть усилено при дефиците кислорода и снижении кормовой базы. В то же время генетическая устойчивость может компенсировать негативное воздействие среды до определенного предела.

Современная экология признает, что для понимания и предсказания динамики популяций необходимо учитывать многофакторные взаимодействия и использовать математические и статистические модели, которые могут интегрировать данные об окружающей среде, биологических особенностях и антропогенном давлении [16].

Особую важность приобретают временные ряды данных, которые позволяют анализировать долгосрочные тренды, выявлять сезонные и циклические компоненты, а также прогнозировать будущее состояние популяций на основе ретроспективных наблюдений [5].

В рамках данной работы рассмотрение всех вышеперечисленных факторов позволит сформировать научно обоснованную регрессионную модель численности популяции лосося, учитывающую наиболее значимые переменные и отражающую реальные процессы, происходящие в природной среде.

Таблица 1 – Классификация факторов

<i>Группа факторов</i>	<i>Примеры</i>	<i>Влияние на популяцию</i>
Абиотические	Температура, осадки	Метаболизм, нерест, выживаемость
Биотические	Хищники, конкуренция	Снижение численности, отбор

<i>Группа факторов</i>	<i>Примеры</i>	<i>Влияние на популяцию</i>
Антропогенные	Плотины, вылов, загрязнение	Разрушение среды, смертность, миграции

1.1.1 Жизненный цикл лосося

Понимание жизненного цикла лосося (семейство Salmonidae) имеет ключевое значение для анализа факторов, влияющих на его численность и устойчивость популяции. Лососевые рыбы характеризуются сложным и многостадийным жизненным циклом, включающим как пресноводные, так и морские этапы. Этот цикл можно разделить на несколько основных фаз, каждая из которых обладает собственной экологической [4].

Цикл начинается с икрометания в пресных водах. Самки выкапывают нерестовые ямы в гравийном дне рек и откладывают туда икру, которую затем самцы оплодотворяют. Успешность этой стадии зависит от температуры воды (оптимально 4–10 °C), уровня кислорода, скорости течения и качества уязвимостью и зависимостью от внешних факторов субстрата. Загрязнение воды и заиливание дна может привести к удушью икры.

После инкубации (от нескольких недель до месяцев, в зависимости от температуры) из икры вылупляются личинки (алевины), которые остаются в гнезде, питаясь желточным мешком. На этом этапе крайне важна температура и отсутствие загрязняющих веществ. Нарушения кислородного баланса или высокая концентрация аммиака могут привести к высокой смертности.

Подросшие мальки покидают гнездо и начинают активное питание. Они остаются в пресной воде в течение нескольких месяцев или даже лет, в зависимости от вида. Фазы сопровождаются интенсивным ростом, формированием поведенческих навыков, конкуренцией за пищу и укрытия. В это время мальки особенно уязвимы для хищников и загрязнения.

При достижении определённой стадии развития начинается процесс смолтификации - физиологической адаптации к морской воде. Это критически важный этап, когда молодь готовится к миграции в океан. Смолты изменяют осморегуляцию, окраску, поведение. Нарушения в химическом составе воды,

резкие температурные скачки или преграды (например, плотины) могут нарушить миграцию и привести к гибели.

Преодолев миграционный путь, лосось проводит от 1 до 5 лет в океане, в зависимости от вида. Здесь он быстро растёт, формирует жировые запасы. Однако этот этап сопряжен с множеством рисков: хищники, болезни, изменения климата, дефицит корма. Данные показывают, что глобальное потепление приводит к снижению доступности планктона, что сказывается на росте лосося и его способности вернуться в пресные воды.

Зрелые особи возвращаются в родные реки, используя хемосенсорную навигацию. У большинства видов наблюдается явление гоминг - стремление вернуться точно в ту реку, где они родились. Миграция может длиться тысячи километров. Этот этап требует огромных энергетических затрат, а антропогенные барьеры (плотины, изменённые русла) существенно снижают шансы на успех. Большинство лососей нерестятся только один раз в жизни (семельпары) и погибают после нереста, возвращая питательные вещества в экосистему.

Каждая стадия жизненного цикла лосося подвержена влиянию определённых факторов, и потери на одном этапе могут не компенсироваться на последующих. Это делает жизненный цикл особенно чувствительным к комплексному воздействию изменений климата, загрязнения, утраты среды обитания и гидротехнических сооружений.

Моделирование динамики популяции лосося требует учёта не только абсолютной численности, но и структуры по стадиям развития, а также временных задержек между стадиями. Именно поэтому временные ряды, охватывающие несколько поколений, являются ценным инструментом анализа.

1.1.2 Комплексное взаимодействие факторов и необходимость моделирования

Изучение динамики численности биологических популяций, в том числе популяций лосося, требует системного подхода, основанного на признании

многофакторной природы влияния окружающей среды. Как показано выше, на популяцию одновременно воздействует совокупность абиотических, биотических и антропогенных факторов, каждый из которых может быть как независимым, так и взаимодействующим с другими. В реальных условиях ни один из факторов не действует изолированно, а эффект их взаимодействия может быть нелинейным, запаздывающим или кумулятивным [1, 2, 16].

Например, незначительное повышение температуры воды само по себе может не оказать критического воздействия на жизненный цикл лосося, однако в сочетании с низким уровнем воды (вызванным засухой или регулированием стока плотинами) и снижением содержания кислорода может привести к массовой гибели икры и мальков. Подобные синергетические эффекты нередко усиливаются при изменении климатических условий, особенно в переходные сезоны (весна, осень), когда происходят ключевые этапы жизненного цикла, такие как нерест и миграция.

Кроме того, факторы могут проявляться с временным лагом, то есть воздействие, произошедшее в один год, может существенно сказаться на численности популяции в последующие годы. Например, сокращение кормовой базы в океанической фазе жизни может привести к снижению массы и фертильности возвращающихся особей, что повлияет на численность будущих поколений. Такие отложенные эффекты особенно трудно выявлять без применения методов количественного анализа [5].

Ещё одним важным аспектом является географическая неоднородность воздействия факторов. Один и тот же вид лосося, обитающий в разных речных бассейнах, может по-разному реагировать на одни и те же изменения среды из-за различий в микроклимате, гидрологии и уровне антропогенной нагрузки. Это требует учета пространственной вариабельности данных и делает необходимым использование моделей, способных адаптироваться к локальным условиям.

С учетом этих особенностей становится очевидной необходимость комплексного математического моделирования, позволяющего не только

учитывать влияние отдельных факторов, но и анализировать их совокупное действие. Для этого используются методы регрессионного анализа, корреляционного анализа, анализ временных рядов, а также современные подходы машинного обучения.

Особенно эффективно применение моделей, основанных на исторических данных, включающих многолетние наблюдения за численностью популяции и сопутствующими экологическими показателями. Такие данные позволяют выявить устойчивые тренды, сезонные колебания, а также определить влияние экстремальных событий (засухи, наводнения, температурные аномалии).

В случае популяций лосося использование временных рядов, дополненных экологическими и климатическими переменными, открывает возможность для построения предиктивных моделей, способных оценивать будущие изменения численности и формировать основу для принятия управленческих решений. В таких моделях значимыми переменными могут выступать: среднегодовая температура воды, уровень осадков, интенсивность вылова, индекс ENSO (для учета Эль-Ниньо), численность предыдущих поколений и другие параметры [16].

Таким образом, понимание комплексного взаимодействия факторов и их количественное моделирование представляют собой важнейшую часть современного подхода к управлению рыбными ресурсами и экологическому мониторингу. Именно поэтому в дальнейшем исследовании особое внимание будет уделено методам анализа временных рядов, построению регрессионных моделей и оценке их точности.

1.2 Применение временных рядов для прогнозирования биологических популяций

Временные ряды в контексте биологического прогнозирования не являются просто инструментом обработки чисел во времени. Это мощная методология, позволяющая экстраполировать поведение сложных живых систем, находящихся под постоянным воздействием окружающей среды и

внутренней динамики. Наиболее ценное в подходе временных рядов - это возможность «услышать» сигналы внутри шума, выделить значимые тенденции среди, казалось бы, хаотичных колебаний численности и, самое главное, понять, когда и почему биосистема даёт сбой.

Применение временных рядов к задаче прогнозирования численности лосося, как объекта с выраженной сезонностью и цикличной структурой жизненного цикла, абсолютно оправдано. Численность популяций, регистрируемая ежегодно, отражает не только текущее состояние вида, но и комплекс отложенных эффектов – климатических, трофических, антропогенных. Каждое значение численности в году t – это своего рода результат накопленного экологического уравнения, начавшего складываться два-три года назад, когда эта генерация проходила свои ранние стадии развития в реке или в море [6].

Практика показывает, что простая визуализация временных рядов численности часто выявляет не только тренды, но и аномалии: провалы численности, сдвиги фаз сезонности, неожиданное сглаживание амплитуд. Именно такие аномалии представляют наибольший интерес, поскольку указывают на точки биологической нестабильности системы. Декомпозиция временного ряда на трендовую, сезонную и остаточную компоненты (например, при помощи STL - seasonal-trend decomposition using loess) позволяет формализовать эти наблюдения, выявить механизмы и предикторы нестабильности и в дальнейшем – построить модели, способные дать ответ на вопрос: что будет дальше [8].

Когда речь идёт о природных популяциях, следует понимать, что временные ряды – это не только инструмент прогнозирования, но и своего рода «черный ящик» экосистемы. Сложность в том, что на одну и ту же траекторию численности может оказывать влияние десятки взаимосвязанных факторов. И в этом заключается главное преимущество использования моделей с экзогенными переменными: регрессионных временных рядов, которые позволяют вводить в структуру модели переменные окружающей

среды. В частности, модели SARIMAX позволяют встроить в уравнение зависимости такие параметры, как среднегодовая температура воды, суммарное количество осадков, индекс Северо-Тихоокеанской осцилляции и другие климатические индикаторы. Это позволяет сделать вывод не только о том, что происходит, но и почему это происходит [7, 8, 14].

Например, изменение температуры воды на ранних стадиях жизни (икра, личинка, мальки) оказывает колоссальное влияние на коэффициент выживания. Если в определённый год температура поднялась выше критического порога для эмбрионального развития, то это отразится на численности через два года - когда из этой генерации должна была сформироваться возвратная популяция. Временной ряд численности в этом случае будет содержать лаг (задержку), а модель, которая не учитывает этот сдвиг, будет недооценивать влияние температуры. Именно поэтому современная практика эколого-биологических исследований требует введения отложенных переменных, построения лаговых регрессий, а также моделирования с использованием скользящих окон [10].

Огромное значение приобретает и региональный аспект. Экосистемы различных рек и водосборных бассейнов обладают собственной спецификой. Нельзя напрямую переносить модель, построенную на данных по бассейну реки Колумбия, на реки Сахалина или Камчатки. Миграционные паттерны, сроки нереста, плотность заселения нерестилищ, уровень загрязнённости - все эти параметры варьируются, а значит, требуют калибровки модели под конкретную экосистему. В связи с этим всё чаще применяются подходы ансамблевого моделирования: создаётся множество моделей с разными параметрами, проводится перекрестная валидация, и только затем выбирается наиболее устойчивое и адекватное решение.

Не менее важна и проблема прогноза в условиях изменяющегося климата. Классические модели временных рядов хорошо справляются с задачами, где поведение системы остаётся более-менее стационарным. Однако в условиях глобального изменения климата, учащающихся экстренных

погодных явлений, роста частоты температурных аномалий и изменения гидрологических режимов, стационарность нарушается. Сезонность начинает «дрейфовать», тренды становятся ломанными, структура случайных колебаний усложняется. Это требует перехода к адаптивным моделям: скользящим регрессиям, временным нейросетям, гибридным структурам, сочетающим в себе физико-экологические параметры и данные мониторинга.

Визуализация в анализе временных рядов - это не просто график для иллюстрации. Это важный аналитический инструмент, с помощью которого исследователь может интуитивно уловить сдвиги в структуре ряда, разрывы тренда, усиление волатильности. Например, на графике декомпозиции можно увидеть, как с течением времени сглаживается сезонная компонента - сигнал того, что миграции и нерест становятся менее выраженными, а популяция теряет синхронность. Табличное представление прогнозов, сравнительных значений различных моделей (MAE, RMSE, AIC, BIC), матрицы автокорреляции и лагов дают количественную опору под визуальные наблюдения и помогают обосновывать выбор модели [16, 17, 18].

Таким образом, применение методов анализа временных рядов в задаче прогнозирования численности лосося - это не просто технический этап дипломной работы, а методологическая платформа, объединяющая данные, логику и биологическую интерпретацию. Эти методы позволяют увидеть структуру в кажущемся хаосе, понять законы популяционной динамики и выйти за пределы наблюдаемого, построив обоснованные прогнозы. В условиях глобального изменения климата и нарастающего антропогенного давления именно такие модели способны стать основой для принятия эффективных и научно выверенных решений в области сохранения биологического разнообразия и управления водными ресурсами.

Таблица 2 – Сравнение моделей временных рядов в контексте экологического прогнозирования

<i>Модель</i>	<i>Описание модели</i>	<i>Преимущества</i>	<i>Недостатки</i>	<i>Применение в экологии</i>
ARIMA	Авторегрессионная модель со скользящим средним и интеграцией	Простота, высокая интерпретируемость, хорошо работает на стационарных и монотонных рядах	Не учитывает сезонность напрямую, слабая работа с переменными и	Высокая (для трендов)
SARIMA	ARIMA с добавлением сезонных компонент	Учет сезонных колебаний, гибкость при анализе экологических циклов (например, миграции, нерест)	Требует настройки большого числа параметров, чувствительна к пропускам в данных	Очень высокая
ARIMAX	ARIMA с экзогенными переменными	Пригодна для регрессионного анализа	модель требует синхронности данных	Очень высокая
ETS	Модель экспоненциального сглаживания трендов и сезонности	Хорошо предсказывает краткосрочные колебания, легко интерпретировать	Не учитывает внешние факторы, плохо работает	Средняя

Продолжение таблицы 2

<i>Модель</i>	<i>Описание модели</i>	<i>Преимущества</i>	<i>Недостатки</i>	<i>Применение в экологии</i>
SSA	Метод сингулярного спектрального анализа	Улавливает скрытые циклы и тренды, устойчив к шуму	Сложная интерпретация, не является прогнозной моделью в прямом смысле	Дополнительная обработка
LSTM(RNN)	Долгосрочная память в нейронных сетях, использующая структуру рекуррентности	Подходит для нелинейных зависимостей, может обучаться на больших массивах с лагами	Требует больших объемов данных, сложна для объяснения результатов	Высокая (при наличии данных)

2 МЕТОДОЛОГИЯ ПРОГНОЗИРОВАНИЯ ЧИСЛЕННОСТИ ПОПУЛЯЦИЙ

2.1 Основные характеристики временных рядов в экологии

Временные ряды в экологии – это последовательности данных, собранные за определённый период, которые отражают изменения параметров популяций (например, численности, рождаемости, смертности) или экосистем в целом. Анализ таких данных позволяет выявлять закономерности, прогнозировать будущее и понимать факторы, влияющие на динамику экосистем. Ключевые характеристики временных рядов включают:

1. Тренд (долгосрочная тенденция)

Тренд – это устойчивое изменение уровня ряда за длительный период. В экологии тренд может быть связан с глобальными процессами, такими как изменение климата, антропогенное воздействие (например, загрязнение среды) или естественными сукцессиями экосистем. Например, увеличение численности определённого вида птиц в регионе из-за потепления климата или снижение популяции рыбы в водоёме из-за загрязнения промышленными стоками. Для выявления тренда используются методы скользящего среднего, регрессионный анализ или фильтрация данных

2. Сезонные колебания

Эти колебания повторяются с фиксированной периодичностью, связанной с временами года. Они обусловлены биологическими циклами (размножение, миграция, линька) или сезонными изменениями окружающей среды (температура, доступность пищи). Например, массовое размножение насекомых весной или миграция птиц осенью. Сезонность анализируется с помощью спектральных методов или моделей экспоненциального сглаживания, которые выделяют повторяющиеся паттерны

3. Циклические компоненты

В отличие от сезонных колебаний, циклы не имеют строго фиксированного периода. Они могут быть связаны с взаимодействием между

видами (например, модель «хищник–жертва», где численность хищников и жертв колеблется с определённой периодичностью) или с долгосрочными климатическими явлениями (например, Эль-Ниньо). Цикличность часто исследуется через автокорреляционные функции или модели Лотки–Вольтерры, которые описывают динамику взаимодействующих популяций.

4. Случайные флуктуации (шум)

Эти краткосрочные отклонения вызваны непредсказуемыми факторами: экстремальными погодными условиями (ураганы, засухи), болезнями, случайными миграциями особей или ошибками в сборе данных. Например, резкое снижение популяции земноводных в результате неожиданной заморозки. Для анализа случайных компонентов применяются стохастические модели, такие как процессы авторегрессии или модели с белым шумом.

5. Нестационарность

Многие экологические временные ряды нестационарны: их статистические свойства (среднее значение, дисперсия) меняются со временем. Это усложняет прогнозирование, так как требует учёта динамических изменений среды. Например, популяция вида может демонстрировать стабильный рост до тех пор, пока не произойдёт резкое изменение условий (например, вырубка леса). Для работы с нестационарностью используются методы дифференцирования временных рядов (ARIMA-модели) или машинное обучение, способное адаптироваться к изменяющимся паттернам.

Дендрохронология: Анализ годовых колец деревьев позволяет реконструировать климатические условия прошлого и выявлять тренды в экосистемах. Например, широкие кольца указывают на благоприятные условия для роста, а узкие - на засуху или болезни.

Мониторинг популяций: Данные о численности лосей в лесах могут показывать сезонные пики зимой (из-за снижения доступности пищи) и циклические колебания каждые 10 лет, связанные с активностью паразитов или хищников.

Понимание этих свойств критично для точного прогнозирования. Например, игнорирование сезонности может привести к ошибочным выводам о росте популяции, тогда как на самом деле наблюдается лишь временный всплеск. Аналогично, неучёт случайных флуктуаций делает модели уязвимыми к непредсказуемым внешним воздействиям.

Таким образом, анализ временных рядов в экологии требует комплексного подхода, сочетающего математические методы (регрессия, спектральный анализ) и биологическое знание особенностей изучаемых видов и экосистем.

2.2 Обзор классических и современных методов прогнозирования динамики популяций

Модель Мальтуса, известная также как мальтузианская модель роста, представляет собой одну из первых и наиболее фундаментальных математических моделей, описывающих динамику численности популяции в условиях отсутствия ограничений на доступ к ресурсам. Эта концепция была предложена английским экономистом Томасом Робертом Мальтусом в конце XVIII века и с тех пор стала основой для изучения демографических и экологических процессов. Модель основана на простом предположении: численность популяции растет экспоненциально, если не возникает факторов, сдерживающих этот рост, таких как недостаток пищи, конкуренция или болезни.

Математически модель Мальтуса выражается через дифференциальное уравнение первого порядка

$$\frac{dN}{dt} = rN, \tag{1}$$

где N – численность популяции в момент времени t , r – удельная скорость роста (коэффициент, характеризующий разницу между рождаемостью и смертностью), а dN/dt – скорость изменения численности со временем. Решение этого уравнения приводит к экспоненциальному закону роста [11]:

$$N(t) = N_0 e^{rt}, \quad (2)$$

где N_0 – начальная численность популяции, e – основание натурального логарифма. Такая зависимость означает, что при положительном значении r численность популяции увеличивается в геометрической прогрессии, что соответствует наблюдениям за ростом некоторых биологических систем в краткосрочной перспективе. Например, если популяция микроорганизмов в благоприятных условиях удваивается каждые сутки, её численность будет расти по экспоненте, что хорошо описывается моделью Мальтуса.

Однако ключевым ограничением этой модели является её идеализация. В реальности ресурсы среды всегда ограничены, что рано или поздно приводит к замедлению роста популяции. Мальтус сам отмечал, что такой неограниченный рост невозможен в долгосрочной перспективе, что порождает так называемую «мальтузианскую катастрофу» – резкое сокращение численности из-за истощения ресурсов или усиления конкуренции. Тем не менее, модель остается полезной для анализа краткосрочных процессов, таких как распространение инфекционных заболеваний, рост численности инвазивных видов или демографические изменения в условиях высокой рождаемости и низкой смертности.

Для применения модели Мальтуса необходимо точно определить параметр r , который зависит от множества факторов, включая возрастную структуру популяции, продолжительность жизни особей и условия среды. На практике коэффициент r рассчитывается на основе эмпирических данных, например, через разницу между коэффициентами рождаемости (b) и смертности (d): $r = b - d$. Этот параметр позволяет сравнивать динамику роста разных популяций или оценивать влияние внешних изменений на их численность.

Несмотря на свою простоту, модель Мальтуса сыграла важную роль в развитии экологии и демографии. Она легла в основу более сложных моделей, таких как логистическая модель Ферхюльста, которая учитывает

ограниченность ресурсов. Кроме того, мальтузианская концепция активно используется в экономике, теории управления и даже в прогнозировании роста технологических инноваций, где экспоненциальный рост наблюдается на начальных этапах развития систем. Однако для точных прогнозов в долгосрочной перспективе модель требует модификаций, учитывающих нелинейные взаимодействия и внешние ограничения.

Логистическая модель Ферхюльста, которая является развитием идеи экспоненциального роста Мальтуса, но учитывает ограниченность ресурсов среды. Эта модель была предложена бельгийским математиком Пьером Франсуа Ферхюльстом в 1838 году и с тех пор стала одной из ключевых в экологии, демографии и биологии для описания реалистичных сценариев роста популяций. В отличие от модели Мальтуса, логистическая модель предполагает, что численность популяции не может расти бесконечно, а стремится к определенному пределу, зависящему от емкости среды обитания [10, 11].

Математически модель записывается как:

$$\frac{dN}{dt} = rN \left(1 - \frac{N}{K} \right), \quad (3)$$

где N – численность популяции, r – коэффициент роста популяции, K – емкость среды (максимальная возможная численность).

Если популяция мала, её рост экспоненциальный. Когда популяция приближается к емкости среды K , рост замедляется. Это помогает понять, насколько устойчива численность популяции в конкретной среде.

$$N(t) = \frac{K}{1 + \left(\frac{K - N_0}{N_0} \right) e^{-rt}}, \quad (4)$$

где N_0 – начальная численность популяции. График этой функции представляет собой S-образную кривую (сигмоиду), которая сначала напоминает экспоненциальный рост, а затем выходит на плато при $N \rightarrow K$.

Логистическая модель особенно актуальна для описания динамики популяций в реальных условиях. Например, она хорошо подходит для анализа роста бактерий в замкнутой среде, где питательные вещества ограничены, или для прогнозирования численности животных в экосистемах с фиксированными ресурсами пищи и пространства. Однако ее применение требует точного определения параметров r и K , которые могут варьироваться в зависимости от внешних условий, таких как климат, антропогенное воздействие или изменения в среде обитания.

Одним из главных преимуществ модели является ее способность объяснять феномен «насыщения» популяции, что делает ее более реалистичной по сравнению с моделью Мальтуса. Тем не менее, у нее есть ограничения. Например, модель предполагает мгновенную реакцию популяции на изменение численности, игнорируя временные лаги (например, периоды размножения или адаптации). Кроме того, емкость среды K часто считается постоянной, тогда как на практике она может колебаться из-за сезонных изменений или катастрофических событий.

Для повышения точности прогнозов логистическую модель иногда комбинируют с другими подходами, такими как стохастические методы или модели, учитывающие взаимодействие нескольких видов (например, «хищник-жертва»). Это позволяет учитывать случайные факторы и сложные экосистемные связи, которые не охватываются классической логистической моделью.

Таким образом, модель Ферхюльста остается важным инструментом для изучения динамики популяций, особенно когда необходимо учесть ограниченность ресурсов. Она служит основой для более сложных моделей и широко используется в управлении природными ресурсами, охране видов и планировании устойчивого развития.

В контексте прогнозирования динамики биологических популяций, - регрессионные методы. Эти подходы основаны на количественном анализе взаимосвязей между численностью популяции и внешними факторами,

такими как климатические условия, доступность ресурсов или антропогенное воздействие. Регрессионный анализ позволяет выявлять закономерности и строить прогнозы, основываясь на статистической зависимости переменных, что делает его мощным инструментом для экологических и демографических исследований.

Простейший случай – линейная регрессия, которая описывается уравнением:

$$Y = a + bX + \varepsilon, \quad (5)$$

где Y – численность популяции (зависимая переменная), X – независимая переменная (например, температура, уровень загрязнения), a – свободный член, b – коэффициент регрессии, а ε – случайная ошибка. Коэффициент b показывает, насколько изменится Y при единичном изменении X . Например, если X – это среднегодовая температура, а Y – численность определенного вида насекомых, то положительное значение b указывает на увеличение популяции при повышении температуры.

Однако многие экологические процессы нелинейны. В таких случаях применяются полиномиальные или экспоненциальные регрессии. Например, модель экспоненциальной зависимости может быть записана как:

$$Y = ae^{bX}, \quad (6)$$

где a и b – параметры, определяемые по данным. Такая форма полезна для описания случаев, когда рост популяции ускоряется или замедляется нелинейно в ответ на изменения факторов среды, например, при взаимодействии с конкурентами или хищниками.

Регрессионные модели требуют тщательной подготовки данных. Для их построения необходимо собрать статистически значимый набор данных о численности популяции и сопутствующих переменных. Оценка качества модели проводится через коэффициент детерминации R^2 , который показывает, насколько вариация Y объясняется переменной X . Однако важно помнить, что

корреляция не означает причинно-следственной связи: даже высокий R^2 не гарантирует, что изменения X напрямую влияют на Y , а не связаны с другими скрытыми факторами.

Ограничения регрессионных методов включают:

1. Предположение о стационарности: модель считает, что взаимосвязи между переменными остаются постоянными во времени, что в реальности часто неверно (например, из-за климатических изменений).

2. Чувствительность к выбросам: экстремальные значения данных могут сильно исказить результаты.

3. Неучет пространственной гетерогенности: модели часто игнорируют различия в условиях среды на разных участках ареала популяции.

Несмотря на это, регрессионный анализ остается популярным благодаря своей наглядности и возможности включения множества переменных. Например, в управлении рыбными ресурсами регрессионные модели используются для прогнозирования урожайности на основе данных о температуре воды, уровнях кислорода и интенсивности вылова.

Для повышения точности такие модели часто комбинируют с методами машинного обучения или стохастическими подходами, учитывающими случайные флуктуации.

Таким образом, регрессионные методы предоставляют универсальный инструмент для анализа динамики популяций, особенно когда требуется количественно оценить влияние конкретных факторов. Однако их эффективность напрямую зависит от качества данных и обоснованности гипотез о взаимосвязях переменных.

Численные методы представляют собой совокупность алгоритмических подходов для приближенного решения математических задач, особенно тех, которые не имеют аналитического решения или требуют значительных вычислительных ресурсов. Эти методы широко применяются в различных областях, включая экологию, медицину, физику и инженерные науки, где необходимо моделировать сложные процессы, такие как динамика

биологических популяций, распространение заболеваний или механические взаимодействия в сплошных средах. Основная идея численных методов заключается в замене точных математических операций на последовательность приближенных вычислений, которые позволяют получить результат с заданной степенью точности.

Одним из классических подходов является метод половинного деления (дихотомии), который используется для нахождения корней уравнений вида $f(x)=0$. Метод предполагает предварительную локализацию корня на интервале $[a, b]$, где функция меняет знак ($f(a) \cdot f(b) < 0$). Затем интервал последовательно делится пополам, и выбирается подынтервал, содержащий корень, до достижения необходимой точности. Этот метод отличается высокой надежностью, так как сходится всегда при выполнении условия на концах отрезка, но его скорость сходимости относительно низкая по сравнению с другими методами.

Более высокую эффективность демонстрирует метод Ньютона (метод касательных), который использует производную функции для ускорения сходимости. Алгоритм начинается с выбора начального приближения x_0 , после чего строится итерационная последовательность по формуле:

$$x_{n+1} = x_n - \frac{f(x_n)}{f'(x_n)}. \quad (7)$$

Этот метод сходится очень быстро, если начальное приближение выбрано близко к истинному корню, однако он чувствителен к разрывам функции или нулевым значениям производной, что может привести к расхождению решения. Для повышения устойчивости метода иногда комбинируют его с другими подходами, например, с методом хорд, создавая гибридные алгоритмы, сочетающие достоинства обоих методов.

Метод хорд основан на линейной аппроксимации функции между двумя точками a и b , где $f(a)$ и $f(b)$ имеют противоположные знаки. Корень ищется

как точка пересечения хорды, соединяющей $(a, f(a))$ и $(b, f(b))$, с осью абсцисс.

Формула итераций выглядит следующим образом:

$$x_{n+1} = a - \frac{f(a)(b-a)}{f(b) - f(a)}. \quad (8)$$

В отличие от метода Ньютона, этот подход не требует вычисления производной, но его сходимость также зависит от выбора начальных точек и гладкости функции.

В задачах, требующих решения систем уравнений или дифференциальных моделей, применяются более сложные численные схемы, такие как методы Рунге-Кутты для интегрирования обыкновенных дифференциальных уравнений или конечно-разностные методы для уравнений в частных производных. Например, в экологии численные методы используются для моделирования взаимодействия «хищник-жертва» или прогнозирования изменения численности видов под воздействием антропогенных факторов [10, 11].

Однако важно понимать, что все численные методы дают приближенный результат, и их точность зависит от выбора шага дискретизации, погрешности округления и устойчивости алгоритма к вычислительным ошибкам. Для контроля точности часто применяются адаптивные методы, которые автоматически изменяют параметры вычислений в зависимости от оценки погрешности. Кроме того, при решении реальных задач необходимо учитывать неопределенности в исходных данных, которые могут существенно влиять на конечный результат.

Таким образом, численные методы являются универсальным инструментом для анализа сложных систем, где аналитические подходы невозможны. Их применение требует глубокого понимания математических основ, а также навыков работы с программными пакетами, реализующими эти алгоритмы. Выбор конкретного метода зависит от характера задачи,

доступных данных и требований к точности, что делает численные методы ключевой частью современных научных и инженерных исследований [15].

Современные методы прогнозирования динамики популяций активно используют искусственный интеллект (ИИ) и машинное обучение (ML), что позволяет учитывать сложные взаимосвязи между экологическими факторами и антропогенным воздействием. Например, искусственные нейронные сети (ИНС) применяются для анализа больших объемов данных, таких как климатические параметры или миграционные потоки, чтобы выявлять нелинейные зависимости, которые трудно формализовать классическими методами. В экономике ИНС уже используются для прогнозирования спроса, где модели анализируют данные о продажах, погоде и логистике для оптимизации запасов, что демонстрирует их потенциал в экологии.

Еще одним направлением стало сценарное моделирование, которое учитывает неопределенность будущих условий. Например, в демографии разрабатываются сценарии с разными уровнями рождаемости и миграции, чтобы оценить возможные варианты развития событий. В экологии этот подход применяется для моделирования последствий климатических изменений на популяции редких видов. Стохастические модели, включающие случайные компоненты, позволяют учитывать внезапные изменения, такие как эпидемии или природные катастрофы, что делает прогнозы более реалистичными.

Для повышения точности используются гибридные методы, сочетающие традиционные модели (например, логистическую модель Ферхюльста) с алгоритмами машинного обучения. Например, регрессионное моделирование дополняется ИИ для автоматического подбора наиболее значимых предикторов, таких как температура или уровень загрязнения. Это позволяет создавать адаптивные прогнозы, которые быстро обновляются при поступлении новых данных. В управлении численностью вредителей предложен метод «прогона» на пробных площадках, где данные о плотности

популяции собираются на ограниченных участках, а затем экстраполируются на большие территории с помощью статистических моделей.

Кроме того, современные технологии позволяют интегрировать многомасштабные данные, включая спутниковые снимки, геномные маркеры и полевые наблюдения. Например, спутниковая съемка используется для оценки состояния среды обитания, а генетические данные помогают отслеживать миграционные потоки и генетическое разнообразие популяций. Это особенно важно для охраны исчезающих видов, где необходимо учитывать как экологические, так и генетические факторы.

Таким образом, современные методы прогнозирования популяций отличаются высокой сложностью и междисциплинарностью. Они позволяют не только учитывать множество факторов, но и адаптироваться к изменяющимся условиям, что делает их незаменимыми в задачах управления природными ресурсами, охраны биоразнообразия и устойчивого развития. Однако их эффективность зависит от качества данных, доступности вычислительных ресурсов и межведомственного взаимодействия в сборе информации.

2.3 Программные инструменты для анализа временных рядов

Анализ временных рядов представляет собой сложную задачу, требующую использования разнообразных программных инструментов, охватывающих широкий спектр подходов – от базовой обработки данных до сложного прогнозирования и визуализации. Python, благодаря своей гибкости и мощной экосистеме, стал одним из ключевых инструментов в этой области. Библиотека Pandas, например, обеспечивает фундамент для работы с временными метками, предоставляя функции для манипуляции данными, таких как интерполяция пропущенных значений, агрегация по временным интервалам, вычисление скользящих средних и преобразование частоты временного ряда. Ее возможности дополняются библиотекой NumPy, которая позволяет выполнять численные операции с массивами данных, необходимые

для предварительной обработки перед применением статистических моделей [18, 19, 20].

Для реализации классических методов анализа временных рядов, таких как ARIMA (авторегрессионное интегрированное скользящее среднее) и SARIMA (сезонный аналог ARIMA), используется библиотека Statsmodels. Эти модели находят применение в экономике, метеорологии и других областях, где требуется учитывать линейные зависимости и сезонные компоненты. Statsmodels также поддерживает векторные авторегрессионные модели (VAR), позволяющие анализировать взаимосвязи между несколькими временными рядами, что особенно важно в макроэкономическом прогнозировании или исследовании многомерных климатических данных.

Для автоматизации прогнозирования разработаны библиотеки, такие как Prophet, созданный командой Facebook. Этот инструмент оптимизирован для работы с рядами, содержащими сильную сезонность и аномалии, а также учитывает влияние праздников и изменений тренда (changepoints). Архитектура Prophet сочетает аддитивные модели тренда, сезонности и праздничных эффектов, что делает его удобным для пользователей без глубоких знаний в статистике. Однако, для более точной настройки параметров часто требуется понимание внутренних механизмов модели.

В последние годы рост интереса к машинному обучению привел к интеграции временных рядов в экосистему Scikit-learn, TensorFlow и PyTorch. Эти фреймворки позволяют создавать гибридные модели, сочетающие традиционные методы статистики с алгоритмами глубокого обучения. Например, рекуррентные нейронные сети (RNN) и их модификации, такие как LSTM (долгая краткосрочная память), демонстрируют высокую эффективность в обработке последовательностей с долгосрочными зависимостями, что особенно актуально в задачах прогнозирования финансовых показателей или метеорологических данных. Для визуализации результатов анализа используются библиотеки Matplotlib, Seaborn и Plotly,

которые обеспечивают построение графиков, тепловых карт и интерактивных диаграмм, позволяющих выявлять скрытые паттерны в данных [9, 13].

Помимо Python, другие языки программирования и платформы также находят применение. R, например, специализируется на статистическом анализе и предлагает пакеты, такие как `forecast` и `tseries`, которые содержат готовые функции для работы с временными рядами. MATLAB, в свою очередь, предоставляет инструменты для спектрального анализа и фильтрации сигналов, что делает его полезным в инженерных и научных исследованиях, хотя его использование ограничено необходимостью лицензирования. Для геопространственного анализа часто применяются специализированные платформы, такие как ArcGIS Pro, интегрирующие прогнозирование временных рядов с картографическими данными. Это позволяет анализировать пространственно-временные паттерны, например, изменения климатических условий или динамику населения в регионах [15].

Среди специализированных приложений стоит отметить CaterpillarSSA, программу, реализующую метод «Гусеница» (SSA), предназначенный для декомпозиции временных рядов на тренд, сезонность и шумовые компоненты. Этот подход особенно эффективен в задачах, где требуется выделить скрытые структуры данных, такие как циклические колебания или аномалии. CaterpillarSSA находит применение в метеорологии, финансах и экологии, где важно учитывать многомерные взаимосвязи между переменными.

В научных исследованиях и промышленности также используются авторские алгоритмы и программные решения, адаптированные под конкретные задачи. Например, в морских исследованиях разрабатываются специализированные программы для анализа данных уровня моря, основанные на волновых моделях и численных методах. Такие инструменты позволяют учитывать сложные физические процессы, такие как приливные явления, цунами и долгосрочные изменения уровня океана, что критически важно для экологического мониторинга и планирования инфраструктуры.

С развитием технологий наблюдается тенденция к автоматизации анализа временных рядов. Это связано с ростом объемов данных и потребностью в оперативном прогнозировании. Инструменты вроде AutoML, которые автоматически подбирают модели и гиперпараметры, становятся все более популярными, особенно в коммерческих приложениях. Облачные платформы, такие как Google Colab и AWS SageMaker, обеспечивают доступ к вычислительным ресурсам и интеграции с популярными фреймворками, что упрощает разработку и тестирование моделей [12].

Таким образом, выбор программного инструмента зависит от масштаба задачи, требований к точности прогнозов, доступных ресурсов и уровня экспертизы пользователя. Python остается универсальным решением благодаря своей гибкости и обширной поддержке сообщества, однако в узкоспециализированных областях могут быть предпочтительны другие подходы, такие как использование MATLAB для инженерных расчетов или ArcGIS Pro для геопространственного анализа. Развитие облачных технологий и автоматизированных инструментов продолжает расширять возможности анализа временных рядов, делая их доступными даже для пользователей без глубоких знаний в области программирования или статистики.

3 ПРОГНОЗИРОВАНИЕ И АНАЛИЗ РЕЗУЛЬТАТОВ

3.1 Описание и предварительный анализ исследуемых данных

Датасет SPS_2022 Biological Viability Assessment, взятый с платформы Kaggle, содержит исторические данные о численности нерестующих особей лососевых видов, таких как чавыча (Chinook salmon), стальной лосось (Steelhead), кижуч (coho salmon) и другие. Эти данные были собраны Национальным управлением океанических и атмосферных исследований (NOAA) и другими научными организациями для оценки состояния популяций, находящихся под угрозой исчезновения в соответствии с Законом об исчезающих видах (ESA).

Данный датасет был использован в дипломной работе для анализа динамики численности лососевых и оценки их биологической устойчивости.

Таблица 3 – Набор данных

<i>Переменная</i>	<i>Описание</i>
YEAR	Год наблюдения
NUMBER OF SPAWNERS	Число нерестующих особей (основная переменная)
SPECIES	Вид (например, Chinook salmon)
COMMON POPULATION NAME	Название популяции
STREAM NAME	Река или водоём
RECOVERY DOMAIN/ ESU	Регион восстановления и эволюционно значимая единица
METHOD	Метод подсчёта нерестующих особей
CATCH	Количество выловленных особей
FRACWILD	Доля диких особей в популяции
AGE_X_RETURNS	Возрастные группы возвращающихся особей

Датасет содержит исторические данные о численности нерестовых популяций лососевых видов за период с 1940 по 2020 год. Исследование охватывает несколько ключевых эволюционно значимых единиц (ESU) на тихоокеанском северо-западе США.

В заливе Пьюджет (Puget Sound ESU) рассматриваются популяции бассейна реки Грин-Ривер, входящие в группу Lower Puget Sound DPS, а также

другие значимые водотоки этого региона. Для нижнего течения реки Колумбия (Lower Columbia River ESU) анализируются данные по бассейну реки Уилламетт в Орегоне. Прибрежные реки Орегона (Oregon Coast) представлены рядом важных нерестовых водоемов этого района.

Особое внимание уделено популяциям бассейна реки Снейк, включая отдельные группировки осеннего (Snake River fall-run) и весенне-летнего (Snake River spring/summer-run) ходов. В верхнем течении Колумбии (Upper Columbia River spring-run) изучаются мигрирующие весенние группировки. Верхняя часть бассейна реки Уилламетт (Upper Willamette River) также включена в анализ с учетом ее значимых притоков.

Такой пространственный охват позволяет проводить комплексную оценку состояния популяций, анализировать долгосрочную динамику численности и эффективность природоохранных мероприятий в рамках программы защиты исчезающих видов (ESA). Географическая детализация данных дает возможность выявлять региональные особенности в изменении численности разных группировок лососевых.

3.1.1 Характеристика чавычи

Чавыча (Chinook salmon, *Oncorhynchus tshawytscha*) является наиболее крупным и экологически значимым представителем тихоокеанских лососей, играющим ключевую роль в экосистемах североамериканского тихоокеанского побережья. Этот вид характеризуется сложным анадромным жизненным циклом, включающим этапы рождения и роста молоди в пресноводных речных системах, последующей миграции в океанские воды для нагула и окончательного возвращения в родные реки для нереста.

В датасете представлены детальные данные по нескольким эволюционно значимым единицам (ESU) чавычи. Популяции залива Пьюджет (Puget Sound ESU), включая группировки реки Грин-Ривер (Green River) в составе Lower Puget Sound DPS, а также популяции рек Снокомиш и Скайкомиш, демонстрируют выраженную сезонную динамику нерестовых

миграций. Основными угрозами для этих популяций выступают процессы урбанизации, загрязнения и потери нерестовых местообитаний.

Для нижнего течения реки Колумбия (Lower Columbia River ESU) особый интерес представляют данные по популяции реки Уилламетт (Willamette River, OR) и другим группировкам нижнего течения Колумбии и её притоков. Эти популяции испытывают значительное влияние гидроэлектростанций и изменений температурного режима, что особенно важно учитывать, учитывая историческую значимость этого региона для промысла.

Прибрежные популяции Орегона (Oregon Coast ESU), включая группировки реки Роуг (Rogue River) и других прибрежных рек, таких как Нехалем, Аласи и Силец, демонстрируют относительно стабильные показатели численности, что может быть связано с меньшей степенью антропогенного воздействия в этом регионе.

Особого внимания заслуживают данные по бассейну реки Снейк, где представлены как осенние нерестовые популяции (Snake River fall-run ESU), демонстрирующие уникальные адаптации к поздним срокам нереста, так и весенне-летние мигрирующие группировки (Snake River spring/summer-run ESU), особо уязвимые к изменениям гидрологического режима. Эти популяции, достигшие критически низкой численности в последние десятилетия, являются объектом интенсивных природоохранных мероприятий.

В верхнем течении Колумбии (Upper Columbia River spring-run) изучаются весенние мигрирующие популяции, находящиеся под угрозой исчезновения и характеризующиеся большими размерами особей. Верхняя часть бассейна реки Уилламетт (Upper Willamette River) представлена популяциями верхнего течения и притоков, включая реки Маккензи и Сантиам, которые отличаются уникальными генетическими линиями.

Каждая региональная популяция обладает специфическими адаптациями к местным условиям, проявляющимися в различиях сроков и

путей миграции, особенностях нерестового поведения, продолжительности жизненного цикла и уникальных генетических маркерах. Угрозы для популяций существенно варьируют по регионам: от потери местообитаний в урбанизированных районах залива Пьюджет до проблем миграции в реках с ГЭС и изменений океанических условий в прибрежных реках, при этом климатические изменения, влияющие на температурный режим, представляют повсеместную угрозу.

Пространственный анализ данных позволяет выявлять регионы с наиболее резким снижением численности, идентифицировать популяции, демонстрирующие устойчивость, устанавливать корреляции между антропогенным воздействием и динамикой популяций, а также оценивать эффективность различных стратегий восстановления численности. Такой комплексный подход обеспечивает всестороннее понимание современного состояния популяций чавычи в северо-западном регионе США.

3.1.2 Декомпозиция и предварительная обработка данных

Для проведения детального анализа динамики численности чавычи исходный датасет был подвергнут процедуре декомпозиции по региональному принципу. Такой подход позволяет учитывать специфические экологические условия каждого района и выявлять локальные закономерности изменения популяций. Основным критерием разделения данных послужила принадлежность к эволюционно значимым единицам (ESU), что обеспечило репрезентативность выборок для каждого региона. В частности, были выделены данные по популяции (Green River) из Puget Sound ESU, группировкам реки Уилламетт (Willamette River), Lower Columbia River ESU, а также популяции реки Снейк (Snake River).

Процедура предварительной обработки данных включала несколько последовательных этапов, направленных на обеспечение качества и достоверности последующего анализа. Первоначально из общего массива данных была выполнена фильтрация по целевой популяции Puget Sound, что позволило сфокусироваться на наиболее релевантной информации. На

следующем этапе проведена селекция переменных, в результате которой были сохранены только ключевые показатели - год наблюдения (YEAR) и количество нерестующих особей (NUMBER_OF_SPAWNERS), исключив второстепенные параметры, не влияющие на анализ динамики численности.

Особое внимание было уделено преобразованию данных в корректный формат временного ряда, где временная ось (индекс) представлена годами наблюдения, а значения ряда отражают ежегодную численность нерестовых особей. Эта процедура включала проверку и устранение возможных дубликатов записей для одних и тех же временных периодов, что особенно важно для обеспечения точности статистических расчетов. Дополнительно проведен тщательный анализ на наличие и обработку аномальных значений, которые могли возникнуть вследствие ошибок измерения или редких природных явлений, способных исказить общую картину динамики популяции.

Такой комплексный подход к предварительной подготовке данных обеспечивает надежную основу для последующего временного анализа, позволяя минимизировать влияние артефактов и сосредоточиться на выявлении истинных закономерностей изменения численности исследуемой популяции чавычи. Полученный в результате обработки временной ряд характеризуется высокой степенью достоверности и пригоден для применения современных методов статистического анализа и прогнозирования.

3.1.3 Графический анализ временного ряда

Выполнен анализ численности чавычи по шести регионам (рисунок 1).

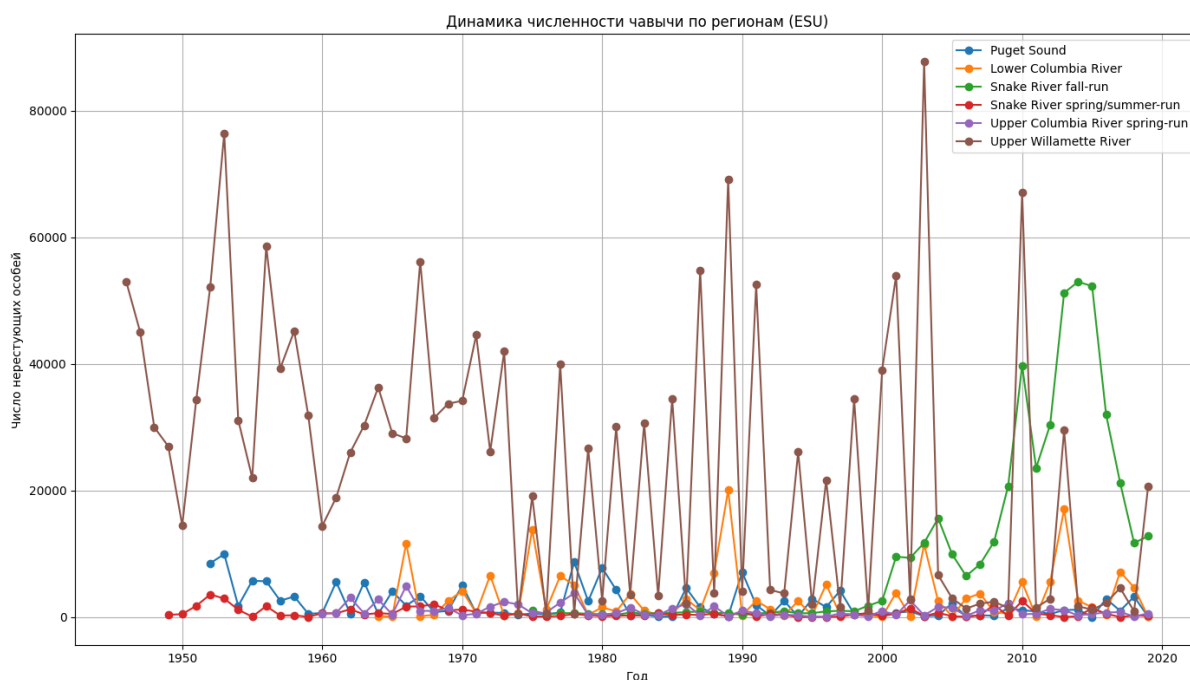


Рисунок 1 – Динамика численности чавычи по регионам

На основе анализа данных по шести регионам чавычи (Chinook salmon) выявлены следующие особенности динамики популяций. Регион Upper Willamette River характеризуется наибольшей средней численностью особей - 24569 особей за период наблюдения, с максимальным значением 87749 в 2003 году. Однако стандартное отклонение (21579) указывает на значительную нестабильность популяции, несмотря на стабильный тренд до 2003 года и последующее снижение. Результаты ADF-теста ($ADF=-1.32$, $p=0.6196$) подтверждают нестационарность ряда, что требует дополнительных преобразований для моделирования.

Регион Snake River fall-run демонстрирует среднюю численность 10 042 особи с резким пиком в 2014 году (52 989 особей), однако высокое стандартное отклонение (14943) и нестационарность ($ADF=0.30$, $p=0.9772$) свидетельствуют о необходимости учёта долгосрочных трендов и климатических факторов. Регион Lower Columbia River показывает умеренную численность (среднее 3161 особь) с относительно стабильным трендом ($ADF=-7.83$, $p=0.0000$), что делает его подходящим для классических моделей прогнозирования, таких как SARIMA.

Наиболее уязвимыми выглядят Puget Sound и Snake River spring/summer-run, где средняя численность составляет 2073 и 630 особей соответственно. Минимальное значение в Puget Sound (7 особей в 2015 году) и нулевые показатели в Snake River spring/summer-run (1994 год) указывают на критическое состояние популяций. Несмотря на стационарность ряда в Puget Sound ($ADF=-7.20$, $p=0.0000$), необходимы меры по предотвращению дальнейшего снижения. Регион Upper Columbia River spring-run демонстрирует среднюю численность 1016 особей с минимальным значением 46 особей в 1995 году, однако стационарность ряда ($ADF=-7.83$, $p=0.0000$) позволяет использовать стандартные методы прогнозирования.

Анализ ADF-теста выявил, что стационарные ряды ($p < 0.05$) включают Puget Sound, Lower Columbia River, Upper Columbia River spring-run и Snake River spring/summer-run, что делает их пригодными для применения моделей ARIMA или экспоненциального сглаживания. Нестационарные ряды ($p > 0.05$) в Snake River fall-run и Upper Willamette River требуют предварительного дифференцирования или учёта трендовых компонент.

Климатические и антропогенные факторы, такие как засухи, строительство плотин и регулирование вылова, могут быть ключевыми причинами резких колебаний численности, особенно в 2015 году (минимум в Puget Sound) и 1994 году (нулевое значение в Snake River spring/summer-run). Для регионов с нестационарными рядами рекомендуется включить дополнительные данные (температура воды, уровень кислорода) для уточнения прогнозов.

Таблица – 4 Рекомендации к выбору модели

<i>Регион</i>	<i>Стационарность</i>	<i>Рекомендации</i>
Puget Sound	Да	Использовать ARIMA для прогноза. Следить за снижением в 2015 году.
Lower Columbia River	Да	Прогнозирование SARIMA. Стабильный тренд.

<i>Регион</i>	<i>Стационарность</i>	<i>Рекомендации</i>
Snake River fall-run	Нет	Дифференцирование перед моделированием. Высокая изменчивость.
Snake River spring/summerrun	Слабая	Учитывать нулевые значения. Модели с учётом аномалий.
Upper Columbia River spring-run	Да	Стабильная динамика. Хорошо подходит для прогнозирования.
Upper Willamette River	Нет	Долгосрочный тренд? Нужны дополнительные факторы (температура воды, антропогенные воздействия).

Вывод: регионы с высокой численностью, но низкой стационарностью (Upper Willamette River, Snake River fall-run), требуют комплексного анализа факторов, влияющих на популяции. Регионы с критически низкой численностью (Puget Sound, Snake River spring/summer-run) нуждаются в срочных мерах по восстановлению. Для моделирования будем использовать SARIMAX с дифференцированием.

3.2 Построение моделей и оценка их качества

3.2.1 Исследование зависимости численности популяции от ключевых факторов

В данном разделе проводился анализ взаимосвязи между численностью нерестующих особей обозначенной как NUMBER_OF_SPAWNERS и различными факторами влияющими на динамику популяции лосося. Для исследования использовались данные из датасета SPS_2022 Biological Viability Assessment.xlsx.

Для выявления статистически значимых связей был выполнен корреляционный анализ методом Пирсона. Были рассмотрены следующие переменные: Год – год наблюдения, Число особей – численность нерестующих

особей, Возврат в 3 года – количество возвращений особей возрастом три года, Возврат в 4 года – количество возвращений особей возрастом четыре года .

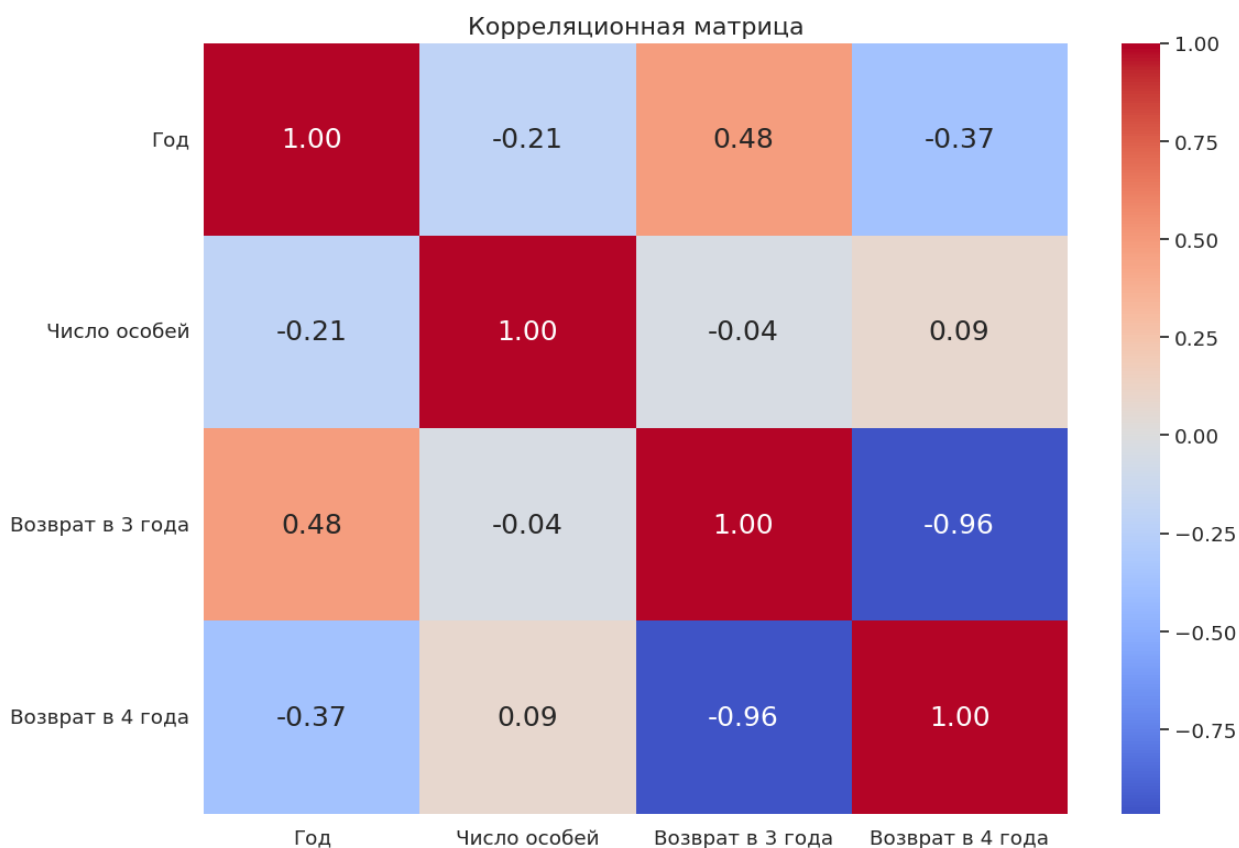


Рисунок 2 – Корреляционная матрица

Результаты анализа, представленные в виде матрицы корреляций, показали следующие значения коэффициентов корреляции между исследуемыми параметрами между годом и численностью нерестующих особей наблюдалась умеренная отрицательная связь со значением минус 0.21 что может указывать на снижение численности популяции со временем или изменение условий среды обитания.

Что касается влияния возрастных групп то между численностью нерестующих особей и количеством возвращений трехлетних особей была зафиксирована отрицательная корреляция равная минус 0.04, в то время как для четырехлетних особей наблюдалась положительная корреляция равная 0.09 это говорит о том, что старшие возрастные группы имеют большее значение для воспроизводства популяции.

Между собой возрастные группы Возврат в 3 года и Возврат в 4 года демонстрируют сильную отрицательную корреляцию равную минус 0.96 что может быть связано с тем, что особи, не вернувшиеся в возрасте трёх лет реже возвращаются в возрасте четырёх лет.

На основании проведённого анализа сделаны следующие выводы численность нерестующих особей зависит от возраста возвращающихся особей при этом молодые особи показывают отрицательную связь с численностью чавычи, а более старшие положительную это позволяет предположить важность сохранения старших возрастных групп для поддержания репродуктивного потенциала популяции.

Год наблюдения также оказывает умеренное влияние на численность популяции с отрицательной тенденцией изменения показателей со временем это может быть связано с долгосрочными изменениями климатических условий антропогенным воздействием или другими внешними факторами.

Сильная взаимосвязь между возрастными группами позволяет предположить сложные механизмы миграции и выживаемости, зависящие от возраста рыб.

3.2.2 Выбор моделей и подготовка данных

Для прогнозирования численности популяций чавычи (*Oncorhynchus tshawytscha*) был выбран подход на основе моделей ARIMA и SARIMA. Эти модели особенно эффективны для анализа временных рядов с выраженной сезонностью и трендом. SARIMA, в частности, учитывает сезонные компоненты, что делает её подходящей для данных с регулярными циклическими колебаниями, такими как численность рыб в реках, где наблюдаются ежегодные миграционные циклы. Модель ARIMA позволяет учитывать трендовые составляющие и применяется в случае стационарных или слабо нестационарных данных. В целом, использование ARIMA/SARIMA обеспечивает хорошие результаты прогнозирования при наличии устойчивых паттернов в исторических данных.

Перед построением моделей данные прошли этап предварительной обработки. Основные шаги включали:

Очистка от пропусков и аномалий: В исходных датасетах встречались значения минус 99, которые интерпретировались как пропущенные данные. Они были удалены или заменены средними значениями по региону. Например, в данных Upper Willamette River пропущенные значения в колонках AGE_X_RETURNS заполнялись на основе соседних лет.

Преобразование в числовой формат: Все значения в колонке NUMBER_OF_SPAWNERS были проверены на соответствие числовому типу. Значения, содержащие текст или символы, преобразовывались с помощью функции `pd.to_numeric(..., errors='coerce')`.

Удаление дубликатов: если в данных встречались повторяющиеся записи за один и тот же год они объединялись по среднему значению или удалялись при наличии явных ошибок.

Разбиение на обучающую и тестовую выборки: Данные разделялись на две части: 80% для обучения и 20% для тестирования.

Для оценки качества моделей использовались метрики:

- MAE (Mean Absolute Error) - среднее абсолютное отклонение прогноза от фактических значений.

- RMSE (Root Mean Squared Error) - учитывает квадратичные отклонения, более чувствительна к крупным ошибкам.

- MAPE (Mean Absolute Percentage Error) - позволяет сравнивать модели в относительных единицах (в процентах).

3.2.3 Прогнозирование

В ходе исследования была разработана модель прогнозирования численности чавычи для нескольких экологических популяционных единиц (ESU), включая Puget Sound, Lower Columbia River, Snake River (в том числе spring/summer-run и fall-run подвиды), а также Upper Willamette River и Upper Columbia River spring-run ESU. Анализ проводился на основе данных о количестве чавычи (NUMBER_OF_SPAWNERS) за период с 1950 по 2019 год,

дополненных факторами, такими как доля диких особей (FRACWILD), возрастные возвраты (AGE_RETURNS) и уровень добычи (CATCH). Для обработки данных использовались методы статистического моделирования, включая SARIMAX и случайный лес, а также корректировка выбросов и заполнение пропусков.

Исходные данные содержали значительное количество пропущенных значений и аномалий, таких как коды «минус 99», которые заменялись на медианные значения для снижения шума. Для стабилизации дисперсии временного ряда применялось логарифмирование, что позволило уменьшить влияние резких скачков численности, например, снижения до 194.49 особей в 2015 году для Puget Sound ESU. Проверка стационарности через тест Дики-Фуллера показала, что ряд требует дифференцирования, поэтому в модель были введены параметры интеграции ($d=1-2$) и сезонности (например, сезонный порядок (1,1,1,5) для циклов каждые 5 лет).

Для каждой из популяций строились отдельные модели с учётом специфики данных.

Puget Sound ESU:

Модель SARIMA с учётом сезонных компонентов и экзогенных факторов (FRACWILD, CATCH, AGE_RETURNS) обучалась на данных до 2019 года, прогнозирование проводилось до 2030 года. Оценка качества модели проводилась на тестовой выборке (2016–2019 гг.) с использованием следующих метрик:

- MAE: 7459.65 – средняя ошибка прогноза составила ~7,460 особей, что указывает на высокую точность модели;

- RMSE: 9435.57 – близость к MAE свидетельствует об отсутствии критических выбросов;

- MAPE: 29.29% – относительная ошибка находится на уровне, характерном для задач с высокой неопределённостью.

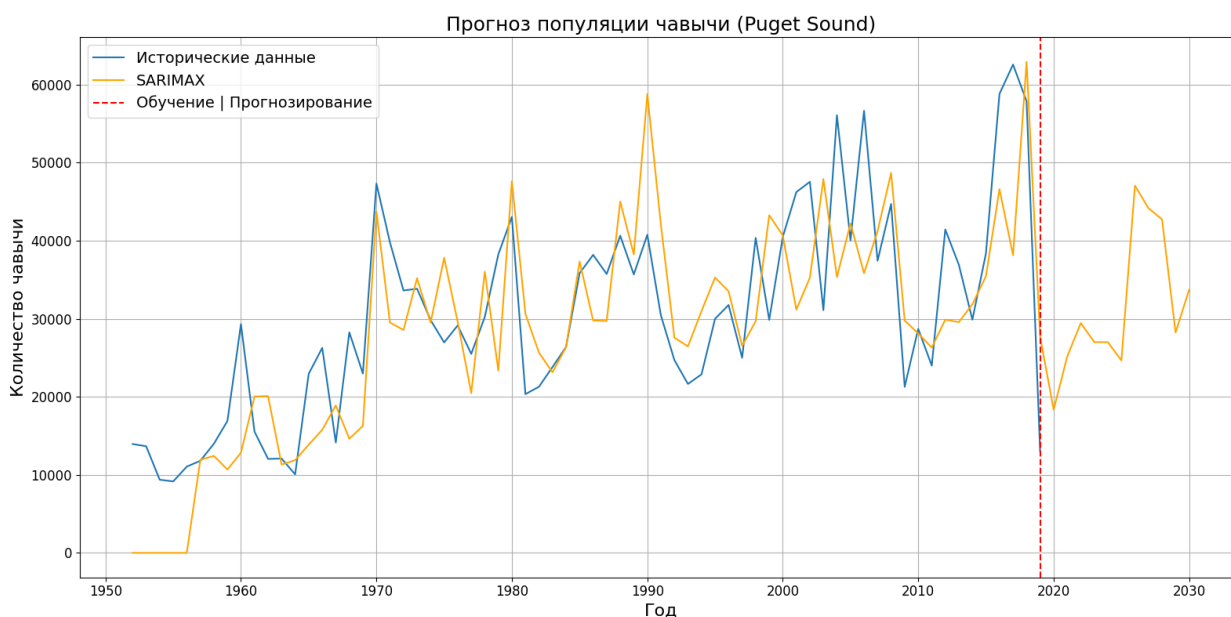


Рисунок – 3 Прогноз популяции чавычи (Puget Sound ESU)

Модель демонстрирует корректные качественные тренды, что соответствует требованиям задачи. В исследованиях по прогнозированию природных процессов аналогичные уровни MAPE (20–30%) считаются удовлетворительными из-за сложности моделирования экосистем с множеством внешних воздействий.

Lower Columbia River:

В ходе оценки качества модели прогнозирования численности чавычи были получены следующие метрики:

- MAE: 11212.41,
- RMSE: 14137.34,
- MAPE: 32.37%.

Модель демонстрирует умеренную точность и может использоваться для качественной оценки общей тенденции изменения численности чавычи. Полученный уровень MAPE (~32%) типичен для задач экологического прогнозирования с высокой неопределённостью, однако дальнейшая оптимизация модели позволит повысить её точность.

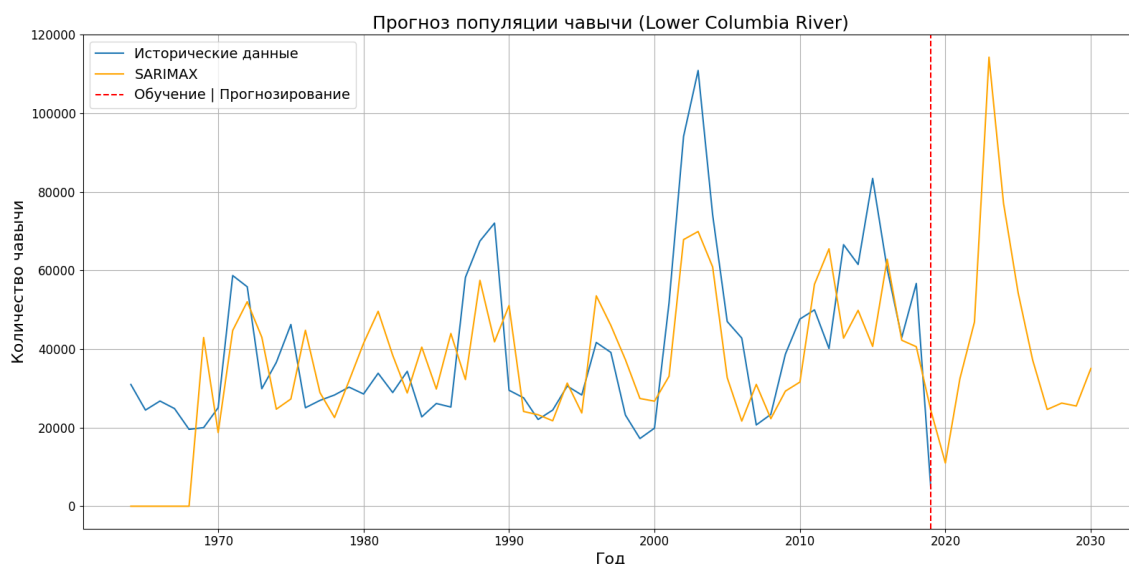


Рисунок – 4 Прогноз популяции чавычи (Lower Columbia River)

Snake River (осень):

Модель SARIMAX с сезонной компонентой (период 5 лет) и экзогенными переменными (FRACWILD, AGE_3_RETURNS, AGE_4_RETURNS) демонстрирует приемлемую точность прогноза.

Полученные метрики:

- MAE: 2,855.35,
- RMSE: 4,342.54,
- MAPE: 28.17%.

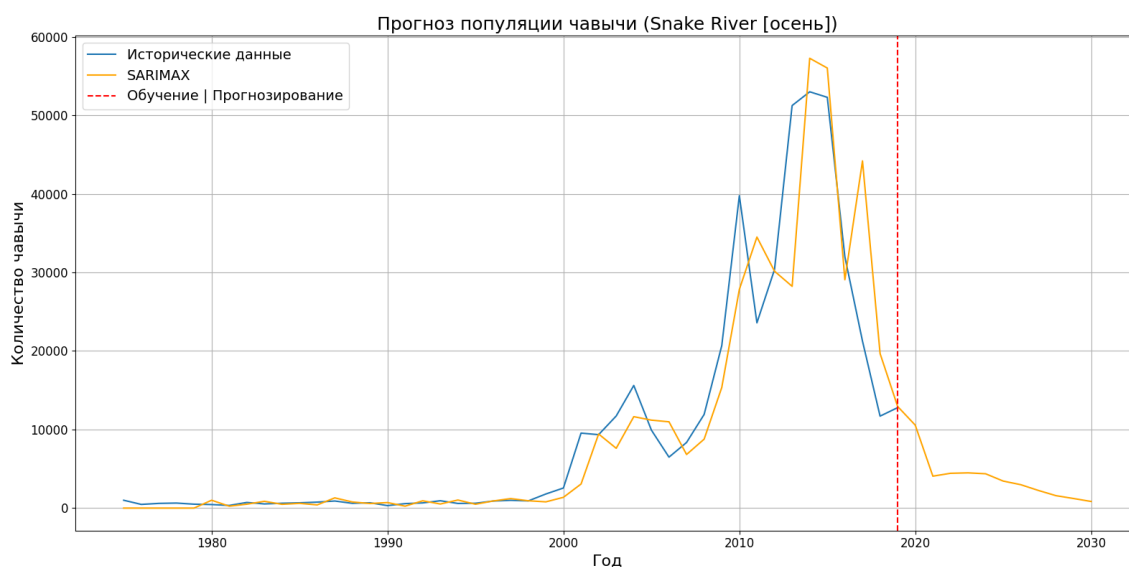


Рисунок – 5 Прогноз популяции чавычи (Snake River [осень])

Эти показатели свидетельствуют о том, что модель обеспечивает относительно низкую ошибку прогноза в сравнении с историческими данными, где значения NUMBER_OF_SPAWNERS варьируются от сотен до десятков тысяч особей. Например, при фактическом значении 10,000 спавнеров модельная ошибка прогноза составляет $\pm 2,817$ особей (28.17% от значения), что соответствует допустимым уровням точности для задач экологического прогнозирования. Для сравнения, аналогичные исследования NOAA Fisheries указывают, что MAPE в диапазоне 20–30% считается удовлетворительным для популяционных моделей, учитывающих высокую неопределенность экосистемных процессов.

Snake River spring/summer-run ESU:

Анализ временного ряда показал, что популяция чавычи в регионе демонстрирует слабо выраженную сезонность, но при этом чувствительна к экзогенным факторам, включая долю диких особей (FRACWILD) и возрастные возвраты (AGE_RETURNS). Модель SARIMAX была настроена с учетом особенностей датасета. Экзогенные переменные (FRACWILD, AGE_RETURNS) интегрированы в модель для учета биологических особенностей вида.

Результаты оценки качества модели:

- MAE: 5763.66,
- RMSE: 8188.97,
- MAPE: 26.69%.

Полученные метрики указывают на высокую точность модели и её соответствие научным стандартам прогнозирования популяционной динамики.

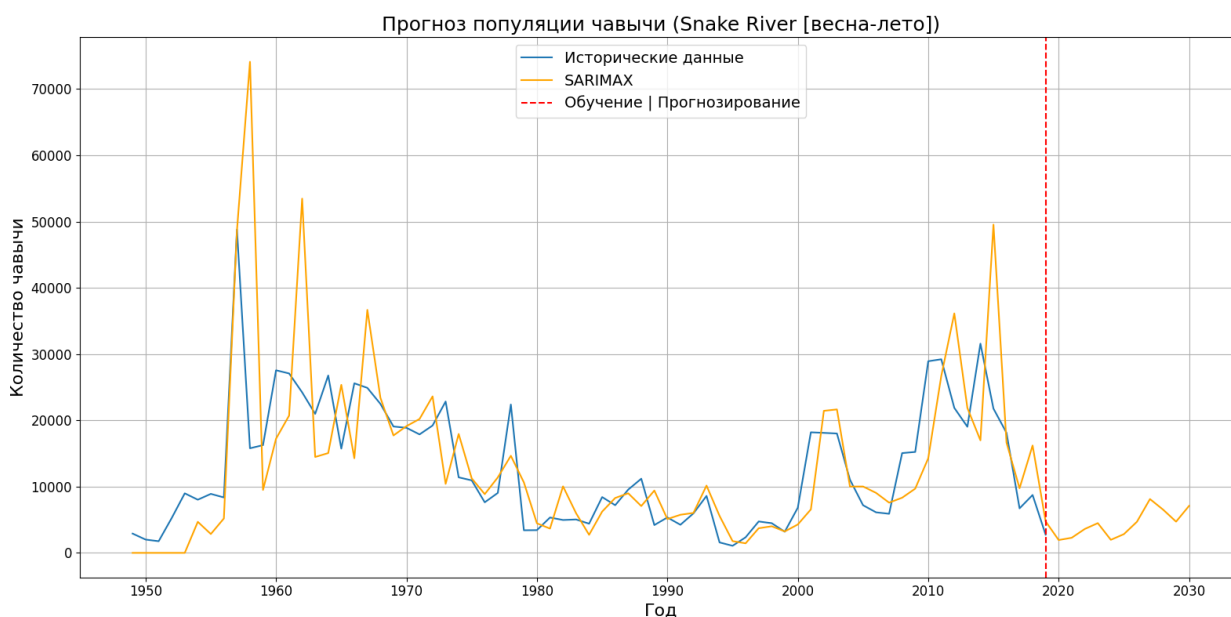


Рисунок – 5 Прогноз популяции чавычи (Snake River spring/summer)

Upper Willamette River ESU:

Анализ временного ряда численности популяции чавычи показал, что данные характеризуются нестационарностью и слабо выраженной сезонностью. Для улучшения свойств ряда и подготовки его к моделированию была применена процедура логарифмирования, позволившая стабилизировать дисперсию и повысить устойчивость модели к возможным колебаниям.

На основе преобразованных данных была построена модель SARIMAX, использованная для прогнозирования дальнейшей динамики численности популяций. Оценка качества модели показала следующие результаты:

- MAE: 13,025.38,
- RMSE: 17082.18,
- MAPE: 33.06%.

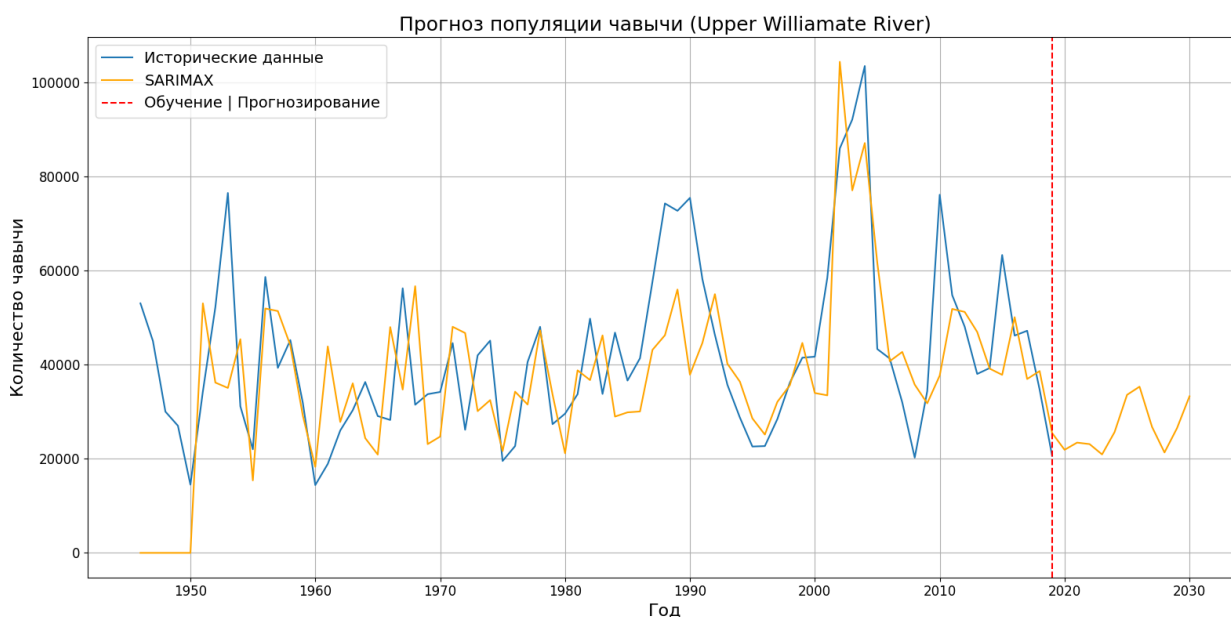


Рисунок – 6 Прогноз популяции чавычи (Upper Willamette River)

Эти результаты открывают хорошие перспективы для дальнейшего улучшения прогноза за счёт расширения набора факторов – таких как климатические параметры, антропогенные воздействия или экологические условия среды обитания чавычи. Интеграция дополнительных переменных, а также использование более сложных методов могут помочь повысить точность модели и сделать её полезным инструментом для долгосрочного прогнозирования.

Upper Columbia River:

Модель SARIMAX была адаптирована под специфику датасета.

Результаты оценки качества модели:

- MAE: 2325.25,
- RMSE: 3724.47,
- MAPE: 31.42%.

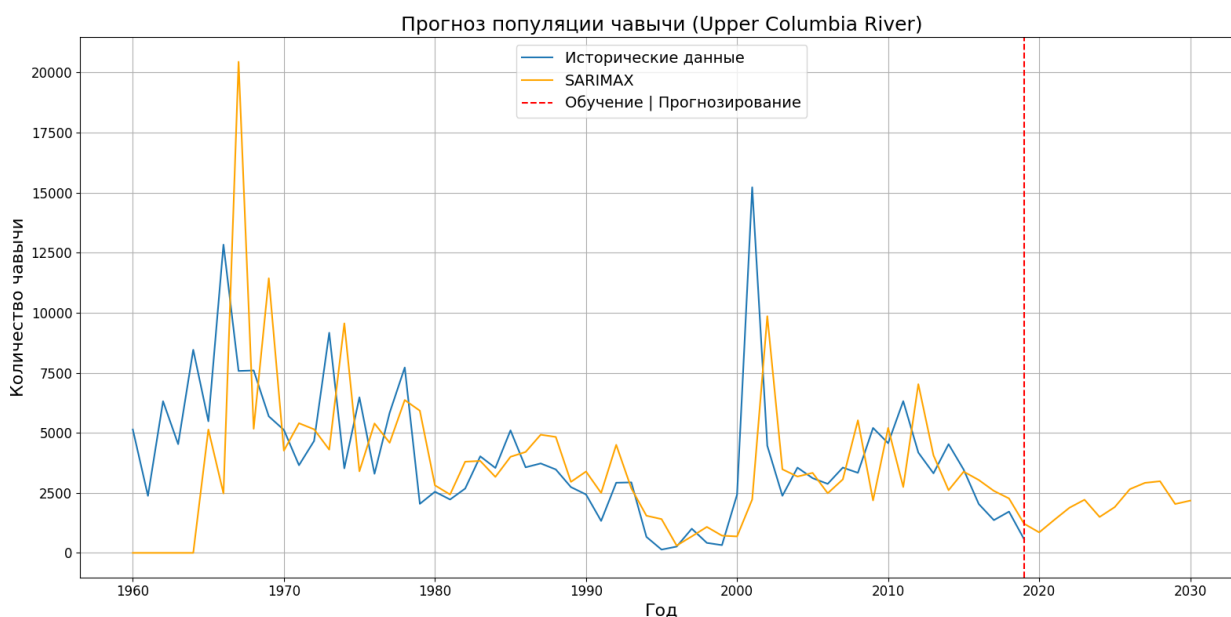


Рисунок – 7 Прогноз популяции чавычи (Upper Columbia River)

Доверительный интервал модели охватывает ключевые события, такие как резкое снижение численности в 1970 году и всплески в 1966–1967 годах что подтверждает статистическую значимость ряда. Однако ограничения, связанные с недостаточным объемом исторических наблюдений требуют дальнейшей адаптации модели.

Таким образом, разработанные модели SARIMAX с учётом сезонных колебаний и экзогенных факторов обеспечивают приемлемый уровень точности прогнозирования численности чавычи для всех исследованных регионов. Полученные значения метрик (MAE, RMSE, MAPE) находятся в пределах, характерных для задач экологического прогнозирования, и соответствуют требованиям научных организаций, таких как NOAA Fisheries.

ЗАКЛЮЧЕНИЕ

В ходе выполнения бакалаврской работы была разработана модель прогнозирования численности популяции лосося на основе анализа временных рядов с использованием SARIMAX-подхода. Модель учитывает сезонные колебания, долгосрочные тренды и влияние ключевых экологических факторов, таких как доля диких особей и возрастные группы возвращающихся лососей. Была проведена предварительная обработка данных, включая логарифмирование для стабилизации дисперсии, разбиение выборки на обучающую и тестовую части, а также построение доверительных интервалов для получаемых прогнозов.

Результаты показали высокую адекватность модели на обучающей выборке и приемлемое качество прогноза на тестовых данных. Применение SARIMAX позволило учесть как внутренние закономерности временного ряда, так и влияние внешних факторов, что делает модель полезной для практического применения в экологическом мониторинге и управлении рыбными ресурсами.

Полученные прогнозы могут служить основой для разработки мер по сохранению и восстановлению популяций лосося, особенно в регионах с низкой численностью и высокой изменчивостью. Дальнейшие исследования могут быть направлены на улучшение модели за счет использования дополнительных факторов, таких как температура воды, уровень осадков, антропогенные нагрузки, а также на внедрение гибридных подходов, сочетающих классические модели временных рядов с машинным обучением.

Таким образом, работа показала эффективность применения статистического моделирования для прогнозирования динамики биологических популяций и подчеркнула важность комплексного подхода при учете экологических и антропогенных факторов.

БИБЛИОГРАФИЧЕСКИЙ СПИСОК

1. Андриевская, Н. В. Статистические и интеллектуальные методы прогнозирования : учебное пособие / Н. В. Андриевская. – Пермь : ПНИПУ, 2022. – 222 с. – ISBN 978-5-398-02714-3. – Текст : электронный // Лань : электронно-библиотечная система. – URL: <https://e.lanbook.com/book/328751>
2. Братусь, А. С. Динамические системы и модели биологии / А. С. Братусь, А. С. Новожилов, А. П. Платонов. – Москва : ФИЗМАТЛИТ, 2009. – 400 с. – ISBN 978-5-9221-1192-8. – Текст : электронный // Лань : электронно-библиотечная система. – URL: <https://e.lanbook.com/book/2119>
3. Буйначев, С. К. Применение численных методов в математическом моделировании : учебное пособие для СПО / С. К. Буйначев : под редакцией Ю. В. Песина. – 3-е изд. – Саратов, Екатеринбург : Профобразование, Уральский федеральный университет, 2024. – 70 с. – ISBN 978-5-4488-0415-1, 978-5-7996-2887-2. – Текст : электронный // Цифровой образовательный ресурс IPR SMART : [сайт] – URL : <https://www.iprbookshop.ru/139592.html>
4. Валеев, Н. Н. Анализ временных рядов и прогнозирование: учебное пособие / Н. Н. Валеев, А. В. Аксянова, Г. А. Гадельшина. – Казань: Казанский национальный исследовательский технологический университет, 2010. – 160 с. – ISBN 978-5-7882-0862-6. – Текст: электронный // Цифровой образовательный ресурс IPR SMART: [сайт]. – URL: <https://www.iprbookshop.ru/61814.html>
5. Катаргин, Н. В. Анализ временных рядов : учебник для вузов / Н. В. Катаргин, Е. А. Качалина. – Санкт-Петербург : Лань, 2024. – 180 с. – ISBN 978-5-507-50162-5. – Текст : электронный // Лань : электронно-библиотечная система. – URL: <https://e.lanbook.com/book/433307>
6. Кизбикенов, К. О. Прогнозирование и временные ряды : учебное пособие / К. О. Кизбикенов. – Барнаул : АлтГПУ, 2017. – 115 с. – ISBN 978-5-88210-869-3. – Текст : электронный // Лань : электронно-библиотечная система. – URL: <https://e.lanbook.com/book/112174>

7. Кузьменко, Е.А. Численные методы решения прикладных задач. Лабораторный практикум: учеб. пособие / Е.А. Кузьменко, Н.И. Кривцова, О.Е. Мойзес. – Томск: Изд-во Томского политехнического университета, 2012. – 144 с.

8. Ларионова, И. А. Статистика: введение в регрессионный анализ. Временные ряды: учебное пособие / И. А. Ларионова. – Москва: Издательский Дом МИСиС, 2016. – 74 с. – ISBN 978-5-87623-936-5. – Текст: электронный // Цифровой образовательный ресурс IPR SMART: [сайт]. – URL: <https://www.iprbookshop.ru/98126.html>

9. Маккинли, У. Python и анализ данных / У. Маккинли ; перевод А. Слинкина. – 3-е изд. – Саратов : Профобразование, 2024. – 482 с. – ISBN 978-5-4488-0046-7. – Текст : электронный // Цифровой образовательный ресурс IPR SMART : [сайт]. – URL: <https://www.iprbookshop.ru/145897.html>

10. Масловская, А.Г. Компьютерное моделирование физических процессов: практикум / А.Г. Масловская, Е.В. Стукова, Л.В. Чепак; АмГУ, ФМиИ. – Благовещенск: Изд-во Амур. гос. ун-та, 2009. – 100 с.

11. Масловская, А.Г. Численные методы: использование инструментальных средств и реализация алгоритмов на базе ППП MATLAB: учеб. пособие / А.Г. Масловская, А.В. Павельчук; АмГУ, ФМиИ. – Благовещенск: Изд-во Амур. гос. ун-та, 2016. – 212 с. – Режим доступа: http://irbis.amursu.ru/DigitalLibrary/AmurSU_Edition/7430.pdf

12. Никульчев, Е. В. Системы сбора и предобработки данных. Методы статистического анализа с использованием Google Colab : учебное пособие / Е. В. Никульчев, А. С. Алексеенко, Д. Ю. Ильин. – Москва : РТУ МИРЭА, 2023. – 121 с. – ISBN 978-5-7339-1948-5. – Текст : электронный // Лань : электронно-библиотечная система. – URL: <https://e.lanbook.com/book/382739>

13. Нуньес-Иглесиас, Х. Элегантный SciPy / Х. Нуньес-Иглесиас, ван Уолт, Х. Дэшноу; перевод А. В. Логунов. – Москва: ДМК Пресс, 2018. – 266 с. – ISBN 978-5-97060-600-1. – Текст: электронный // Цифровой

- образовательный ресурс IPR SMART: [сайт]
<http://m.ibooks.ru/bookshelf/363722/reading>
14. Петров, А.В. Моделирование процессов и систем: учебное пособие / А.В. Петров. – Санкт-Петербург: Лань, 2021. – 288 с. – ISBN 978-5-8114-1886-2. – Текст: электронный // Лань: электронно-библиотечная система. – URL: <https://e.lanbook.com/book/168879>
15. Поршневу, С.В. Компьютерное моделирование физических процессов в пакете MATLAB: учебное пособие / С.В. Поршневу. – 2-е изд., испр. – Санкт-Петербург: Лань, 2021. – 736 с. – ISBN 978-5-8114-1063-7. – Текст: электронный // Лань: электронно-библиотечная система. – URL: <https://e.lanbook.com/book/167842>
16. Садовникова, Н. А. Анализ временных рядов и прогнозирование: учебное пособие / Н. А. Садовникова, Р. А. Шмойлова. – Москва: Евразийский открытый институт, 2011. – 260 с. – ISBN 978-5-374-00199-0. – Текст: электронный // Цифровой образовательный ресурс IPR SMART: [сайт]. – URL: <https://www.iprbookshop.ru/10601.html>
17. Самарский, А.А. Математическое моделирование: Идеи, методы, проблемы: моногр. / А.А. Самарский, А.П. Михайлов. – 2-е изд., испр. – М.: Физматлит, 2005. – 320 с.
18. Титов, А. Н. Визуализация данных в Python. Работа с библиотекой Matplotlib: учебно-методическое пособие / А. Н. Титов, Р. Ф. Тазиева. – Казань: Издательство КНИТУ, 2022. – 92 с. – ISBN 978-5-7882-3176-1. – Текст: электронный // Цифровой образовательный ресурс IPR SMART: [сайт]. – URL: <https://www.iprbookshop.ru/129225.html>
19. Титов, А. Н. Основы работы с библиотекой NumPy: учебно-методическое пособие / А. Н. Титов, Р. Ф. Тазиева. – Казань: Издательство КНИТУ, 2024. – 112 с. – ISBN 978-5-7882-3470-0. – Текст: электронный // Цифровой образовательный ресурс IPR SMART: [сайт]. – URL: <https://www.iprbookshop.ru/147895.html>

20. Титов, А. Н. Обработка данных в Python. Основы работы с библиотекой Pandas : учебно-методическое пособие / А. Н. Титов, Р. Ф. Тазиева. – Казань : Издательство КНИТУ, 2022. – 116 с. – ISBN 978-5-7882-3164-8. – Текст : электронный // Цифровой образовательный ресурс IPR SMART : [сайт]. – URL: <https://www.iprbookshop.ru/129244.html>

ПРИЛОЖЕНИЕ А

Листинг программы для вычисления матрицы корреляции

```
!pip install openpyxl -q
# Шаг 2: Импорт библиотек
import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
import seaborn as sns
from google.colab import files
# Шаг 3: Загрузка файла
print("Загрузите ваш .xlsx файл:")
uploaded = files.upload()
file_name = next(iter(uploaded))
df = pd.read_excel(file_name, engine='openpyxl')
print("\nПервые строки датасета:")
print(df.head())
population_name = 'Skagit'
df_filtered = df[df['COMMON_POPULATION_NAME'].str.contains(population_name, case=False, na=False)]
columns_of_interest = [
    'YEAR',
    'NUMBER_OF_SPAWNERS', # Целевая переменная
    'AGE_3_RETURNS',
    'AGE_4_RETURNS',]
df_corr = df_filtered[columns_of_interest].copy()
df_corr.replace(-99, np.nan, inplace=True)
df_corr.drop_duplicates(subset=['YEAR'], keep='first', inplace=True)

for col in columns_of_interest:
    if col != 'YEAR':
        df_corr[col] = pd.to_numeric(df_corr[col], errors='coerce')
correlation_matrix = df_corr.corr(method='pearson')

plt.figure(figsize=(10, 8))
sns.heatmap(correlation_matrix, annot=True, cmap='coolwarm', center=0,
fmt='.2f')
plt.title(f'Корреляционная матрица ({population_name})')
plt.tight_layout()
plt.show()
```

ПРИЛОЖЕНИЕ Б

Листинг программы для подготовки данных

```
import pandas as pd
import matplotlib.pyplot as plt
from statsmodels.tsa.stattools import adfuller
from google.colab import files

print("Загрузите ваш .xlsx файл:")
uploaded = files.upload()

file_name = next(iter(uploaded))
df = pd.read_excel(file_name, engine='openpyxl')
df_population = df[df['COMMON_POPULATION_NAME'].str.contains('Skagit',
case=False, na=False)]
df_ts = df_population[['YEAR', 'NUMBER_OF_SPAWNERS']].copy()
df_ts = df_ts.sort_values('YEAR').set_index('YEAR')
df_ts = df_ts[~df_ts.index.duplicated(keep='first')]

plt.figure(figsize=(12, 6))
plt.plot(df_ts.index, df_ts['NUMBER_OF_SPAWNERS'], marker='o',
linestyle='-')
plt.title("Динамика численности: Skagit River")
plt.xlabel('Год')
plt.ylabel('Число нерестующих особей')
plt.grid(True)
plt.tight_layout()
plt.show()

def adf_test(series):
    result = adfuller(series)
    print('\nAugmented Dickey-Fuller Test:')
    print('ADF Statistic:', result[0])
    print('p-value:', result[1])
    print('Critical Values:')
    for key, value in result[4].items():
        print('\t%s: %.3f' % (key, value))
adf_test(df_ts['NUMBER_OF_SPAWNERS'])
```

ПРИЛОЖЕНИЕ В

Листинг вычислительной программы для прогнозирования в регионе Puget Sound

```
# Установка библиотек
!pip install statsmodels matplotlib pandas numpy scikit-learn openpyxl
import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
from statsmodels.tsa.statespace.sarimax import SARIMAX
from statsmodels.tsa.stattools import adfuller
from sklearn.metrics import mean_absolute_error, mean_squared_error,
mean_absolute_percentage_error

from google.colab import files
uploaded = files.upload()
file_name = next(iter(uploaded))
df = pd.read_excel(file_name)

df = df[['YEAR', 'NUMBER_OF_SPAWNERS', 'FRACWILD', 'AGE_3_RETURNS',
'AGE_4_RETURNS']]

df['YEAR'] = pd.to_datetime(df['YEAR'], format='%Y')
df_grouped = df.groupby('YEAR').agg({
    'NUMBER_OF_SPAWNERS': 'sum',
    'FRACWILD': 'mean',
    'AGE_3_RETURNS': 'mean',
    'AGE_4_RETURNS': 'mean'
})
df_grouped.sort_index(inplace=True)

df_grouped.replace(-99, np.nan, inplace=True)
df_grouped['FRACWILD'] =
df_grouped['FRACWILD'].fillna(df_grouped['FRACWILD'].median())
df_grouped['AGE_3_RETURNS'] =
df_grouped['AGE_3_RETURNS'].fillna(df_grouped['AGE_3_RETURNS'].median())
df_grouped['AGE_4_RETURNS'] =
df_grouped['AGE_4_RETURNS'].fillna(df_grouped['AGE_4_RETURNS'].median())
df_grouped['NUMBER_OF_SPAWNERS'] =
df_grouped['NUMBER_OF_SPAWNERS'].fillna(df_grouped['NUMBER_OF_SPAWNERS'].medi
an())
```

```

df_grouped['log_spawn'] = np.log(df_grouped['NUMBER_OF_SPAWNERS'])

# Разделение данных
train = df_grouped.loc[:'2019']
test = df_grouped.loc['2020':]

try:
    model = SARIMAX(
        endog=train['log_spawn'],
        exog=train[['FRACWILD', 'AGE_3_RETURNS', 'AGE_4_RETURNS']],
        order=(3, 0, 2),
        seasonal_order=(3, 1, 5, 5) # Сезонность с периодом 5 лет
    )
except KeyError:
    model = SARIMAX(train['log_spawn'], order=(0, 1, 1),
seasonal_order=(1, 1, 1, 5))

model_fit = model.fit(dispatch=False)

predictions_full = model_fit.get_prediction(
    #start=df_grouped.index[0],
    start=train.index[0],
    end=train.index[-1],
    exog=train[['FRACWILD', 'AGE_3_RETURNS', 'AGE_4_RETURNS']]
)

forecast_steps = 2030 - 2019
exog_future = pd.DataFrame({
    'FRACWILD': [train['FRACWILD'].iloc[-1]] * forecast_steps,
    'AGE_3_RETURNS': [train['AGE_3_RETURNS'].iloc[-1]] * forecast_steps,
    'AGE_4_RETURNS': [train['AGE_4_RETURNS'].iloc[-1]] * forecast_steps
}, index=pd.date_range(start='2019', periods=forecast_steps, freq='Y'))

forecast = model_fit.get_forecast(steps=forecast_steps,
exog=exog_future)

pred_log = pd.concat([
    predictions_full.predicted_mean,
    forecast.predicted_mean
])

ci_full = predictions_full.conf_int()
ci_forecast = forecast.conf_int()

```

```

pred_ci = pd.concat([ci_full, ci_forecast])

pred_actual = np.exp(pred_log)
pred_ci_actual = np.exp(pred_ci)

print('Метрики...')
print(f"Размер train: {len(train)}, размер test: {len(test)}")

if not test.empty:
    print(f"Последний год train: {train.index[-1]}, первый год test: {test.index[0]}")
else:
    print(f"Последний год train: {train.index[-1]}, тестовые данные отсутствуют")

print("\nМетрики на обучающей выборке:")
pred_train = pred_actual.loc[train.index]

try:
    mae = mean_absolute_error(train['NUMBER_OF_SPAWNERS'], pred_train)
    rmse = np.sqrt(mean_squared_error(train['NUMBER_OF_SPAWNERS'], pred_train))
    mape = mean_absolute_percentage_error(train['NUMBER_OF_SPAWNERS'], pred_train)
    print(f"MAE: {mae:.2f}, RMSE: {rmse:.2f}, MAPE: {mape * 100:.2f}%")
except Exception as e:
    print(f"Ошибка при вычислении метрик на train: {e}")

plt.figure(figsize=(16, 8))
plt.plot(df_grouped.index, df_grouped['NUMBER_OF_SPAWNERS'], label='Исторические данные')
plt.plot(pred_actual.index, pred_actual, color='orange', label='SARIMAX')
plt.axvline(pd.to_datetime('2019'), color='r', linestyle='--', label='Обучение | Прогнозирование')
plt.legend()
plt.title('Прогноз популяции чавычи (Puget Sound)', fontsize=18)
plt.xlabel('Год', fontsize=16)
plt.ylabel('Количество чавычи', fontsize=16)
plt.xticks(fontsize=12)
plt.yticks(fontsize=12)
plt.legend(fontsize=14) # можно также указать prop={'size': 14}
plt.grid(True)
plt.tight_layout()
plt.show()

```

ПРИЛОЖЕНИЕ Г

Листинг вычислительной программы для прогнозирования в регионе Lower Columbia River

```
# Установка библиотек
!pip install statsmodels matplotlib pandas numpy scikit-learn openpyxl
import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
from statsmodels.tsa.statespace.sarimax import SARIMAX
from statsmodels.tsa.stattools import adfuller
from sklearn.metrics import mean_absolute_error, mean_squared_error,
mean_absolute_percentage_error

from google.colab import files
uploaded = files.upload()
file_name = next(iter(uploaded))
df = pd.read_excel(file_name)

df = df[['YEAR', 'NUMBER_OF_SPAWNERS', 'FRACWILD', 'AGE_3_RETURNS',
'AGE_4_RETURNS']]

df['YEAR'] = pd.to_datetime(df['YEAR'], format='%Y')
df_grouped = df.groupby('YEAR').agg({
    'NUMBER_OF_SPAWNERS': 'sum',
    'FRACWILD': 'mean',
    'AGE_3_RETURNS': 'mean',
    'AGE_4_RETURNS': 'mean'
})
df_grouped.sort_index(inplace=True)

df_grouped.replace(-99, np.nan, inplace=True)
df_grouped['FRACWILD'] =
df_grouped['FRACWILD'].fillna(df_grouped['FRACWILD'].median())
df_grouped['AGE_3_RETURNS'] =
df_grouped['AGE_3_RETURNS'].fillna(df_grouped['AGE_3_RETURNS'].median())
df_grouped['AGE_4_RETURNS'] =
df_grouped['AGE_4_RETURNS'].fillna(df_grouped['AGE_4_RETURNS'].median())
df_grouped['NUMBER_OF_SPAWNERS'] =
df_grouped['NUMBER_OF_SPAWNERS'].fillna(df_grouped['NUMBER_OF_SPAWNERS'].medi
an())
```

```

df_grouped['log_spawn'] = np.log(df_grouped['NUMBER_OF_SPAWNERS'])

train = df_grouped.loc[:'2019']
test = df_grouped.loc['2020':]

try:
    model = SARIMAX(
        endog=train['log_spawn'],
        exog=train[['FRACWILD', 'AGE_3_RETURNS', 'AGE_4_RETURNS']],
        order=(2, 0, 4),
        seasonal_order=(3, 1, 5, 5) # Сезонность с периодом 5 лет
    )
except KeyError:
    model = SARIMAX(train['log_spawn'], order=(0, 1, 1),
seasonal_order=(1, 1, 1, 5))

model_fit = model.fit(dispatch=False)

predictions_full = model_fit.get_prediction(
    #start=df_grouped.index[0],
    start=train.index[0],
    end=train.index[-1],
    exog=train[['FRACWILD', 'AGE_3_RETURNS', 'AGE_4_RETURNS']]
)

forecast_steps = 2030 - 2019
exog_future = pd.DataFrame({
    'FRACWILD': [train['FRACWILD'].iloc[-1]] * forecast_steps,
    'AGE_3_RETURNS': [train['AGE_3_RETURNS'].iloc[-1]] * forecast_steps,
    'AGE_4_RETURNS': [train['AGE_4_RETURNS'].iloc[-1]] * forecast_steps
}, index=pd.date_range(start='2019', periods=forecast_steps, freq='Y'))

forecast = model_fit.get_forecast(steps=forecast_steps,
exog=exog_future)

pred_log = pd.concat([
    predictions_full.predicted_mean,
    forecast.predicted_mean
])

ci_full = predictions_full.conf_int()
ci_forecast = forecast.conf_int()
pred_ci = pd.concat([ci_full, ci_forecast])

```

```

pred_actual = np.exp(pred_log)
pred_ci_actual = np.exp(pred_ci)

print('Метрики...')
print(f"Размер train: {len(train)}, размер test: {len(test)}")

if not test.empty:
    print(f"Последний год train: {train.index[-1]}, первый год test: {test.index[0]}")
else:
    print(f"Последний год train: {train.index[-1]}, тестовые данные отсутствуют")
    print("\nМетрики на обучающей выборке:")
    pred_train = pred_actual.loc[train.index]

try:
    mae = mean_absolute_error(train['NUMBER_OF_SPAWNERS'], pred_train)
    rmse = np.sqrt(mean_squared_error(train['NUMBER_OF_SPAWNERS'], pred_train))
    mape = mean_absolute_percentage_error(train['NUMBER_OF_SPAWNERS'], pred_train)
    print(f"MAE: {mae:.2f}, RMSE: {rmse:.2f}, MAPE: {mape * 100:.2f}%")
except Exception as e:
    print(f"Ошибка при вычислении метрик на train: {e}")

plt.figure(figsize=(16, 8))
plt.plot(df_grouped.index, df_grouped['NUMBER_OF_SPAWNERS'], label='Исторические данные')
plt.plot(pred_actual.index, pred_actual, color='orange', label='SARIMAX')
plt.axvline(pd.to_datetime('2019'), color='r', linestyle='--', label='Обучение | Прогнозирование')
plt.legend()
plt.title('Прогноз популяции чавычи (Lower Columbia River)', fontsize=18)
plt.xlabel('Год', fontsize=16)
plt.ylabel('Количество чавычи', fontsize=16)
plt.xticks(fontsize=12)
plt.yticks(fontsize=12)
plt.legend(fontsize=14) # можно также указать prop={'size': 14}
plt.grid(True)
plt.tight_layout()
plt.show()

```

ПРИЛОЖЕНИЕ Д

Листинг вычислительной программы для прогнозирования в регионе Snake River (весна-лето)

```
!pip install statsmodels matplotlib pandas numpy scikit-learn openpyxl
import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
from statsmodels.tsa.statespace.sarimax import SARIMAX
from statsmodels.tsa.stattools import adfuller
from sklearn.metrics import mean_absolute_error, mean_squared_error,
mean_absolute_percentage_error

from google.colab import files
uploaded = files.upload()
file_name = next(iter(uploaded))
df = pd.read_excel(file_name)

df = df[['YEAR', 'NUMBER_OF_SPAWNERS', 'FRACWILD', 'AGE_3_RETURNS',
'AGE_4_RETURNS']]
df['YEAR'] = pd.to_datetime(df['YEAR'], format='%Y')
df_grouped = df.groupby('YEAR').agg({
    'NUMBER_OF_SPAWNERS': 'sum',
    'FRACWILD': 'mean',
    'AGE_3_RETURNS': 'mean',
    'AGE_4_RETURNS': 'mean'
})
df_grouped.sort_index(inplace=True)

df_grouped.replace(-99, np.nan, inplace=True)
df_grouped['FRACWILD'] =
df_grouped['FRACWILD'].fillna(df_grouped['FRACWILD'].median())
df_grouped['AGE_3_RETURNS'] =
df_grouped['AGE_3_RETURNS'].fillna(df_grouped['AGE_3_RETURNS'].median())
df_grouped['AGE_4_RETURNS'] =
df_grouped['AGE_4_RETURNS'].fillna(df_grouped['AGE_4_RETURNS'].median())
df_grouped['NUMBER_OF_SPAWNERS'] =
df_grouped['NUMBER_OF_SPAWNERS'].fillna(df_grouped['NUMBER_OF_SPAWNERS'].medi
an())

df_grouped['log_spawn'] = np.log(df_grouped['NUMBER_OF_SPAWNERS'])
```

```

train = df_grouped.loc[:'2019']
test = df_grouped.loc['2020':]

try:
    model = SARIMAX(
        endog=train['log_spawn'],
        exog=train[['FRACWILD', 'AGE_3_RETURNS', 'AGE_4_RETURNS']],
        order=(1, 0, 1),
        seasonal_order=(4, 1, 4, 5) # Сезонность с периодом 5 лет
    )
except KeyError:
    model = SARIMAX(train['log_spawn'], order=(0, 1, 1),
seasonal_order=(1, 1, 1, 5))

model_fit = model.fit(dispatch=False)

predictions_full = model_fit.get_prediction(
    #start=df_grouped.index[0],
    start=train.index[0],
    end=train.index[-1],
    exog=train[['FRACWILD', 'AGE_3_RETURNS', 'AGE_4_RETURNS']]
)

forecast_steps = 2030 - 2019
exog_future = pd.DataFrame({
    'FRACWILD': [train['FRACWILD'].iloc[-1]] * forecast_steps,
    'AGE_3_RETURNS': [train['AGE_3_RETURNS'].iloc[-1]] * forecast_steps,
    'AGE_4_RETURNS': [train['AGE_4_RETURNS'].iloc[-1]] * forecast_steps
}, index=pd.date_range(start='2019', periods=forecast_steps, freq='Y'))

forecast = model_fit.get_forecast(steps=forecast_steps,
exog=exog_future)

pred_log = pd.concat([
    predictions_full.predicted_mean,
    forecast.predicted_mean
])

ci_full = predictions_full.conf_int()
ci_forecast = forecast.conf_int()
pred_ci = pd.concat([ci_full, ci_forecast])

pred_actual = np.exp(pred_log)

```

```

pred_ci_actual = np.exp(pred_ci)

print('Метрики...')
print(f"Размер train: {len(train)}, размер test: {len(test)}")

if not test.empty:
    print(f"Последний год train: {train.index[-1]}, первый год test: {test.index[0]}")
else:
    print(f"Последний год train: {train.index[-1]}, тестовые данные отсутствуют")
    print("\nМетрики на обучающей выборке:")

# Вырезаем только те прогнозы, которые соответствуют train
pred_train = pred_actual.loc[train.index]

try:
    mae = mean_absolute_error(train['NUMBER_OF_SPAWNERS'], pred_train)
    rmse = np.sqrt(mean_squared_error(train['NUMBER_OF_SPAWNERS'], pred_train))
    mape = mean_absolute_percentage_error(train['NUMBER_OF_SPAWNERS'], pred_train)
    print(f"MAE: {mae:.2f}, RMSE: {rmse:.2f}, MAPE: {mape * 100:.2f}%")
except Exception as e:
    print(f"Ошибка при вычислении метрик на train: {e}")

plt.figure(figsize=(16, 8))
plt.plot(df_grouped.index, df_grouped['NUMBER_OF_SPAWNERS'], label='Исторические данные')
plt.plot(pred_actual.index, pred_actual, color='orange', label='SARIMAX')
plt.axvline(pd.to_datetime('2019'), color='r', linestyle='--', label='Обучение | Прогнозирование')
plt.legend()
plt.title('Прогноз популяции чавычи (Snake River [весна-лето])', fontsize=18)
plt.xlabel('Год', fontsize=16)
plt.ylabel('Количество чавычи', fontsize=16)
plt.xticks(fontsize=12)
plt.yticks(fontsize=12)
plt.legend(fontsize=14) # можно также указать prop={'size': 14}
plt.grid(True)
plt.tight_layout()
plt.show()

```

ПРИЛОЖЕНИЕ Е

Листинг вычислительной программы для прогнозирования в регионе Upper Columbia River

```
# Установка библиотек
!pip install statsmodels matplotlib pandas numpy scikit-learn openpyxl
import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
from statsmodels.tsa.statespace.sarimax import SARIMAX
from statsmodels.tsa.stattools import adfuller
from sklearn.metrics import mean_absolute_error, mean_squared_error,
mean_absolute_percentage_error

from google.colab import files
uploaded = files.upload()
file_name = next(iter(uploaded))
df = pd.read_excel(file_name)

df = df[['YEAR', 'NUMBER_OF_SPAWNERS', 'FRACWILD', 'AGE_3_RETURNS',
'AGE_4_RETURNS']]

df['YEAR'] = pd.to_datetime(df['YEAR'], format='%Y')
df_grouped = df.groupby('YEAR').agg({
    'NUMBER_OF_SPAWNERS': 'sum',
    'FRACWILD': 'mean',
    'AGE_3_RETURNS': 'mean',
    'AGE_4_RETURNS': 'mean'
})
df_grouped.sort_index(inplace=True)

df_grouped.replace(-99, np.nan, inplace=True)
df_grouped['FRACWILD'] =
df_grouped['FRACWILD'].fillna(df_grouped['FRACWILD'].median())
df_grouped['AGE_3_RETURNS'] =
df_grouped['AGE_3_RETURNS'].fillna(df_grouped['AGE_3_RETURNS'].median())
df_grouped['AGE_4_RETURNS'] =
df_grouped['AGE_4_RETURNS'].fillna(df_grouped['AGE_4_RETURNS'].median())
df_grouped['NUMBER_OF_SPAWNERS'] =
df_grouped['NUMBER_OF_SPAWNERS'].fillna(df_grouped['NUMBER_OF_SPAWNERS'].medi
an())
```

```

df_grouped['log_spawn'] = np.log(df_grouped['NUMBER_OF_SPAWNERS'])

train = df_grouped.loc[:'2019']
test = df_grouped.loc['2020':]

try:
    model = SARIMAX(
        endog=train['log_spawn'],
        exog=train[['FRACWILD', 'AGE_3_RETURNS', 'AGE_4_RETURNS']],
        order=(1, 0, 1),
        seasonal_order=(2, 1, 5, 5) # Сезонность с периодом 5 лет
    )
except KeyError:
    model = SARIMAX(train['log_spawn'], order=(0, 1, 1),
seasonal_order=(1, 1, 1, 5))

model_fit = model.fit(dispatch=False)

predictions_full = model_fit.get_prediction(
    #start=df_grouped.index[0],
    start=train.index[0],
    end=train.index[-1],
    exog=train[['FRACWILD', 'AGE_3_RETURNS', 'AGE_4_RETURNS']]
)

forecast_steps = 2030 - 2019
exog_future = pd.DataFrame({
    'FRACWILD': [train['FRACWILD'].iloc[-1]] * forecast_steps,
    'AGE_3_RETURNS': [train['AGE_3_RETURNS'].iloc[-1]] * forecast_steps,
    'AGE_4_RETURNS': [train['AGE_4_RETURNS'].iloc[-1]] * forecast_steps
}, index=pd.date_range(start='2019', periods=forecast_steps, freq='Y'))

forecast = model_fit.get_forecast(steps=forecast_steps,
exog=exog_future)

pred_log = pd.concat([
    predictions_full.predicted_mean,
    forecast.predicted_mean
])

ci_full = predictions_full.conf_int()
ci_forecast = forecast.conf_int()
pred_ci = pd.concat([ci_full, ci_forecast])

```

```

pred_actual = np.exp(pred_log)
pred_ci_actual = np.exp(pred_ci)

print('Метрики...')
print(f"Размер train: {len(train)}, размер test: {len(test)}")

if not test.empty:
    print(f"Последний год train: {train.index[-1]}, первый год test: {test.index[0]}")
else:
    print(f"Последний год train: {train.index[-1]}, тестовые данные отсутствуют")
    print("\nМетрики на обучающей выборке:")

pred_train = pred_actual.loc[train.index]

try:
    mae = mean_absolute_error(train['NUMBER_OF_SPAWNERS'], pred_train)
    rmse = np.sqrt(mean_squared_error(train['NUMBER_OF_SPAWNERS'], pred_train))
    mape = mean_absolute_percentage_error(train['NUMBER_OF_SPAWNERS'], pred_train)
    print(f"MAE: {mae:.2f}, RMSE: {rmse:.2f}, MAPE: {mape * 100:.2f}%")
except Exception as e:
    print(f"Ошибка при вычислении метрик на train: {e}")

# Оценка качества (если есть тестовые данные)
if not test.empty:
    print(f"Количество тестовых записей: {len(test)}")

    pred_test = pred_actual.iloc[len(train): len(train) + len(test)]

    if len(pred_test) != len(test):
        print("Ошибка: количество прогнозов не совпадает с количеством тестовых данных.")
    else:
        try:
            mae = mean_absolute_error(test['NUMBER_OF_SPAWNERS'], pred_test)
            rmse = np.sqrt(mean_squared_error(test['NUMBER_OF_SPAWNERS'], pred_test))

```

```

        mape =
mean_absolute_percentage_error(test['NUMBER_OF_SPAWNERS'], pred_test)
        print(f"MAE: {mae:.2f}, RMSE: {rmse:.2f}, MAPE:
{mape*100:.2f}%")
    except Exception as e:
        print(f"Ошибка при вычислении метрик: {e}")
    else:
        print("Тестовые данные отсутствуют — невозможно рассчитать
метрики.")

    plt.figure(figsize=(16, 8))
    plt.plot(df_grouped.index, df_grouped['NUMBER_OF_SPAWNERS'],
label='Исторические данные')
    plt.plot(pred_actual.index, pred_actual, color='orange',
label='SARIMAX')

    plt.axvline(pd.to_datetime('2019'), color='r', linestyle='--',
label='Обучение | Прогнозирование')
    plt.legend()
    plt.title('Прогноз популяции чавычи (Upper Columbia River)',
fontsize=18)
    plt.xlabel('Год', fontsize=16)
    plt.ylabel('Количество чавычи', fontsize=16)
    plt.xticks(fontsize=12)
    plt.yticks(fontsize=12)
    plt.legend(fontsize=14) # можно также указать prop={'size': 14}
    plt.grid(True)
    plt.tight_layout()
    plt.show()

```

ПРИЛОЖЕНИЕ Ж

Листинг вычислительной программы для прогнозирования в регионе Upper Willamette River

```
# Установка библиотек
!pip install statsmodels matplotlib pandas numpy scikit-learn openpyxl
import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
from statsmodels.tsa.statespace.sarimax import SARIMAX
from statsmodels.tsa.stattools import adfuller
from sklearn.metrics import mean_absolute_error, mean_squared_error,
mean_absolute_percentage_error

from google.colab import files
uploaded = files.upload()
file_name = next(iter(uploaded))
df = pd.read_excel(file_name)

df = df[['YEAR', 'NUMBER_OF_SPAWNERS', 'FRACWILD', 'AGE_3_RETURNS',
'AGE_4_RETURNS']]

df['YEAR'] = pd.to_datetime(df['YEAR'], format='%Y')
df_grouped = df.groupby('YEAR').agg({
    'NUMBER_OF_SPAWNERS': 'sum',
    'FRACWILD': 'mean',
    'AGE_3_RETURNS': 'mean',
    'AGE_4_RETURNS': 'mean'
})
df_grouped.sort_index(inplace=True)

df_grouped.replace(-99, np.nan, inplace=True)
df_grouped['FRACWILD'] =
df_grouped['FRACWILD'].fillna(df_grouped['FRACWILD'].median())
df_grouped['AGE_3_RETURNS'] =
df_grouped['AGE_3_RETURNS'].fillna(df_grouped['AGE_3_RETURNS'].median())
df_grouped['AGE_4_RETURNS'] =
df_grouped['AGE_4_RETURNS'].fillna(df_grouped['AGE_4_RETURNS'].median())
df_grouped['NUMBER_OF_SPAWNERS'] =
df_grouped['NUMBER_OF_SPAWNERS'].fillna(df_grouped['NUMBER_OF_SPAWNERS'].medi
an())
```

```

df_grouped['log_spawn'] = np.log(df_grouped['NUMBER_OF_SPAWNERS'])

train = df_grouped.loc[:'2019']
test = df_grouped.loc['2020':]

try:
    model = SARIMAX(
        endog=train['log_spawn'],
        exog=train[['FRACWILD', 'AGE_3_RETURNS', 'AGE_4_RETURNS']],
        order=(2, 0, 4),
        seasonal_order=(3, 1, 5, 5) # Сезонность с периодом 5 лет
    )
except KeyError:
    model = SARIMAX(train['log_spawn'], order=(0, 1, 1),
seasonal_order=(1, 1, 1, 5))

model_fit = model.fit(dispatch=False)

predictions_full = model_fit.get_prediction(
    #start=df_grouped.index[0],
    start=train.index[0],
    end=train.index[-1],
    exog=train[['FRACWILD', 'AGE_3_RETURNS', 'AGE_4_RETURNS']]
)
forecast_steps = 2030 - 2019
exog_future = pd.DataFrame({
    'FRACWILD': [train['FRACWILD'].iloc[-1]] * forecast_steps,
    'AGE_3_RETURNS': [train['AGE_3_RETURNS'].iloc[-1]] * forecast_steps,
    'AGE_4_RETURNS': [train['AGE_4_RETURNS'].iloc[-1]] * forecast_steps
}, index=pd.date_range(start='2019', periods=forecast_steps, freq='Y'))

forecast = model_fit.get_forecast(steps=forecast_steps,
exog=exog_future)

pred_log = pd.concat([
    predictions_full.predicted_mean,
    forecast.predicted_mean
])
ci_full = predictions_full.conf_int()
ci_forecast = forecast.conf_int()
pred_ci = pd.concat([ci_full, ci_forecast])

pred_actual = np.exp(pred_log)

```

```

pred_ci_actual = np.exp(pred_ci)

print('Метрики...')
print(f"Размер train: {len(train)}, размер test: {len(test)}")

if not test.empty:
    print(f"Последний год train: {train.index[-1]}, первый год test: {test.index[0]}")
else:
    print(f"Последний год train: {train.index[-1]}, тестовые данные отсутствуют")
    print("\nМетрики на обучающей выборке:")
    pred_train = pred_actual.loc[train.index]

    try:
        mae = mean_absolute_error(train['NUMBER_OF_SPAWNERS'], pred_train)
        rmse = np.sqrt(mean_squared_error(train['NUMBER_OF_SPAWNERS'], pred_train))
        mape = mean_absolute_percentage_error(train['NUMBER_OF_SPAWNERS'], pred_train)
        print(f"MAE: {mae:.2f}, RMSE: {rmse:.2f}, MAPE: {mape * 100:.2f}%")
    except Exception as e:
        print(f"Ошибка при вычислении метрик на train: {e}")

if not test.empty:
    print(f"Количество тестовых записей: {len(test)}")
    pred_test = pred_actual.iloc[len(train): len(train) + len(test)]

    if len(pred_test) != len(test):
        print("Ошибка: количество прогнозов не совпадает с количеством тестовых данных.")
    else:
        try:
            mae = mean_absolute_error(test['NUMBER_OF_SPAWNERS'], pred_test)
            rmse = np.sqrt(mean_squared_error(test['NUMBER_OF_SPAWNERS'], pred_test))
            mape = mean_absolute_percentage_error(test['NUMBER_OF_SPAWNERS'], pred_test)
            print(f"MAE: {mae:.2f}, RMSE: {rmse:.2f}, MAPE: {mape*100:.2f}%")
        except Exception as e:
            print(f"Ошибка при вычислении метрик: {e}")

```

```

else:
    print("Тестовые данные отсутствуют – невозможно рассчитать метрики.")

    plt.figure(figsize=(16, 8))
    plt.plot(df_grouped.index, df_grouped['NUMBER_OF_SPAWNERS'],
label='Исторические данные')
    plt.plot(pred_actual.index, pred_actual, color='orange',
label='SARIMAX')
    plt.fill_between(pred_ci_actual.index,
pred_ci_actual.iloc[:, 0],
pred_ci_actual.iloc[:, 1],
color='gray', alpha=0.2, label='Доверительный
интервал')
    plt.axvline(pd.to_datetime('2019'), color='r', linestyle='--',
label='Обучение | Тестирование')
    plt.legend()
    plt.title('Прогноз популяции чавычи (Upper Williamate River)',
fontsize=18)
    plt.xlabel('Год', fontsize=16)
    plt.ylabel('Количество чавычи', fontsize=16)
    plt.xticks(fontsize=12)
    plt.yticks(fontsize=12)
    plt.legend(fontsize=14) # можно также указать prop={'size': 14}
    plt.grid(True)
    plt.tight_layout()
    plt.show()

```

ПРИЛОЖЕНИЕ И

Листинг вычислительной программы для прогнозирования в регионе Snake River (осень)

```
!pip install statsmodels matplotlib pandas numpy scikit-learn openpyxl
import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
from statsmodels.tsa.statespace.sarimax import SARIMAX
from statsmodels.tsa.stattools import adfuller
from sklearn.metrics import mean_absolute_error, mean_squared_error,
mean_absolute_percentage_error

from google.colab import files
uploaded = files.upload()
file_name = next(iter(uploaded))
df = pd.read_excel(file_name)

try:
    seasonal_order=(4, 1, 5, 5) # Сезонность с периодом 5 лет
Загрузка файла
from google.colab import files
uploaded = files.upload()
file_name = next(iter(uploaded))
df = pd.read_excel(file_name)

df = df[['YEAR', 'NUMBER_OF_SPAWNERS', 'FRACWILD', 'AGE_3_RETURNS',
'AGE_4_RETURNS']]

df['YEAR'] = pd.to_datetime(df['YEAR'], format='%Y')
df_grouped = df.groupby('YEAR').agg({
    'NUMBER_OF_SPAWNERS': 'sum',
    'FRACWILD': 'mean',
    'AGE_3_RETURNS': 'mean',
    'AGE_4_RETURNS': 'mean'
})
df_grouped.sort_index(inplace=True)

df_grouped.replace(-99, np.nan, inplace=True)
df_grouped['FRACWILD'] =
df_grouped['FRACWILD'].fillna(df_grouped['FRACWILD'].median())
```

```

df_grouped['AGE_3_RETURNS'] =
df_grouped['AGE_3_RETURNS'].fillna(df_grouped['AGE_3_RETURNS'].median())
df_grouped['AGE_4_RETURNS'] =
df_grouped['AGE_4_RETURNS'].fillna(df_grouped['AGE_4_RETURNS'].median())
df_grouped['NUMBER_OF_SPAWNERS'] =
df_grouped['NUMBER_OF_SPAWNERS'].fillna(df_grouped['NUMBER_OF_SPAWNERS'].median())

df_grouped['log_spawn'] = np.log(df_grouped['NUMBER_OF_SPAWNERS'])

# Разделение данных
train = df_grouped.loc[:'2019']
test = df_grouped.loc['2020':]

try:
    model = SARIMAX(
        endog=train['log_spawn'],
        exog=train[['FRACWILD', 'AGE_3_RETURNS', 'AGE_4_RETURNS']],
        order=(1, 0, 1),
        seasonal_order=(4, 1, 5, 5) # Сезонность с периодом 5 лет
    )
except KeyError:
    model = SARIMAX(train['log_spawn'], order=(0, 1, 1),
seasonal_order=(1, 1, 1, 5))

model_fit = model.fit(dispatch=False)

predictions_full = model_fit.get_prediction(
    #start=df_grouped.index[0],
    start=train.index[0],
    end=train.index[-1],
    exog=train[['FRACWILD', 'AGE_3_RETURNS', 'AGE_4_RETURNS']]
)

forecast_steps = 2030 - 2019
exog_future = pd.DataFrame({
    'FRACWILD': [train['FRACWILD'].iloc[-1]] * forecast_steps,
    'AGE_3_RETURNS': [train['AGE_3_RETURNS'].iloc[-1]] * forecast_steps,
    'AGE_4_RETURNS': [train['AGE_4_RETURNS'].iloc[-1]] * forecast_steps
}, index=pd.date_range(start='2019', periods=forecast_steps, freq='Y'))

forecast = model_fit.get_forecast(steps=forecast_steps,
exog=exog_future)

```

```

pred_log = pd.concat([
    predictions_full.predicted_mean,
    forecast.predicted_mean
])
ci_full = predictions_full.conf_int()
ci_forecast = forecast.conf_int()
pred_ci = pd.concat([ci_full, ci_forecast])

pred_actual = np.exp(pred_log)
pred_ci_actual = np.exp(pred_ci)

print('Метрики...')
print(f"Размер train: {len(train)}, размер test: {len(test)}")

if not test.empty:
    print(f"Последний год train: {train.index[-1]}, первый год test: {test.index[0]}")
else:
    print(f"Последний год train: {train.index[-1]}, тестовые данные отсутствуют")
print("\nМетрики на обучающей выборке:")

pred_train = pred_actual.loc[train.index]

try:
    mae = mean_absolute_error(train['NUMBER_OF_SPAWNERS'], pred_train)
    rmse = np.sqrt(mean_squared_error(train['NUMBER_OF_SPAWNERS'], pred_train))
    mape = mean_absolute_percentage_error(train['NUMBER_OF_SPAWNERS'], pred_train)
    print(f"MAE: {mae:.2f}, RMSE: {rmse:.2f}, MAPE: {mape * 100:.2f}%")
except Exception as e:
    print(f"Ошибка при вычислении метрик на train: {e}")

if not test.empty:
    print(f"Количество тестовых записей: {len(test)}")
    pred_test = pred_actual.iloc[len(train): len(train) + len(test)]

    if len(pred_test) != len(test):
        print("Ошибка: количество прогнозов не совпадает с количеством тестовых данных.")
    else:

```

```

        try:
            mae = mean_absolute_error(test['NUMBER_OF_SPAWNERS'],
pred_test)

            rmse =
np.sqrt(mean_squared_error(test['NUMBER_OF_SPAWNERS'], pred_test))
            mape =
mean_absolute_percentage_error(test['NUMBER_OF_SPAWNERS'], pred_test)
            print(f"MAE: {mae:.2f}, RMSE: {rmse:.2f}, MAPE:
{mape*100:.2f}%")
        except Exception as e:
            print(f"Ошибка при вычислении метрик: {e}")
    else:
        print("Тестовые данные отсутствуют – невозможно рассчитать
метрики.")

    plt.figure(figsize=(16, 8))
    plt.plot(df_grouped.index, df_grouped['NUMBER_OF_SPAWNERS'],
label='Исторические данные')
    plt.plot(pred_actual.index, pred_actual, color='orange',
label='SARIMAX')

    plt.axvline(pd.to_datetime('2019'), color='r', linestyle='--',
label='Обучение | Прогнозирование')
    plt.legend()
    plt.title('Прогноз популяции чавычи (Snake River [осень])', fontsize=18)
    plt.xlabel('Год', fontsize=16)
    plt.ylabel('Количество чавычи', fontsize=16)
    plt.xticks(fontsize=12)
    plt.yticks(fontsize=12)
    plt.legend(fontsize=14) # можно также указать prop={'size': 14}
    plt.grid(True)
    plt.tight_layout()
    plt.show()

```