

Казахский Национальный Университет им. аль-Фараби

Институт Информационных и Вычислительных Технологий КН МНВО РК

УДК 004.9:502(043)

На правах рукописи

**НУРЖАНОВ ЧИНГИЗ АСКАРОВИЧ**

**Проектирование информационной системы для прогнозирования и  
принятия решений в процессе очистки почвы, содержащей токсичные эле-  
менты**

6D060200 – Информатика

Диссертация на соискание учёной степени  
доктора философии (PhD)

Научный руководитель:  
д.т.н., ассоц. профессор Найзабаева Л.К.

Зарубежный руководитель:  
Full Professor Pidlisnyuk V.  
Университет Яна Евангелиста  
Пуркине, Чешская Республика

Республика Казахстан  
Алматы, 2023

## ОГЛАВЛЕНИЕ

### ОБОЗНАЧЕНИЯ И СОКРАЩЕНИЯ

<b>ВВЕДЕНИЕ</b>	5
Глава 1 Обзор литературы	12
1.1 Информационная система экологического контроля химических и физических факторов на базе технологий хранилищ данных	12
1.2 Математическое (компьютерное) моделирование в области экологии	21
1.2.1 Математические модели процесса переноса токсических элементов (ТЭ) из почвы в органы растения	25
1.2.2 Математическое и компьютерное моделирование биологических методов рекультивации	29
1.2.3 Модели продуктивности растений на загрязненной ТЭ почве	32
Глава 2 Моделирование продуктивности биоэнергетического растения на почвах, загрязненных токсичными элементами в зависимости от климатических условий среды	35
2.1 Создание модели MiscanCalc для прогнозирования динамики формирования продуктивности растения на загрязненной токсичными элементами почве в зависимости от климатических условий среды	36
2.2 Создание математической модели методом самоорганизации на продуктивность растений на загрязненной токсичными элементами почве в зависимости от климатических условий среды	52
2.3 Прогнозирование продуктивности растения на загрязненной ТЭ почве в зависимости от климатических условий среды с помощью технологий машинного обучения	60
2.4 Выводы по разделу, постановка задачи	91
Глава 3 Создание базы данных о землях, загрязненных токсичными элементами и базы данных, содержащей информацию о видах растений, способных к восстановлению почв	93
3.1 Создание базы данных о землях, загрязненных токсичными элементами, с учетом географического месторасположения территории, количество токсичных элементов и концентрации их в почве	95
3.2 Создание базы данных, содержащей информацию о древесных и травянистых видах растений, способных к восстановлению почв	106
3.3 Разработка программного приложения для управления специализированным корпоративным хранилищем данных	115
3.4 Выводы по разделу, постановка задачи	122

Глава 4	Математическая модель поглощения ТЭ растением и перемещение их из почвы в растение в зависимости от условий среды	125
4.1	Выводы по разделу, постановка задачи	133
Глава 5	Разработка информационной системы для прогнозирования и принятия решений в процессе очистки почвы от токсичных элементов	134
5.1	Выводы по разделу, постановка задачи	147
<b>ЗАКЛЮЧЕНИЕ</b>		149
<b>СПИСОК ИСПОЛЬЗОВАННЫХ ИСТОЧНИКОВ</b>		150
<b>ПРИЛОЖЕНИЕ А. Таблицы вспомогательных данных</b>		169
<b>ПРИЛОЖЕНИЕ Б. Авторские свидетельства</b>		187
<b>ПРИЛОЖЕНИЕ В. Акты внедрения</b>		190

## ОБОЗНАЧЕНИЯ И СОКРАЩЕНИЯ

БД	– база данных
ИС	– информационная система
КБП	– коэффициент биологического поглощения
МГУА	– метод группового учёта аргументов
МНК	– метод наименьших квадратов
МО	– машинное обучение
НСИ	– нормативно-справочная информация
ОДУ	– обыкновенное дифференциальное уравнение
СВ	– сухое вещество
Сервер	– обработчик поступающих в систему заявок
СУБД	– система управления базами данных
ТЭ	– токсические микроэлементы
ЭВМ	– электронно-вычислительная машина
API	– application programming interface (интерфейс прикладного программирования)
AutoML	– автоматизированное машинное обучение
DL	– deep learning (глубокое обучение)
DSS	– decision support system (система поддержки принятия решений)
ENN	– метод расширенного ближайшего соседа
JSON	– JavaScript object notation; простой формат обмена данными, основанный на подмножестве синтаксиса JavaScript
MAE	– среднее абсолютное отклонение
MAPE	– средняя абсолютная ошибка в процентах
MSE	– среднеквадратическая ошибка
MSPE	– среднеквадратичная ошибка в процентах
$R^2$	– коэффициент дифференциации
RFECV	– рекурсивное устранение признаков с перекрестной проверкой
RMSE	– ошибка корневого квадрата
RMSLE	– среднеквадратичная логарифмическая ошибка
SelectKBest	– одномерный отбор признаков
SHAP	– shap feature important
Soft computing	– модель мягких вычислений
SQL	– structured query language; описание структуры БД, управления данными, правами доступ и транзакциями

## ВВЕДЕНИЕ

Технологический прогресс влечет за собой ухудшение экологической ситуации в мире. Казахстан с огромной территорией и большими природными ресурсами не является исключением. В стране 8,9 млн. га деградированных земель, из которых 200 тыс. га загрязнены различными токсическими микроэлементами (ТЭ) [1]. Накоплен многочисленный научный материал о содержании в почве тяжелых металлов, пестицидов, углеводов и других токсических загрязнителей среды [2]. Поскольку широкомасштабная химизация в сельском хозяйстве привела не только к загрязнению, но и к снижению урожайности сельскохозяйственных культур, повышение рентабельности растениеводства зависит от экологичности производимой продукции. В связи с этим, цифровизация сельского хозяйства для отслеживания урожайности сельскохозяйственных культур в режиме реального времени, производства экологически чистой продукции занимает ведущее место в аграрном секторе [3,4].

Цифровая технология открыла дополнительные возможности для изучения изменений, происходящих в природе, в результате хозяйственной деятельности человека. В настоящее время машинное обучение (МО), используя статистические методы регрессии и классификации, находит применение на современном уровне в разных сферах науки, например, в области информационной науки [5-7], экологии [8, 9], сельского хозяйства [10-13], а также используются при моделировании климата и аналитики [14, 15]. С помощью методов математического моделирования и вычислительной техники на основе прорывных технологий создаются новые методы, модели, алгоритмы и технологии для решения глобальных экологических проблем взаимодействия человека и природы [16,17].

Для обеспечения экологической безопасности разрабатываются информационные системы экологического мониторинга, за счет сбора больших данных о состоянии воздуха, почвы и воды [17,18-20]. Эффективность автоматизированной системы осуществляется за счет непрерывного сбора, обработки и хранения данных, наряду с традиционными способами с использованием систем контроля базами данных (БД) [19]. Большие данные получают через [21,22]: 1) дистанционное зондирование; 2) распределенные сети недорогих автоматизированных GPS датчиков; 3) БД по биоразнообразию и инвентаризации; 4) ресурсы данных, курируемые сообществом; 5) гражданскую науку; 6) национальные и международные станции долгосрочных экологических исследований и мониторинга. Благодаря большому объему генерации данных и, как следствие, более доступному хранению этих данных, они создаются в социальных сетях или датчиках, либо хранятся для преобразования в полезную информацию [22]. Реализация задачи больших данных, соответственно, опирается на прорывных технологиях, такие как облачные вычисления, Интернет вещей и аналитика данных [14].

Интерес к математическому моделированию процесса рекультивации загрязненных ТЭ экосистемы растет из года в год. Создаются алгоритмы для

быстрой нейтрализации источника загрязнения [23,24] и математические модели способствуют выявить оптимальный способ, описывающие процессы в загрязненной ТЭ почве, влияния их на растительный и животный мир, выбора оптимальной стратегии рекультивации [24]. Системный анализ с привлечением математического моделирования направлены на моделирование процесса восстановления загрязненных земель, в основном углеводородов [25]. При этом нет теории и модели, описывающих поведение других основных загрязнителей среды в почвенно-грунтовой толще, процесс их миграции в системе «почва-растение», которые могли бы стать основой разработки эффективной технологии восстановления загрязненной почвы/воды.

До сих пор в Казахстане отсутствует единая компьютерная информационная БД для решения рекультивационных мероприятий, более того, до сих пор нет учета уровня загрязнения антропогенно-нарушенных экосистем и их отходов [26]. В связи с этим, необходимо создать информационную систему, содержащую данные о накоплении и поведении различных ксенобиотиков в почвенно-грунтовой толще, миграции в системе «почва-растение», которые могли бы быть основой разработки эффективной технологии очистки загрязненных земель. Важность создания единой БД позволит правительству страны решать стратегически важные вопросы, связанные с экологией, например, строительство высокотехнологичных сооружений для сжигания отходов промышленного производства, а главное оценить реальный уровень техногенного загрязнения, возможный риск для окружающей среды и здоровья населения.

В связи с этим, разработка математической модели для технологии восстановления загрязненных ксенобиотиками территорий страны, создание единой БД о загрязненных территориях и БД о растениях, способных к восстановлению, станет основой для решения важных экологических проектов и задач в области охраны окружающей страны, улучшения благосостояния населения, проживающего в экологически опасных регионах.

**Цель** – разработка интеллектуальной информационной системы для обработки данных о почвах, загрязненных токсичными элементами для прогнозирования и принятия решений об очистке земель Республики Казахстан.

Для достижения поставленной цели необходимо было решить следующие **задачи**:

1 Изучение математического моделирования продуктивности биоэнергетического растения на почвах, загрязненных токсичными элементами.

2 Создание баз данных: о древесных и травянистых видах растений, способных к восстановлению почв; о землях, загрязненных токсичными элементами, с учетом географического месторасположения территории; о количестве устаревших пестицидов и концентрации их в почве.

3 Разработка информационной системы для прогнозирования и принятия решений в процессе очистки почвы от токсичных элементов.

4 Применение методов машинного обучения для прогнозирования продуктивности растений, поглощающих токсичные элементы из почвы.

**Объектом исследования** являются данные о почвах и растениях, загрязненных токсичными элементами, о климатических условиях (на примере Алма-тинской области).

**Предметом исследования** являются климатические данные AlmatyWeatherDataSet.csv, данные за 2015-2022 гг., продуктивности растений на почвах, загрязненных токсичными элементами.

**Методы исследования.** Для решения поставленных задач использовались методы машинного обучения, многорядный эвристический метод самоорганизации для построения регрессионных уравнений, методы моделирования, методы регрессионного и дисперсионного многофакторного анализа.

При разработке комплексного подхода к созданию информационной системы были применены теория проектирования информационных систем, методы проектирования баз данных, процессно-ориентированный подход

**Научная новизна** заключается в разработке и получении следующих выводов:

Предложена усовершенствованная модель MiscanCalc на основе MiscanMod для прогнозирования биомассы растений на загрязненной ТЭ почве с учетом климатических данных.

Адаптирован метод многорядной самоорганизации, по прогнозирующим свойствам превосходящий в точности регрессионные модели, обеспечивающий автоматический отбор информативных входных переменных и выбор структуры регрессионной модели оптимальной сложности.

Расширена модель рекультивации почвы на основе методов машинного обучения с помощью интегрированного подхода библиотеки XGBoost, обладающей высокой производительностью и устойчивостью к переобучению.

**Теоретическая и практическая значимость исследования** – фундаментальные аспекты информационной технологии и математического моделирования в задачах природопользования.

**Прикладная ценность** – в оперативности информировать государственные органы, осуществляющим управление земельными фондами, при формировании рекультивационных мероприятий; в возможности использования опытными хозяйствами, занимающимися внедрением информационных технологий, а также организациям, осуществляющих агроэкологический мониторинг.

**Уровни организация исследований:** база данных, математические модели, программное обеспечение Jupyterlab Anaconda Environment, Microsoft Visual Studio, SQL Server Management Studio, языки программирования (Python, C# и Delphi) и прикладные аспекты математического моделирования.

**Связь с планом основных научных работ:** диссертационная работа были выполнена в Институте информационной и вычислительной технологии КН МНВО РК, в рамках грантового финансирования МНВО РК по научным и (или) научно-техническим проектам на 2023-2025 годы AP19678926 «Разработка интеллектуальной системы для исследования и решения экологических проблем загрязнения почвы и воздуха с помощью методов науки о данных».

**Личный вклад автора.** Все основные результаты, описанные в диссертации, выполнены и собраны автором. Кроме того, основные результаты исследований, анализы, модели, программы созданы автором, выводы сделаны на основе результатов, полученных от работы и исследования PhD докторанта.

**Объем и структура диссертации.**

Диссертация состоит из обозначения и сокращения, введения, трех глав, заключения и списка использованных источников. Полный объем диссертации составляет 191 страницу, 62 рисунка и 18 таблиц. Список литературы содержит 268 наименований.

**Основное содержание работы.**

**Во введении** обоснована актуальность темы диссертационной работы, цель, объект, предмет и задачи научно-исследовательской работы. Описаны результаты исследований, научная новизна, теоретическая и практическая значимость, уровни организация исследований. Приведены результаты апробации основных результатов диссертационной работы.

**В первой главе** диссертационной работы рассматривается литературный обзор об информационной системе экологического контроля факторов окружающей среды на основе технологий хранилищ информации, методах моделирования в области экологии, сельского хозяйства и экологической биотехнологии.

**Во второй главе** для оценки продуктивности растений (на примере энергетического сельскохозяйственного биоэнергетического вида мискантус), произрастающих на загрязненной токсичными элементами почве, в зависимости от климатических условий (на примере Алматинской области) рассмотрены три подхода: 1) использование регрессионных моделей машинного обучения для точного прогноза продуктивности растений и выбора гиперпредикторов; 2) создание модели Miscancalc на основе модификации модели MiscanMod для процесса формирования биомассы растений; 3) разработка математической модели «Многорядный эвристический метод самоорганизации для построения регрессионных уравнений» для отбора предикторов.

Предложена усовершенствованная модель Miscancalc на основе Miscanmod, используя C# в Visual Studio для прогнозирования продуктивности биомассы мискантуса на загрязненных ТЭ почвах, непригодных для производства продуктов питания:

- ввода в блок дополнительного алгоритма путем расчета различий между эмпирическими месячными данными биомассы культур, выращенных на загрязненных и незагрязненных почвах, методом наименьших квадратов;

- ввода в блок входных данных путем расчета фотосинтетически активной солнечной радиации на основании информации региональных коэффициентов и данных интегральной радиации Казгидрометеорологической службы по солнечной радиации;

- ввода в блок входных данных путем расчета потенциального испарения влаги почвы на основании температуры, индекса тепла и средней продолжительности дня.



Разработанная модель в течение вегетационного периода проводит расчет индекса листовой поверхности в зависимости от дефицита влаги в почве; оценивает эффективность использования излучения и температуры; количества максимальных дней в период активного накопления биомассы с учетом температуры и фотопериода в период цветения и отмирания; даты завершения стадии цветения; количества максимальных дней в период активного накопления биомассы с учетом температуры и фотопериода; определяет сроки начала уборки культуры; прогнозирует урожайности в  $\text{г м}^{-2}$  сухой биомассы, как на техногенно-загрязненной, так и незагрязненной почвах.

Используя Jupyter Notebook, Sklearn Libraries, провели оценку, выбор моделей регрессии и информативных признаков с помощью 13 регрессионных моделей, используя климатические данные с 2015 по 2021 год (осадки, температура воздуха, потенциальное испарение влаги из почвы и активное фотосинтетическое излучение) и выявили лидер-модель машинного обучения для прогноза продуктивности растений. Выявлено, что наиболее подходящими для прогнозирования продуктивности растений в зависимости от климатических метаданных является модель H2OAutoML ( $R^2 - 0,99$ ; MSE – 15,3; MAE– 2,91; MAPE – 0.0), а для отбора признаков – из моделей scikit-learn 3 модели (Stacking Regressor, Gradient Boosting Regressor и H2OAutoML). С помощью подхода SHAP установлено, что гиперпараметрами, влияющие на продуктивность являются месяц (период произрастания), ветер и облачность. Кросс валидация 3-х моделей (H2O Stacked Ensemble Estimator; H2O Stacked Ensemble Estimator, Regression GLM) выявила ансамблевый модель регрессии Stacked Ensemble от H<sub>2</sub>O с высокой математической точностью  $R^2 - 0.99$ , MSE– 9.54, MAE – 7.26. RMSE – 9.77, RMSLE – 5.38.

Разработан многорядный метод самоорганизации для анализа связи биомассы растения на загрязненной ТЭ почве в зависимости от климатических условий среды. Показано, что уравнение математической модели, описывающее зависимость роста биомассы от испарения влаги почвы, фотосинтетической активной радиации, осадков и температуры, имеет нелинейный вид. Установлено, что параметры испарение влаги почвы, фотосинтетическая активная радиация и осадки оказывают влияние на продуктивность растений на загрязненной ТЭ почве. Разработанный многорядный метод самоорганизации автоматически отбирает входные переменные и проводит выбор структуры регрессионной модели оптимальной сложности.

**В третьей главе** представлены результаты разработки комплексного подхода к созданию информационной системы на базе данных о загрязненных токсичными элементами территориях и растениях, способных к восстановлению техногенных ландшафтов. Созданы две базы данных о загрязненных токсичными элементами территориях и растениях казахстанской флоры, способных к очистке техногенно-загрязненной почвы. Приведены функциональные требования, этапы создания программы и их построение. Определены структуры таблиц базы данных и их взаимосвязь. Разработан графический интерфейс пользователя приложения. На базе СУБД SQL Server Management Studio реализована интерфейсная часть в среде Microsoft Visual Studio на языке

С#. Спроектирована информационная система для формирования отчетности об образовании, утилизации и хранения отходов и рекультивации почв. Применен подход «Сущность – Связь» при построении концептуального проекта информационной системы. Приведены основные программные компоненты, выявлены классы с указанием атрибутов, методов, и связей между ними.

**В четвертой главе** представлены результаты создания математической модели. Математическая модель описывает зависимость накопления ТЭ в вегетативных органах растений от типа почвы и экологии среды. Установлено, что поглощение ТЭ растением зависит от типа почвы: на песчаной загрязненной почве спектр поглощения элементов выше, чем у растений, произрастающих на светло-каштановой загрязненной почве. Показано, что предикторами, оказывающие наибольшее влияние на поглощение ТЭ растением из загрязненной среды являются кислотность и влажность почвы.

**В пятой главе** представлены данные о создании информационной системы для прогнозирования процесса очистки 1 га почвы с глубины залегания 0-20 см, с помощью растений от ТЭ. В основу концепции информационной системы нами был предложен интегрированный подход включая: сбор, передачу, накопление и обработка измерительных данных, данные БД, модели продуктивности растений, модели фитотоксичности почвы, модели поглощения ТЭ растением и содержания их в почве.

В качестве прогнозируемых значений использовали экспериментальные данные образцов, созданных моделей фитотоксичности и поглощения ТЭ растением, что позволяет получить достоверный прогноз периода очистки почв для конкретного вида.

Для тренировки модели XGBoost использовали данные продуктивности растений, в результате разработки моделей по отношению к климатическим условиям за 2022 год Проведены расчеты с помощью модели оптимальное количество растений на 1 га земли, используя данные площади и нормы посева для прогнозирования связи между количеством растений и временем очистки почвы. Получили графическое изображение снижения ТЭ в почве из года в год, за счет их поглощения растением. Установлено, что длительность периода очистки почвы 1 га почвы с глубины залегания 0-20 см с помощью растения мискантус для разных элементов различна: период очистки особо токсичных элементов таких как цинк (Zn), и медь (Cu) составляет 7 лет, стронций (Sr) – 4 года, свинца (Pb) – 12 лет.

Данный подход создания информационной системы может быть использован для принятия решений по восстановлению загрязненной ТЭ почвы с помощью других видов растений.

**В заключении** изложены основные результаты работы, выводы диссертации и будущие шаги исследования данного направления.

**Апробация исследования.** Основные результаты по теме диссертации изложены в 17 работах: 4 статьи в журналах, рекомендованных Комитетом по контролю в сфере образования и науки МНВО РК, 5 статьи в международных научных изданиях с импакт-факторами, входящих базу данных Thomson

Reuters и Scopus, и 8 статей в материалах международных зарубежных конференций: International multidisciplinary conference “Sustainable technologies and the legal/economic aspects of agricultural production”, 27-28 April 2016, Kyiv, Ukraine; международная научная конференция ИИВТ МОН РК «Современные проблемы информатики и вычислительных технологий», 28-29 июня 2016; международная конференция «Информатика и прикладная математика», посвященная 25-летию Независимости Казахстана и 25-летию ИИВТ МОН РК, 21-24 сентября 2016; международная научно-практическая конференция «Математические методы и информационные технологии макроэкономического анализа и экономической политики», посвященная празднованию 80-летнего юбилея академика НАН РК А.А. Ашимова, 11-12 апреля 2017; II International Conference on Modern Problems of Computer Science and Computer Technology, National Academy of Sciences of the CS MES RK, 27-30 September 2017, Almaty, Kazakhstan; круглый стол «Цифровизация промышленности – основа четвертой промышленной революции», 20 апреля 2018, Алматы, Казахстан; 15<sup>th</sup> International Phytotechnology conference, 1-5 October 2018, University of Novi Sad, Serbia; международная научная конференция в области информационных технологий, посвященная 75-летию профессора У.А. Тукеева, 8 октября 2021, Алматы, Казахстан; международный симпозиум "Устойчивая энергетика и энергомашиностроение- 2021: SUSE-2021". Казань, 18–20 февраля 2021 года; международная конференция «Цифровые двойники: создание, применение, эффективность», 26 мая 2022, РУДН, Москва, Российская Федерация; 2023 International Conference on Smart Information Systems and Technologies 4-6 May, 2023, Astana.

Получены 3 свидетельства о государственной регистрации прав на объект авторского права.

*Свидетельства о государственной регистрации прав на объект авторского права:*

Найзабаева Л., Нуржанов Ч. Свидетельство о государственной регистрации прав на объект авторского права № 971 от 13 декабря 2018 «База данных о загрязненных пестицидами территориях» (программа для ЭВМ), Министерство Юстиции, 2018, Астана, Казахстан;

Найзабаева Л., Нуржанов Ч. Свидетельство о государственной регистрации прав на объект авторского права № 976 от 13 декабря 2018 «База данных о растениях фиторемедиантах» (программа для ЭВМ), Министерство Юстиции, 2018, Астана, Казахстан;

Мазаков Т.Ж., Джомартова Ш.А., Мазакова А.Т., Нуржанов Ч.А., Жакып Б.М., Амирханов Б.С., Шорманов Т.С. Свидетельство о государственной регистрации прав на объект авторского права № 9968 от 19 мая 2020 «Многорядный Эвристический Метод Самоорганизации для построения регрессионных уравнений» (программа для ЭВМ), Министерство Юстиции, 2020, Астана, Казахстан.

## ГЛАВА 1. ОБЗОР ЛИТЕРАТУРЫ

### 1.1 Информационная система экологического контроля химических и физических факторов на базе технологий хранилищ данных

Экологическая информационная система необходима для управления данными о природных ресурсах и обеспечивает эффективный и точный доступ к информации об окружающей среде и уровне их загрязнения [27]. Поэтому совершенствование информационной системы в сфере экологии с возрастанием уровня загрязнения почвы/воды различными ТЭ является одним из основных критериев здоровья человека. Экологическая информационная система собирает данные для достижения следующих целей [16]:

- систематического сбора исторических данных и обеспечения их доступности для углубленного анализа природных явлений, т.е. системы обработки и анализа;
- системы отчетности в режиме реального времени;
- предсказания и предупреждения потенциальной экологической безопасности.

К основным задачам экологической информационной системы относятся: сбор данных, управление и оповещение экологической информации в режиме реального времени, а также разработка персонализированных услуг, обмен информацией.

Существуют два пути создания информационных систем на основе интегрированных баз данных (БД) [16]:

1) основанный на глубоком анализе бизнес-процессов и существующей системы документооборота и создании на этой основе структурированной единой БД на сервере. Разработка на этой основе информационной системы для пользователей.

2) анализ потребностей администрации в информационных ресурсах. На сервере создается административная сетевая информационная система на основе БД и разработки клиентского интерфейса

Экологическая информационная система разработана во многих странах мира, например, система ТИМ в Венгрии, которая содержит информацию о состоянии, химии и биологии почвы из 1236 точек измерения, система Air Quality обеспечивает доступ к информации о качестве воздуха крупных европейских городов, система WISE – о качестве воды, а система CORINE) и Natura 2000 – о состоянии флоры и фауны [27,28]. Технология Agent моделирует управление водными ресурсами, лесами, пожарами [20]. В статье Athanasiadis и Mitkas [19] на основе технологии Agent разработана методология для оказания цифровых экологических услуг для обработки данных и принятия решений.

Продуктивность контроля осуществляется информационными системами для непрерывного сбора и хранения данных. Инструментом информационного обеспечения процессов в области экологии является экологический мониторинг [29]. Основная задача мониторинга – информационное обеспечение и

поддержка процедур принятия решений в области природоохранной деятельности и экологической безопасности [29,30]. Мониторинг окружающей среды и природных ресурсов способствует решению следующих задач [29]:

- получение достоверной информации о состоянии экосистем, биоразнообразия и об источниках загрязнения, возникшие в результате антропогенной деятельности негативно влияющие на здоровье человека;
- оценка и прогноз состояния среды обитания;
- обеспечение данными для обеспечения экологической безопасности и принятия управленческих решений.

Эффективность систем контроля осуществляется за счет непрерывного сбора, обработки и хранения данных [31]. Оценка окружающей среды также связана с управлением природными ресурсами, которая осуществляется на основе информации, полученной в результате наблюдений и/или данных о состоянии поверхности Земли и атмосферы. Сети экологического мониторинга создаются по всему миру, в основном в регионах, где существуют потенциальные проблемы загрязнения. Специальные службы наблюдают за ними и регистрируют параметры окружающей среды. Через эти сети собираются огромные объемы необработанных данных. Скрининг природных ресурсов и окружающей среды проводится с помощью спутниковых снимков и метеорологической службы. По спутниковым снимкам оценивают биофизические свойства растительности (например, биомассу, индекс площади листьев, содержание воды), состояние лесного покрова (классы лесного покрова, видовой состав, продуктивность лесов, сожженные участки и участки возникновения пожаров), а также состояние засухи сельскохозяйственных культур [32, 33].

В Казахстане за окружающей средой наблюдает республиканское государственное предприятие «Казгидромет». Предмет деятельности предприятия – оценка состояния окружающей среды, метеорологический и гидрологический мониторинг. В каждом городе существует сеть метеорологических станций. Ежемесячно «Казгидромет» выпускает бюллетень «Информационные бюллетени о состоянии окружающей среды» об уровне загрязнения атмосферы и поверхностных вод. Цели этой деятельности, согласно уставу, заключаются в получении достоверной информации о повышении качества прогнозов состояния окружающей среды, а также в сборе и обобщении данных.

Традиционно оценка загрязнения основана на статистическом сравнении результатов мониторинга с предельно-допустимыми концентрациями. Данный подход утвержден в законодательном порядке. Следует отметить, что система мониторинга не включает деятельность по управлению качеством среды, просто содержит информацию и является источником для принятия экологически значимых решений. Полученная информация используется для принятия оптимальных решений при планировании конкретных экологических задач, с учетом социально-экономических критериев и допустимых нагрузок на окружающую среду [34,35].

Системный подход к вопросу экологического мониторинга позволяет контролировать состояние воздуха, воды и почвы и здоровье человека со стороны государства. Мониторингом является [35]: 1) конкретизация объекта

наблюдения; 2) исследование объекта наблюдения; 3) определение составных частей информационной модели объекта наблюдения; 4) составление плана проведения измерений; 5) определение состояния объекта наблюдения, его оценка и определение степени адекватности его составленной информационной модели; 6) прогноз трансформаций состояния объекта наблюдения через промежутки времени; 7) удобный интерфейс для получения информации и возможность доведения ее до пользователя.

В статье М.Б. Жумадиловой с соавторами [36] отмечено, что в Казахстане планируется на основе принятых законодательных актов, создание информационной системы государственного мониторинга окружающей среды и природных ресурсов. Основой информационной системы является БД для быстрого доступа необходимой информации с минимальными расходами, максимальной надежностью, точностью и объективностью. Используя БД, можно не только обеспечить легкий доступ, но и облегчить эффективную работу с хранимыми данными. Это позволяет устанавливать правила и отношения внутри хранилища, определять поведение пространственно связанных географических объектов и классов объектов, а также поддержку целостности данных [37,38]. В качестве посредника между пользователем и БД выступают системы управления БД, обеспечивающие пользователям не только простой доступ к информации, но и позволяющие создавать, удалять, редактировать записи, обеспечивать безопасность и т.д. [38].

СУБД состоит из интегрированного набора компьютерного программного обеспечения, который позволяет пользователям взаимодействовать с одной или несколькими БД, а также осуществлять доступ и блокировку ко всем данным; изменять схему данных; определять логическую структуру БД. Различные СУБД используют свои SQL-диалекты, расширяющие исходный язык запросов [39]. Основная функция программного комплекса СУБД – управление транзакциями, управление оперативной памятью, хранение данных, поддержка языков БД и журнализация. Критериями выбора СУБД – тип проекта, тип БД, объем хранилища, нагрузка, масштабируемость, безопасность и отказоустойчивость. При выборе СУБД учитывают функциональные возможности, документальную поддержку, наличие литературы, масштабируемость и интеграцию с другими программными продуктами. Самые популярные СУБД, поддерживаемые инструментами: MySQL Server, Oracle, DB2, Informix, PostgreSQL, MySQL, MS Access, SQLite, MS SQL Server и SQL Developer, которые позволяют просматривать объекты БД, запускать различные SQL инструкции, создавать и редактировать объекты БД, импортировать и экспортировать данные, а также создавать всевозможные отчеты [40]. Программа Oracle SQL Developer Data Modeler представляет собой универсальный, полностью автономный инструмент, с поддержкой логического, реляционного, многомерного моделирования и моделирования типов данных. С помощью пакета пользователи могут создавать, расширять и модифицировать модели данных, а также создавать приложения, используя различные технологии Java, J2EE, XML, PL/SQL и Web services. Пакет выполняет следующие функции:

визуальное моделирование взаимосвязей между сущностями; разделение реляционных и физических моделей; создание одной реляционной модели для разных версий БД; преобразование типов данных в реляционной модели [16].

Эффективность систем мониторинга зависит от используемой информационной технологии для непрерывного сбора, обработки и хранения данных. Информационные технологии на основе хранилищ данных позволяют обеспечить долговременное и надежное хранение информации, реализовать высокоэффективные информационно-поисковые системы с целью подготовки управляющих решений, сохраняя преемственность при изменении форм носителей информации. Основой информационной системы является БД.

*Хранилище данных.* Хранилище данных – система хранения из разных источников данных, предназначена для отчетности и анализа, поддержки принятия административных решений [37].

Хранилище данных (рисунок 1) состоит из [41]:

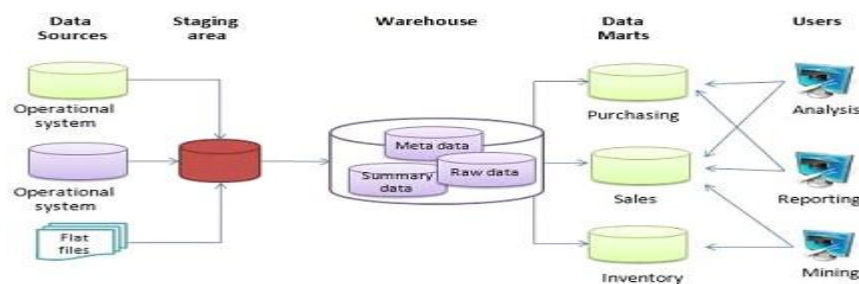


Рисунок 1.1 – Базовая архитектура хранилища данных [37]

- БД, которая образует единый источник для обработки данных для принятия решений;
- средств загрузки: процедур и технологий загрузки, очистки, структуризации данных для сбора качественных данных;
- средств анализа: OLAP, Data Mining, сводные отчеты.

Источниками хранилища данных могут быть: традиционные системы регистрации операций; отдельные документы; наборы данных. Операции с данными: извлечение (перемещение информации от источников данных в отдельную БД, приведение их к единому формату); преобразование (подготовка информации к хранению в оптимальной форме для реализации запроса необходимого для принятия решений); загрузка (помещение данных в хранилище); анализ (OLAP, Data Mining, сводные отчеты).

Общие принципы организации хранилища [39]:

- Проблемно-предметная ориентация. Информация и хранилище результатов в соответствии с основными аспектами деятельности организации, что отличает их от реляционной БД, где данные организованы в соответствии с процессами. Предметная организация данных в хранилище способствует упрощению анализа, и повышению скорости их выполнения

- **Неизменяемость.** Данные не изменяются в отличие от реляционной БД. Стабильность данных облегчает их анализ.

- **Интегрированность.** Исходные данные извлекаются из реляционной БД, проверяются, очищаются, приводятся к единому виду, в нужной степени агрегируют и загружаются в хранилище.

- **Привязка ко времени.** Данные в хранилище связаны с определенным периодом времени, что позволяет анализировать и наблюдать тенденции в развитии организации.

Используют методику компании Price Waterhouse (анализ процессов и данных; выбор модели данных и программного обеспечения; проектирование структуры хранилища данных; проектирование хранилища метаданных; загрузка хранилища данных) при разработке хранилища [42].

*База данных «Отходы».* В последние годы в странах СНГ, в частности в Республике Казахстан, накоплен многочисленный экспериментальный материал, связанный с отходами производства и потребления, накопленных в период единой государственной деятельности СССР. До сих пор многие города, районные и сельские населенные пункты находятся в зоне воздействия накопившихся отходов. Среди них особое место занимают крупнотоннажные отходы, которые образовались в результате многолетней промышленной и сельскохозяйственной деятельности предприятий. Во многих странах СНГ создаются региональные БД об отходах, образованные в результате агро-промышленной деятельности. Например, в Приморье, приступили к разработке электронной региональной БД об отходах производства и потребления. Создана электронная региональная БД «Отходы» [43], куда заносится полная и объективная информация об отходах: места, где складывается мусор; объемы; источники образования; применяемые предприятиями технологии использования и утилизации отходов; лицензии на обращение с отходами. В Белоруссии создана БД о расположении складов с непригодными пестицидами, их количестве, наличии СОЗ-пестицидов и смесей, с указанием даты их ликвидации и информации по их захоронению [44]. В Тюмени создана БД участков, подлежащих рекультивации вследствие загрязнения нефтью [45]. В Молдавии создана БД о загрязненных пестицидами территориях (Новости – Молдова, 3 января 2009) на базе мониторинга образцов из 1000 участков, инвентаризации бывших складов и прилегающей местности, для оптимальных способов решения проблем очистки и утилизации. В России разработана электронная модель схемы обращения с отходами на платформе ArcGIS 10.1, позволяющей управлять базами пространственных данных растровых и векторных форматов, формировать выборку из баз геоданных, на основе SQL-запросов с их графической визуализацией [46]. Полученная информация позволяет правительству этих стран решать стратегически важные вопросы, связанные с экологией. Организации, отвечающие за проведение рекультивации нарушенных земель, путем электронного обмена могут передавать поступившую информацию управлениям земельного ресурса для кадастрового учета, оценки и регистрации, а также налоговым органам и др. [42]. Например, база геоданных ArcGIS 9.3



применяется для земельных кадастров, систем инженерных коммуникаций, геодезии, недропользования и в других областях [47].

Сбор и обработка данных о состоянии воздуха, атмосферы и воды, а также об экосистемах окружающей среды проводится на основе больших данных, разработки моделей и компьютерного моделирования экологических информационных систем [11]. Разработана модель HDFS для хранения данных; модель MapReduce – обработки данных; Apache Pig – очистки и нормализации данных; Apache Hive – анализа; Cloudera Hue – визуализации данных [20]. Достижения в аналитических, сенсорных и вычислительных инструментариях позволили собрать огромное количество разнородных данных из разных дисциплин в области охраны окружающей среды. Доступность этих данных впервые позволила химикам-экологам оценить угрозу загрязнителей для здоровья человека и окружающей среды [48]. Для оценки загрязнения подземных вод используют модели мягких вычислений (soft computing). Эти исследования направлены на оценку приложений различных моделей мягких вычислений для моделирования загрязнения подземных вод нитратами. Считается, что автоматизация устройств мониторинга может повысить точность и скорость анализа окружающей среды [49].

Понятие «Big data» относится к набору разнообразных данных, поступающие с высокой скоростью для сбора, обработки и управления данными в реальном времени, т.е. обозначение структурированных и неструктурированных данных огромных объёмов, и используются для обработки больших, в сравнении со «стандартными» объемами, данных; для работы с быстро поступающими данными в очень больших объемах; а также для работы со структурированными и плохо структурированными данными [50,51]. Различные системы «больших данных» имеют следующие характеристики, три V: volume (количество данных), velocity (скорость данных), и variety (диапазон типов данных и источников). В дальнейшем критерии «больших данных» были дополнены еще пятью V: veracity (достоверность данных), viability (жизнеспособность данных), value (ценность данных), variability (переменчивость данных), и visualization (визуализация данных). Компания IDC, например, использует именно четвёртый критерий V – value (ценность данных), подчеркивая экономическую целесообразность обработки больших объёмов данных [52].

Размер «больших данных» варьируется от терабайтов до эксабайтов в зависимости от исследовательской области и наборов данных, программного и аппаратного обеспечения, используемых в изучаемых областях [53]. Благодаря большому объёму генерации данных и, как следствие, более доступному хранению этих данных, «большие данные» создаются в различных социальных сетях или датчиках для преобразования в полезную информацию [23].

Существуют две известные концепции, характеризующие «размер» данных [54-57]:

1) Первая концепция оценивает данные относительно способности человека или дисциплины управлять и анализировать данные. Объём «больших

данных» варьируется в зависимости от дисциплины, а для некоторых дисциплин они могут быть неуправляемо велики и будут зависеть от возможностей объекта, пытающегося его проанализировать.

2) Вторая концепция использует популярную и гибкую структуру «Четыре против» для оценки «больших данных». В этой структуре набор данных описывается по четырем характеристикам: объем, разнообразие, достоверность и скорость.

При создании приложений, работающих с «большими данными» ученые сталкиваются со следующим кругом проблем: чрезмерно большой объем данных; интенсифицированные потоки данных; существенное сокращение допустимого времени анализа данных; предел времени принятия решений при любом количестве данных; возрастание морфологической сложности моделей; возрастание структурной сложности моделей и систем; возрастание вычислительной сложности; относительный рост слабоструктурированной исходной информации; относительный рост нечеткой информации; рост потребностей в параллельных вычислениях и др. [23].

Работу с «большими данными» условно разбивают на три этапа: сбор данных (могут быть открытыми и внутренними источниками, для передачи информации используются стандартные интерфейсы приложений и протоколов); интеграция данных (специальные системы преобразовывают их в формат, подходящий для хранения, или непрерывно отслеживают для важных триггеров); обработка и анализ данных (операции выполняются в режиме реального времени).

*Сбор данных.* Увеличению размеров данных способствует деятельность IT-компаний, информационные данные в научных центрах и других организациях в самых различных областях человеческой деятельности. Предполагается, что выявление закономерностей в больших массивах данных становится основным инструментом исследования и получения новых знаний. Увеличению объемов данных способствуют не только деятельность IT-компаний, но и получение информационных данных в научных центрах и других организациях в самых различных областях человеческой деятельности. Большой объем данных обусловлен тем, что организации собирают информацию из различных источников, в том числе с помощью мультимедийных технологий и глобальной сети Интернет [58]. Активно разрабатываются системы для открытого обмена данными, методами, стандартами и программным обеспечением, включая среды совместной разработки, такие как GitHub и Open Science Framework, интерфейсы прикладного программирования (API), которые предоставляют ресурсы научных данных [4,55].

Для сбора данных используют следующие платформы:

1) DataONE, которая является организационным звеном для обнаружения данных из ресурсов для обучения обмена данными, этике и информатике;

2) EcoData Retriever – программная среда, которая автоматизирует задачи обнаружения, загрузки и переформатирования файлов экологических данных для хранения в локальных файлах или в реляционной БД. Автоматизация этих

задач значительно экономит время исследователей и существенно снижает вероятность ошибок, возникающих в результате ручных манипуляций с данными и сложности отдельных наборов данных. EcoData Retriever состоит из трех элементов:

- 1) основное приложение, которое управляет всеми стандартными задачами, связанными с загрузкой, предварительной обработкой и структурированием данных;
- 2) набор механизмов СУБД, которые позволяют Retriever взаимодействовать с различными видами программного обеспечения БД;
- 3) хранение информации, необходимой для получения данных.

Темпы создания, агрегирования и интерпретации экологических данных растут по многим направлениям с быстрым ростом объемов данных, методов сбора данных и новых аналитических и вычислительных подходов. Области больших и растущих потоков экологических данных включают [58,59]:

- 1) постоянную доставку петабайт данных, с помощью удаленных датчиков в системах наблюдения Земли;
- 2) объединение индивидуальных научных наблюдений и экспериментов в более крупные ресурсы данных сообщества (например, AmeriFlux, Глобальный информационный фонд по биоразнообразию, NutNet и Neotoma);
- 3) инвестиции в долгосрочные сети экологического мониторинга от национального до континентального масштаба (например, LTER, NEON и CZO);
- 4) развертывание автоматизированных и недорогих сенсорных сетей (например, фенологических камер, ловушек для съемки дикой природы и регистраторов температуры);
- 5) инициативы в области гражданской науки.

Рост и агрегация этих потоков данных впечатляют, потому что они создают новые возможности для изучения экологических систем с высоким разрешением и в широком масштабе, лучшего понимания лежащих в основе процессов и улучшения экологического прогнозирования.

*Интеграция данных.* Первоначально хранение «больших данных» было проблемой для ученых, так как обычные БД не подходят для их хранения. Появление экономичных и гибких инструментов для обработки и хранения данных типа Apache Hadoop и Apache Spark упростили ситуацию. Скорость обработки стала соответствовать современным требованиям обработки «больших данных» в реальном времени [60]. Apache Hadoop – исходный фреймворк с открытым исходным кодом для обработки и анализа больших наборов данных в кластерах. Для набора и построения сложных структур данных используют язык программирования Python, что позволяет эффективно обрабатывать необработанные данные любой степени сложности [61]. В 2009 г. Duncan et al. [62] разработал программу под названием FFI (Feat/Firemon Integrated) путем интеграции двух систем мониторинга последствий пожара, используемых в США: Firemon и Fire Ecology Assessment. FFI предназначена помочь менеджерам различных компаний в осуществлении сбора, хранения и анализа экологической информации, поддержания масштабируемого (от проекта до масштаба ландшафта) мониторинга на полевом и исследовательском уровнях, а

также поощрения совместного, межведомственного управления данными и обменом информацией [62].

*Анализ и обработка данных.* Для анализа «больших данных» используют различные методы [20,23,63-65], а именно:

- Регрессионный анализ (Regression analysis) – построение параметрической функции, описывающей изменение указанной числовой величины в указанный промежуток времени. Эта функция строится на основе известных данных, а затем используется для предсказания дальнейших значений этой же величины. С точки зрения обработки данных при регрессионном анализе ключевыми операциями являются вычисление текущей суммарной разницы и корректировка значений параметров;

- Классификация (Classification) – построение и использование зависимости одной переменной от нескольких других. Разработаны множество методов классификации, такие как функция Байеса, нейронные сети, машины поддерживающих векторов, деревья решений и т.д., например, используя нейронную сеть, моделируют накопление загрязняющих веществ в корнях растений, что улучшает наше понимание поглощения и перемещения загрязнителей окружающей среды растениями [24];

- Кластеризация (Cluster) – разделение больших групп на более мелкие, так, что члены одной группы более похожи друг на друга. В качестве критерия схожести используется функция – расстояние, на вход которой поступают две сущности, а на выход – степень их схожести. Известно множество различных способов кластеризации (иерархические, сети Кохонена и др.);

- Смешение и интеграция данных (Data fusion и data integration) – данные из разнообразных источников для глубинного анализа;

- Машинное обучение (МО) или искусственный интеллект (Machine learning) – преследует цель создания алгоритмов самообучения на базе статистического анализа данных или МО для получения комплексных прогнозов. Нейронные сети являются структурой МО и состоят из множества искусственных нейронов, которые при обучении образуют связи и потом могут анализировать информацию. Нейронные сети работают по единому алгоритму: получают на входе данные, «прогоняют» их через сеть своих нейронов и на выходе выдают результаты;

- Генетические алгоритмы (Genetic algorithms) – стохастический алгоритм глобальной поисковой оптимизации;

- Визуализация аналитических данных (Visualization of analysis data) – перевод информации в визуальный контекст, в виде карт, рисунков, графиков, схем;

- Краудсорсинг (Crowd sourcing) – информация от большого количества людей, которые предоставляют результаты через Интернет или социальные сети;

- А/В тестирование (A/B testing) – методология сравнения 2-х версий веб-страницы или приложения друг с другом, чтобы оценить их достоверность и определить какая из них работает лучше;

• Множественный поток команд или множественный поток данных (Multiple Instruction, Multiple Data) – вычислительная система со множественным потоком команд и множественным потоком данных.

Таким образом, в данной главе рассматриваются фундаментальные работы о результатах разработки информационной системы мониторинга факторов окружающей среды на базе больших данных и технологии хранилищ данных.

## **1.2 Математическое (компьютерное) моделирование в области экологии**

Диапазон и масштаб моделируемых процессов крайне велик – от глобальной экологии до прогнозирования динамики отдельных компонентов агроценозов, поэтому при классификации экологических моделей используются различные подходы [66,67]. Различают математические, детерминантные, стохастические и имитационные модели [67,68].

*Математические модели* представляют собой программу для ЭВМ, предназначенную для определения различных параметров изучаемых процессов, объектов и явлений, основанных на данных наблюдений и/или сформулированных математически либо статистически на законах развития процессов и явлений.

*Детерминантные модели* базируются на уравнениях, описывающих фундаментальные законы материального мира (законы сохранения энергии, массы, движения, переноса, трансформации и др.). Например, моделирование влагопереноса в почвенном профиле орошаемого поля, которое базируется на законах движения почвенной влаги.

*Стохастические модели* основаны на установлении закономерностей функционирования экологических систем путем статистического выявления взаимосвязей в этих системах или объектах. Разработка подобных моделей базируется на математической статистике, алгоритмах и расчете компьютерными средствами статистических данных [68].

Имитационное моделирование – это частный случай математического моделирования. Данный метод, позволяет строить модели, описывающие процессы так, как они проходили бы в действительности. К имитационному моделированию прибегают, когда: дорого или невозможно экспериментировать на реальном объекте; невозможно построить аналитическую модель; в системе есть время и причинные связи, последствия, нелинейность, стохастические (случайные) переменные; необходимо имитировать поведение системы во времени. Цель имитационного моделирования состоит в воспроизведении поведения исследуемой системы, на основе результатов анализа наиболее существенных взаимосвязей между её элементами [69].

По способам реализации модели в экологии подразделяются на регрессионные и эколого-физиологические [66]. Первые базируются на массовом экспериментальном материале. Вторые основаны на представлении о механизме функционирования модели целого объекта. Их оценка производится на основе

экспериментальных данных, что позволяет оценить достоверность гипотезы [70].

*Регрессионные модели.* МО нацелено на решение задач оптимизации, в эту схему входит техника регрессии. Математическая аппроксимация сводится к переопределенной и недоопределенной задаче оптимизации для линейных систем. Аппроксимация кривой, как простая техника регрессии, полиномиальная и экспоненциальная аппроксимация решается линейной системой уравнений [71]. Основная цель регрессионного анализа – предсказать значение зависимой переменной и оценить влияние некоторой объясняющей переменной на зависимую переменную:  $Y$  наблюдаемого бъекта и величинами  $x_1, x_2, \dots, x_n$ , которые обуславливают и объясняют изменения  $Y$ . Задача регрессионного анализа состоит в сопоставлении входного значения ( $x$ ) с непрерывной выходной переменной ( $Y$ ). Задача классификации сопоставляет входное значение  $x$  с дискретной выходной переменной  $Y$  [72]. Математические модели, описывающие статистические зависимости между величинами, изменение которых во времени обусловлено факторами, скрытыми от наблюдателя, являются инструментом исследования сложных систем процессов, протекающих в окружающем мире. В большом количестве случаев исходная информация, используемая для построения данных моделей и оценки их адекватности, представляет собой временные ряды (последовательности результатов измерений текущих значений одного или нескольких параметров, проводимых в упорядоченные моменты времени). Традиционные методы прогнозирования были предназначены в основном для линейных временных рядов, однако последние десятилетия начали активно развиваться методы прогнозирования нелинейных временных рядов [73]. Считают, что нелинейные модели более сложны, при этом перспективны и быстро прогрессируют в плане открытия новых принципов при описании сложных явлений [74]. Правильный выбор переменных существенно влияет на точность получения результатов прогнозирования и на время, необходимое для определения модели. При малой величине выборки, когда при построении регрессионных моделей, в случае невозможности применения Гауссовского применяют индуктивный подход [75]. Достижение глобального минимума внешнего критерия при порождении моделей означает, что модель, доставляющая такой минимум, является искомой [75].

Последние десятилетия одними из наиболее главных наработок в прогнозировании временных рядов являются комбинации математических и гибридных моделей. Гибридные математические модели и методы используются в различных приложениях. Математические модели и методы, использующие нейронные сети, нечеткую логику и эволюционные алгоритмы, являются наиболее часто используемыми, именно они вносят значительный вклад в улучшение понимания проблем распознавания и контроля сложных нелинейных систем и, при недостаточной информации, объекта управления [76-79]. Для нечетких нейронных сетей доказаны теоретические результаты, заключающиеся в том, что при определенных условиях данные объекты являются универсальными аппроксиматорами для любых непрерывных функций, заданных на компактном множестве [80]. Эволюционные алгоритмы, нейронные сети и

нечеткая логика имеют общую методологическую основу в виде формализации эвристических представлений о развитии биологических систем [77].

Среди гибридных математических моделей выделяются модели, объединяющие искусственные нейронные сети с моделями ARIMA [80]; метод группового учёта аргументов с искусственно нейронной сетью (МГУА) [81-83]; МГУА с различными преобразованиями [84,85]; МГУА с LSSVM (least squares support vector machines) [86]; основного компонента с МГУА [87]; гибридные модели [88].

Традиционный МГУА основан на исходном предположении, что данные могут быть смоделированы с использованием аппроксимации Вольтерра или Колмогорова-Габора многочлен. Система мониторинга и диагностики разработана на основе Artificial Neural Network – ANN methodologies, and applied to the Ipen research Reactor IEA-R1 [89,90]. Модель МГУА широко используется для решения широкого спектра задач, например, при классификации, прогнозировании, идентификации сложных систем, нахождении эмпирических зависимостей и кластеризации в разных областях науки. Например, для прогнозирования параметров волн для портовых, прибрежных и морских сооружений [83]; для оценки дорожного покрытия на дорогах [91].

МГУА – основан на рекурсивном селективном отборе моделей, на основе которых строятся более сложные модели. Точность моделирования на каждом следующем шаге рекурсии увеличивается за счет усложнения модели. МГУА был разработан российским кибернетиком А.Г. Ивахненко в 1966 г. для сложной системы моделирования, прогнозирования, идентификации и аппроксимации многофакторных систем для диагностики, распознавания образов и кластеризации выборки данных. Сущность МГУА состоит в том, что все детали разбиваются на классы по видам обработок. Каждый класс разбивается на группы, за основу которых берется не отдельная деталь, а группа сходных деталей. МГУА в модели представлены в виде линейного многочлена. Сложность структуры модели определяется количеством слагаемых в полиноме. Правило итерации увеличивается на один член при переходе к следующей строке, которая обеспечивает поиск уравнений. Для сокращения времени вычислений разработан пороговый алгоритм, который предварительно оценивает эффективность входных переменных на информационном уровне и ищет кандидатов-моделей на основе наиболее эффективных входных переменных (аргументов или функций). МГУА выстраивает полиномиальную степень в сотнях, тогда как стандартная множественная регрессия увязает в вычислениях и линейной зависимости. Метод А.Г. Ивахненко, как отмечает Farlow [91] является эмпирическим по своей природе. Модель МГУА основана на селективном отборе моделей [92].

*Эколого-физиологические модели.* Цель анализа заключается в том, чтобы определить величину и условия проявления максимума процесса, который является показателем реакции растения или сообщество растений на действие среды [93]. Математический анализ моделей позволяет исследовать взаимное и раздельное влияние факторов на изучаемый процесс, определять потенци-

альные возможности устойчивости растений к их отклонениям и другим эколого-физиологические показатели. Все это позволяет рассматривать регрессионные модели зависимости реакций растений на условия среды, полученных в многофакторных планируемых экспериментах, как эколого-физиологические характеристики генотипа сорта, популяции и вида [94].

В учебном пособии Куликовой и Лебедевой [95] отмечено, что в настоящее время наиболее широкое распространение получили полуэмпирические модели (имитационные), одновременно использующиеся при описании поведения вещества в почве, как формулы, выражающие фундаментальные законы природы, так и эмпирические показатели отдельных почвенных процессов с помощью уравнения Дарси и Ричардса для изучения и прогнозирования потоков воды и транспортировки растворенных веществ в различных почвах, посевах и климатических условиях [96, 97]. Данные подходы, включенные в программное обеспечение многих пакетов статистической обработки данных, получили развитие в виде пакета прикладных программ или библиотеки процедур [98,99]. Например, из-за многосторонности уравнение Ричардса применяется в программе SWAP (soil – water – atmosphere – plant). Программное обеспечение SWAP написано на языке программирования ФОРТРАН 77 [99]. Модели Hydrus используются для моделирования потока воды в пористых срезах почвенного субстрата [100]. Достоинством моделей является возможность прогнозирования транспорта веществ в условиях выраженного рельефа поверхности. Модели Hydrus (I, II) используются для имитации процесса инфильтрации и последующего передвижения влаги в рассматриваемых гидротермических условиях. Для моделирования Hydrus используется также нелинейное дифференциальное уравнение Ричардса для описания движения воды в ненасыщенных грунтах. Модель Leachm, разработанная Hutson and Wagenet [101] в 2014 г., является одномерной моделью воды и движения растворенного вещества, химических реакций и поглощения растениями в зоне аэрации. Модель моделирует движение потока воды и растворения веществ в почве в конкретных климатических условиях и основана на решении конвекционной дисперсии. Одним из результативных методов прогнозирования являются модели, основанные на физических законах передвижения влаги в почве, способные с заданной точностью воспроизвести во взаимной связи все режимобразующие процессы в почве [102]. Особый класс математических моделей представляют искусственные нейронные сети. Нейронные сети используют в качестве альтернативного способа решения широкого диапазона гидрологических задач [103]. На сегодняшний день, лидирующие позиции в исследовании моделирования в области экологии занимают Агрофизический научно-исследовательский институт Россельхозакадемии, Днепропетровский национальный университет, Кубанский государственный технологический университет, Казанский университет, и др. В России основные исследования направлены на моделирование процесса биодegradации углеводов, моделирования загрязнения почвы при разрыве нефтепровода для автоматизации процесса расчёта загрязнения почвы, переноса загрязнителей в почвенной структуре и их



взаимодействия с почвенной биотой, изменения концентрации нефти в загрязненных почвах под действием сорбентов и микроорганизмов, и т.д. [104,105].

В Казахстане компьютерное моделирование успешно используется для решения вопросов в области агробиоценоза, например, прогнозирования урожайности в зависимости от климатических и агротехнических условий. Это определяется практической значимостью таких моделей для оптимизации агрокультуры и тепличного хозяйства. Для аридных районов Северной Евразии казахстанские ученые совместно с российскими разработали модель для прогнозирования продуктивности пшеницы в зависимости от климатических факторов региона [106].

### **1.2.1 Математические модели процесса переноса токсических элементов (ТЭ) из почвы в органы растения**

Математическими моделями миграции и трансформации ТЭ, в частности тяжелых металлов занимались многие ученые мира: Фрид, Рубин, Грановский, Ефремов, Ху, Reihera и др. Актуальность создания математических моделей миграции и трансформации микроэлементов и ТЭ связана в первую очередь с тем, что изучение химического состава почвы связано с оценкой ее плодородия, выявлением токсичных химикатов и их метаболитов. Кроме того, известно, что ТЭ в почве могут превращаться в более опасные, чем исходное вещество [105]. Существующие модели миграции ТЭ разделяют на следующие группы: статистические, балансовые, вероятностные, конвективно-диффузионный перенос ТЭ в почвенном грунте [67,104]. Считают, что эти модели полностью не отражают процессы их накопления и перенос их из почвы в органы растений. Корень определяют эффективность поглощения растением воды и питательных веществ. Представление этих процессов в численных моделях имеют решающее значение, поскольку они оказывают глубокое влияние на глобальные запасы углерода и обратную связь с растением [107,108]. Тем не менее, авторы отмечают, что около корневая система изучалась в основном независимо от физики и химии почвы, микробиологии и физиологии растений. Сопутствующее воздействие гидрологических, геохимических и биологических процессов на потоки питательных веществ и углерода через корни из почвы было определено, как критический пробел в знаниях о них [109,110].

Существует два основных подхода при моделировании биологических процессов. Первый – это формализация известных фактов, полученных экспериментальным путем. Например, уравнения, связывающие в логарифмической форме ПДК вредных веществ в воде и атмосфере с учетом их физико-химическим свойств (молекулярной массой, растворимостью и т.д.). Другой подход к моделированию называют кибернетическим, так как процессы, происходящие в живых системах, можно рассматривать, как процессы управления обратной связью. Эти модели могут быть реализованы, как в виде уравнений функциональной зависимости, так и в виде программ для ЭВМ [105]. Исторически модели поглощения были разработаны для питательных веществ, т.е. переход растворенных питательных веществ из почвы [111]. Эти модели были использованы для обеспечения количественного знания о влиянии растений

на концентрацию элементов в почве [112,113]. Модели поглощения растением питательных веществ и ТЭ эволюционировали от простой модели на основе диффузии и массового расхода до более сложных моделей [114,115]. Доступность питательных веществ имеет решающее значение для роста растений и продуктивности [116]. В статье Hinsinger et al. [117] подчеркнута, что многокомпонентные модели реактивного транспорта, разработанные в геохимии, должны использоваться для учета многочисленных одновременных биогеохимических процессов, которые взаимодействуют с доступностью питательных веществ для растений. Здесь термин «многокомпонентный реактивный транспорт» означает, что в нескольких реакциях может быть один и тот же ион. Это может привести к взаимодействию между ионами и конкуренции за общие ионы, которые систематически контролируют концентрации. Например, изменения рН отдельного обменного катиона может влиять на концентрацию всех обменных катионов [118]. Для моделирования поглощения растворенных веществ корнями растений используют два подхода: микроскопический и макроскопический. При микроскопическом подходе корень растения, рассматривают в виде цилиндра, поэтому поглощение растворенного вещества корнем рассматривается в цилиндрической системе координат. Общее поглощение растворенного вещества корнем рассматривается, как сумма поглощения растворенного вещества одним корнем. Уравнение Михаэлиса-Ментона используется для расчета потребления питательных веществ растениями различными исследователями, учитывая ионы элементов, которые движутся к корням массовым потоком, диффузией, скорость поглощения, которая зависит от их концентрации. Например, уравнение Михаэлиса-Ментона было использовано для описания поглощения кадмия корнями [119]. При макроскопическом подходе в моделях извлечения вещества корневой зоной имитируют влияние засоленности почвы на поглощение воды растением, напор почвы и воды, давление корня и осмотический напор в системе «почва-вода» [120]. Большинство существующих моделей, которые имитируют поглощение растворенного вещества корнями имеют микроскопическую природу и используются для выявления факторов, которые влияют на механизм поглощения. Например, Gérard et al. [121] первым создал модель, где связаны модель архитектуры корневой системы с многокомпонентной моделью реактивного транспорта с использованием макроскопического подхода. Они продемонстрировали модель через приложение для поглощения фосфора растением в щелочной почве. Тем не менее, ученые игнорировали наличие потока воды в корне и почве при их моделировании. Для изучения гидрологических процессов в околокорневой зоне были разработаны многочисленные модели различной сложности, включая двустороннюю обратную связь между сосудами корневой системы и почвой. Некоторые модели используют аналогию схемы потока воды через сеть сосудов корневой системы [122-124]. Другие представляют собой полностью связанные корневые и почвенные системы, либо в одном [125], либо в трех измерениях с учетом корневой структуры [126-128]. Считают, что трехмерные (3D) модели могут лучше разрешать пространственное распределение воды и питательных веществ в результате взаимодействия между отдельными корнями и

гидро-биогеохимией почвы [124]. Потребность в 3D модели заключается в возможности объединения в системе «корень-почва»: 1) цикла поглощения воды; 2) поглощения и выделения воды корнями; 3) поглощения питательных веществ; 4) переменного насыщенного потока воды из почвы в сосуды корня; 5) адвекции, рассеивания и диффузии ионов в почве; 6) многокомпонентных реакций. Следует отметить, что эти модели предназначены в основном для прогнозирования потребления питательных веществ или загрязнения сельскохозяйственной продукции. Основываясь на моделях поглощения растений, первоначально разработанных для питательных веществ, разрабатывают модели поглощения ТЭ растением из загрязненной почвы. Многие исследователи отмечают, что математические модели, построенные на основе экспериментальных данных, не позволяют раскрыть механизмы, происходящие в системе, поэтому используют имитационное моделирование [129]. В таких моделях обычно используется значительное количество компонентов, учитывающих процессы трансформации химических соединений, диффузии, сорбции и др. Считают, что имитационное моделирование позволяет интегрировать значительный объем информации о физико-химических процессах, происходящих в системе, что позволяет интегрировать процессы миграции токсиканта в почвенно-растительной системе. Например, S.Trapp and J.McFarlane [130] разработали модель PLANTX для имитации поглощения ТЭ растением. Хотя модель можно рассматривать как полезный инструмент, для прогнозирования загрязнения растений ТЭ, авторы отмечают, что входные растительные параметры сложно получить. Далее была разработана модель CTSPAC (Coupled Transport of water, heat, and solutes in the Soil–Plant Atmosphere Continuum), как одномерная математическая модель, для совместного переноса воды и растворенных веществ в системе «почва–растение–атмосфера» [131]. Далее Ouyang [132] апробировал модель CTSPAC для учета очистки почвы, загрязненной 1,4-диоксаном, черенками тополя. Результаты этого исследования показали, что CTSPAC не подходит для применения к данным, полученным в полевых условиях. Mathur [133] разработал численную модель для определения поглощения кадмия растениями, выращенными для восстановления загрязненных почв. В числовой модели использовал одномерное уравнение Ричардса, связанное с уравнением переноса растворенного вещества. Кинетика поглощения вещества растением была определена уравнением Михаэлиса-Ментона, так как поглощение кадмия происходит по кинетической схеме. Далее в результате моделирования процессов взаимодействия в системе «почва-растение» использовали стационарное распределение в Марковских цепях. На основе стационарных марковских цепей в работе Ефремова [104] предложена система уравнений для оценки риска загрязнения компонентов системы:

$$P_{\Pi} = \frac{\lambda_2 \cdot \lambda_3}{\lambda_1 \cdot \lambda_2 + \lambda_2 \cdot \lambda_3 + \lambda_3 \cdot \lambda_1} \quad (1.1)$$

$$P_{\text{H}} = \frac{\lambda_1 \cdot \lambda_2}{\lambda_1 \cdot \lambda_2 + \lambda_2 \cdot \lambda_3 + \lambda_3 \cdot \lambda_1} \quad (1.2)$$

$$P_K = \frac{\lambda_1 \cdot \lambda_3}{\lambda_1 \cdot \lambda_2 + \lambda_2 \cdot \lambda_3 + \lambda_3 \cdot \lambda_1} \quad (1.3)$$

где  $P_n$ ,  $P_k$ ,  $P_h$  – вероятность загрязнения почвы, надземной и корневой части растений;  $I = \lambda_1 \cdot \lambda_2 + \lambda_2 \cdot \lambda_3 + \lambda_1 \cdot \lambda_3$  – интегральный показатель, характеризующий почвенно-растительный комплекс. Интенсивность переходов элементов  $\lambda_i$  определяли, как отношение концентраций элементов в соответствующих компонентах почвенно-растительных комплексов.

Наряду с математическими методами, созданы различные прикладные программы (Themodel, CTSPAC, PDM, Mass-Hoffman и Номее), которые оценивают способность поглощения ТЭ растением, на основе уравнения Ричардса и кинетической модели транспорта токсиканта в надземную биомассу, рассчитанной с помощью MATLAB 7.9 [134,135]. Так, в работе Sung et al. [136] предложена модель Themodel для прогнозирования восстановления почв, загрязненных тринитротолуолом и полибромированными бифенилами. Themodel – количественная модель, способная моделировать движение органических веществ через почву и воду, а также поглощение органических химических веществ. Моделирование предсказывает снижение концентрации загрязняющих веществ во времени и глубине почвы в полевых условиях. Далее Ouyang et al. [134] также моделировал очистку почвы, загрязненной тринитротолуолом, с помощью тополя. В качестве модели использовали модифицируемую модель CTSPAC, которая состоит из подмодели почвы и подмодели растения. Для имитации восстановления использовали вводные данные: диффузию тринитротолуола на границе почва-корень; скорость проникновения загрязнителя в ксилему и флоэму (проводящие ткани); время хранения тринитротолуола в производственных помещениях. Калибровку модели осуществляли путем корректировки этих трех значений методом проб и ошибок до тех пор, пока модельные предсказания не совпали с экспериментальными измерениями. В статье Motesharezadeh et al. [137] использовали модели Mass-Hoffman и Номее для моделирования очистки разных типов почв, загрязненных кадмием, в условиях засоления. Модель Mass-Hoffman использовалась для очистки почв, загрязненных кадмием, с помощью подсолнуха, а модель Номее – для моделирования солевого стресса. В своей статье Canales-Pastrana and Paredes [135] использовали динамическую модель очистки для имитирования процесса улетучивания хлорида ртути через биомассу. Данная модель оценивала скорость потока загрязняющих веществ и их концентрацию в растительной ткани.

Таким образом, моделирование процессов в системе «почва-растение» последние годы приобретает особую актуальность в области биорекультивации, как экономически выгодные по сравнению с другими дорогостоящими физическими и химическими технологиями.

### **1.2.2 Математическое и компьютерное моделирование биологических методов рекультивации**

Последние годы особое внимание уделяется вопросам рекультивации нарушенных экосистем биотехнологическими методами (восстановление почв/воды с помощью микроорганизмов и растений) и оценке риска ТЭ для окружающей среды и человека. Интерес к математическому и компьютерному моделированию процесса рекультивации загрязненных ТЭ почв/воды растет из года в год, в связи с высоким уровнем загрязнения среды ТЭ и разработкой технологий рекультивации. Исследование сложных механизмов рекультивации требует создания математических моделей, отражающих реальные процессы, протекающие в загрязненной почве [139-141].

*Биоремедиация – рекультивация почвы/воды с помощью микроорганизмов.* Для моделирования рекультивации почвы, загрязненной в основном углеводородами, с помощью микроорганизмов создан пакет прикладных программ [142,143]. Так, в работе Kuyukina et al. [144] предложена математическая модель стационарного процесса нефтеотмывания загрязненного грунта под действием микроорганизма *Rhodococcus* и сорбента. Разработанная модель основана на теории фильтрации жидкости через пористые материалы и была построена на основе экспериментальных данных по проникновению биосурфактанта в модельной почвенной колонке. Процесс фильтрации описывали уравнениями без инерционного течения идеальной несжимаемой жидкости под действием силы тяжести, объемной силы сопротивления, пропорциональной скорости и силы сопротивления, сосредоточенной на фронте проникновения и зависящей от скорости. Авторами предложено адекватное прогнозирование промышленного процесса отмывания, загрязненного почвогрунта в полевых условиях. В статьях Korostina с соавторами [142,143] показано, что авторы создали прикладные программы для решения вопроса восстановления почвы, загрязненной нефтью и тяжелыми металлами с помощью препарата на основе вторичных метаболитов микроорганизма «*Rhodococcus*». Программа состоит из 3-х блоков: 1-ый блок «Расчет времени отмывания биосурфактантом загрязненной нефтью почвы, находящейся в накопительном резервуаре». 2-ой блок «Определение параметра  $k$  фильтрации биосурфактанта в почве, загрязненной солями тяжелых металлов». 3-ий блок «Расчет динамики сорбции и десорбции солей тяжелых металлов в почве при фильтрации в ней раствора биосурфактанта». Последний блок прогнозирует эффективность очистки с помощью используемого препарата. В статье Э.В. Чеботаревой [145] построены математические модели изменения концентрации нефти в загрязненных почвах под действием сорбентов и микроорганизмов. При построении математической модели изменения концентрации нефти в почве, с помощью нефтеразлагающих микроорганизмов использовали модель Моно. Модель описывает процесс изменения количества нефтепродуктов под действием микроорганизмов. Далее, используя нестандартные численные методы, разработали модель биоремедиации трихлорэтилена, как распространенного загрязнителя грунтовых вод. Установлено, что нестандартные методы хорошо работают в условиях, где есть транспортные и нелинейные реакции. Модель включала поток

воды, транспорт трихлорэтилена и питательных веществ, рост бактерий и деградацию загрязнителя в воде. Авторы описали модель биodeградации загрязнителя в воде с помощью микроорганизмов [146].

*Фиторемедиация – рекультивация почвы/воды с помощью растений.* Несмотря на всеобщее признание фиторемедиации, как экономически эффективной и экологически безопасной технологии, ее эффективность трудно предсказать в отношении времени и из-за сложности биогеохимии почвы и физиологии растений в присутствии токсичных веществ [147]. Фиторемедиация в большей степени, чем другие методы очистки, требует конкретного исследования и точного планирования. Некоторые виды растений могут не подходить для восстановления загрязненных почв/воды. При моделировании очистки почвы/воды с помощью растений, отмечают трудности из-за многочисленных параметров, определяющих восстановительную значимость растительного организма [104,105]. В связи с этим, используют один параметр – коэффициент биологического поглощения (КБП). КБП, как основной показатель поглощения растением ТЭ из почвы, оценивается как отношение содержания ТЭ в вегетативных органах к их содержанию в почвенной среде [105]. Например, в статье Григорьева с соавт. [148] применяли имитационную модель, в основу которой положен КБП. Авторы оценили зависимость накопления в период вегетации различными органами растения тяжелых металлов (свинца, кадмия, ртути, мышьяка) от их содержания в почве. В качестве объекта использовали топинамбур. Авторы пришли к выводу о том, что практическое применение модели, позволит разработать технологию фиторемедиации загрязненных земель с помощью возделывания топинамбура для возврата маргинальных деградированных земель в сельскохозяйственный оборот. В работе Behrendt et al. [149] на основе КБП разработали математическую модель, основываясь на взаимодействии пестицидов и сельскохозяйственных растений. Используя математические модели, Wagenet and Rao [150] разработали модель поглощения пестицидов корнями в зависимости от движения воды в почве и их плотности. Аналогичные исследования показаны в работе Paterson and Schnoor [151], ученые представили компартментальную модель, описывающую накопление пестицидов в листьях растений по скорости переноса в системе «почва – корень – побеги – листья», а в исследовании Narayanan [152] представлена математическая модель, учитывающая влияние почвенной биоты в ассоциации с растением деградацию ТЭ в почве. В работе Kamalu et al. [153] моделировали кинетику транспорта тяжелых металлов в ксилему и флоэму, используя уравнение Ричардса и обыкновенное дифференциальное уравнение (ОДУ). ОДУ флоэмы и ксилемы подсчитывали с помощью программы MATLAB 7.9. Авторы установили, что с помощью ОДУ можно рассчитать кинетику транспорта тяжелых металлов по ксилеме и флоэме у растений-гипераккумуляторов, что способствует пониманию механизма накопления. Авторы отмечают, что при использовании ОДУ, экспериментальные данные из научных работ совпадают до 99.8-99.9%, а при использовании уравнения Ричардса эти совпадения намного ниже. В статье G.Thoma et al. [154] представлена математическая модель для

фитоочистки нефте-загрязненных земель, которая учитывает рост и старение корней в системе. Авторы представили математическую модель рекультивации загрязненных неподвижными органическими загрязнителями почв. Модель управляется форсированием функций, которые описывают скорость роста и распада биомассы корня, как функцию времени. Учитывая наблюдения за различными условиями среды (удобрение, выбор видов и т.д.), модель может служить руководством для выбора соответствующих стратегий, которые повысят скорость миграции загрязняющих веществ в полевых условиях. Модель может использоваться для выделения важных переменных из тех, которые имеют только незначительные эффекты, и экстраполировать результаты из одного географического региона на другой. Последние годы активно разрабатываются модели, учитывающие влияние динамики микробной биомассы на деградацию загрязняющих веществ в корневой зоне [132,136], и модели, связанные с внесением хелатирующих агентов для снижения загрязнителя в среде [155], а также модели трансформации ксенобиотика в корневой зоне [140,141]. В работе Lugli et al. [138] моделировали фитоочистку воды, загрязненной ионами  $Cd^{2+}$ ,  $Pb^{2+}$  и  $Zn^{2+}$ , с учетом типа культуры, плотности посева, системы полива, образования почвенной корки и длины корневой системы, используя численную модель Hydrus-1D. Модель Hydrus представляет собой набор Windows, которую используют для анализа расхода воды, тепла и транспорта растворенного вещества в разно насыщенных пористых средах (например, почв). Авторы установили, что взаимодействие загрязнителя в почве влияет на эффективность миграции элементов из почвы в растение. Из-за низкой подвижности  $Pb^{2+}$  в почве, данный элемент плохо поддается очистке, а для загрязнителей, характеризующимся высокой мобильностью (например ионы  $Zn^{2+}$ ), очистка более эффективна. Эти исследования также показали о возможности оптимизации процесса очистки почв путем выбора растений, с определенной длиной корневой системы.

Таким образом, литературный анализ показал, что компьютерное моделирование, как эффективный инструмент, может прогнозировать последствия нарушения почвенных процессов и проводить выбор оптимальной стратегии их восстановления.

### **1.2.3 Модели прогнозирования продуктивности растений на загрязненной ТЭ почве**

Последние годы среди растений особый интерес вызывает многолетние виды с высокой урожайностью. В основном это растения, которые относятся к энергетическим видам, такие тополь, ива, сахарный тростник, мискантус, эвкалипт, просо и др., которые используются для получения биотоплива, в частности биоэтанола и твердых видов топлива (пеллеты). Это связано с тем, что стремительно расширяющееся производство биотоплива из продовольственных культур ведет к уничтожению природных экосистем и утере биологического разнообразия [156].

Для выбора культур с высоким производством биомассы с учетом климатических условий, агротехнических свойств почвы и технических возможностей фермеров с учетом экономических, экологических и социальных рисков используют метод MCDM (multi-criteria decision making). Метод MCDM определяет цели, выбирает критерии отбора объекта, отражает их относительную важность с помощью математических методов [157-159]. Следующая модель АНР (The analytic hierarchy process) разработанная в 1980 г. Saaty [160], используя концепцию парных сравнений с помощью математических решений, позволяет легко проводить синтез качественных и количественных оценок в процессе принятия решений от выбора энергетических видов до выбора поставщиков. В статье Cobuloglu and Büyüktaktakın [161], используя методологию АНР с помощью стохастических решений, в штате Канзас (США) установили, что по экономическим, экологическим и социальным критериям просо является доминирующим видом для производства биотоплива, на втором месте находится мискантус, а пшеница и кукуруза получили низкие баллы. Авторы отмечают, что методы MCDM и АНР можно использовать в различных решениях, связанных с возобновляемой энергией и источниками биомассы. Используя методологию MCDM и АНР, разрабатываются другие модели и математические методы для отбора вида растения в зависимости от поставленной цели. Например, используя математическую модель TOPSIS, в Таиланде выбрали древесную стружку, как наиболее подходящее для страны по критериям эффективности, цены, простоты в эксплуатации [162,163]. В Италии разработали модель LUMP для выявления спроса на землю для выращивания целлюлозных энергетических культур для 28 государств Евросоюза [164]. В рамках модели LUMP была создана карта земель для этих стран, пригодных для выращивания двух основных групп энергетических культур: травянистых (мискантус, просо, гигантский тростник, тростниковая канареечная трава и артишок) и древесных (ива, тополь и эвкалипт), с использованием методов MCDM. Для моделирования местности учитывали биофизические факторы: климат, свойства почвы и топографические аспекты. С помощью модели LUMP имитировали ежегодные изменения в землепользовании на основе биофизических и социально-экономических факторов от 2010 до 2050 гг. Для прогнозирования урожайности биомассы энергетических видов и производства целлюлозной биомассы для конкретной местности с учетом климатических условий, гидрологических и биогеохимических свойств почвы разрабатываются другие модели, используя методологию MCDM [165]. 19 моделей были оценены в соответствии с критериями MCDM, как подходящие модели для оценки урожайности энергетических видов растений – это модели FASSET, ECOSSE, ANIMO, DNDC, DayCent, Expert-N, Ecosys, WNMM и CERES-NOE. Предпочтение отдают моделям, протестированным и разработанным для соответствующих мест и культур. Например, CERES-NOE не был проверен для многолетних растений, что может сделать его менее применимым [165]. Следует отметить, что универсальной модели нет, которая бы учитывала все виды энергетических культур в условиях конкретного региона. В связи с этим, раз-



рабатываются модели для прогнозирования урожайности энергетических травянистых видов в зависимости от условия произрастания для конкретных регионов, с учетом гидрологических и биогеохимических свойств почвы [166,167]. Например, модель DRAINMOD-GRAS разработана для прогнозирования урожайности энергетических травянистых видов (мискантус и просо) в равнинных регионах восточной части Северной Каролины, США, с учетом гидрологических и биогеохимических свойств почвы [166]. В обзоре Rui et al. [168] описано, что 14 моделей, из разработанных 23 моделей, внедрены в Европе для прогнозирования продуктивности биоэнергетических культур [169].

Разработанные модели делятся на 3 типа [167,170]. 1-ый тип – основан на солнечной радиации, их называют радиационными моделями: EPIC, ALMANAC, APSIM, APEX, ISAM, MISCANMOD и MISCANFOR [171-174]. Модели APSIM, MISCANMOD, MISCANFOR, WIMOVAC, Agro-IBIS, Agro-BGC, 3PG и SECRETS моделируют наряду с продуктивностью биомассы различные фенологические стадии энергетических культур (например, цветение), тогда как модели EPIC, ALMANAC, AUSCANE и LPJmL – только площадь листа, продолжительность вегетационного периода, снижение листового индекса в период старения растений [166,175-178]. Модели EPIC, ALMANAC, APSIM, ISAM и MISCANMOD основаны на эффективности использования ФАР, метеорологических параметров (температура, осадки), максимального индекса площади листа, максимальной глубины укоренения, методов управления посевами культур. Наряду с перечисленными моделями разрабатываются новые модели, что свидетельствует о широком интересе к этой области [179]. Например, в работе Jing et al. [167] провели расчет общего выхода биомассы путем прогнозирования годового выхода биомассы мискантуса, ивы и эвкалипта, с помощью добавления новых модулей в модель LINTUL (Light Interception and Utilization simulator): 1) индекса площади листа, моделируемого независимо от конкретной площади листа; 2) ежедневной эффективности использования света, температуры и интенсивности света в физиологических диапазонах сельскохозяйственных культур. Модель LINTUL имитирует рост, развитие (включая динамику индекса площади листа) и биомассу различных видов (деревья/кустарники и травы). Возможность LINPAC генерировать сайт-специфический вход параметров моделирования роста культур делает его применимым в региональных и глобальных масштабах. Модели типа AquaCrop моделируют урожайность энергетических культур с учетом водного баланса почвы в корневой зоне и интенсивности транспирации в вегетационный период [177]. Данные модели не позволяют получить удовлетворительное моделирование в стрессовых ситуациях, особенно при стрессе воды в поздних стадиях роста растений. Третий тип – интегрированная модель, которая учитывает биохимические (поглощение углерода и ассимиляцию) и физиологические процессы (фотосинтез и дыхание) в клетках растений при формировании биомассы. Типичные модели, использующие биохимический подход: Agro-IBIS и Agro-BGC, WIMOVAC и SECRETS [167].

Таким образом, большой коммерческий интерес к моделированию урожайности энергетических растений, в частности мискантуса, связано с высокой урожайностью биомассы и способностью произрастать на деградированных землях. Наряду с этим, растение обладает способностью рекультивировать загрязненные ТЭ и органическими поллютантами почвы [180-185]. Последнее десятилетие в Казахстане возрос интерес к возобновляемым источникам энергии в связи с увеличением спроса на энергию, нестабильностью цен на природные ресурсы и неуверенности в поставках из-за геополитических разногласий. Республика Казахстан для вхождения в пятерку мировых лидеров по производству альтернативных видов топлива проявляет интерес к развитию биотопливных технологий. В стране для производства биотоплива используют продовольственные сельскохозяйственные культуры, такие как рапс, соя, подсолнечник, кукуруза и пшеница третьего сорта. Отмечают, что непродовольственное растение мискантус является перспективной культурой для выращивания на нашей территории, в качестве источника целлюлозы [186].

## **ГЛАВА 2. МОДЕЛИРОВАНИЕ ПРОДУКТИВНОСТИ БИОЭНЕРГЕТИЧЕСКОГО РАСТЕНИЯ НА ПОЧВАХ, ЗАГРЯЗНЕННЫХ ТОКСИЧНЫМИ ЭЛЕМЕНТАМИ В ЗАВИСИМОСТИ ОТ КЛИМАТИЧЕСКИХ УСЛОВИЙ СРЕДЫ**

В последние годы были реализованы различные математические методы и модели машинного обучения (МО) для точного прогнозирования урожайности различных культур, которые позволяют оценить широкий спектр факторов высокой точностью. Математические методы, используемые для прогнозирования и анализа производственных результатов, обычно представляет собой модель линейной регрессии [187], наиболее успешными методами МО являются искусственные нейронные сети, регрессия опорных векторов, деревья регрессии [188]. Они различаются методами управления почвой и источниками питательных веществ [189]. В то же время широкомасштабная химизация в сельском хозяйстве привела не только к ухудшению качества почвы, но и к снижению урожайности сельскохозяйственных культур [190]. Потенциальный риск снижения и потери продуктивности культуры возникает не только из-за химизации, но также из-за погодных изменений [56, 57, 60, 61, 191].

Предметом нашего исследования является управление факторами риска, к которым относятся климатические условия и уровень загрязнения почвы токсичными элементами (ТЭ), на примере Алматинской области. В этом исследовании использовали климатические переменные, которые состоят из температуры воздуха, скорости ветра, максимальные и минимальные температуры и осадки, и расчеты потенциальное испарение почвенной влаги и активное фотосинтетическое излучение, уровень загрязнения почвы ТЭ, средний уровень загрязнения почвы составляет 20.0 мг/кг. Целью данной главы является оценка проблемы и потенциала использования больших данных (на примере климатических данных) для моделирования продуктивности биоэнергетического растения мискантус на загрязненных ТЭ землях.

При создании модели продуктивности растений на загрязнённой ТЭ почве рассмотрели три подхода:

- создание модели MiscanCalc на основе модификации модели MiscanMod для процесса формирования биомассы растений, как на незагрязненной, так и загрязненной ТЭ почве в зависимости от климатических условий среды;

- оценка регрессионных моделей МО для прогноза продуктивности растений, произрастающих на загрязненной ТЭ почве в зависимости от климатических условий среды;

- разработка математической модели для анализа связи между динамикой биомассы растения, произрастающих на загрязненной ТЭ почве и климатическими параметрами с помощью многоуровневого метода самоорганизации.

## 2.1 Создание модели Miscancalc для прогнозирования динамики формирования продуктивности растения на загрязненной токсичными элементами почве в зависимости от климатических условий среды

Для прогнозирования динамики формирования продуктивности энергетического растения мискантус на деградированных землях стран Европы, используя модель Miscanmod. Модель Miscanmod, разработана ирландскими учёными во главе с J.Clifton-Brown, на языке SimileXMLv3 в 2000 г. для моделирования роста и урожая мискантуса для стран Европы. Название файла исходного кода ProdMod (для EuClim) v7.1.xls. Данная модель была внедрена и апробирована для моделирования продуктивности биомассы мискантуса во Франции, Польше и Германии, так как эти страны имеют наиболее высокий потенциал для ее производства. Планируют 10% сельскохозяйственных земель использовать для выращивания мискантуса, как альтернативного источника энергии [192]. Язык Simile XMLv3 – предназначен для использования в качестве стандартного представления XML моделей Simile с прямыми отношениями «один к одному» между частями предложений Prolong и элементами XML для обеспечения преобразования в любом направлении.

*Комментарии по использованию модели Simile.* Пользователи сами указывают контролируемую температуру за вегетационный период. Первый день в вегетационном периоде вводится в переменную Simile «gs start day», а последний день вегетационного периода – «gs end day».

Miscanmod является непрерывной имитационной моделью блочного типа, работает на языке Microsoft Excel и состоит из 13 взаимосвязанных рабочих листов (включая 1 лист заметок). Модель рассчитывает входные параметры растительного организма ежедневно, что позволяет проводить сравнение с экспериментальными данными. Дата начала уборки при расчетах определяется по двум критериям: по прекращению наращивания биомассы растений и по падению листового индекса в ноль.

Модель Miscanmod, прогнозирует:

- расчет индекса площади листа в зависимости от дефицита влажности почвы, т.е. в результате водного стресса;
- эффективность использования радиации и температуры;
- количество максимальных дней в период активного накопления биомассы с учетом температуры и фотопериода в период цветения;
- дату завершения стадии цветения;
- количество максимальных дней в период активного накопления биомассы с учетом температуры и фотопериода в стадии развития;
- окончание сезона созревания, как начало уборки;
- прогнозирует расчет урожайности в  $\text{г м}^{-2}$  сухой биомассы.

Модель Miscanmod оценивает продуктивность растения на маргинальных деградированных землях без учета уровня загрязнения почвы. Высокие концентрации токсичных веществ в почве, образующихся в результате химизации в сельскохозяйственных и промышленных регионах, создает необходимость учитывать влияние их на продуктивные показатели растений. В связи с этим,

усовершенствовали модель Miscanmod с включением дополнительного блока «входных данных» путем расчета фотосинтетически активной солнечной радиации на основании информации региональных коэффициентов и данных интегральной радиации Казгидрометеорологической службы; блока входных данных путем расчета потенциального испарения влаги почвы на основании температуры, индекса тепла и средней продолжительности дня; ввода в блок метода путем дополнительного расчета разницы между эмпирическими месячными данными биомассы растений, выращенных на загрязненной и незагрязненной почвах (индекс загрязнения).

Модификационная модель Miscancalc была создана на языке C# в Visual Studio 2013, Windows 10 для оценки динамики формирования продуктивности растения на загрязненной ТЭ почве в зависимости от климатических условий среды на примере Алматинской области [193-196].

*Сравнение моделей Miscancalc и Miscanmod.* В отличие от модели Miscanmod, созданная нами модель *Miscancalc* (1) прогнозирует продуктивность биомассы мискантуса на загрязненных ТЭ землях, непригодных для выращивания агрокультур, (2) на основе конкретного изменения метода. Метод был создан путем дополнительного расчета разницы между эмпирическими месячными данными биомассы растений, выращенных на загрязненной и незагрязненной почвах, с помощью МНК. Коэффициенты детерминации были статистически значимы и близки к 1, что подтвердило статистическую достоверность уравнений в целом. Имитированный индикатор: так называемый индекс загрязнения  $F(x)$  был включен в блок модели Miscanmod.

$$F(x) = -0.0001x^3 + 0.023x^2 + 0.6689x - 4.842 \quad (2.1)$$

$$R^2 = 0.9993$$

где  $F(x)$  – индекс загрязнения (кг м<sup>-2</sup>),  $x$  – период роста (дни).

Математическую модель прогнозирования продуктивности мискантуса с помощью программы Miscancalc рассчитывали по следующей формуле:

$$W_y = S * \varepsilon_1 * \varepsilon_0 * \eta lk = S_t * K_p * l * \varepsilon_1 * \varepsilon_0 * \eta lk - F(x) \quad (2.2)$$

где  $W_y$  – урожайность (кг м<sup>-2</sup>);  $S$  – фотосинтетически активная радиация (ФАР) в районе в период роста (МДж м<sup>-2</sup>);  $\varepsilon_1 * \varepsilon_0$  – отношение перехваченного излучения культурой и преобразования его в энергию;  $\eta$  – эффективность уборки урожая;  $k$  – содержание энергии биомассы (МДж кг<sup>-1</sup>);  $S_t$  – общая годовая падающая радиация;  $K_p$  – отношение инцидентов ФАР к суммарному падающему излучению за вегетационный период;  $l$  – вегетационный период в течение года (дни);  $F(x)$  – уравнение индекса загрязнения (кг м<sup>-2</sup>);  $X$  – период роста растений (дни).

Математическую модель прогнозирования продуктивности мискантуса с помощью программы Miscanmod рассчитывали по следующей формуле:

$$W_y = S * \varepsilon_1 * \varepsilon_0 * \eta l k = S_t * K_p * l * \varepsilon_1 * \varepsilon_0 * \eta l k \quad (2.3)$$

где  $W_y$  – урожайность (кг м<sup>-2</sup>);  $S$  – фотосинтетически активная радиация (ФАР) в районе в период роста (МДж м<sup>2</sup>);  $\varepsilon_1 * \varepsilon_0$  – отношение перехваченного излучения культурой и преобразования его в энергию;  $\eta$  – эффективность уборки урожая;  $k$  – содержание энергии биомассы (МДж кг<sup>-1</sup>);  $S_t$  – общая годовая падающая радиация;  $K_p$  – отношение инцидентов ФАР к суммарному падающему излучению за вегетационный период;  $l$  – вегетационный период в течение года (дни).

*Постановка задачи. Входные метеорологические данные.* Использовали следующие климатические данные: 1 – испарение влаги почвы (РЕ/ПИ), 2 – фотосинтетическую активность радиации (PAR/ФАР), 3 – осадки (Rainfall), 4 – температуру (Temperature).

В файл разместили следующую информацию:

PE

PAR

Rainfall

Temperature.

Необходимое значение данных фонового излучения, осадков и температуры воздуха для Алматинской области (где проводился эксперимент) было предоставлено стационарной метеостанцией Казгидрометеорологической службы РК. Запрашиваемые значения ПИ и ФАР были математически рассчитаны, поскольку эти показатели не были измерены метеорологической службой, при модифицировании блоков модели Miscanmod: – блока входных данных путем расчета ФАР на основании информации региональных коэффициентов и данных интегральной радиации Казгидрометеорологической службы по солнечной радиации; блока входных данных путем расчета потенциального испарения влаги почвы на основании температуры, индекса тепла и средней продолжительности дня.

Потенциальное испарение влаги почвы оценивали по разнице между температурой, полученной из Казгидрометеорологической службы, индекса тепла и средней продолжительностью дня (час/мес) по формуле С.Thornthwaite [197]:

$$PET = 16 \left(\frac{L}{12}\right) \left(\frac{N}{30}\right) \left(\frac{10T_a}{I}\right)^a \quad (2.4)$$

где PET – потенциальное испарение (мм мес<sup>-1</sup>);  $T_a$  – средняя дневная температура (°C);  $N$  – количество дней в месяце;  $L$  – средняя продолжительность дня (часы) в месяце;  $I$  – индекс тепла, который зависит от 12 среднемесячных температур и подсчитывается по формуле:

$$I = \sum_{i=1}^{12} \left(\frac{T_a}{5}\right)^{1.514} \quad (2.5)$$

Приход ФАР (фотосинтетическая активная радиация). Среднесуточная плотность радиоактивных выпадений в приземном слое атмосферы на территории области колебалась в пределах 0,6-3,2 Бк м<sup>-2</sup>. Излучение солнца -163,3 кДж см<sup>-2</sup>, а средние значения радиационного гамма-фона приземного слоя атмосферы по населенным пунктам области находились в пределах 0,11-0,25 мкЗв ч<sup>-1</sup>.

Приход ФАР в течение вегетации (март-октябрь) рассчитывали по формуле Х.Тооминга [198], с использованием информации региональных коэффициентов и данных интегральной радиации Казгидрометеорологической службы:

$$Q_{\text{ФАР}} = 0.41 \sum S + 0.62 \sum D \quad (2.6)$$

где  $Q_{\text{ФАР}}$  – сумма прихода фотосинтетической активности радиации на посев культуры (МДж м<sup>-2</sup>);  $\sum S$  – сумма прямой радиации (МДж м<sup>-2</sup>);  $\sum D$  – сумма рассеянной солнечной радиации (МДж м<sup>-2</sup>).

Общей приход ФАР на посев культуры в период вегетации сельскохозяйственных растений с 1 марта по 30 сентября составил 1939 МДж м<sup>-2</sup>.

В начале вегетации мискантуса в апреле средняя температура составила 9,3°C (максимум - 15,6°C; минимум - 4,2°C), средняя температура в период вегетации растений – 17,3°C. Количество осадков с марта по сентябрь было 403 мм в год; потенциальное испарение влаги из почвы составило 397 мм. Максимальный дефицит влажности почвы в августе был равен 36,4 мм. Среднесуточная плотность радиоактивных выпадений в приземном слое атмосферы на территории области колебалась в пределах 0,6-3,2 Бк м<sup>-2</sup>; излучение солнца -163,3 кДж см<sup>-2</sup>, а средние значения радиационного гамма-фона приземного слоя атмосферы по населенным пунктам области было в пределах 0,11-0,25 мкЗв ч<sup>-1</sup>.

*Описание модели MiscanCalc.* Структура модели имеет интерфейс диалогового окна, состоящего из 6 страниц: «Входные данные» (Блок 1); «Физиостат» (Блок 2); «Индекса площади листа» (Блок 3); «Результаты биомассы на загрязненной почве» и «Продуктивность на незагрязненной почве» (Блок 4); «Продуктивность на техногенно-загрязненной почве» (Блок 5); «Результаты на графике» (Блок 6).

Диаграмма классов и их взаимосвязь, и диаграмма потока данных MiscanCalc представлены на рисунках 2.1.1 и 2.1.2.

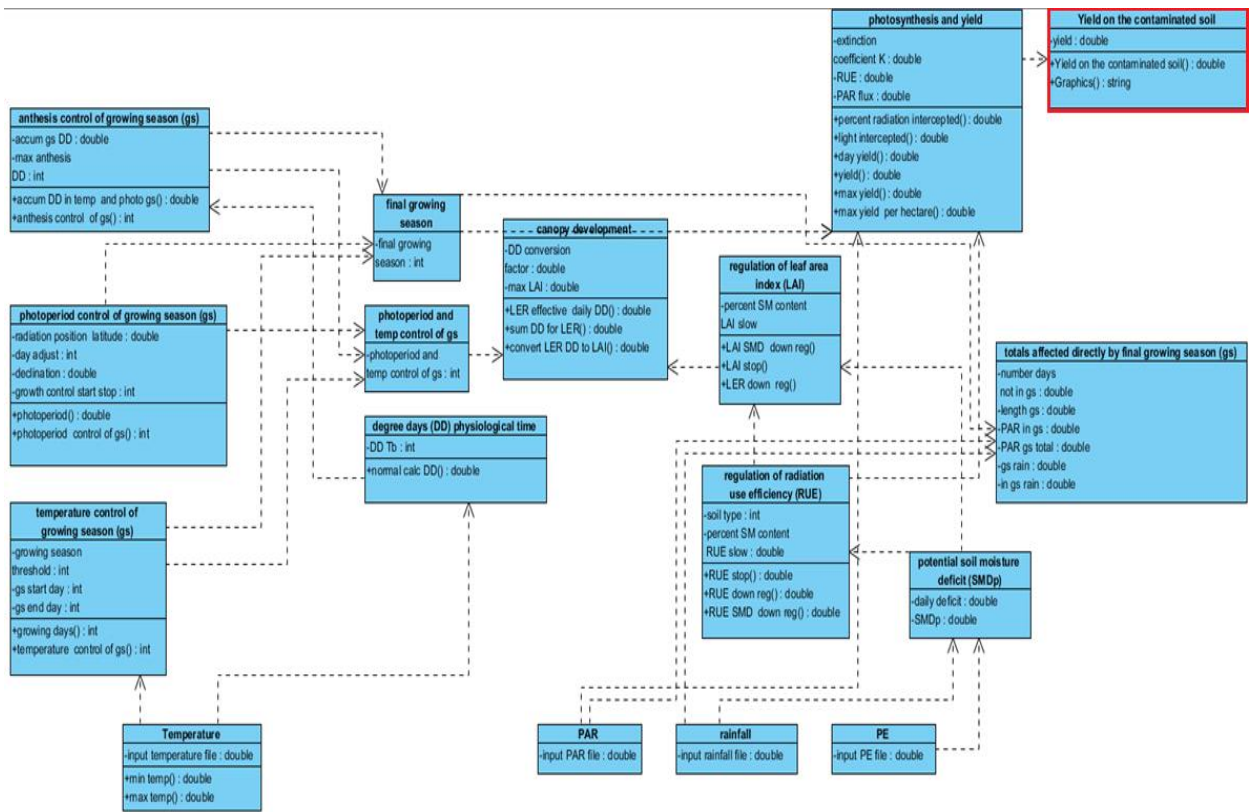


Рисунок 2.1.1 – Диаграмма классов Miscanalc

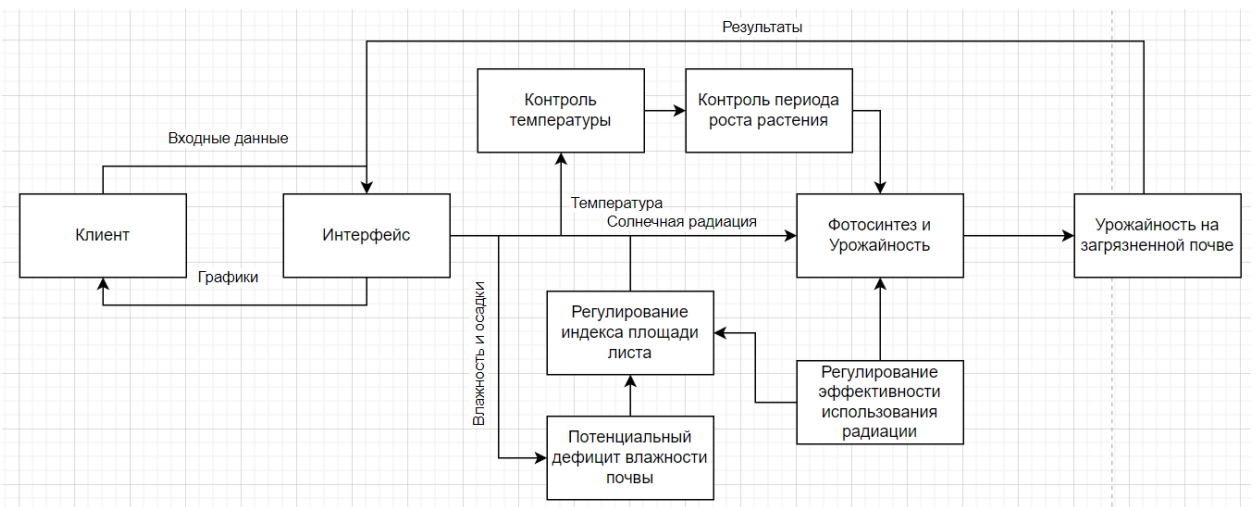


Рисунок 2.1.2 – Диаграмма потока данных Miscanalc

Первый блок «Входные данные» состоит из 4-х субмоделей. В программу вводятся исходные данные расчета ФАР (1 субмодель); исходные годовые данные количества осадков (2 субмодель); исходные годовые данные расчета интенсивности транспирации (3 субмодель); исходные годовые данные температуры воздуха (4 субмодель).

В данном блоке автоматически рассчитываются следующие показатели: максимальная и минимальная температуры; минимальная суточная температура; потенциальный дефицит влажности почвы; зависимость дефицита влаги



в почве от эффективности использования излучения; влияние водного стресса на коэффициент индекса площади листа.

*Математические расчеты модели.* Методы модели рассчитывали максимальную и минимальную суточную температуру по следующим формулам (°C):

$$T_{max} = T + (t * 0.2) \quad (2.7)$$

$$T_{min} = T - (t * 0.2) \quad (2.8)$$

где T – суточная температура (°C).

*Расчет зависимость дефицита влаги в почве от эффективности использования излучения (ФАР).* Процентное содержание влаги в почве при котором эффективность использования излучения (ФАР) начинает замедляться.  $RUE_{dr}$  методы модели рассчитывает:

$$a = \frac{\sqrt{RUESMD_{dr}}}{\sqrt{LAI_{st}} - \sqrt{RUESMD_{dr}}} \quad (2.9)$$

$$b = \frac{1}{(\sqrt{LAI_{st}} - RUESMD_{dr})} \quad (2.10)$$

$$RUE_{dr} = \begin{cases} 0; & SMD_p < RUI_{st} \\ 1; & SMD_p > RUI_{st} \\ 1 + a + b * SMD_p; & SMD_p < RUESMD_{dr} \\ 1; & SMD_p > RUESMD_{dr} \end{cases} \quad (2.11)$$

Вычисляется масштабный коэффициент (от 0 до 1, включительно), который регулирует расчет выхода с использованием эффективности излучения (ФАР) в зависимости от потенциального дефицита влажности почвы (SMD<sub>p</sub>), то есть степени стресса воды.

Потенциальный дефицит влажности почвы оценивается, когда эффективность использования излучения (RUE), коэффициент масштабирования становится равным нулю  $RUESMD_{dr}$  (мм):

$$RUESMD_{dr} = RUI_{st} * \left( \frac{\text{percentSMcontentRUEslow}}{100} \right) \quad (2.12)$$

где percentSMcontentRUEslow = 75 – процентное содержание влаги в почве (SMC), при котором эффективность использования излучения (RUE) начинает замедляться равен 75, а  $RUI_{st} = 162$  – равен 162.

Потенциальный дефицит влажности почвы SMD<sub>p</sub> (мм):

$$SMD_p = 0; \text{rainfall} - PE > 0; \text{rainfall} - PE < 0 \quad (2.13)$$

где Rainfall - осадки (мм), PE- потенциальное испарение (мм).

*Расчет влияния водного стресса на коэффициент индекса площади листа.* Влияние водного стресса на коэффициент индекса площади листа определяет:

$$LAI_{st} = RUI_{st} - (RUI_{st} / 10) \quad (2.14)$$

Потенциальный дефицит влажности почвы ( $SMD_p$ ) (мм), когда коэффициент масштабирования индекса площади листа ( $LAI$ ) становится равным нулю:

$$LAISMD_{dr} = LAI_{st} \left( \frac{\text{percentSMcontentLAIslow}}{100} \right) \quad (2.15)$$

где percentSMcontentLAIslow = 50 – процентное содержание влаги в почве (SMC), при котором индекс площади листа ( $LAI$ ) начинает замедляться равен 50%.

Далее оценивается масштабный коэффициент (от 0 до 1, включительно), который регулирует расчет индекса площади листа ( $LAI$ ) (через уменьшенный  $LER$ ) в зависимости от потенциального дефицита влажности почвы ( $SMD_p$ ), то есть степень водного стресса ( $LER_{dr}$ ) (мм):

$$a = \frac{\sqrt{LAISMD_{dr}}}{\sqrt{LAI_{st}} - \sqrt{LAISMD_{dr}}} \quad (2.16)$$

$$b = \frac{1}{(\sqrt{LAI_{st}} - LAISMD_{dr})} \quad (3.17)$$

$$LER_{dr} = \begin{cases} 0; & SMD_p < LAI_{st} \\ 1; & SMD_p > LAI_{st} \\ 1 + a + b * SMD_p; & SMD_p < LAISMD_{dr} \\ 1; & SMD_p < LAISMD_{dr} \end{cases} \quad (2.18)$$

*Расчет дефицита влажности почвы:*

Daily defecit = if (rainfall - PE) > 0 then 0 else rainfall - PE.

Расчет зависимости дефицита влаги в почве от эффективности использования излучения (RUE):

RUE stop = if soil\_type == -9999 then 0 else - 1 \* soil\_type.

Расчет если RUE=0. RUE SMD down reg = RUE\_stop \* (percent\_SM\_content\_RUE\_slow / 100)

Расчет если RUE < 1. Percent SM content RUE slow = 75

Процентное содержание влаги в почве (SMC), при котором эффективность использования излучения (RUE) начинает замедляться.

$$RUE\ down\ reg = a = \frac{|RUE\ SMD\ down\ reg|}{|(RUE\_stop| - |RUE\_SMD\_down\_reg|)}, b$$

$$= \frac{1}{|RUE\_stop| + |RUE\_SMD\_down\_reg|} \quad (2.19)$$

$$if\ (SMDp < RUE\_stop)\ then\ 0\ elseif\ (SMDp < RUE\_SMD\_down\_reg)\ then\ 1 + a + b * SMDp\ else\ 1. (74) \quad (2.20)$$

Расчет влияния водного стресса на коэффициент индекса площади листа. Регулирование индекса площади листа в зависимости от влажности почвы, т.е. от степени водного стресса, рассчитывался:

$$LAI\ stop = RUE\_stop - \left(\frac{RUE\_stop}{10}\right) \quad (2.21)$$

В случае, когда потенциальный дефицит влажности почвы (SMDp) (мм), когда индекс площади листа (LAI) = 0.

$$LAI\ SMD\ down\ reg = \frac{LAI\_stop * (percent\_SM\_content\_LAI\_slow)}{100} \quad (2.22)$$

В случае, когда потенциальный дефицит влажности почвы (SMDp) (мм), когда индекс площади листа (LAI) < 1.

$$LER\ down\ reg = a = \frac{|LAI\_SMD\_down\_reg|}{|(LAI\_stop| - |LAI\_SMD\_down\_reg|)}, b$$

$$= \frac{1}{|LAI\_stop| + |LAI\_SMD\_down\_reg|}, if\ (SMDp < LAI\_stop)\ then\ 0\ elseif\ (SMDp < LAI\_SMD\_down\_reg)\ then\ 1 + a + b * SMDp\ else\ 1 \quad (2.23)$$

Процентное содержание влаги в почве (SMC), при котором индекс площади листа (LAI) начинает замедляться составляет 50%.

$$Percent\ SM\ content\ LAI\ slow = 50.$$

Второй блок «Физиостат» состоит из 3-х субмоделей. Программа рассчитывают следующие показатели: количество физиологически активного периода роста при высокой и пороговой температуре в дневной период (субмодель 1); количество дней в году, когда температура (взятая отдельно) фактически способствовала росту растений (субмодель 2); количество дней в период активного накопления биомассы с учетом температуры и фотопериода в период вегетации (субмодель 3).

Программы и математические расчеты во втором блоке рассчитывают следующие показатели: количество физиологически активного периода роста при высокой и пороговой температуре в дневной период; количество дней в

году, когда температура фактически способствовала росту растений; количество дней в период активного накопления биомассы с учетом температуры и фотопериода в период вегетации.

*Математические расчеты.* Модель рассчитывает день начала роста растений в зависимости от температуры (т.е. первый день в блоке) – 121-ый день (расчет дней ведется с начала года),  $startd = 121$ , и день завершения роста растений в зависимости от температуры (т.е. последний день в блоке) – 275-ый день с начала года,  $endd = 275$ .

Оценивается среднесуточный температурный порог ( $^{\circ}\text{C}$ ), который потенциально допускает рост при равном или превышенном температурном контроле в течение вегетационного периода  $TC_{gs}$ .

$$TC_{gs} = \begin{cases} 0; & t \geq startd \cup t \leq endd \\ 1 & \end{cases} \quad (2.24)$$

Оценивается влияние температуры на рост растений  $T_b$  (день  $^{\circ}\text{C}$ ) :

$$calcDD = \begin{cases} 0; & T_{max} < T_D \\ \frac{(T_{max} - T_D)}{4}; & \frac{(T_{min} + T_{max})}{2} < T_D \\ \frac{(T_{max} - T_D)}{2} - \frac{(T_D - T_{min})}{2}; & \frac{(T_{min} + T_{max})}{2} > T_D \\ \frac{(T_{min} + T_{max})}{2}; & T_{max} > T_D \end{cases} \quad (2.25)$$

*Третий блок «Физиостат»* состоит из 2-х субмоделей. В данном блоке на основании модели рассчитываются следующие показатели: процент снижения ФАР, влияющего на урожайность, с учетом индекса площади листа; количество фотосинтетически активных дней в период роста.

*Математические расчеты.* Географическая широта местности в градусах составляет 52,40 (radiation position latitude = 52.40). Склонение солнца определяются:

$$D = 0.1303 * \pi * \cos\left(2 * \pi * \frac{(dayadj + 10)}{365}\right) \quad (2.26)$$

Скорректированное число дней, необходимое для расчетов склонения солнца подсчитывает:

$$dayadj = \begin{cases} a = t + 183; & a > 365 \\ a - 365; & a < 365 \\ a; & a = 365 \end{cases} \quad (2.27)$$

Фотопериод (дневной свет) с учетом широты и склонения солнца рассчитывается:

$$P = \begin{cases} a = -1 * \frac{\tan(L_{rp} * \pi)}{180} \\ b = \tan(D) \end{cases} \quad (2.28)$$

Модель учитывает количество дней в зависимости от фотопериода, т.е. период активного роста растений:

$$P_{gs} = \begin{cases} 0; P > 0 \\ 1; P < 0 \end{cases} \quad (2.29)$$

Эффективность использования радиации  $RUE$  (моль фотосинтетически активных квантов – 1 восстановленного  $CO_2$ , г мДж<sup>-1</sup>) составляет 2,35. На основании эффективности  $RUE$  далее рассчитывается процент падающего фотосинтетически активного излучения (PAR) на растение, на основании этих данных программа рассчитывает индекс площади листа (LAI):

$$PRI = (1 - e^{-1 * K * LER_{DD}}) * 100 \quad (2.30)$$

где  $K$  – коэффициент отмирания равен 0.68 .

$$LER_{DD} = LER_{dr} * calcDD \quad (2.31)$$

photoperiod control of gs= if photoperiod >growth\_control\_start\_stop then 0 else 1.

На основании расчета физиологически активного периода роста растений при высокой и пороговой температурах в дневной период прогнозируется период начала и окончания цветения, активного роста и накопления биомассы.

Модель вычисляет влияние температуры на рост растений  $Tb$  (день °C):

$$\begin{aligned} Normal\ calc\ DD &= if (max\_temp < DD\_Tb) then 0 elseif ((min\_temp + max\_temp)/2 < DD\_Tb) then (max\_temp - DD\_Tb)/4 elseif (min\_temp < DD\_Tb) then ((max\_temp - DD\_Tb) / 2) - ((DD\_Tb - min\_temp) / 4) else ((min\_temp + max\_temp) / 2) - DD\_Tb \end{aligned} \quad (2.32)$$

Далее модель определяет пороговую дневную температуру (°C):

$$DD\ Tb = 10$$

Оценивает период цветения в вегетационный период:

$$Accum\ DD\ in\ temp\ and\ photo\ gs = 0$$

Учитывает количество дней в период накопления биомассы (DD, в дни °C) в зависимости от температуры и фотопериода (гс):

$$Max\ anthesis\ DD = 1800$$

Оценивает количество дней в период цветения в зависимости от температуры и фотопериода, следовательно, прогнозирует окончание вегетационного периода:

*Anthesis control of gs = if accum\_DD\_in\_temp\_and\_photo\_gs  
> max\_anthesis\_DD then 1 else 0*

Переключатель указывает время завершения периода цветения:

*Accum gs DD = current\_max = prev(1), if accum\_DD\_in\_temp\_and\_photo\_gs  
> current\_max then accum\_DD\_in\_temp\_and\_photo\_gs else current\_max*

Переключатель, определяющий дни, когда температура и фотопериод способствуют росту растений («температурный и фотопериодический вегетационный период»).

*Photoperiod and temp control of gs*

*= if (photoperiod\_control\_of\_gs + temperature\_control\_of\_gs) =  
= 0 then 0 else 1*

Переключатель определяет завершение вегетационного периода:

*final growing season = sum\_of\_control\_switches = anthesis\_control\_of\_gs +  
temperature\_control\_of\_gs +  
photoperiod\_control\_of\_gs, if sum\_of\_control\_switches == 0 then 0 else 1.*

Расчета прихода ФАР на культуру, учитывают его текущий индекс площади листа (LAI):

*percent radiation intercepted*

*= (1 - exp(-1 \* extinction\_coefficient\_K \* convert\_LER\_DD\_to\_LAI))  
\* 100*

*Влияние ФАР на накопление надземной биомассы в период вегетации в зависимости от температуры и фотопериода.*

*Light intercepted = PAR \* (percent\_radiation\_intercepted/100)*

Программа определяет количество ФАР, влияющего на накопление надземной биомассы в дневное время:

*day yield = if final\_growing\_season*

*< 1 then light\_intercepted \* RUE\_down\_reg \* RUE else 0*

вычисляет коэффициент вымирания К (0.68).

*extinction coefficient*

Далее программа вычисляет максимальный урожай, в г сухого вещества (СВ) м<sup>-2</sup> площади земли:

*max yield = current\_max = prev(1), if yield*

*> current\_max then yield else current\_max*

а также максимальный урожай, т СВ га<sup>-1</sup>.

*Max yield per hectare = max\_yield/100*

*Расчета показателей температуры и солнечной радиации, которые непосредственно влияют на рост и накопление надземной биомассы в течение вегетационного периода. Алгоритмы модели проводят температурный контроль в течение вегетационного периода и рассчитывают:*

1. Начало роста растений в зависимости от температуры:

*gs start day = 121*

2. Окончание роста растений в зависимости от температуры:

*gs end day = 275*

3. Коммутатор устанавливает 0, если средняя температура дня потенциально может привести к росту растения, если нет, то 1:

$$\text{Growing days} = \text{if} ((\text{max\_temp} + \text{min\_temp})/2) < \text{growing\_season\_threshold} \text{ then } 1 \text{ else } 0.$$

4. Проводит учет дней года, когда температура (взятая отдельно) фактически способствовала росту растений:

$$\text{Temperature control of gs} = \text{today} = \text{time}(), \text{if} (\text{today} > \text{gs\_start\_day}) \text{ and } (\text{today} \leq \text{gs\_end\_day}) \text{ then } 0 \text{ else } 1.$$

*Контроль за фотопериодом в течение вегетационного периода*

5. Солнечное излучение в период вегетационного периода:

$$\text{Photoperiod} = a = -1 * \tan((\text{radiation\_position\_latitude} * \pi()) / 180), b = \tan(\text{declination}), 24 * (\text{acos}(a * b) / \pi())$$

6. Приход солнечной радиации:

$$\text{photoperiod control of gs} = \text{if photoperiod} > \text{growth\_control\_start\_stop} \text{ then } 0 \text{ else } 1.$$

7. Количество дней активного солнечного излучения в году, которое способствовало росту растений:

$$\text{Declination} = 0.1303 * \pi() * \cos((2 * \pi() * (\text{day\_adjust} + 10)) / 365)$$

*Алгоритмы расчета периода завершения роста растений оценивает:*

1. Суммирование количества дней в году, не входящих в вегетационный период:

$$\text{number days not in gs} = 0.$$

2. Суммирование прихода ФАР в конце вегетационного периода:

$$\text{PAR in gs} = 0;$$

3. Суммирование количества осадков (мм) в конце вегетационного периода:

$$\text{gs rainfall total} = 0.$$

Время уборки урожая растения программа оценивала по 2-м параметрам: 1) период окончания накопления биомассы; 2) падение листового индекса до 0. На рисунке 2.1.3 представлены результаты расчета влияния смоделированной засухи на снижение роста индекса площади листа мискантуса в период вегетации, а период активного накопления биомассы мискантуса в зависимости от ФАР в течение вегетационного периода изображен на рисунке 2.1.4.

Расчеты физиологически активного периода роста растений выявили, что физиологически активный рост мискантуса начинается с 121 дня и завершается на 275 день от начала года. Вегетационный период составляет 154 дня. Растение начинает прорастать в конце марта, когда средняя температура воздуха в третьей декаде достигает 14,6°C. Начало роста по литературным данным в странах Европы также инициируется при температуре воздуха 10°C [137]. В конце сентября активный рост растений завершается, происходит постепенное отмирание листьев при температуре 9,8°C.

*Четвертый блок «Продуктивность на незагрязненной почве» имитирует продуктивность биомассы с учетом температуры и фотопериода в период вегетации на незагрязненной деградированной почве, используя математические расчеты.*

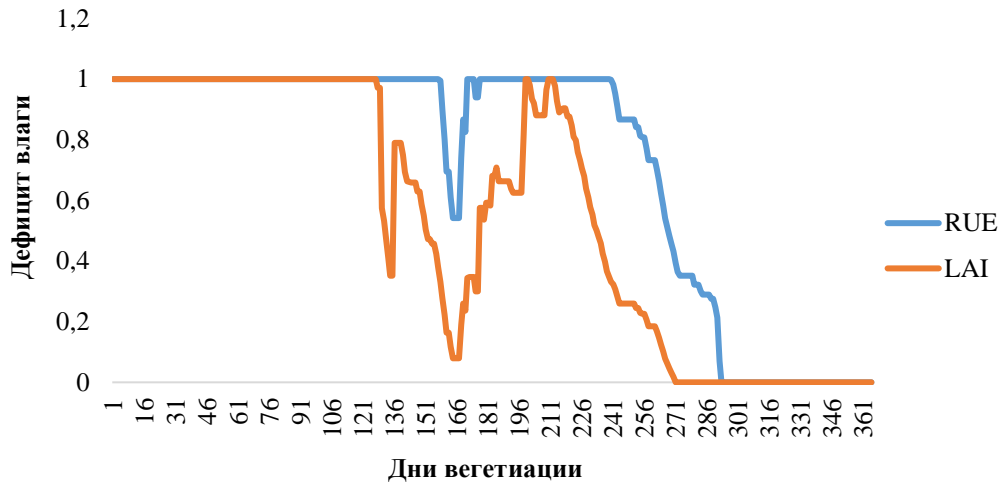


Рисунок 2.1.3 – Смоделированный дефицит влаги в почве (RUE) из-за эффективности использования радиации и влияния водного стресса на коэффициент индекса площади листа (LAI)

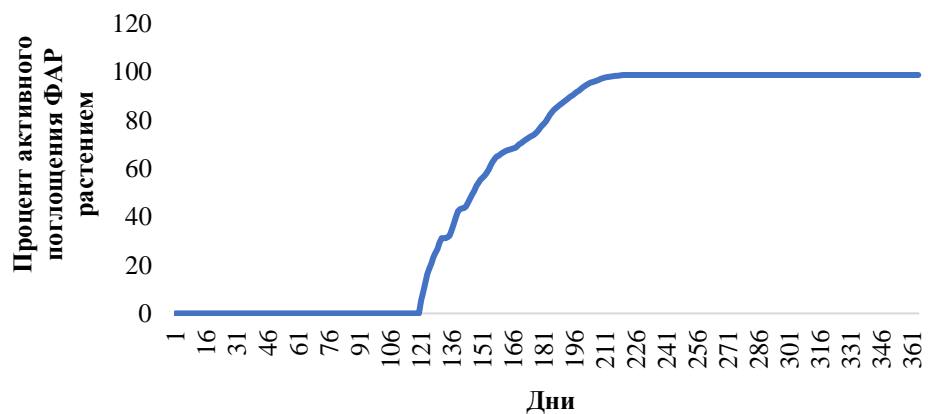


Рисунок 2.1.4 – Период активного накопления биомассы мискантуса в зависимости от ФАР в течение вегетационного периода

*Математические расчеты.* Продуктивность биомассы с учетом температуры и фотопериода в период вегетации на незагрязненной деградированной почве. Максимальный урожай (кг СВ га<sup>-1</sup>):

$$\sum B_{max} = \frac{\sum B_{day}}{100} \quad (2.33)$$

Биомасса (г СВ м<sup>-2</sup>):

$$B_{day} = \begin{cases} LI * RUE_{dr} * RUE; & FGS < 1 \\ 0; & FGS > 1; \end{cases} \quad (2.34)$$



Перехваченный света (%):

$$LI = PAR * \left(\frac{PRI}{100}\right) \quad (2.35)$$

Пятый блок «Продуктивность на техногенно-загрязненной почве на основании имитированного метода оценивает продуктивность мискантуса с учетом температуры и дневного света в период вегетации на загрязненной ТЭ почве.

Математические расчеты:

$$\sum B_p = \sum B_{day} - F(x) \quad (2.36)$$

$$F(x) = -0.0001x^3 + 0.023x^2 + 0.6689x - 4.842 \quad (2.37)$$

$$R^2 = 0.9993$$

Шестой блок «Результаты на графике» – графически изображаются данные динамики формирования биомассы растительного организма в течение вегетационного периода на незагрязненной и загрязненной ТЭ почвах (г СВ м<sup>-2</sup>; рисунок 2.1.5).

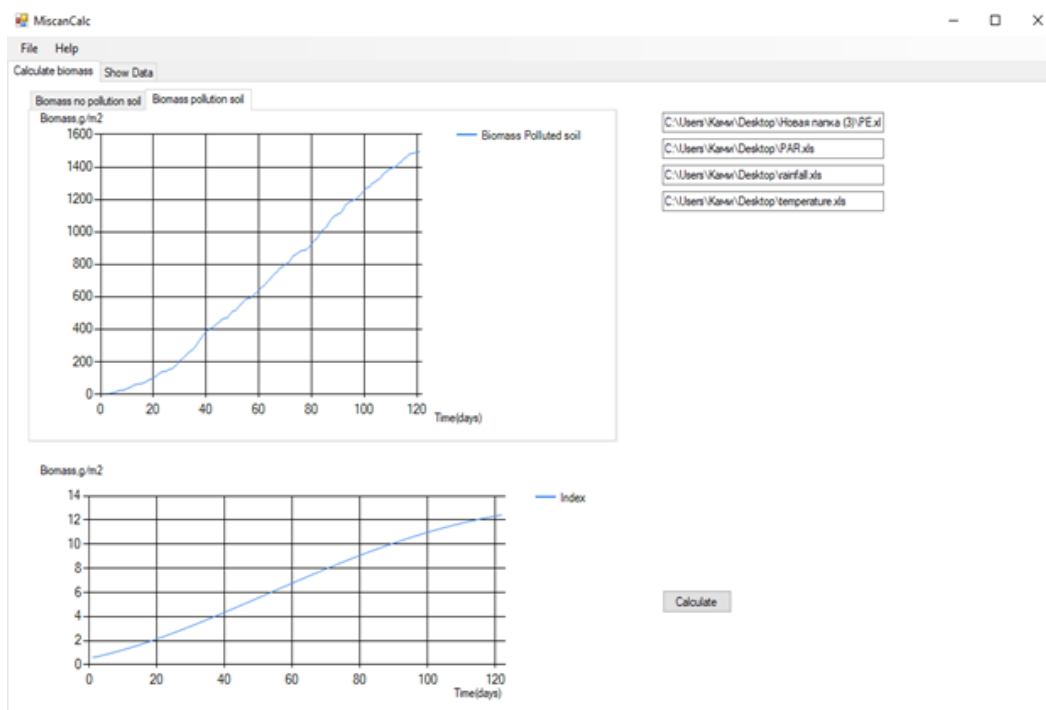


Рисунок 2.1.5 – Десктоп приложение модели *MiscanCalc*

Расчеты показали, что максимальный урожай на незагрязненной почве составлял 1743,8 г СВ м<sup>-2</sup>, а на загрязненной ТЭ почве – 1484,1 г СВ м<sup>-2</sup>, т.е. на 15% ниже, чем на незагрязненной почве.

Сравнение экспериментальных данных с данными модели *Miscanalc*. Сравнительные экспериментальные результаты и модельные данные представлены на рисунках 2.1.6 и 2.1.7.

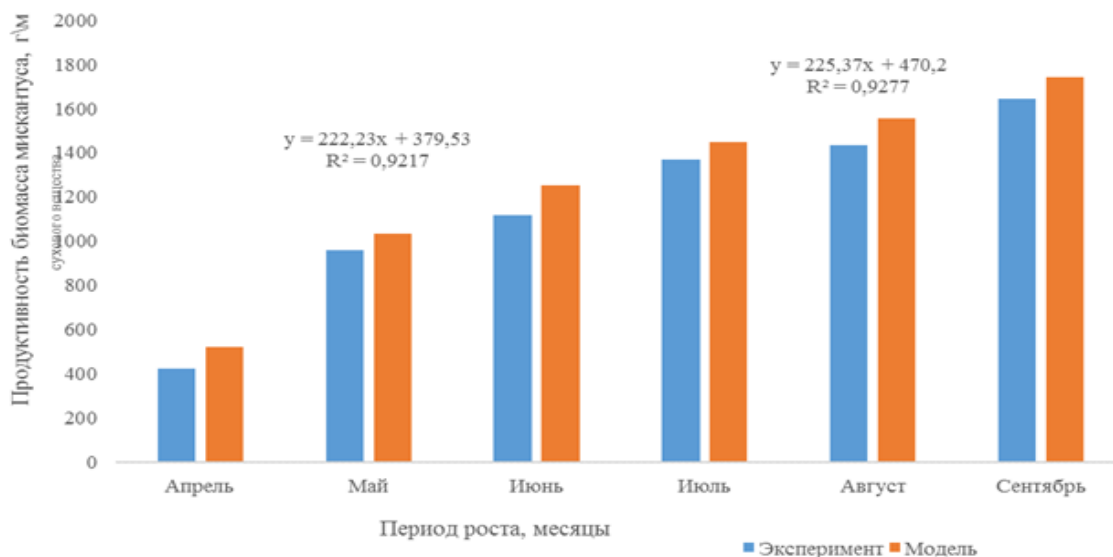


Рисунок 2.1.6 – Сравнительные данные (экспериментальные и модельные) продуктивности мискантуса, произрастающего на незагрязненной почве

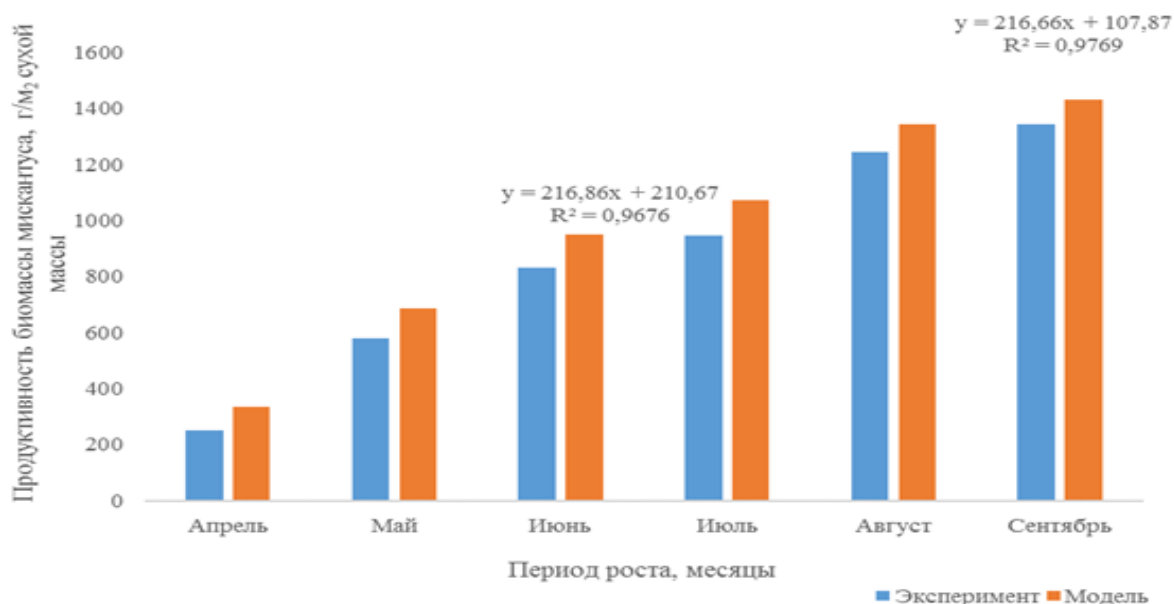


Рисунок 2.1.7 – Сравнительные данные (экспериментальные и модельные) продуктивности мискантуса, произрастающего на загрязненной почве

При проведении оценки адекватности и адаптации разработанной модели к экспериментальным данным установлено, что модель практически совпадает

с экспериментальными данными. Высокие значения коэффициентов детерминации подтверждают статистическую достоверность уравнений в целом, что доказало возможность использования разработанной модели MiscanCalc для прогнозирования продуктивности мискантуса при возделывании на загрязненной ТЭ почве.

Резюмируя данные раздела, следует отметить, что созданная модель MiscanCalc, как системное программное обеспечение MiscanMod, в качестве входных данных также использует: количество осадков, температуру воздуха, потенциальное испарение влаги почвы и фотосинтетическую активную радиацию. Однако в отличие от модели программного обеспечения системы MiscanMod, созданная модель MiscanCalc оценивает динамику формирования биомассы мискантуса на двух типах почвы: на загрязненной ТЭ почве и незагрязненной почве. За счет усовершенствования блоков модели MiscanMod, была создана модель MiscanCalc: блока входных данных путем расчета ФАР на основании информации региональных коэффициентов и данных интегральной радиации Казгидрометеорологической службы по солнечной радиации; блока входных данных путем расчета потенциального испарения влаги почвы на основании температуры, индекса тепла и средней продолжительности дня; ввода в блок метода путем дополнительного расчета разницы между эмпирическими месячными данными биомассы растений, выращенных на загрязненной и незагрязненной почвах (индекс загрязнения) для оценки параметров биомассы мискантуса при выращивании на загрязненной ТЭ почве. Показано, что разработанная модель в течение вегетационного периода проводит расчет индекса листовой поверхности в зависимости от дефицита влаги в почве; оценивает эффективность использования излучения и температуры; количества максимальных дней в период активного накопления биомассы с учетом температуры и фотопериода в период цветения и отмирания; дату завершения стадии цветения; количества максимальных дней в период активного накопления биомассы с учетом температуры и фотопериода на стадии развития; определяет сроки начала уборки культуры; прогнозирует урожайности в  $\text{г м}^{-2}$  сухой биомассы, как на ТЭ-загрязненной, так и незагрязненной почвах, используя ежедневные климатические данные.

Созданная модель может быть предложена для определения основных параметров продуктивности биомассы мискантуса (период биологически активного роста, период сбора урожая, объем биомассы) в течение вегетационного периода при выращивании на техногенно-загрязненной почве. Кроме того, модель может использоваться для прогнозирования урожайности биомассы мискантуса при выращивании в других регионах с различными климатическими условиями и агрохимическими характеристиками почвы, в том числе загрязненными ТЭ, и может быть успешно применена при формировании стратегии развития биоэнергетической отрасли.

## **2.2 Создание математической модели методом самоорганизации на продуктивность растений на загрязненной токсичными элементами почве в зависимости от климатических условий среды**

Для объяснения причинно-следственной связи между продуктивными показателями и окружающей средой использовали задачи синтеза модели временного ряда. Модели временных рядов прогнозируются МГУА [199,200]. МГУА отображают схему селективного отбора, синтезируют и отбирают члены обобщенного полинома Колмогорова-Габора. В анализе временных рядов выделяют две основные задачи: задача идентификации и задачи прогноза [201,202].

*Программная реализация. Принцип программной реализации* согласно МГУА составляется уравнения решающих функций из парных комбинации исходных признаков, затем находят значения их параметров. Из набора решающих функций отбирается частная функция и данная функция, как промежуточные переменные служит исходным аргументом для аналогичного синтеза новых решающих функций. Селективный синтез продолжается до достижения экстремума оценки качества решающей функции. Используя методы самоорганизации, можно решать различные сложные задачи, например, распознавание образов, прогнозирование случайных процессов, идентификацию многоэкстремальных статических и динамических характеристик и управления сложными объектами. Например, методом самоорганизации можно получить модель, выражающую закон изменения отклика  $Y$  в зависимости от конкретных значений независимых переменных  $X$  [203].

*Задача моделирования:* выявление формы зависимости и выбор факторов, оказывающих существенное влияние на зависимую переменную. Структура модели в отличие от регрессионного анализа заранее не фиксирует, а выбирает из множества вариантов по критерию абсолютной ошибки. Используемая математическая модель «Многорядный эвристический метод самоорганизации построения регрессионных уравнений» предназначена для построения регрессионных уравнений на основе обработки экспериментальных данных. Построение математических моделей на основе данных реализуется автоматически. В процессе моделирования ставится задача выявления формы зависимости и выбора факторов, оказывающих существенное влияние на зависимую переменную. При этом структура модели в отличие от регрессионного анализа, заранее не фиксируется, а выбирается из множества вариантов по критерию абсолютной ошибки. Суть предлагаемого метода многорядного эвристического метода самоорганизации заключается в том, что он воспроизводит селекционную схему.

*Методы исследования. Регрессионное уравнение:*

$$y = f(x_1, x_2, \dots, x_n) \quad (2.38)$$

Регрессионное уравнение заменяется несколькими рядами частных описаний:

Первый ряд

$$\begin{aligned}
z_1 &= a_{11} * x_1 + b_{11} * x_2 + c_{11} * x_1 * x_2, z_2 \\
&= a_{12} * x_2 + b_{12} * x_3 + c_{12} * x_2 * x_3, \dots, z_k \\
&= a_{1k} * x_{n-1} + b_{1k} * x_n + c_{1k} * x_{n-1} * x_n
\end{aligned} \tag{2.39}$$

где  $k = n*(n + 1)/2$

Второй ряд

$$\begin{aligned}
\phi_1 &= a_{21} * z_1 + b_{21} * z_2 + c_{21} * z_1 * z_2, \phi_2 \\
&= a_{22} * z_2 + b_{22} * z_3 + c_{22} * z_2 * z_3, \dots, \phi_k \\
&= a_{2k} * z_{n-1} + b_{2k} * z_n + c_{2k} * z_{n-1} * z_n
\end{aligned} \tag{2.40}$$

Для линейной модели коэффициенты  $c_{ij}$  принимаются равными нулю. Каждое частное описание является функцией только двух переменных. Поэтому коэффициенты такого регрессионного уравнения могут быть легко определены даже по небольшому числу наблюдений с помощью МНК.

При использовании нелинейных частных описаний можно получить модели практически любой сложности, так как на каждом ряду селекции степень полинома удваивается [204, 205]. Степень полного описания объекта повышается с каждым рядом селекции и в результате можно определить числовые значения коэффициентов сколь угодно сложного полного описания по малому числу натуральных (экспериментальных данных).

В процессе селекции отбираются переменные в соответствии с критериями минимизации функционала абсолютной ошибки:

$$F = \sum_{t=1}^M (y_t - \hat{y}_t)^2 \tag{2.41}$$

где  $y_t$  – значение показателя в момент времени  $t$ ;  $\hat{y}_t$  – прогнозируемое значение показателя в момент времени  $t$ .

В соответствии с МГУА после каждого ряда селекции выбираются  $K$  уравнений регрессии вида:

первый ряд  $z = f(x_i, x_j)$ ,

второй ряд  $\phi = f(z_i, z_j)$ ,

третий ряд  $v = f(\phi_i, \phi_j)$ ,

четвертый ряд и т.д.

После выполнения первого ряда селекции выбираются первые  $N$  уравнений, имеющих наименьшую ошибку. Полученные при этом уравнения регрессии обозначают:  $z_k = f(x_i, x_j)$ ,  $k = \overline{1, N}$ . Вторые и последующие ряды селекции рассчитываются аналогично первому. На каждом шаге селекции коэффициенты регрессионного уравнения определяют на основе МНК. Функция  $Y = Y(U, V)$  заданная таблицей рассчитывает зависимость между ними по формуле:

$$Y(U, V) = a * U + b * V + c * U * V \tag{2.42}$$

где  $a, b, c$  – неизвестные параметры.

Далее подбираем значения этих параметров так, чтобы была наименьшей сумма квадратов уклонений опытных данных  $Y_i$  и теоретических:

$$\begin{aligned} Y_i^{\wedge} &= a * U_i + b * V_i + c * U_i * V_i \\ \sigma &= Cntgt \end{aligned} \quad (2.43)$$

где величина  $\sigma$  является функцией трех переменных  $a, b, c$ .

Необходимым и достаточным условием существования минимума этой функции является равенство нулю частных производных функции  $\sigma$  по всем переменным:

$$\frac{\partial \sigma}{\partial a} = 0, \frac{\partial \sigma}{\partial b} = 0, \frac{\partial \sigma}{\partial c} = 0 \quad (2.44)$$

так как

$$\begin{aligned} \frac{\partial \sigma}{\partial a} &= -2 \sum_{i=1}^n (Y_i - aU_i - bV_i - cU_iV_i)U_i \\ \frac{\partial \sigma}{\partial b} &= -2 \sum_{i=1}^n (Y_i - aU_i - bV_i - cU_iV_i)V_i \\ \frac{\partial \sigma}{\partial c} &= -2 \sum_{i=1}^n (Y_i - aU_i - bV_i - cU_iV_i)U_iV_i \end{aligned} \quad (2.45)$$

то система для нахождения  $a, b, c$  будет иметь следующий вид:

$$\begin{aligned} a \sum_{i=1}^n U_i^2 + b \sum_{i=1}^n U_iV_i + c \sum_{i=1}^n U_i^2V_i &= \sum_{i=1}^n Y_iU_i \\ a \sum_{i=1}^n U_iV_i + b \sum_{i=1}^n V_i^2 + c \sum_{i=1}^n V_i^2U_i &= \sum_{i=1}^n Y_iV_i \\ a \sum_{i=1}^n U_i^2V_i + b \sum_{i=1}^n U_iV_i^2 + c \sum_{i=1}^n U_i^2V_i^2 &= \sum_{i=1}^n Y_iU_iV_i \end{aligned} \quad (2.46)$$

Принцип метода Крамера заключается в решении систем линейных алгебраических уравнений с числом уравнений равным числу неизвестных с ненулевым главным определителем матрицы коэффициентов системы [206,207]. Используя метод Крамера, в соответствии с методом самоорганизации, после каждого ряда селекции выбираются  $K$  уравнений регрессии вида:

первый ряд  $z = f(x_i, x_j)$ ,  
второй ряд  $\varphi = f(z_i, z_j)$ ,  
третий ряд  $v = f(\phi_i, \phi_j)$ ,  
четвертый ряд  $\omega = f(v_i, v_j)$ , и т.д.

После выполнения первого ряда селекции выбираются первые  $N$  уравнений, имеющих наименьшую ошибку. Полученные при этом уравнения регрессии обозначают, как:  $z_k = f(x_i, x_j)$ ,  $k = \overline{1, N}$ . Дальнейшие ряды селекции строятся аналогично первому.

Остановка моделей на последующих рядах происходит в том случае, когда с увеличением номера слоя, т.е. с усложнением моделей, не происходит уменьшения внешнего критерия лучшей модели. Усложнение модели прекращается в том случае, если дальнейшее улучшение качества селекции не будет превышать параметра  $\varepsilon$ .

*Программное приложение.*

*Язык программирования:* Delphi используется для написания прикладного программного обеспечения, является структурированным, высокоуровневым языком программирования со строгой статистической типизацией переменных [208-210].

*Среда программирования:* Delphi.

*Тип реализующей ЭВМ. Рекомендуемые требования:* Pentium/AMD, тактовая частота – не ниже 2 ГГц, ОЗУ – не менее 1 ГБ, объем свободного места на жестком диске – не меньше 1 ГБ, операционные системы - Windows XP/7/8. *Минимальные требования:* Pentium/AMD, тактовая частота – не ниже 600 МГц, ОЗУ – не менее 256 МБ, объем свободного места на жестком диске – не меньше 500 МБ, операционные системы - Windows XP/7/8.

*Применение программного продукта.* Для построения математической модели влияния климатических показателей среды на биомассу мискантуса, произрастающего на загрязненной ТЭ почве, использовали программу самоорганизации. С помощью данной программы оценили зависимость биомассы растения-очистителя, обладающего способностью извлекать ТЭ из загрязненной почвы погодных параметров среды (температуры, влажности почвы, кислотности среды, фотосинтетической активности радиации) с учетом величины достоверности аппроксимации [210]. Актуальность данного исследования связана с тем, что энергетическое растение мискантус широко используется в биоэнергетике (получают твердое биотопливо из соломы) с одной стороны, а с другой, данный вид обладает способностью очищать почву, загрязненную ксенобиотиками различного происхождения, в частности ТЭ и нефтепродуктами. Поглощение ТЭ коррелирует с биомассой растения. Кроме этого, ТЭ ограничивают рост растений, тем самым понижая эффективность очистки [193].

Использовали многорядный метод самоорганизации для анализа связи между динамикой биомассы растения в период роста и развития (Biomass) и следующими климатическими параметрами: 1 – испарение влаги почвы (PE),

2 – фотосинтетическая активность радиации (PAR), 3 – осадки (Rainfall), 4 – температура (Temperature).

В файле FisxP.txt разместили следующую информацию:

4 100 2

Biomass

PE

PAR

Rainfall

Temperature.

В файле FisxD.txt размещаются построчно экспериментальные данные (Приложение А, Таблица А 2.2.1). В строке первое число указывает на значение результирующего параметра, далее перечисляются значения вспомогательных параметров. Параметрические исходные данные: число аргументов – 4, количество точек – 100.

В таблице 2.2.1 представлены средние данные переменных. В первой строке первое число указывает на количество параметров, второе число – на количество данных, третье число - на тип модели (линейная – 1, нелинейная – 2).

Таблица 2.2.1 – Средние, максимальные и минимальные данные переменных

Переменные	Максимальное	Минимальное	Среднее
Y	1264,565	1,050	544,753
X(1)	4,807	0,000	2,291
X(2)	14,030	1,460	7,782
X(3)	38,100	0,000	5,684
X(4)	111,000	6,500	21,242

*Результаты расчета с помощью программы Самоорганизации.* Результаты расчета с помощью программы 1-ой, 2-ой, 3-ей, 4-ой и 5-ой селекций представлены в таблицах 2.2.2–2.2.6.

*Алгоритмы расчета 1-ой селекции:*

$$\begin{aligned}
 z1(1) &= 0,687 * X(1) + 0,640 * X(4) - 1,078 * X(1) * X(4) \\
 z1(2) &= 0,687 * X(1) + 0,657 * X(3) - 1,020 * X(1) * X(3) \\
 z1(3) &= 0,791 * X(1) + 0,764 * X(2) - 1,305 * X(1) * X(2) \\
 z1(4) &= 0,715 * X(2) + 0,855 * X(3) - 1,360 * X(2) * X(3)
 \end{aligned}
 \tag{2.47}$$

Таблица 2.2.2 – Результаты расчета 1-ой селекции



Ряды	Ошибка	Коэффициент 1	Коэффициент 2	Коэффициент 3
3(1-4)	16,442745	0,687433	0,639918	-1,078038
2(1-3)	16,641309	0,686506	0,656516	-1,019773
1(1-2)	16,738916	0,790674	0,763545	-1,305119
4(2-3)	17,036969	0,715034	0,855330	-1,359676
5(2-4)	17,412309	0,649833	0,670954	-1,035383
6(3-4)	17,424851	0,742313	0,770697	-1,275642

*Алгоритмы расчета 2 ряда селекции:*

$$\begin{aligned}
 z2(1) &= 1,619 * Z1(1) + 1,361 * Z1(3) - 9,062 * Z1(1) * Z1(3) \\
 z2(2) &= 1,020 * Z1(1) + 1,879 * Z1(4) - 8,893 * Z1(1) * Z1(4) \\
 z2(3) &= 1,464 * Z1(3) + 0,954 * Z1(4) - 3,065 * Z1(3) * Z1(4) \\
 z2(4) &= 1,095 * Z1(2) + 1,006 * Z1(4) - 2,759 * Z1(2) * Z1(4)
 \end{aligned}
 \tag{2.48}$$

Таблица 2.2.3 – Результаты расчета 2-ой селекции

Ряды	Ошибка	Коэффициент 1	Коэффициент 2	Коэффициент 3
3(1-3)	14,452810	1,618656	1,360954	-9,061547
4(1-4)	14,871161	1,020455	1,879339	-8,893387
7(3-4)	15,046972	1,463690	0,953528	-3,064651
6(2-4)	15,327753	1,094837	1,005669	-2,758885
5(2-3)	16,314729	1,077392	1,105448	-2,506673
1(1-2)	16,442745	1,000000	0,000000	0,000000
2(1-2)	16,571151	1,459143	1,204295	-4,596856

*Алгоритмы расчета 3-ей селекции:*

$$\begin{aligned}
 z3(1) &= 0,909 * Z2(3) + 0,999 * Z2(4) - 1,490 * Z2(3) * Z2(4) \\
 z3(2) &= 0,784 * Z2(2) + 0,993 * Z2(4) - 1,172 * Z2(2) * Z2(4) \\
 z3(3) &= 0,736 * Z2(1) + 1,110 * Z2(4) - 1,440 * Z2(1) * Z2(4) \\
 z3(4) &= 1,027 * Z2(2) + 1,107 * Z2(3) - 2,175 * Z2(2) * Z2(3)
 \end{aligned}
 \tag{2.49}$$

Таблица 2.2.4 – Результаты расчета 3-ей селекции

Ряды	Ошибка	Коэффициент 1	Коэффициент 2	Коэффициент 3
7(3-4)	11,940320	0,909004	0,998533	-1,489588
6(2-4)	12,143334	0,783708	0,993105	-1,172293
4(1-4)	12,239301	0,735509	1,109891	-1,440375
5(2-3)	12,419342	1,026507	1,106805	-2,174982
2(1-2)	12,480062	1,197777	1,237831	-3,162864
3(1-3)	12,616615	0,925663	1,116772	-2,057170
1(1-2)	14,452810	1,000000	0,000000	0,000000

*Алгоритмы расчета 4-ой селекции:*

$$\begin{aligned}
 z4(1) &= 0,909 * Z3(3) + 0,999 * Z3(4) - 1,490 * Z3(3) * Z3(4) \\
 z4(2) &= 1,000 * Z3(1) + 0,000 * Z3(2) + 0,000 * Z3(1) * Z3(2) \\
 z4(3) &= 0,784 * Z3(2) + 0,993 * Z3(4) - 1,172 * Z3(2) * Z3(4)
 \end{aligned}
 \tag{2.50}$$

$$z_4(4) = 0,736 * Z_3(1) + 1,110 * Z_3(4) - 1,440 * Z_3(1) * Z_3(4)$$

Таблица 2.2.5 – Результаты расчета 4-ой селекции

Ряды	Ошибка	Коэффициент 1	Коэффициент 2	Коэффициент 3
1(1-2)	11,940320	1,000000	0,000000	0,000000
4(1-4)	12,802728	1,009403	1,719933	-4,853063
3(1-3)	12,880619	1,218947	1,385676	-4,595693
6(2-4)	13,302587	1,158373	1,204906	-3,611096
7(3-4)	14,603084	0,890151	0,937029	-1,608685
5(2-3)	15,269376	0,872656	1,045008	-2,020681
2(1-2)	15,358824	1,041783	1,226586	-3,371989

*Алгоритмы расчета 5-ой селекции:*

$$\begin{aligned}
 z_5(1) &= 0,909 * Z_4(3) + 0,999 * Z_4(4) - 1,490 * Z_4(3) * Z_4(4) \\
 z_5(2) &= 1,000 * Z_4(1) + 0,000 * Z_4(2) + 0,000 * Z_4(1) * Z_4(2) \\
 z_5(3) &= 1,000 * Z_4(1) + 0,000 * Z_4(2) + 0,000 * Z_4(1) * Z_4(2) \\
 z_5(4) &= 0,784 * Z_4(2) + 0,993 * Z_4(4) - 1,172 * Z_4(2) * Z_4(4)
 \end{aligned}
 \tag{3.51}$$

Таблица 2.2.6 – Результаты расчета 5-ой селекции

Ряды	Ошибка	Коэффициент 1	Коэффициент 2	Коэффициент 3
1(1-2)	11,940320	1,000000	0,000000	0,000000
4(1-4)	14,851743	1,078460	1,163604	-3,101076
2(1-2)	14,876404	1,179823	1,011738	-2,886278
3(1-3)	14,884729	1,115208	1,117868	-3,023926
6(2-4)	15,421893	1,077140	1,047707	-2,720931
5(2-3)	15,452147	1,139382	0,944632	-2,586491
7(3-4)	16,021664	1,080676	0,910812	-2,341144

В таблице 2.2.7 представлены результаты расчета алгоритмов селекции.

Таблица 2.7 – Расчеты алгоритмов селекции

Ряды	Ошибка	Коэффициент 1	Коэффициент 2	Коэффициент 3
7(3-4) = 3	11,940320	0,909004	0,998533	-1,489588
1(1-2) = 4	11,940320	1,000000	0,000000	0,000000
1(1-2) = 5	11,940320	1,000000	0,000000	0,000000
6(2-4) = 3	12,143334	0,783708	0,993105	-1,172293

*Расчеты алгоритма селекции:*

$$\begin{aligned}
 Y(1) &= 0,909 * Z_2(3) + 0,999 * Z_2(4) - 1,490 * Z_2(3) * Z_2(4) \\
 Y(2) &= 1,000 * Z_3(1) + 0,000 * Z_3(2) + 0,000 * Z_3(1) * Z_3(2)
 \end{aligned}
 \tag{2.52}$$

$$\begin{aligned}
Y(3) &= 1,000 * Z4(1) + 0,000 * Z4(2) + 0,000 * Z4(1) * Z4(2) \\
Y(4) &= 0,784 * Z2(2) + 0,993 * Z2(4) - 1,172 * Z2(2) * Z2(4)
\end{aligned}$$

Конец селекции.

На каждом шаге итерации отбирали по 4 оптимальные модели-претендента. Наилучшая модель для прогноза роста биомассы была получена на 3-м ряду селекции и основывается на 3 исходных параметрах из 4. Исключен параметр X4 – температура. Как видно оптимальный результат получен на третьем ряду селекции:

$$Y(1) = 0,909 * Z2(3) + 0,999 * Z2(4) - 1,490 * Z2(3) * Z2(4) \quad (2.53)$$

Анализ показал, что уравнение, описывающее зависимость роста биомассы от испарения влаги почвы, фотосинтетической активной радиации, осадков и температуры имеет нелинейный вид. При этом температура воздуха не оказывает существенного влияния на исследуемый процесс. Далее переписываем уравнение (2.54) в переменных Y, X1, X2, X3, X4:

$$\begin{aligned}
y &= 1.804x_1 + 2.355x_2 + 2.319x_3 - 7.927x_1x_2 - 8.257x_1x_3 - \\
&10.77x_2x_3 - 1.298x_1^2 - 3.452x_2^2 - 3.469x_3^2 + 38.77x_1x_2x_3 + \\
&12.017x_1x_2^2 + 13.283x_1x_3^2 + 17.624x_2x_3^2 + 6.423x_1^2x_2 + \\
&7.043x_1^2x_3 + 17.51x_2^2x_3 - 1.056x_1^2x_2^2 - 12.601x_1^2x_3^2 - \\
&31.067x_2^2x_3^2 - 51.256x_1x_2^2x_3 - 66.02x_1x_2x_3^2 - 34.498x_1^2x_2x_3 + \\
&37.15x_1^2x_2x_3^2 + 111.269x_1x_2^2x_3^2 + 56.245x_1^2x_2^2x_3 - \\
&99.157x_1^2x_2^2x_3^2 + 1.794x_2^3 + 1.886x_3^3 - 7.714x_1x_3^3 - \\
&10.624x_2x_3^3 - 10.052x_2^3x_3 - 6.443x_1x_2^3 + 39.674x_1x_2x_3^3 - \\
&11.814x_1^2x_2x_3^3 + 7.428x_1^2x_3^3 + 19.486x_2^2x_3^3 - 67.499x_1x_2^2x_3^3 + \\
&18.784x_2^3x_3^3 - 63.404x_1x_2^3x_3^2 + 23.597x_1x_2^3x_3 + 38.159x_1x_2^3x_3^3 - \\
&11.707x_2^3x_3^3 + 5.772x_1^2x_2^3 + 53.491x_1^2x_2^3x_3^2 - 30.527x_1^2x_2^3x_3 + \\
&57.817x_1^2x_2^3x_3^3 - 31.02x_1^2x_2^3x_3^3
\end{aligned} \quad (2.54)$$

Как видно из формулы связь между динамикой объема биомассы Мискантуса (Biomass) и климатическими параметрами: 1 – испарением влаги почвы; 2 – фотосинтетической активной радиацией; 3 – осадками, описывается полиномом 8-й степени. Показателями, оказывающие наибольшее влияние на биомассу при произрастании на загрязненной ТЭ почве, являются испарение влаги почвы, фотосинтетическая активная радиация и осадки.

Таким образом, разработана программа для построения регрессионных уравнений на основе обработки экспериментальных данных. Данная программа применена для вывода уравнений математической модели влияния различных факторов среды на биомассу растений, произрастающих на загрязненной ТЭ почве, с учетом величины достоверности аппроксимации. Проведенный анализ показал, что показателями влияющие на урожайность растений, произраставших на загрязненной ТЭ почве, являются испарение влаги почвы, фотосинтетическая активная радиация и осадки.

Программа «Многорядный эвристический метод самоорганизации построения регрессионных уравнений» может быть использована при создании математической модели биомассы растений, произрастающих на загрязненной ТЭ почве, в зависимости от условия среды в различных регионах страны.

Получено свидетельство о государственной регистрации прав на объект авторского права (Приложение Б).

### **2.3 Прогнозирование продуктивности растения на загрязненной ТЭ почве в зависимости от климатических условий среды с помощью технологий машинного обучения**

Использование машинного обучения (МО) для отслеживания урожайности сельскохозяйственных культур в режиме реального времени в зависимости от окружающих климатических условий для фермерских хозяйств является приоритетным направлением, так как продуктивность сельскохозяйственных растений зависит от климатических условий, тем самым затрудняя проектирование точных бизнес-процессов. Системы прогнозирования на основе интеллекта использовались в США, Великобритании, Финляндии и других странах. Нидерланды сообщили о значительном снижении ошибки прогнозирования [212-214]. С модернизацией сельского хозяйства метаданные о климатических условиях получают с помощью сетей недорогих автоматизированных датчиков, дистанционного зондирования и систем GPS [22]. Подходы на основе дистанционного зондирования и систем GPS требуют обработки огромных объемов данных с разных платформ, поэтому в настоящее время больше внимания уделяется методам МО. Это связано со способностью систем на основе МО обрабатывать большое количество входных данных и решать нелинейные задачи.

Одним из основных преимуществ методов МО является то, что они способны автономно решать большие нелинейные задачи с использованием наборов данных из нескольких взаимосвязанных источников [215]. В последние годы были реализованы различные методы МО для точного прогнозирования урожайности различных культур. Наиболее успешными методами МО были искусственные нейронные сети, регрессия опорных векторов, деревья регрессии [216].

Регрессионный анализ является основным методом анализа больших данных, используют при построении параметрической функции, которая строится на основе известных данных [64, 217].

Цель – прогнозирование продуктивности сельскохозяйственных культур при произрастании на загрязненной ТЭ почве в зависимости от метаданных климатических условий среды с помощью программного обеспечения Jupyterlab Anaconda Environment.

Для решения поставленной цели были выполнены следующие задачи:

- Оценка и выбор моделей регрессии для прогноза продуктивности растений;
- Отбор информативных признаков с помощью регрессионных моделей, влияющих на продуктивность растений (Feature selection);

•Создание модели глубокого обучения для прогнозирования продуктивности растения, используя Tensorflow keras.

*Ход работы Jupyterlab Anaconda Environment модуля Python.*

*# Библиотека.*

```
import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
%matplotlib inline
import seaborn as sns
import time
```

*# Разделение данных (Data split).*

```
from sklearn.model_selection import ShuffleSplit
```

*# Предварительная обработка (Preprocessing).*

```
from sklearn.impute import SimpleImputer, KNNImputer
```

```
from sklearn.compose import make_column_transformer, ColumnTransformer
```

```
from sklearn.preprocessing import OneHotEncoder, PolynomialFeatures, StandardScaler, PolynomialFeatures
```

*# Источники информации (Pipeline)ю*

```
from sklearn.pipeline import make_pipeline, Pipeline
```

```
from sklearn import set_config
```

*# Показатели (Metrics).*

```
from sklearn.metrics import mean_squared_error, mean_absolute_error, r2_score
```

*# Разработка функций (Feature Engineering)*

```
from sklearn.feature_selection import f_regression, mutual_info_regression, SelectKBest, RFE, RFECV
```

```
from sklearn.inspection import permutation_importance
```

```
import shap
```

*# Линейные модели (Linear Models)*

```
from sklearn.linear_model import LinearRegression, Lasso, Ridge, RidgeCV, ElasticNet, SGDRegressor
```

*# Модели на основе дерева (Tree-based models)*

```
from sklearn.tree import DecisionTreeRegressor
```

```
from sklearn.ensemble import GradientBoostingRegressor, RandomForestRegressor
```

```
import xgboost as xgb
```

*# Модель стейкинга (Staking model)*

```
from sklearn.ensemble import StackingRegressor
```

*# Модели нейронной сети (NN)*

```
import os
```

```
os.environ["KERAS_BACKEND"] = "plaidml.keras.backend"
```

```
import tensorflow as tf
```

```
from tensorflow import keras
```

```
set_config(display="diagram")
```

```
import warnings
warnings.filterwarnings ("ignore", category=FutureWarning)
from sklearn.exceptions import DataConversionWarning
warnings.filterwarnings(action='ignore', category=DataConversionWarning)
Чтение данных
df = pd.read_csv ("AlmatyWeatherDataSet.csv") # загрузка файла с данными
в pandas DataFrame
```

#### *Программная реализация*

Используя программное обеспечение Jupyterlab Anaconda Environment модуля Python при построении моделей МО, использовали Scikit-learn для предварительной обработки данных, уменьшения размерности, выбора модели, регрессии, классификации и кластерного анализа, а также предоставляет несколько наборов данных, которые можно использовать для тестирования моделей [218].

*Описание изученных моделей.* Для анализа метаданных использовали различные линейные и ансамблевые модели [63,64,200].

Регрессионный анализ (Regression analysis) – одна из важнейших областей статистики и машинного обучения. Регрессия ищет отношения между переменными. Для решения регрессионных задач необходимо найти  $f(x) = y$ , где  $x$  – векторы признаков,  $y$  – цель. Для каждого нового объекта, чтобы спрогнозировать значение целевой переменной, необходимо найти метод [219]. Обучение модели регрессии производится с помощью функции *fit*, а предсказание с помощью обученной модели – с помощью функции *predict*. Эти функции существуют для любых моделей регрессии и классификации, поддерживаемых библиотекой SKLearn [220].

#### *Линейные модели.*

Линейная регрессия (Linear Regression) – линейная модель, используемая в статистике, моделируют линейную связь между входными данными и непрерывной числовой выходной переменной.

Лассо (Lasso) – линейная модель, выполняющая выбор переменных более эффективно, чем регрессия байесовского гребня.

Эластичная сеть (Elastic Net) – линейная модель регрессии с двумя регуляризаторами, которая представляет собой комбинацию двух наиболее популярных регуляризованных вариантов линейной регрессии: ребра (Ridge) и лассо. Ridge использует штраф L2, а Lasso использует штраф L1.

Ридж-регрессия или гребневая регрессия (Ridge) регулирует признаки с низкими прогнозирующими результатами, уменьшая их коэффициенты к нулю. Может быть использована для классификации или регрессии. Применяется для борьбы с избыточностью данных, когда независимые переменные коррелируют друг с другом, вследствие чего проявляется неустойчивость оценок коэффициентов многомерной линейной регрессии.

Стохастический градиентный спуск (SGDRegress) – SGD означает стохастический градиент потерь, оценивает и обновляет коэффициенты на каждой итерации, делает прогноз для обучающего экземпляра, вычисляется ошибка, и

модель обновляется, чтобы уменьшить ошибку для следующего прогнозирования. Этот процесс повторяется для фиксированного числа итераций. Делает вычисления быстрее, чем градиентный спуск и пакетный градиентный спуск.

Пассивно-агрессивный регрессор (Passive Aggressive Regressor) – линейная модель, семейство методов для крупномасштабного обучения. Она похожа на перцептрон в том, что не требует скорости обучения. Однако, в отличие от перцептрона, включают параметр регуляризации.

*Ансамблевые модели* или *древесные модели* обеспечивают визуальный инструмент для обследования данных, для получения представления о том, какие переменные важны и как они связаны друг с другом. В отличие от линейных моделей, деревья имеют способность обнаружить скрытые регулярности, соответствующие сложным взаимодействиям в данных.

Деревья регрессии (Decision Tree Regressor) – модель, в которых целевая переменная может принимать непрерывные значения. Модели древо решений создают правила принятия решений для функций для получения прогнозов. Ее можно использовать для классификации или регрессии.

Случайный лес (Random Forest Regressor) – ансамбльный метод обучения, который объединяет результаты нескольких деревьев решений, основным преимуществом является высокая точность по сравнению с другими моделями.

Градиентный бустинг (Gradient Boosting Regressor) – техника МО для задач классификации и регрессии, которая строит модель предсказания в форме ансамбля. На каждом этапе дерево регрессии аппроксимируется отрицательным градиентом заданной функции потерь. Градиентный бустинг использует бустинг для создания прогностических моделей из ансамбля слабых прогнозируемых обучающихся.

Дерево решений (Extra Trees Regressor) является моделью контролируемого МО, используемой для прогнозирования цели путем обучения правил принятия решений из функций.

Адаптивный бустинг регрессии (Ada Boost Regressor) – метаоценщик, который начинает с подгонки регрессора к исходному набору данных, а затем подбирает дополнительные копии регрессора к тому же набору данных, но где веса экземпляров корректируются в соответствии с ошибкой текущего прогноза. Таким образом, последующие регрессоры больше фокусируются на сложных случаях.

Бэггинг-регрессор (Bagging Regressor) – ансамблевая метаоценка, которая подбирает базовые регрессоры для каждого из случайных подмножеств исходного набора данных, а затем объединяет их индивидуальные прогнозы для формирования окончательного прогноза.

Байесовская регрессия автоматического определения релевантности (ARDR Regression) основана на методе байесовского вывода. API Scikit-learn предоставляет класс ARDRRegression для соответствия модели регрессии с использованием метода ARD. ARDRRegression рассматривает веса модели как распределенные по Гауссу и оценивает параметры лямбда и альфа посредством итерации.

Обобщенные линейные модели ( $H_2O$  Generalized Linear Estimator, GLM) оценивают модели регрессии для результатов после экспоненциального распределения Гаусса, Пуассона, биномиального и гамма-распределения. Каждый из них служит разным целям и, в зависимости от распределения и выбора функции связи, может использоваться либо для предсказания, либо для классификации.

В комплект GLM входят:

- Гауссова регрессия (Gaussian regression)
- Регрессия Пуассона (Poisson regression)
- Биномиальная регрессия (классификация) (Binomial regression)
- Дробная биномиальная регрессия (Fractional binomial regression)
- Квазбиномиальная регрессия (Quasibinomial regression)
- Полиномиальная классификация (Multinomial classification)
- Гамма-регрессия (Gamma regression)
- Порядковая регрессия (Ordinal regression)
- Отрицательная биномиальная регрессия (Negative Binomial regression)
- Распределение Твиди (Tweedie distribution)

*Оценка и выбор моделей регрессии для прогноза продуктивности растений.* Процесс моделирования кодом Python начали с выбора источника данных, загрузкой данных, отображением структуры данных и постановки критерия оценки. На рисунке 2.3.1 проиллюстрирована схема выбора источника данных, процесс исследования данных, ETL (сбор, очистка, объединение данных), EDA (визуализация данных), очистка данных (preprocessing), выбор модели и проверка метрики (оценка результата) по критериям, которые мы заранее выставили.

*Выбор данных.* Базовым уровнем анализа данных является сбора данных. Данные были получены из различных источников: метеорологическими факторами (относительные влажность, облачность, осадки и температура воздуха), продуктивные показатели растений и факторы окружающей среды (период роста растений, уровень загрязнения почвы).

1. Основная информация: метаданные о погоде (температура, влажность, облачность, осадки и т.д.) были получены через сайт `AlmatyWeatherDataSet.csv` за 2015-2021 гг.



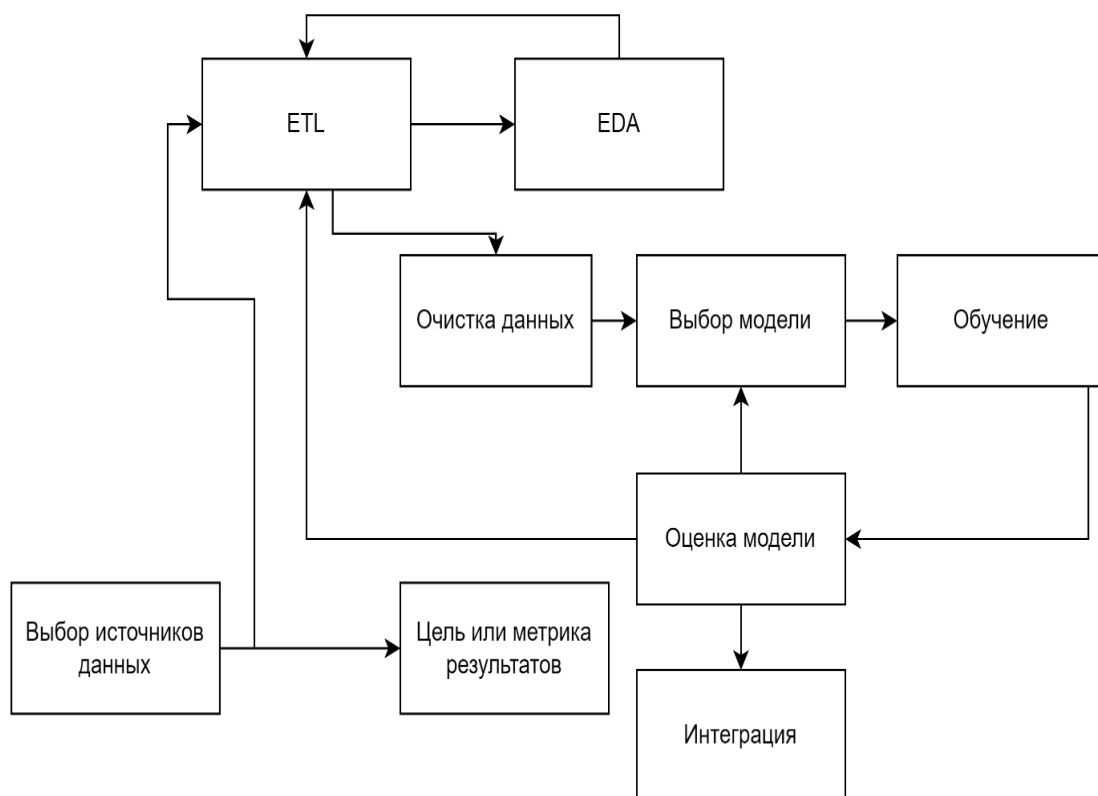


Рисунок 2.3.1 – Схема выбора источника данных и критерии. ETL - сбор, очистка, объединение данных; EDA - визуализация данных

## 2. Дополнительная информация:

yield – продуктивность на техногенно-загрязненной почве (данные 2015-2021 гг., предоставленные Институтом биологии и биотехнологии растений);  
count – дни (с начала посадки растений);

Информацию о наборе данных 37 переменных получили через сайт visualcrossing (Приложение А, Таблица А 2.3.1):

27 – численная информация (int, float);

7 – информация о типе объекта (object);

3 – информация типа дата (DateTime).

В результате предварительной обработки (PreProcessing) данных находили и заполняли пустые значения, масштабировали числовые признаки и объединяли категориальные признаки. Для нахождения и заполнения пустых значений использовали Simpleimpute r(strategy = 'most\_frequent'), затем заменяли отсутствующие признаки, используя наиболее часто встречающиеся значения в каждом столбце. Использовали OneHotEncoding при кодировании категориальных признаков, где каждый класс был представлен в виде отдельного столбца, а значение 0/1 указывало на их наличие. Гистограмма о сводках числовых атрибутов, выполненная с помощью команды df.describe(), представлена на рисунке 2.3.2.

	count	mean	std	min	25%	50%	75%	max
tempmax	847.0	28.851948	5.067322	11.600000	25.500000	29.100000	33.000000	40.000000
tempmin	847.0	14.967296	4.108986	1.000000	12.500000	15.000000	18.000000	26.700000
temp	847.0	22.158323	4.367481	6.500000	19.150000	22.600000	25.500000	31.600000
feelslikemax	847.0	27.780756	4.296138	11.600000	25.500000	28.100000	30.800000	37.400000
feelslikemin	847.0	14.825030	4.408077	-1.300000	12.500000	15.000000	18.000000	26.500000
feelslike	847.0	21.807084	4.138732	4.900000	19.100000	22.400000	25.000000	30.300000
dew	847.0	9.109917	3.017431	-0.900000	7.000000	9.300000	11.300000	16.300000
humidity	847.0	47.944168	13.096656	19.030000	37.355000	46.910000	56.280000	90.540000
precip	847.0	2.105821	8.152752	0.000000	0.000000	0.000000	0.770000	178.880000
precipprob	0.0	NaN	NaN	NaN	NaN	NaN	NaN	NaN
precipcover	847.0	2.436517	3.195834	0.000000	0.000000	0.000000	4.170000	12.500000
preciptype	0.0	NaN	NaN	NaN	NaN	NaN	NaN	NaN
snow	0.0	NaN	NaN	NaN	NaN	NaN	NaN	NaN
snowdepth	497.0	1.597465	2.204296	1.000000	1.000000	1.000000	1.000000	19.410000
windgust	838.0	27.540811	21.103330	7.200000	14.400000	18.000000	36.000000	172.800000
windspeed	847.0	18.994451	6.817675	6.900000	14.400000	18.000000	21.600000	50.400000
winddir	847.0	160.683117	40.147966	62.400000	129.050000	161.000000	192.700000	253.300000
sealevelpressure	847.0	1008.251122	5.613642	993.700000	1003.900000	1007.600000	1011.800000	1030.600000
cloudcover	847.0	57.354900	23.384586	1.300000	40.250000	59.400000	77.400000	97.600000
visibility	847.0	10.700354	1.056596	6.400000	10.000000	10.600000	11.400000	19.600000
solarradiation	0.0	NaN	NaN	NaN	NaN	NaN	NaN	NaN
solarenergy	0.0	NaN	NaN	NaN	NaN	NaN	NaN	NaN
uvindex	0.0	NaN	NaN	NaN	NaN	NaN	NaN	NaN
severerisk	0.0	NaN	NaN	NaN	NaN	NaN	NaN	NaN
moonphase	847.0	0.504357	0.304297	0.000000	0.265000	0.500000	0.750000	1.000000
count	847.0	61.000000	34.949136	1.000000	31.000000	61.000000	91.000000	121.000000
yield	847.0	721.352464	506.197174	1.035633	226.894796	688.730798	1185.116925	1603.156351

Рисунок 2.3.2 – Гистограмма о сводках числовых атрибутов

С помощью команды `df.describe()`.Т выявили числовые значения. Каждая строка представлял круг, который состоял из 29 атрибутов, информацию о численных переменных изучаемых признаков: количество столбцов, среднее значение, минимальное значение, значение на 25%, значение на 50%, значение на 75%, максимальное значение, где каждая строка образует одно наблюдение со значениями переменных в столбцах, с помощью команды `df.describe()`. В каждом атрибуте 18 столбцов с 847 признаками, 1 столбец с 497 признаками, 1 столбец с 837 признаками. В атрибутах выявили 7 отсутствующих данных: 'precipprob', 'precip type', 'snow', 'solar radiation', 'solar energy', 'uv index', 'severe risk'. Всего в столбцах 16581 ячеек информации. Пустые признаки были удалены из DataFrame, используя следующие функции:

```
plt.figure(figsize= (15,10))
sns.heatmap(df.isna(),
cmap="YlGnBu",
cbar_kws={'label': 'Missing Data'})
plt.show()
```

В МО используют 3 техники избавления от пустых значений: заполнение пустых значений средним значением; заполнение пустых значений значением медианы; удалением пустых значений. Опираясь на структуру в столбцах, выбрали технику заполнения отсутствующих данных средними значениями, используя функцию `SimpleImputer()`, модуля `sklearn.impute`. Далее отобрали `min` `max` масштабирование в диапазоне от 0 до 1. Для этого использовали функцию модуля экспериментальной библиотеки `DataFrameMapper sklearn_pandas`, который разработан для преодоления разрыва между реальными пандами и `scikit-learn`. Функция `DataFrameMapper` позволяет получать `DataFrame` модуля `pandas`, что позволяет лучше наблюдать за изменениями модели. Данная модуль позволяет сгруппировать все этапы предварительной обработки данных и сохранить их в одном объекте [221].

С помощью данной функции разделили информацию о 847 признаков на функции и цели, провели преобразование их в массивы, чтобы предсказать, какие параметры могут влиять на прогноз продуктивности растений (рисунок 2.3.3).

```

Для манипулирования данными использовали функции:
categorical_features = X_train.select_dtypes(include=['object','category']).columns
numerical_features = X_train.select_dtypes(include=['int64', 'float64']).columns
boolean_features = X_train.select_dtypes(include=['bool']).columns
cat = [(c, [SimpleImputer(strategy='most_frequent'), OneHotEncoder(
sparse=False, handle_unknown='ignore')]) for c in categorical_features]
num = [(n, [SimpleImputer(strategy='mean'), MinMaxScaler()]) for n in numerical_features]
boolean = [(b, [OneHotEncoder(sparse=False, handle_unknown='ignore')])
for b in boolean_features]
ct = DataFrameMapper(num + cat + boolean, input_df=True, df_out=True

```

Функция `DataFrameMapper` позволила получить функцию `DataFrame` модуля `Pandas`, что улучшает наблюдение за изменениями модели. Проверка показала, что присутствие пустого значения признака «Snowdepth» (в гистограмме признак отмечен в виде тёмных пятен) свидетельствует об отсутствии данных этого признака. Признак «Snowdepth» означает глубину снега, отсутствие признака естественно, так как растение произрастает с марта месяца. При проверке на дубликаты с помощью функции `df.drop_duplicates(inplace=True)`, дубликаты не обнаружены.

Для разделения набора данных на обучающие и тестовые использовали функцию `ShuffleSplit` модуля `Train_test_split` из библиотеки `Sklearn model_selection` (рисунок 2.3.4). На рисунке представлена основная информация о числовой функции и их распределении. Данная функция выполняет `train_test_split`, кросс-валидацию (`cv`) в функции `cross_val_score`, `RFECV`.

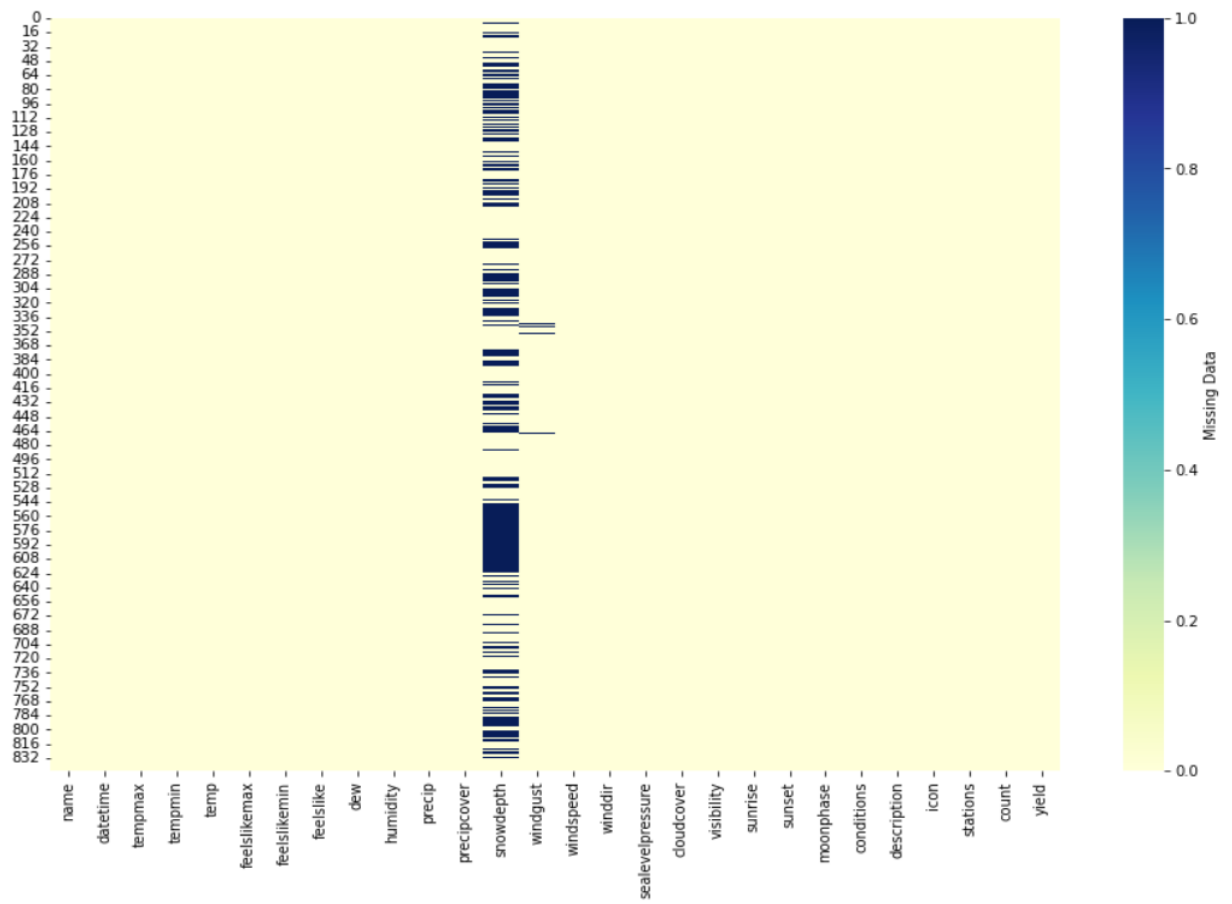


Рисунок 2.3.3 – Проверка пустых значений

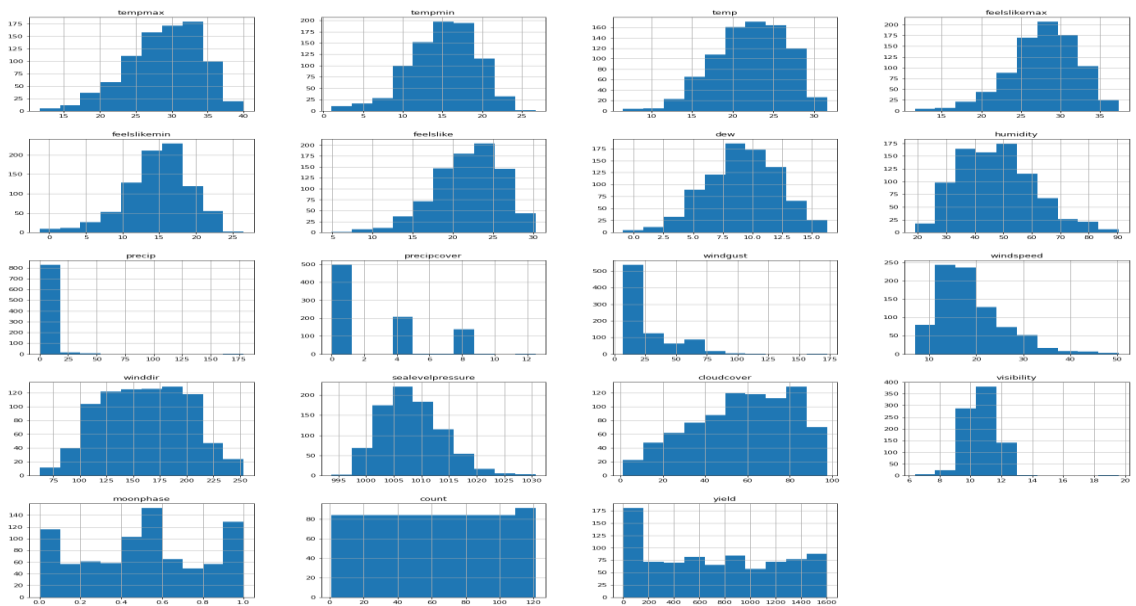


Рисунок 2.3.4 – Гистограмма для числовой функции и их распределение

Модуль `Train_test_split` перемешивает данные и делит их на выборки `train` и `test`: 80% для обучения модели регрессии и 20% для оценки качества. Для разделения данных использовали следующие функции:

```
X = df.copy(deep = True)
X = X.drop('yield', axis=1)
y = df['yield']
from sklearn.model_selection import ShuffleSplit
sss = ShuffleSplit(n_splits=1, train_size=0.3, random_state=0)
train_index, val_index = next(iter(sss.split(X, y)))
X_train, X_test = X.iloc[train_index], X.iloc[val_index]
y_train, y_test = y.iloc[train_index], y.iloc[val_index]
```

с помощью следующих функции:

```
X = df.copy(deep = True)
X = X.drop('yield', axis=1)
y = df['yield']
from sklearn.model_selection import ShuffleSplit
sss = ShuffleSplit(n_splits=1, train_size=0.3, random_state=0)
train_index, val_index = next(iter(sss.split(X, y)))
X_train, X_test = X.iloc[train_index], X.iloc[val_index]
y_train, y_test = y.iloc[train_index], y.iloc[val_index]
```

*Создание корреляционной матрицы.* Корреляционная матрица отображает взаимосвязь между всеми возможными парами признаков, а также для выявления и визуализации закономерностей данных. Корреляционная матрица представляет собой квадратную матрицу размерности  $M \times M$ , где  $M$  – количество числовых признаков. Матрица состоит из 19 строк и 19 столбцов переменных. Каждая ячейка таблицы содержит информацию коэффициента корреляции (рисунок 2.3.5).

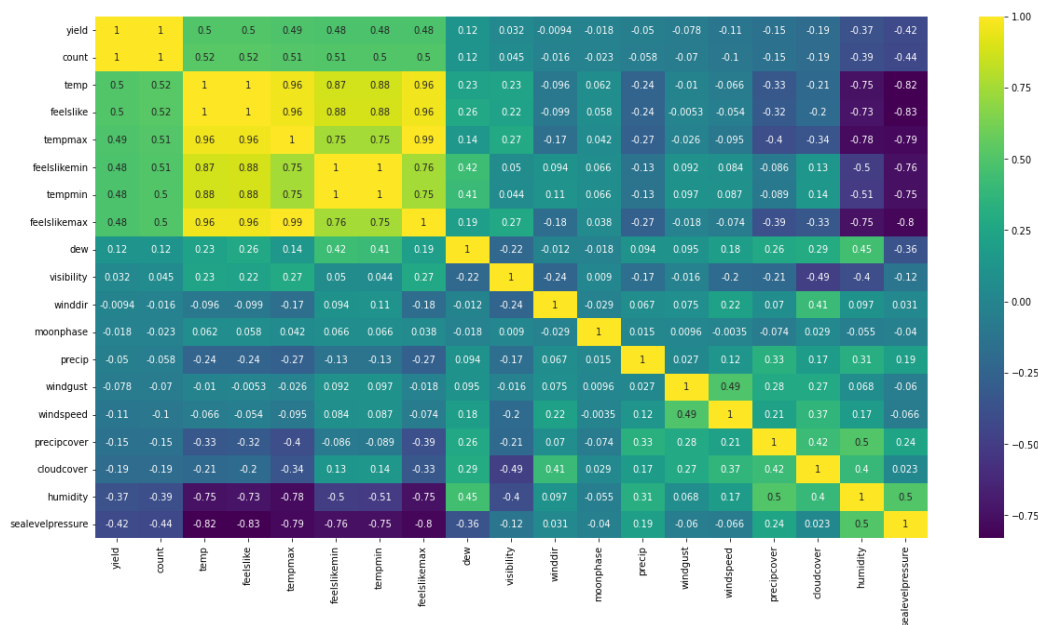


Рисунок 2.3.5 – Корреляционная матрица для переменных

Положительная корреляция представляет собой связь, при которой увеличение одной переменной связано с увеличением другой переменной. Если у двух переменных при возрастании значений одной из них, значения другой убывают, это означает отрицательную корреляцию. В процессе корреляции было отобрано 100 предикторов с положительной корреляцией для дальнейшего исследования ( $k = 100 \#$ ). Интерес представляли признаки со значением корреляции  $< 0.75$ .

Особый акцент был направлен на атрибут «Temp», который отвечает за температуру `df[df.dtypes[(df.dtypes=="float64")|(df.dtypes=="int64")].index.values] = [15,10]`, так как данный признак связан с признаком «продуктивность».

Пять признаков с отрицательной корреляцией 'feels like', 'temp min', 'temp max', 'feel slike max', 'feel slike min' были удалены с помощью функции:

```
cols = df.corr().nlargest(k, 'yield')['yield'].index
cm = df[cols].corr()
plt.figure(figsize=(20,10))
sns.heatmap(cm, annot=True, cmap = 'viridis').
```

*Визуализация средних значений.* Боксплот визуализирует средние значения и выбросы. Выбросы данных были удалены с помощью межквартильного размаха (IQR). Для расчета IQR набор данных делили на квартили, с помощью линейной интерполяции. Эти квартили обозначали, как Q1 (нижний квартиль), Q2 (медиана) и Q3 (верхний квартиль). Нижний квартиль соответствует 25-му перцентилю, а верхний квартиль – 75-му перцентилю,  $IQR = Q3 - Q1$ , так как IQR средним спредом 50% или Н-раскладом определяется, как разница между 75-м и 25-м перцентилеями данных. При визуализации средних значений выявлено 11 строк выбросов (рисунок 2.3.6).

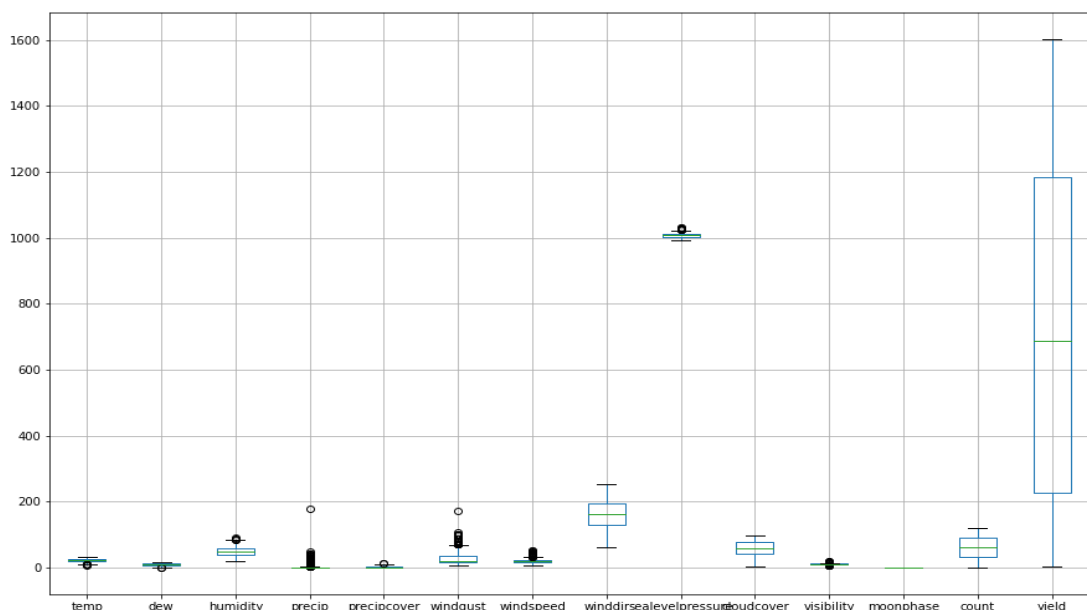


Рисунок 2.3.6 – Визуализация средних значений и выбросов данных

Используемая функция при визуализации средних значений:

```

from collections import Counter
def detect_outliers(df,n,features):
    outlier_indices = []
    # итерация по признакам (столбцам) (# iterate over features(columns))
    for col in features:
        # 1st quartile (25%)
        Q1 = df[col].quantile(0.25)
        # 3rd quartile (75%)
        Q3 = df[col].quantile(0.75)
        # Interquartile range (IQR)
        IQR = Q3 - Q1
        # шаги выброса (outlier step)
        outlier_step = 1.5 * IQR
        # Определить список индекса выброса для столбца функции (Determine a
        list of indices of outliers for feature col) outlier_list_col = df[(df[col] < Q1 - outlier_step) | (df[col] > Q3 + outlier_step)].index.
        # добавить найденные индексы выбросов в столбцы (append the found outlier indices for col to the list of outlier indices)
        outlier_indices.extend(outlier_list_col).
        # выбрать наблюдения, содержащие более 2 выброса (select observations containing more than 2 outliers) outlier_indices = Counter(outlier_indices) multiple_outliers = list (k for k, v in outlier_indices.items () if v > n) return multiple_outliers.
    # Выявление выбросов (detect outliers 'tempmax', 'tempmin', 'temp', 'feelslikemax', 'feelslikemin','feelslike', 'dew', 'humidity', 'precip', 'precipprob', 'precipcover',
    #'preciptype', 'snow', 'snowdepth', 'windgust', 'windspeed', 'winddir',
    #'sealevelpressure', 'cloudcover', 'visibility', 'solarradiation',
    #'solarenergy', 'uvindex', 'severerisk', 'moonphase', 'count', 'yield'
    Outliers_to_drop = detect_outliers(df, 2, numerical_features) df = df.drop(Outliers_to_drop, axis = 0).reset_index(drop=True).

```

После удаления выбросов, рассмотрели признак «datetime» (дата), который находится в Object. Поменяли его формат на datetime 64[ns] и создали дополнительно 4 признака: datetime\_0 (год), datetime\_1 (дни), datetime\_2 (месяц), datetime\_3 (неделя).

*Выбор модели.* Используя различные модели линейной и ансамблевой регрессии, производили расчёты в scikit-learn с помощью функции cross\_val\_score, модуля model\_selection. Задача регрессии состоит в том, чтобы найти  $f(x) = y$ , где  $x$  – векторы признаков, а  $y$  – переменная. Для каждого нового объекта находили методы, чтобы спрогнозировать значение целевой переменной.

Выполнены следующие этапы работ:

1. Достоверность 9 видов моделей без тюнинга и без выбора признаков: 5 моделей линейной регрессии (Linear Regression, Lasso, Elastic Net, Ridge, SGD Regression) и 4 модели ансамблевой регрессии (Random Forest Regressor,

Decision Tree Regressor, Gradient Boosting Regressor, XGB Regressor) оценивали по коэффициенту детерминации.  $R^2$  является еще одним показателем, который используется для оценки модели, и он тесно связан с среднеквадратической ошибкой (Приложение А, Таблица А 2.3.2).

2. Результативность 15 видов моделей без тюнинга и без выбора признаков: 5 моделей линейной регрессии (Linear Regression, Lasso, Elastic Net, Ridge, SGD Regression) и 10 моделей ансамблевой регрессии (Random Forest Regressor, Decision Tree Regressor, Gradient Boosting Regressor, XGB Regressor, Stacking Regressor without Linear Regression, Elastic Net, Stacking Regressor, Voting Regressor without Linear Regression, Elastic Net, Voting Regressor, Keros и H<sub>2</sub>OAuto ML) оценили, используя следующие метрики ошибок: среднеквадратическая ошибка (MSE), среднее абсолютное отклонение (MAE), средняя абсолютная ошибка в процентах (MAPE) и  $R^2$  (Приложение А, Таблица А.2.3.3).

3. Достоверность 9 видов моделей с тюнингом и без выбора признаков: 5 моделей линейной регрессии (Linear Regression, Lasso, Elastic Net, Ridge, SGD Regression) и 4 модели ансамблевой регрессии (Random Forest Regressor, Decision Tree Regressor, Gradient Boosting Regressor, XGB Regressor) оценивали по коэффициенту детерминации (Приложение А, Таблица А 2.3.4).

4. Результативность 15 видов моделей с тюнингом и без выбора признаков: 5 моделей линейной регрессии (Linear Regression, Lasso, Elastic Net, Ridge, SGD Regression) и 10 моделей ансамблевой регрессии (Random Forest Regressor, Decision Tree Regressor, Gradient Boosting Regressor, XGB Regressor, Stacking Regressor without Linear Regression, Elastic Net, Stacking Regressor, Voting Regressor without Linear Regression, Elastic Net, Voting Regressor, Keros и H<sub>2</sub>OAuto ML) оценили, используя следующие метрики ошибок: MSE, MAE, MAPE и  $R^2$  (Приложение А, Таблица А 2.3.5).

Для вычисления результативности модели использовали функции `model_performance` from `sklearn.model_selection` import `cross_val_score` def `model_performance` (X,y,models):

```
for name, model in models: cv = cross_val_score(model,X,y,cv=sss, n_jobs=-1) print(f'{name !r:30}: {np.mean(cv)}') %time model_performance (X_train,y_train,estimators).
```

Для тюнинга моделей использовали следующие метод моделей:

```
model_performance from sklearn.model_selection import cross_val_score
```

```
def model_performance(X,y,models): for name, model in models:
```

```
cv = cross_val_score(model,X,y,cv=sss,n_jobs=-1) print (f'{name !r:30} : {np.mean(cv)}')%time model_performance(X_train,y_train,estimators
```

Полученные результаты позволили сделать вывод, что результативность зависит от модели регрессии. Низкое значение  $R^2$  выявлено только у модели Elastic Net:  $R^2 \leq 1$ ; MAPE составляет 8,7%; MAE – 344,2. Результативность остальных 14 моделей регрессии была достоверна и составила  $R^2 \geq 1$ ; MAPE был в пределах 0,2-1,6; MAE – в пределах 4-45. Для улучшения показателей качества изученных регрессионных моделей (в том числе Elastic Net) провели



тюнинг моделей. После тюнинга коэффициент достоверности модели Elastic Net возрос ( $R^2 \geq 1$ ).

На основе анализа эффективности и точности регрессионных моделей (с тюнингом и без отбора признаков) с использованием метрики ошибок было выявлено, что 10 моделей (Decision Tree Regressor, Gradient Boosting Regressor, XGB Regressor, Stacking Regressor (without Linear Regression, Elastic Net), Stacking Regressor, Voting Regressor without Linear Regression, Elastic Net, Voting Regressor, Keros и H<sub>2</sub>OAuto ML) являются наиболее подходящими для прогнозирования продуктивности растений в зависимости от климатических метаданных: MAPE  $\leq 1\%$ , а  $R^2 \geq 1$ .

*Отбор признаков.* В фазах предварительной обработки одной из основных задач является отбор атрибутов (признаков) 1 и 2 из набора данных, которые больше всего влияют на производительность модели. Аналогично, как и при выборе моделей, была проведена оценка достоверности моделей регрессии.

Выполнены следующие этапы работ:

1. Достоверность 9 видов моделей без тюнинга, но с отбором атрибутов (Linear Regression, Lasso, Elastic Net, Ridge, SGD Regression, Random Forest Regressor, Decision Tree Regressor, Gradient Boosting Regressor, XGB Regressor) оценивали по  $R^2$  (Приложение А, Таблицы 2.3.6 и 2.3.7).

2. Результативность 15 видов моделей без тюнинга, но с отбором атрибутов (Linear Regression, Lasso, Elastic Net, Ridge, SGD Regression, Random Forest Regressor, Decision Tree Regressor, Gradient Boosting Regressor, XGB Regressor, Stacking Regressor without Linear Regression, Elastic Net, Stacking Regressor, Voting Regressor without Linear Regression, Elastic Net, Voting Regressor, Keros и H<sub>2</sub>OAuto ML) оценили по показателям MSE, MAE, MAPE и  $R^2$  (Приложение А, Таблицы А 2.3.8 и А2.3.9).

3. Достоверность 9 видов моделей с тюнингом и отбором атрибутов (Linear Regression, Lasso, Elastic Net, Ridge, SGD Regression, Random Forest Regressor, Decision Tree Regressor, Gradient Boosting Regressor, XGB Regressor) оценивали по  $R^2$  (Приложение А, Таблицы А 2.3.10 и А 2.3.11).

4. Результативность 15 видов моделей с тюнингом и отбором атрибутов (Linear Regression, Lasso, Elastic Net, Ridge, SGD Regression, Random Forest Regressor, Decision Tree Regressor, Gradient Boosting Regressor, XGB Regressor; Stacking Regressor without Linear Regression, Elastic Net, Stacking Regressor, Voting Regressor without Linear Regression, Elastic Net, Voting Regressor, Keros и H<sub>2</sub>OAuto ML) оценили по MSE, MAE, MAPE и  $R^2$  (Приложение А, Таблицы А 2.3.12 и А.2.3.13).

Полученные результаты позволили сделать вывод, что низкое значение  $R^2$  ( $R^2 \leq 1$ ) при отборе атрибутов было характерно для модели Elastic Net, после тюнинга результативность модели улучшилась.

При отборе признаков 1 и 2 более эффективными и достоверными (MAPE  $\leq 1\%$ ,  $R^2 \geq 1$ ) были следующие 9 моделей: Decision Tree Regressor, Gradient Boosting Regressor, XGB Regressor, Stacking Regressor without Linear Regres-

sion, Stacking Regressor, Voting Regressor without Linear Regression, Voting Regressor, Keros и H<sub>2</sub>OAuto ML. Модель H<sub>2</sub>OAuto ML показала наилучшие результаты при оценке продуктивности в зависимости от климатических условий после выбора признаков и настройки гиперпараметров. Модельный расчет получен с высокой точностью:  $R^2 = 0,99$ ; MSE – 35,6; MAE – 4,38. Из моделей scikit-learn StackingRegressor показывает наилучший результат после выбора функций и настройки гиперпараметров. Мы видим, что  $R^2 = 0,99$ ; MSE составила 425,89; MAE – 15,58 (Приложение А, Таблица А.2.3.14).

При отборе признаков использовали модели «Случайный лес» (Random Forest Regressor). Из ансамблевых моделей по показателям метрики ошибок модель показала наилучший результат после выбора функций и настройки гиперпараметров:  $R^2 = 0,99$ ; MAE – 11,3; RMSE (ошибка корневого квадрата) – 180,8. Значение RMSE не более 180, низкое значение MAE и  $R^2 \geq 1$ , считаются хорошим баллом при выборе точности модели.

*Отбор информативных признаков с помощью регрессионных моделей, влияющих на продуктивность растений (Feature selection).* Преобразование необработанных наблюдений в желаемые признаки с использованием подходов МО называется проектирование признаков. Отбор информативных признаков является важным этапом построения МО для улучшения точности классификации и снижения времени обработки данных, так как неинформативные или слабо информативные признаки понижают эффективность модели. Задача выбора переменных состоит из трех частей: повышение эффективности прогнозирования признаков, предоставление более быстрых и экономичных предикторов и обеспечение лучшего понимания основного процесса, который генерирует данные [222].

Для уточнения пространства информационных признаков использовали методы автоматизированного отбора признаков Python и библиотеки scikit-learn: одномерный отбор признаков (SelectKBest); рекурсивное устранение признаков с перекрестной проверкой (RFECV); отбор с использованием ансамблевого на основе «случайный лес» (Random Forest Regressor) и «градиентный бустинг»; последовательный выбор признаков (Sequential Feature Selector); подхода SHAP (Shapley Additive exPlanations).

*Одномерный отбор признаков.* Одномерный выбор признака – признаки, имеющие наиболее выраженную взаимосвязь с целевой переменной, могут быть отобраны с помощью статистических критериев. Библиотека scikit-learn содержит класс SelectKBest, реализующий одномерный отбор признаков, который работает путем выбора лучших признаков на основе одномерных статистических тестов. Scikit-learn предоставляет процедуры выбора функций. Этот класс можно применять совместно с различными статистическими критериями для отбора заданного количества признаков с использованием MSE, MAE, MAPE и  $R^2$ , чтобы отобрать предикторы, влияющие на продуктивность. SelectKBest удаляет все признаки, кроме (k) функций с самым высоким рейтингом [223]. Результаты показали, что метод расширенного ближайшего соседа (ENN), основанный на оценивании сходства объектов лучше, чем тради-

ционные нейронные сети с обратным распространением и множественным регрессионным анализом. Метод, основанный на ENN, имеет гораздо более низкую частоту ошибок, чем традиционные нейронные сети с обратным распространением, а множественный регрессионный анализ имеет частоту ошибок 12,4%. В данной работе предложенный метод анализа данных может применяться в сельском хозяйстве для прогноза продуктивности растений.

С помощью одномерного выбора признаков с помощью статистических критериев отобрано 19 признаков-предикторов:

Index (['temp' (температура); 'dew' (роса); 'humidity' (влажность); 'windgust' порывы ветра); 'windspeed' (скорость ветра); 'winddir' (направление ветра); 'sea level pressure' (давление над уровнем моря); 'cloudcover' (облачность); 'visibility' (видимость); 'datetime\_1' (дни); 'datetime\_2' (месяцы); 'conditions\_x0\_Clear' (безоблачный день); 'description\_x0\_Clear conditions throughout the day' (описание безоблачного периода в течение дня); 'description\_x0\_Partly cloudy throughout the day with afternoon rain' (описание переменной облачности в течение дня в период дождя); 'description\_x0\_Partly cloudy throughout the day with rain' (описание переменной облачности в течение дня в период дождя); 'icon\_x0\_clear-day' (количество дождей).

```
stations_x0_36883099999,36885099999,UAAA,36872099999,36889099999,36870099999',
```

```
'stations_x0_36883099999,UAAA,36872099999,36870099999',
```

```
'stations_x0_UAAA,36872099999,36870099999'],
```

```
dtype='object')
```

*Рекурсивное устранение признаков (RFE) с помощью модели «Случайный лес».* Рекурсивное устранение признаков с перекрестной проверкой (RFECV) проводит ранжирование признаков с рекурсивным устранением признаков. Для вычисления оптимальных количеств признаков была выбрана модель – случайный лес (random\_state =0, n\_jobs =-1). После каждого вычисления результаты отбора оптимального количества признаков фиксируются графически (рисунок 2.3.7).

В процессе ранжирования признаков с рекурсивным устранением осталось только 17 (из 37) признаков-предикторов, влияющих на продуктивность растений:

Index (['temp' (температура); 'dew' (роса); 'humidity' (влажность); 'precip' (осадки); 'windgust' (порывы ветра); 'windspeed' (скорость ветра); 'winddir' (направление ветра); 'sealevelpressure' (давление над уровнем моря); 'cloudcover' (облачность); 'visibility' (видимость); 'moonphase' (фаза луны); 'datetime\_1' (дни); 'datetime\_2' (месяцы); 'description\_x0\_Clear conditions throughout the day with early morning rain' (описание безоблачного периода после утреннего дождя).

```
'stations_x0_36883099999,36871099999,AAA,36872099999,36870099999',
```

```
stations_x0_36883099999,36885099999,36871099999,
```

```
UAAA,36872099999,36889099999,36870099999',
```

```
'stations_x0_36883099999, UAAA,36872099999,36870099999'],dtype='ob-
```

```
ject')
```

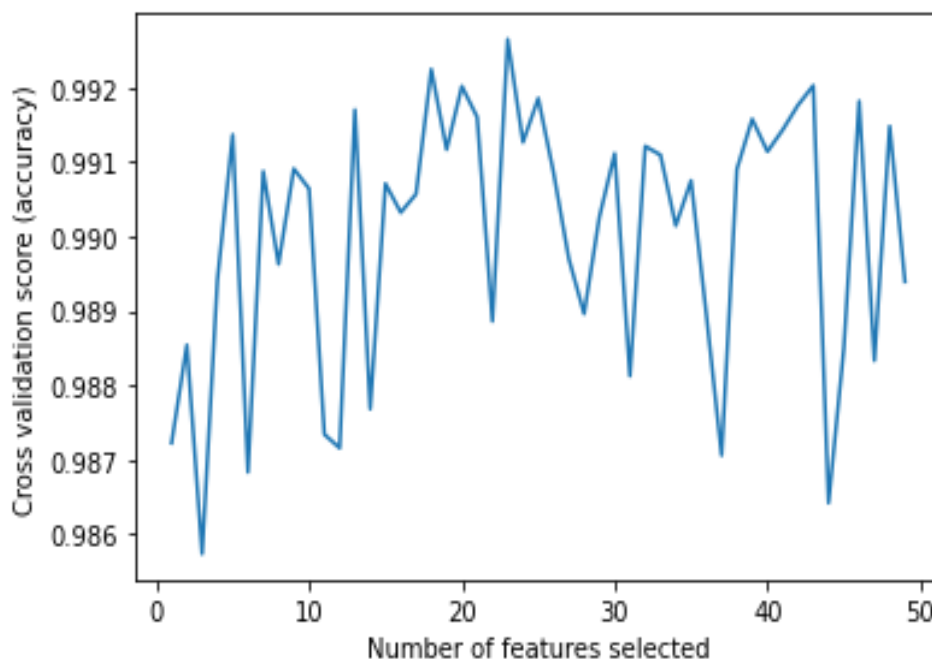


Рисунок 2.3.7 – Результаты вычисления оптимальных признаков с помощью модели случайный лес (Random Forest Regressor)

*Последовательный отбор информативных признаков (Sequential Feature Selector).* Для последовательного отбора признаков применили прямой отбор (Forward selection) и обратный отбор (Backward selection) с использованием модели «Случайный лес» (random\_state =0).

*Прямой отбор.* С помощью прямого отбора выделено 8 признаков: Index ([ 'dew' (роса); 'humidity' (влажность); 'precip' (осадки); 'cloudcover' (облачность); 'visibility' (видимость); 'moonphase' (фаза луны); 'datetime\_1' (дни); 'datetime\_2' (месяцы)], dtype='object'), - оказывающих наибольшее влияние на продуктивность растений.

С помощью обратного отбора - 9 признаков: Index ([ 'temp' (температура); 'humidity' (влажность); 'windspeed' (скорость ветра); 'cloudcover' (облачность); 'visibility' (видимость); 'moonphase' (фаза луны); 'datetime\_1' (дни); 'datetime\_2' (месяцы); 'conditions\_x0\_Overcast' (пасмурные дни)], dtype='object').

*Методы отбора признаков с помощью модели Градиентный бустинг.* Градиентный бустинг используется для создания прогностических моделей. Выявление оптимальных признаков с помощью регрессионных моделей способствует избавлению ненужных признаков либо признаков, незначительно влияющих на вычисления. Оптимальные признаки определяются на основе общего снижения критериев, характерных данному признаку. Модель «Градиентный бустинг» из рассмотренных 37 признаков выявил 23 показателя, влияющих на продуктивность растений (значение > 0), а не оказывающих влияние 14 признака (значение 0). Они, как ненужные признаки были удалены (рисунок 2.3.8).

Вывод, из 23 показателей самое большое влияние оказывало только 2 признака: `datetime_2` (месяцы) и `datetime_1` (дни), т.е. период роста растений. Признаки `'precipcover'` (осадки), `'icon_x0_cloudy-day'` (количество ясных дней) и `'description_x0_Partly cloudy throughout the day'` (описание переменной облачности в течение дня) оказывали слабое влияние.

*Оценка важности признаков путем перестановок [BRE].* Оценка важности перестановки признака рассчитывается следующим образом: базовая метрика, определяемая путем подсчета очков, оценивается в наборе данных, определенном  $X$ ; столбец признаков из набора проверки переставляется, и метрика оценивается снова. Важность перестановки определяется как разница между базовой метрикой и метрикой перестановки признака. При оценке важности признаков путем перестановок, значения каждого столбца признаков изменяются случайным образом, далее оценивается модель. При выборе признаков на основе фильтра показатели вычисляются перед созданием модели. Обусловлено тем, что при оценке важности признаков путем перестановок не измеряется связь между признаком и целевым значением. Вместо этого вычисляется, насколько сильно каждый признак влияет на прогнозы модели [224]. Для оценки важности признаков путем перестановок используя подход «SHAP» (Sharp feature importance), которая поддерживается для моделей типа «ансамбль деревьев».

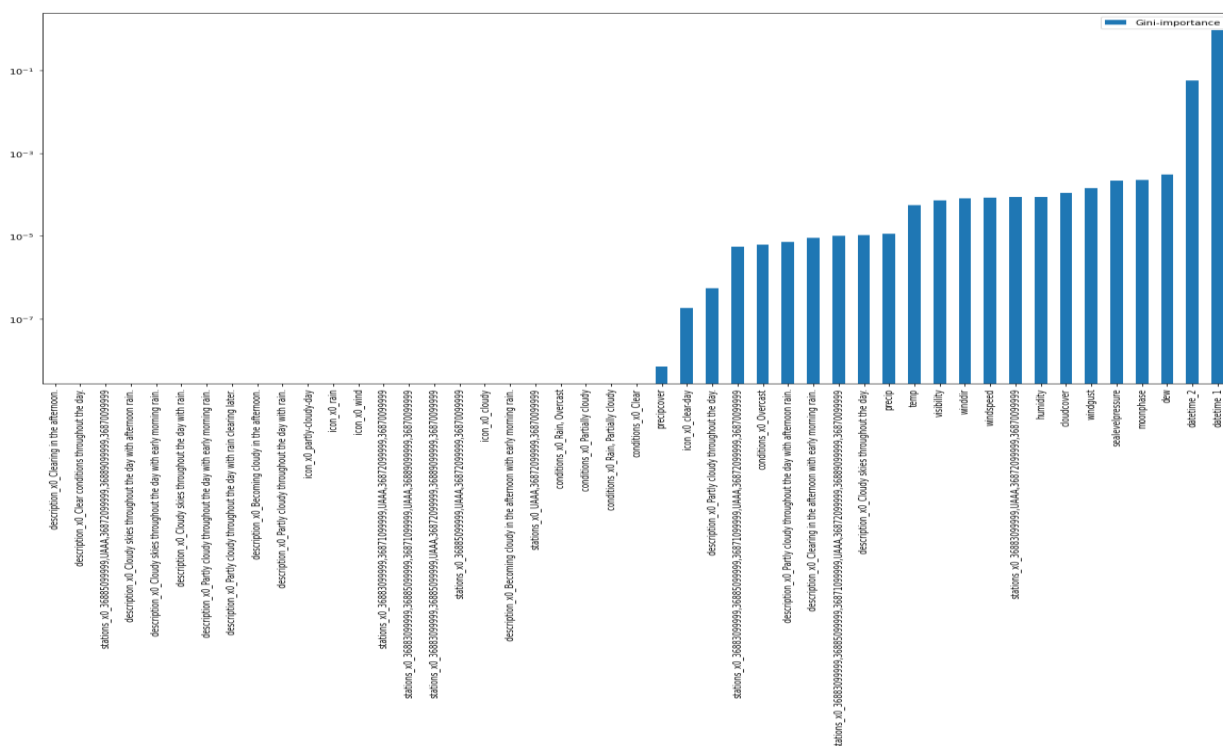


Рисунок 2.3.8 –Отбор признаков с помощью модели Градиентный бустинг (Gradient Boosting Regressor)

SHAP – теоретико-игровой подход к объяснению выходных данных любой модели МО. Он связывает оптимальное распределение с локальными объяснениями, используя классические значения Шепли из теории игр и их соответствующих расширений. Сводные графики SHAP дают общее представление о важности функций и о том, что ими движет. График состоит из множества точек. Каждая точка имеет три характеристики: вертикальное расположение показывает, какую функцию оно изображает; цвет показывает, был ли этот признак высоким или низким для этой строки набора данных; горизонтальное положение показывает вызвало ли влияние этого значения более высокий или более низкий прогноз. Полученный график интерпретируется следующим образом [225]: значения слева от центральной вертикальной линии – negative-класс (0), справа – positive (1) по матрице ошибок предиктивной МО-модели; толщина линии прямо пропорциональна количеству точек наблюдения; чем краснее точки, тем большее значение имеет признак в данной точке (рисунок 2.3.9).

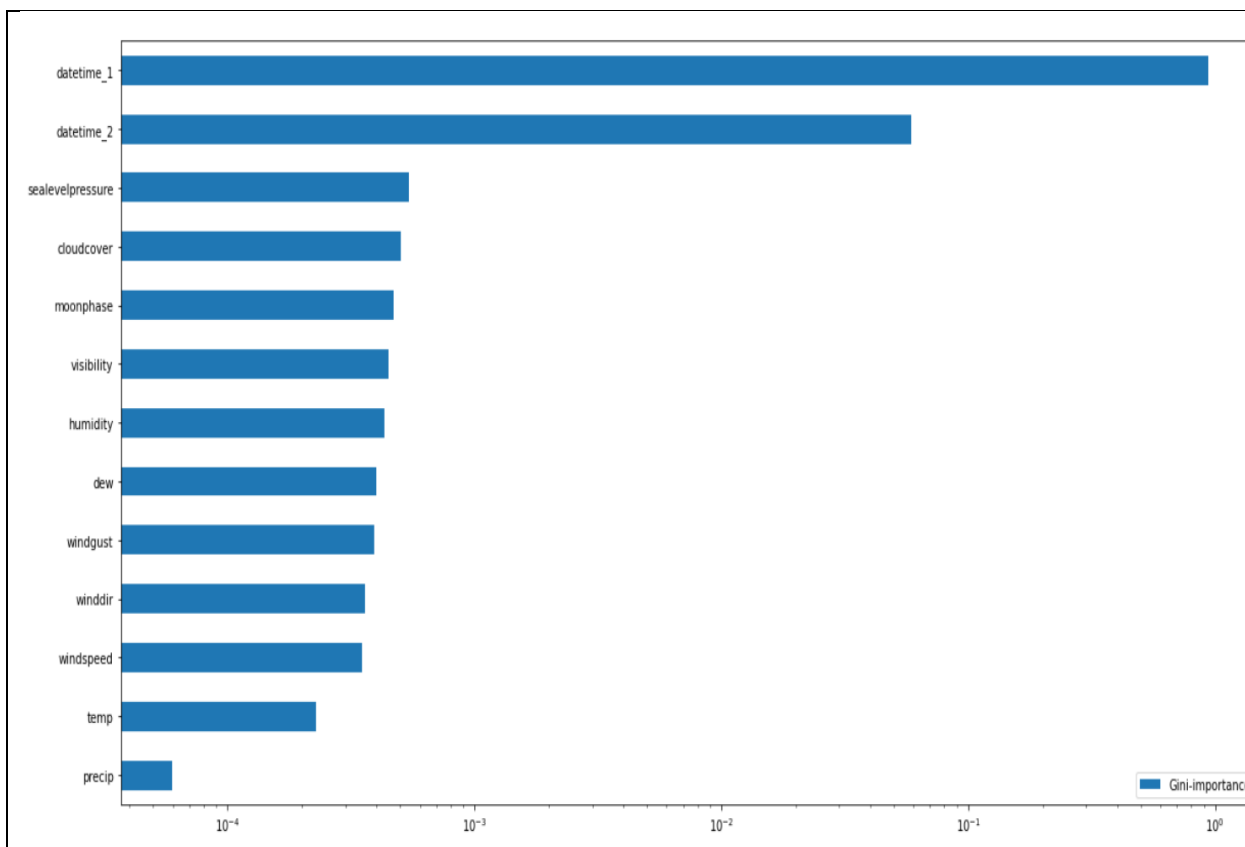


Рисунок 2.3.9 – Сводные графики SHAP важных признаков

С помощью подхода SHAP выявлено 13 признаков: Temp (температура); dew (роса); humidity (влажность); precip (осадки); windgust (порывы ветра); windspeed (скорость ветра); winddir (направление ветра); sealevelpressure (давление над уровнем моря); cloudcover (облачность); visibilit (видимость); moonphase (фаза луны); datetime\_1 (месяц); datetime\_2 (дни); temp (температура).

Из графика можно сделать следующие выводы, что 2 признака: `datetime_1` (дни) и `datetime_2` (месяцы), т.е. длительность периода роста растений, являются наиболее эффективными, оказывающее влияние на прогноз продуктивности, а остальные 14 параметров можно рассматривать как побочные.

На рисунке 2.3.10 показана степень влияния разных предикторов на прогноз продуктивности растений. В графике каждая линия предсказывает влияние признаков на прогноз продуктивности.

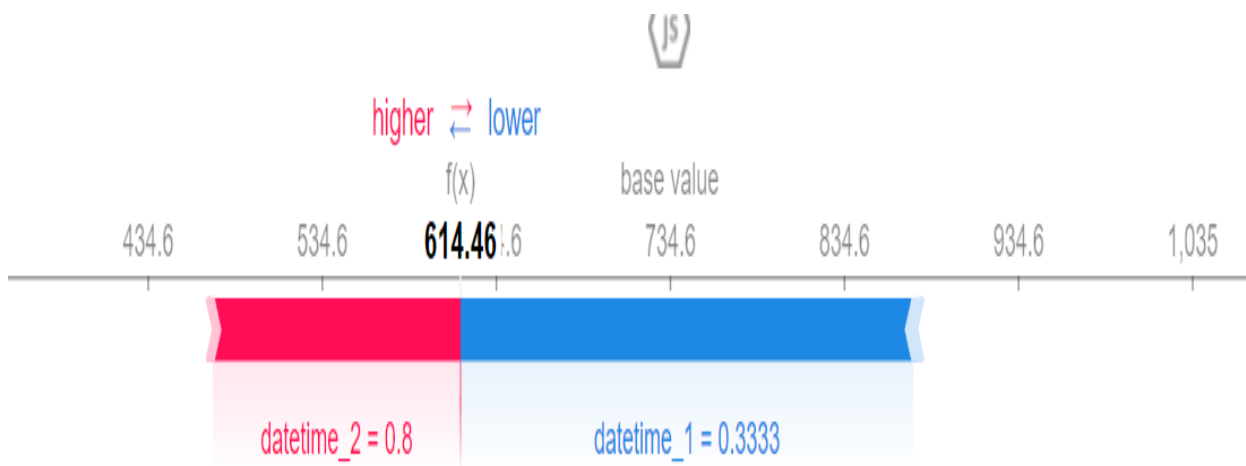


Рисунок 2.3.10 – Степень влияния разных предикторов на продуктивность растений

Высокий уровень розового (красного) цвета означает смещение прогноза к значению 614,46 (т.е. увеличиваются выходные значения модели), тогда как синие значения смещают его к значению 0 (т.е. уменьшаются выходные значения модели). Установлено, что признак `datetime_2` (месяцы) оказывает наибольшее влияние на прогноз продуктивности, чем признак `datetime_1` (дни). Далее, используя совокупные значения SHAP, провели визуализацию модели для прогноза продуктивности (рисунок 2.13.11).

После визуализации количество признаков, оказывающих влияние на прогноз продуктивности растений, увеличилось наряду с признаком `Temp` (температура); `dew` (роса); `humidity` (влажность); `precip` (осадки); `windgust` (порывы ветра); `windspeed` (скорость ветра); `winddir` (направление ветра); `sealevelpressure` (давление над уровнем моря); `cloudcover` (облачность); `visibilit` (видимость); `moonphase` (фаза луны); `datetime_1` (месяц); `datetime_2` (дни); `temp` (температура).

*Интерпретация показателей взаимной информации (Interpreting Mutual Information Scores).* Совместная информация между двумя величинами означает насколько значение одной величины снижает неопределенность в отношении другой. Взаимная информация является логарифмической величиной, чем меньше значение  $x$ , тем меньше связь, когда совместная информация равно 0, то величины независимы. На рисунке 2.3.12 видно, что признаки `datetime_1` (дни) и `datetime_2` (месяцы) взаимосвязаны между собой.

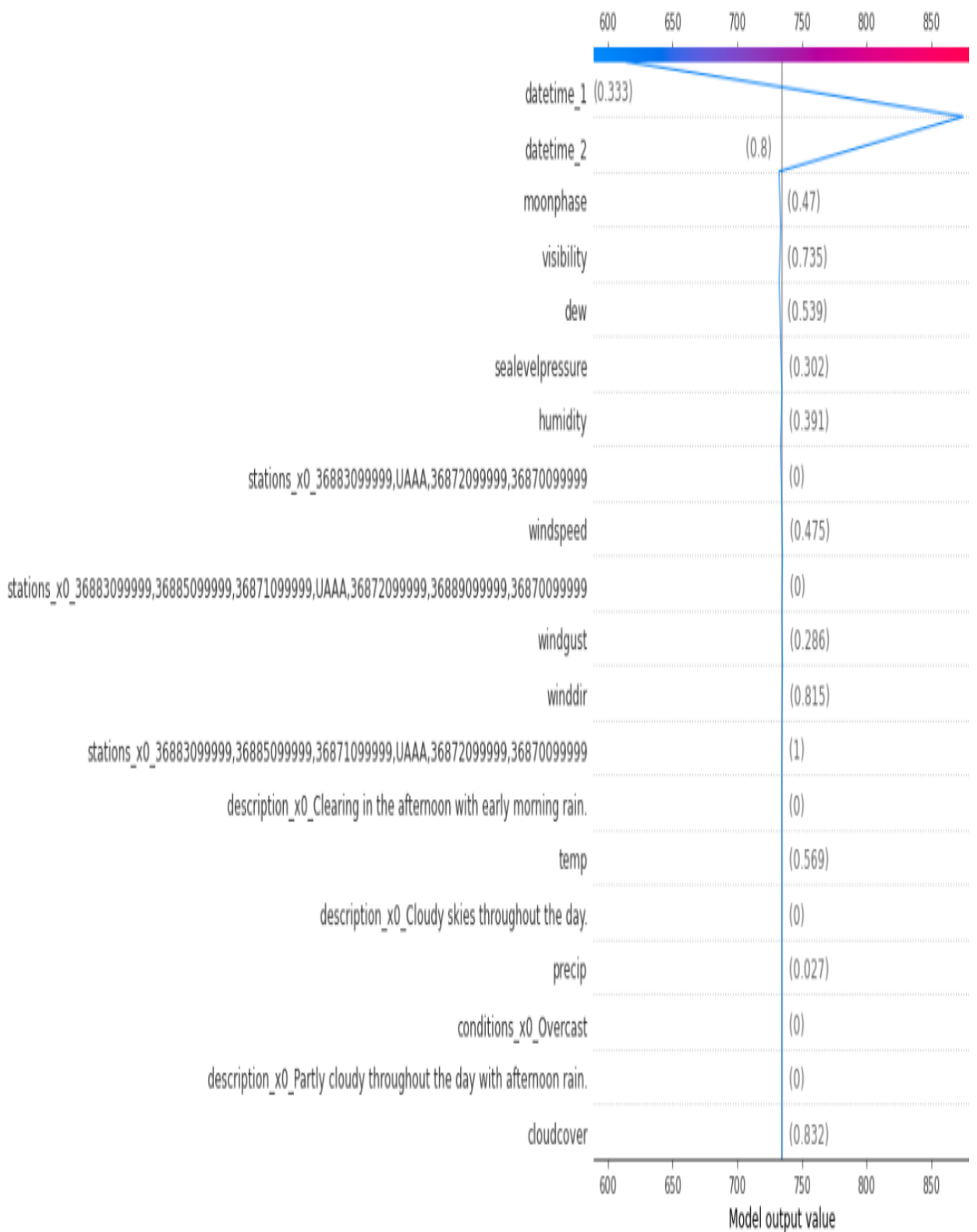


Рисунок 2.3.11 – Визуализация прогноза модели с помощью совокупности значения SHAP



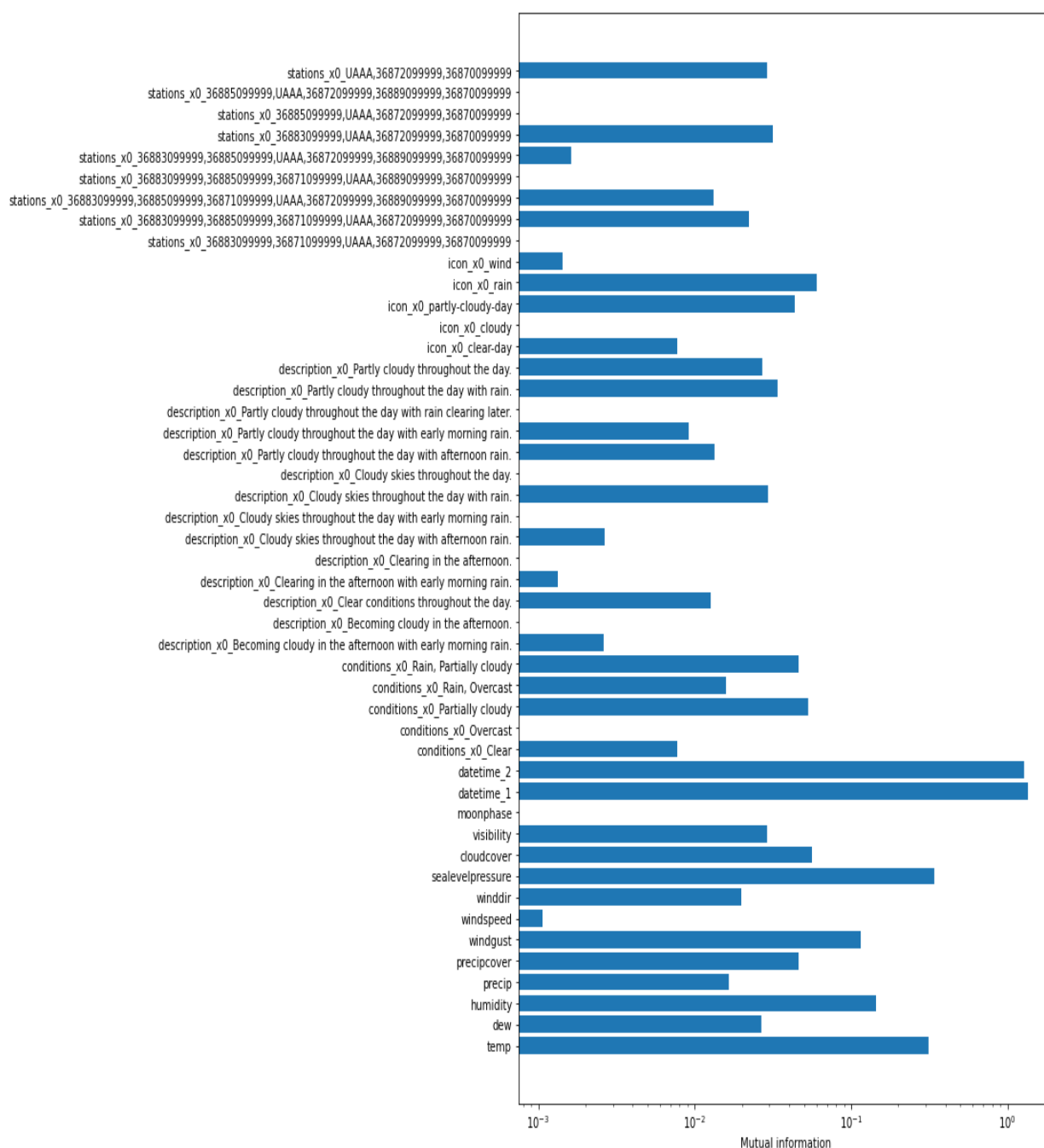


Рисунок 2.3.12 – Взаимная информация признаков, влияющих на прогноз продуктивности растений

Создание модели глубокого обучения для прогноза продуктивности растений в зависимости от климатических условий, используя модель *Tensorflow keras*. Установлено, что большое количество наборов данных приводит к более высокой точности, которая обеспечивает точное прогнозирование продуктивности сельскохозяйственных растений [200]. Алгоритмы МО для моделирования высокоуровневых абстракций с применением многочисленных нелинейных преобразований рассматриваются в моделях глубокого обучения. В системах глубокого обучения автоматизируется процесс выбора и настройки признаков [226,227].

При создании модели для прогноза биомассы растений при произрастании на ТЭ загрязненной почве использовали Keras - высокоуровневый API-интерфейс, выпущенный в составе платформы TensorFlow, которая позволяет сохранять модель. Сохраненную модель используют для прогнозирования функций. TensorFlow – сквозная платформа с открытым исходным кодом для МО [236].

*Преимущества модели Keras:*

- Простота – снижает когнитивную нагрузку на разработчика, чтобы вы могли сосредоточиться на действительно важных частях проблемы;
- Гибкость – использует принцип постепенного раскрытия сложности;
- Мощность – обеспечивает лучшую в отрасли производительность и масштабируемость; ее используют организации и компании, в том числе NASA, YouTube и Waymo.

Keras – библиотека для Python, позволяет легко и быстро создавать нейронные сети [227]. Нейронные сети являются структурой МО и состоят из множества искусственных нейронов и является системой, которая учится делать прогнозы, выполнив следующие шаги: получение входных данных; делать прогноз; сравнивать прогноз с конкретным результатом [228,229].

Keras включает в себя большое количество слоев, создает, последовательно добавляя слои, модель, используя метод add():

```
model = keras.models.Sequential()
model.add(keras.layers.Dense(835,
input_dim=input, kernel_initializer='normal', activation='relu'))
model.add(keras.layers.Dense(417, kernel_initializer='normal', activation=
'relu'))
model.add(keras.layers.Dense(213, kernel_initializer='normal', activation =
'relu'))
model.add(keras.layers.Dense(1, kernel_initializer='normal'))
from tensorflow.keras.utils import plot_model
plot_model(model, to_file='model.png', show_shapes=True)
from IPython.display import Image
Image("model.png").
```

Каждый слой представляет собой функцию, которая принимает на вход результат работы предыдущего слоя. Структуру такой модели можно вывести наглядно со всеми зависимостями слоев друг от друга (рисунок 2.3.13).

Основная структура нейронной сети состоит из трех типов слоев – нейронов (взаимосвязанных узлов). Первый слой является входным уровнем состоит из 49 нейронов (входные данные). Второй слой является скрытым уровнем, в котором обрабатываются данные для получения необходимых промежуточных данных и для определения окончательного решения, и состоит из 835 слоев, а третий слой является выходным слоем, где производятся результаты, которые предсказывают прогноз продуктивности растений.

Общее количество признаков: 49

Изучаемые признаки: 49

Неизучаемых признаков: 0

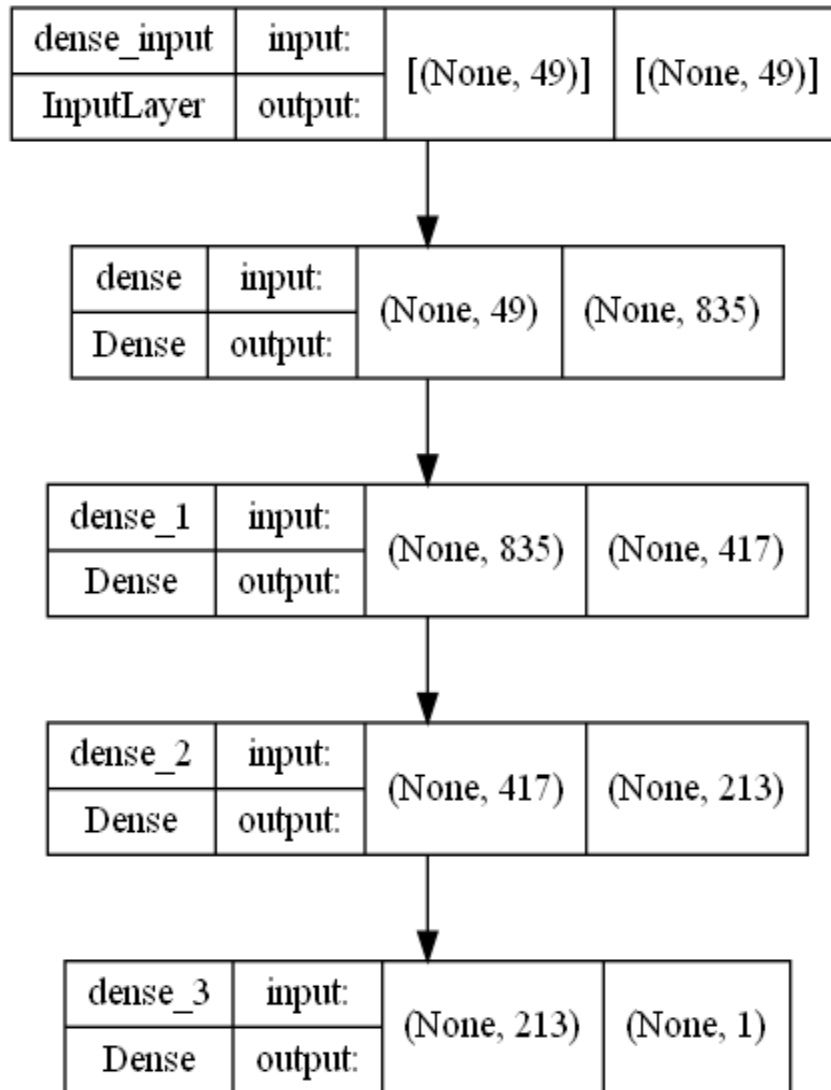


Рисунок 2.3.13 – Структура модели Keras

Создание нейронной сети осуществляли с помощью платформы Tensorflow keras. В данной программе столбцы выборки  $x_1 - x_{49}$  - входные данные информативных признаков, а столбец  $Y$  – выход сети (продуктивность) (рисунок 2.3.14).

Для создания нейронной сети использовали следующие методы:

```
optimizer = tf.keras.optimizers.RMSprop(learning_rate=0.001, rho=0.9, epsilon=1e-08, decay=0.0)
model.compile(loss='mean_squared_error',
optimizer=optimizer, metrics=['mae','mse'])
learning_rate_reduction = keras.callbacks.ReduceLROnPlateau(monitor='-mse', patience=4, verbose=1, factor=0.9, min_lr=0.0000000001, batch_size = 128
epochs = 300
fig, ax = plt.subplots(3,1)
ax[0].plot(history.history['loss'], color='b', label="Training loss")
```

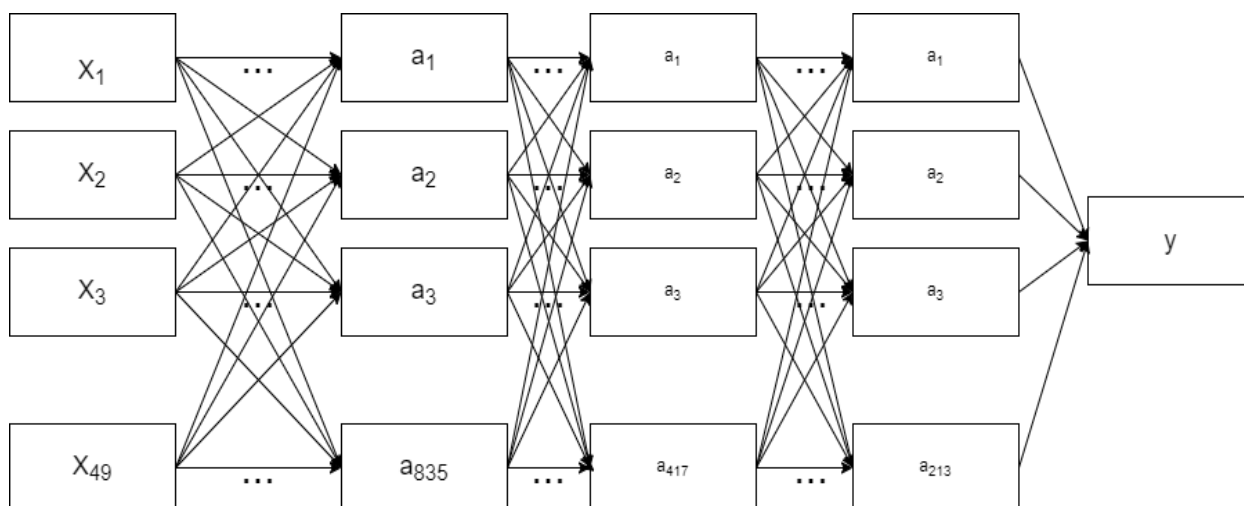


Рисунок 2.3.13 – Модель Keras

```
ax[0].plot(history.history['val_loss'], color='r', label="validation loss", axes
=ax[0])
legend = ax[0].legend(loc='best', shadow=True)
ax[1].plot(history.history['mae'], color='b', label="Training mae")
ax[1].plot(history.history['val_mae'], color='r',label="Validation val_mae")
legend = ax[1].legend(loc='best', shadow=True)
ax[2].plot(history.history['mse'], color='b', label="Training mse")
ax[2].plot(history.history['val_mse'], color='r',label="Validation val_mse")
legend = ax[2].legend(loc='best', shadow=True)
```

Гистограмма результатов валидации модели Keras представлена на рисунке 2.3.14. Таким образом, создана нейронная сеть с помощью платформы Tensorflow keras. Модель Keras показала неплохие результаты после выбора признаков и настройки гиперпараметров. Модельный расчет получен с высокой точностью:  $R^2 - 0,996$ ;  $MSE - 774,78$ ;  $MAE - 21,63$ .

*Создание модели глубокого обучения для прогноза продуктивности растений в зависимости от климатических условий, используя H<sub>2</sub>O AutoML.* AutoML (автоматизированное МО) разработана для простых унифицированных интерфейсов для различных методов МО, таких как предварительная обработка данных, настройка гиперпараметров, выбор и оценка модели. Основной код H<sub>2</sub>O написан на Java. Внутри H<sub>2</sub>O хранилище распределенного ключа/значения, которые используется для доступа и ссылки на данные, модели, объекты и т.д. на всех узлах и машинах. Модель H<sub>2</sub>O позволяет получить доступ ко всем возможностям из внешней программы или скрипта через JSON из HTTP. Rest API используется веб-интерфейсом H<sub>2</sub>O (Flow UI), привязкой R (H<sub>2</sub>O-R) и привязкой Python (H<sub>2</sub>O-Python). Скорость, качество, простота использования и развертывание моделей для различных передовых контролируемых и неконтролируемых, таких как Deep Learning, Tree Ensembles и GLRM,

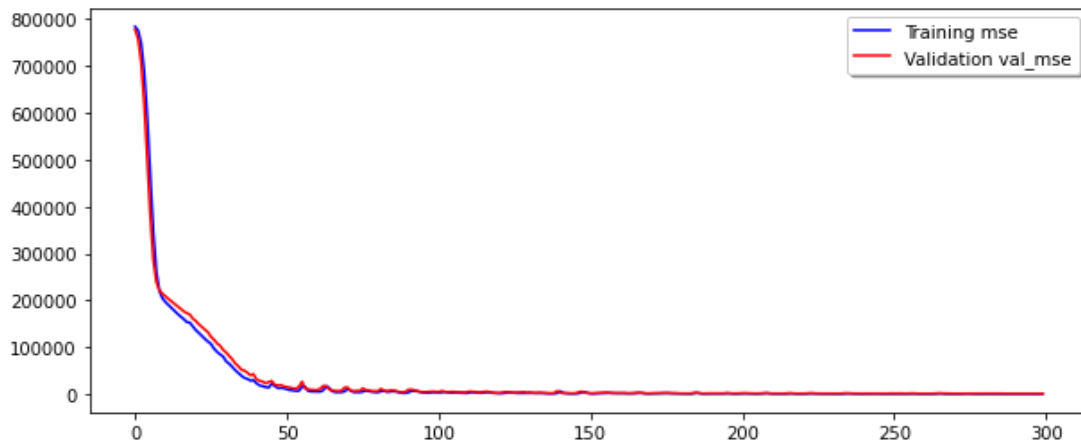
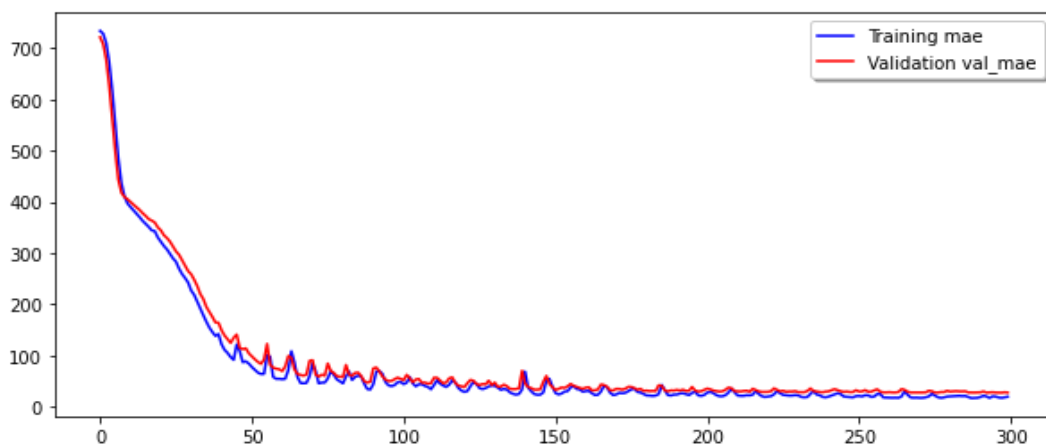
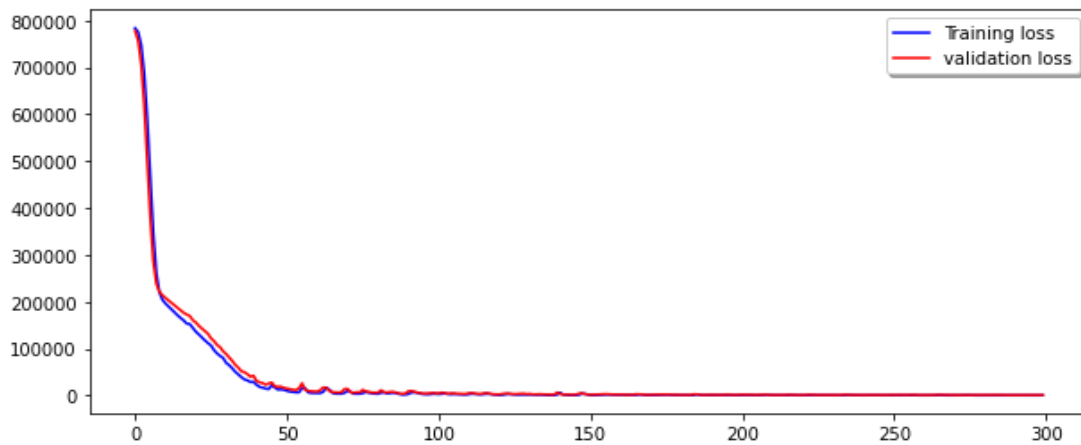


Рисунок 2.3.14 – Гистограмма валидации модели Keras

делают H<sub>2</sub>O востребованным API для обработки больших данных и распространяется под лицензией Apache, версия 2. Модель H<sub>2</sub>O Map/Reduce используют платформу Java Fork/Join. Данные считываются параллельно, распределяются по кластеру и хранятся в памяти в виде столбцов в сжатом виде. Парсер данных H<sub>2</sub>O имеет встроенную интеллектуальную функцию для угадывания схемы входящего набора данных и поддерживает получение данных из нескольких источников в различных форматах [230]. Перекрестная проверка позволяет более точно и надежно оценить производительность модели. Версия

AutoML перебирает различные модели и параметры, пытаясь найти наилучшие, и выполняет 5-кратную перекрестную проверку достоверности моделей. Модели ранжируются по метрике в зависимости от типа проблемы. В задачах бинарной классификации AUC (Area Under Curve) является наиболее важным параметром оценки эффективности любой модели классификации; в задачах мультиклассовой классификации метрикой является вычисление среднего метрики по всем классам, т.е. средняя ошибка для каждого класса; а в задачах регрессии метрикой сортировки по умолчанию являются RMSE, MSE,  $R^2$  и RMSLE (среднеквадратичная логарифмическая ошибка). Версия AutoML обучает и перекрестно проверяет следующие методы: трех предварительно заданных моделей и случайной сетки Градиентный бустинг (Gradient Boosting Machine); фиксированной сетки обобщенной линейной модели (H<sub>2</sub>O GLM); модели Случайный лес (Random Forest Regressor) по умолчанию; пяти предварительно заданных моделей Машины повышения градиента (Gradient boosting machine, H<sub>2</sub>O GLM); глубокой нейронной сети по умолчанию и случайной сетки; модели Чрезвычайно рандомизированные деревья (Extremely Randomized Forest, XRT).

Настройку модели H<sub>2</sub>O AutoML проводили, используя следующие функции `L`: `aml = H2O AutoML(max_models=20, seed=1) aml.train (x=x, y='yield', training_frame = X)`.

Провели перекрестную проверку следующих моделях H<sub>2</sub>O Stacked Ensemble Estimator; H<sub>2</sub>O Stacked Ensemble Estimator, Regression GLM.

Model Details

=====

H<sub>2</sub>O Stacked Ensemble Estimator: Stacked Ensemble

Model Key: StackedEnsemble\_AllModels\_1\_AutoML\_1\_20220516\_83001

Model Details

=====

H<sub>2</sub>O Stacked Ensemble Estimator: Stacked Ensemble

Model Key: Stacked

Ensemble\_AllModels\_1\_AutoML\_2\_20220517\_120432

No model summary for this model

Model Metrics Regression GLM: stackedensemble

\*\* Reported on train data. \*\*

MSE: 30.43278632433223

RMSE: 5.516591912071458

MAE: 3.907953055039384

RMSLE: NaN

$R^2$ : 0.9998800970339042

Mean Residual Deviance: 30.43278632433223

Null degrees of freedom: 835

Residual degrees of freedom: 828

Null deviance: 212186655.55627948

Residual deviance: 25441.809367141745  
AIC: 5245.8403862196465

Model Metrics Regression GLM: stackedensemble

\*\* Reported on cross-validation data. \*\*

MSE: 331.7076154507837

RMSE: 18.212842047598823

MAE: 12.984645302239619

RMSLE: NaN

R<sup>2</sup>: 0.9986930961054555

Mean Residual Deviance: 331.7076154507837

Null degrees of freedom: 835

Residual degrees of freedom: 828

Null deviance: 212391405.2940222

Residual deviance: 277307.56651685515

AIC: 7242.8214917138

Cross-Validation Metrics Summary:

По сводкам показателей 5-ти перекрестной проверки модели, Regression GLM не является наилучшей моделью, показатели метрики ошибки повысились после валидации, так значение MSE от 30,43 до 331,70; RMSE – от 5,51 до 18,21; MAE – от 3,90 до 12,98; RMSLE – в пределах NaN; R<sup>2</sup> – в пределах 0,99.

Наилучшие результаты показала модель Stacked Ensemble\_AllModels\_1\_AutoML\_1\_20220516\_83001 – лидер Stacked Ensemble. Модельный расчет Stacked Ensemble от H<sub>2</sub>O после 5-ти валидации был получен с высокой точностью: R<sup>2</sup> – 0,99; MSE – 9,54; MAE – 7,26; RMSE – 9,77; RMSLE – 5,38. Следует заметить, что после 1-го и 4-го показатели перекрестной проверки были низкими, что доказывает эффективность модели (Таблица 2.3.1).

Модель Stacked Ensemble от H<sub>2</sub>O – ансамблевый модели МО, который находит оптимальную комбинацию набора моделей прогнозирования с использованием процесса, называемого суммированием. Метод поддерживает регрессию и бинарную классификацию [230].

Таким образом, обобщая использование МО для прогнозирования продуктивности растения на загрязненной ТЭ почве, в зависимости от климатических условий среды следует отметить:

1. Проанализировали 15 различных моделей регрессии, используя климатические данные с 2015 по 2021 год (осадки, температура воздуха, потенциальное испарение влаги из почвы и активное фотосинтетическое излучение), результаты сухой биомассы в течение 7 лет мискантуса, произрастающего на загрязненной ТЭ почве (суммарный уровень загрязнения почвы – 20,0 мг кг<sup>-1</sup>), используя Jupyter Notebook, Sklearn Libraries. Оценку результата проводили по критериям MSE, MAE, MAPE, коэффициента детерминации и RMSLE.

Таблица 2.3.1 – Сводка показателей перекрестной проверки (валидации) модели H<sub>2</sub>O Stacked Ensemble

Метрика ошибок	MEAN	STD	CV_1_VALID	CV_2_VALID	CV_3_VALID	CV_4_VALID	CV_5_VALID
MAE	7,160792e+00	3,403506e-01	7,397977e+00	7,144588e+00	6,584667e+00	7,415339e+00	7,261392e+00
Mean_residual_deviance	9,643430e+01	9,060707e+00	1,080191e+02	9,736732e+01	8,273301e+01	9,858855e+01	9,546354e+01
MSE	9,643430e+01	9,060707e+00	1,080191e+02	9,736732e+01	8,273301e+01	9,858855e+01	9,546354e+01
NULL_deviance	4,247828e+07	2,709452e+06	4,237122e+07	4,487712e+07	4,202360e+07	4,485400e+07	3,826546e+07
R <sup>2</sup>	9,996178e-01	4,856041e-05	9,995533e-01	9,996392e-01	9,996809e-01	9,996257e-01	9,995899e-01
Residual_deviance	1,616048e+04	2,012801e+03	1,890334e+04	1,616298e+04	1,332001e+04	1,676005e+04	1,565602e+04
RMSE	9,811240e+00	4,661896e-01	1,039322e+01	9,867488e+00	9,095769e+00	9,929176e+00	9,770544e+00
RMSLE	7,456394e-02	3,700514e-02	NaN	NaN	1,172878e-01	5,258190e-02	5,382213e-02

Примечание: RMSE – ошибка корневого квадрата; MSE – средняя квадратная ошибка; R<sup>2</sup> – коэффициент детерминации; RMSLE – среднеквадратичная логарифмическая ошибка.



2. На основе анализа эффективности и точности регрессионных моделей с использованием метрики ошибок было выявлено 10 моделей (Decision Tree Regressor, Gradient Boosting Regressor, Random Forest Regressor, XGB Regressor, Stacking Regressor (without Linear Regression, Elastic Net), Stacking Regressor, Voting Regressor (without Linear Regression, Elastic Net), Voting Regressor, Keras, H<sub>2</sub>O AutoML) наиболее подходящими для прогноза продуктивности растений в зависимости от климатических метаданных.

3. С помощью подхода SHAP выявлены 13 признаков для прогноза продуктивности растений: непосредственное влияние оказывают 2 признака (datetime\_2 = 0,8 (месяцы, оставшиеся параметры оказывают вспомогательное действие).

4. Лидер-моделью является ансамблевый алгоритм MO, модель Stacked Ensemble от H<sub>2</sub>O ( $R^2 = 0,99$ ; MSE – 9,54; MAE – 7,26; RMSE – 9,77; RMSLE – 5,38).

## 2.4 Выводы по разделу, постановка задачи

Урожайность сельскохозяйственных растений зависит от климатических условий, что затрудняет проектирование точных бизнес-процессов. Поэтому внедрение IT-технологий позволяет спрогнозировать будущую урожайность сельскохозяйственных растений в зависимости от климатических условий за счет использования «больших данных», инструментов анализа и визуализации.

Целью данной главы является оценить проблемы и потенциал использования больших данных (на примере климатических данных) для моделирования продуктивности биоэнергетического растения мискантус на загрязненных ТЭ землях.

Для решения поставленной задачи при оценке продуктивности растений (на примере энергетического сельскохозяйственного вида мискантус), произрастающих на ТЭ-загрязненной и незагрязненной почве рассмотрены три подхода: использование регрессионных моделей МО для прогноза продуктивности растений и выбора гиперпредикторов; создание модели MiscanCalc на основе модификации модели MiscanMod для процесса формирования биомассы растений в течение вегетационного периода; разработка математической модели «Многорядный эвристический метод самоорганизации для построения регрессионных уравнений» для отбора предикторов.

1. Усовершенствованная модель MiscanCalc на основе Miscanmod на языке C# в Visual Studio для прогнозирования продуктивности биомассы мискантуса на загрязненных ТЭ почвах, непригодных для производства продуктов питания:

- ввода в блок дополнительного алгоритма путем расчета различий между эмпирическими месячными данными биомассы культур, выращенных на загрязненных и незагрязненных почвах, методом наименьших квадратов;

- ввода в блок входных данных путем расчета фотосинтетически активной солнечной радиации на основании информации региональных коэффициентов и данных интегральной радиации Казгидрометеорологической службы по солнечной радиации;

- ввода в блок входных данных путем расчета потенциального испарения влаги почвы на основании температуры, индекса тепла и средней продолжительности дня.

Разработанная модель в течение вегетационного периода проводит расчет индекса листовой поверхности в зависимости от дефицита влаги в почве; оценивает эффективность использования излучения и температуры; количества максимальных дней в период активного накопления биомассы с учетом температуры и фотопериода в период цветения и отмирания; даты завершения стадии цветения; количества максимальных дней в период активного накопления биомассы с учетом температуры и фотопериода; определяет сроки начала уборки культуры; прогнозирует урожайности в г м<sup>-2</sup> сухой биомассы, как на техногенно-загрязненной, так и незагрязненной почвах.

2. Используя Jupyter Notebook, Sklearn Libraries провели оценку, выбор моделей регрессии и информативных признаков с помощью 15 регрессионных моделей, используя климатические данные с 2015 по 2021 год (осадки, температура воздуха, потенциальное испарение влаги из почвы и активное фотосинтетическое излучение) и выбор модели МО для прогноза продуктивности растений, произрастающих на ТЭ-загрязненной почве, в зависимости от климатических условий среды.

Выявлено, что наиболее подходящими для прогнозирования продуктивности растений в зависимости от климатических метаданных является модель H2OAutoML ( $R^2 = 0,99$ ; MSE – 15,3; MAE– 2,91; MAPE – 0.0), а для отбора признаков – из моделей scikit-learn 3 модели (Stacking Regressor, Gradient Boosting Regressor и H2OAutoML). Модельный расчет H2OAutoML получен с высокой точностью  $R^2 = 0,99$ , MSE – 35,6, MAE– 4,38.

С помощью подхода SHAP выявлено 13 признаков, ответственных за продуктивность растений на загрязненной ТЭ почве. Отмечена высокая степень влияния параметра `datetime_2` (`datetime_2 = 0,8`) на продуктивность растений. После визуализации количество признаков, оказывающих влияние на прогноз продуктивности растений, увеличилось наряду с признаком `datetime_2`: Temp (температура); dew (роса); humidity (влажность); precip (осадки); windgust (порывы ветра); windspeed (скорость ветра); winddir (направление ветра); sealevelpressure (давление над уровнем моря); cloudcover (облачность); visibilit (видимость); moonphase (фаза луны).

Создана нейронная сеть с помощью платформы Tensorflow *keras* после выбора признаков и настройки гиперпараметров. Модельный расчет получен с математической точностью  $R^2 = 0,996$ , MSE– 774,78, MAE – 21,63. Кросс валидация 3-х моделей (H2O Stacked Ensemble Estimator; H2O Stacked Ensemble Estimator, Regression GLM) выявила лидер-модель – ансамблевый модель регрессии Stacked Ensemble от H<sub>2</sub>O с высокой достоверностью:  $R^2 = 0.99$ , MSE– 9.54, MAE – 7.26. RMSE – 9.77, RMSLE – 5.38.

3. Разработан многорядный метод самоорганизации для анализа связи биомассы растения в зависимости от климатических условий среды. Уравнение математической модели, описывающее зависимость роста биомассы от испарения влаги почвы, фотосинтетической активной радиации, осадков и температуры имеет нелинейный вид. Проведенный анализ показал, что показателями, влияющие на урожайность растений, произрастающих на техногенно-загрязненной почве, являются испарение влаги почвы, фотосинтетическая активная радиация и осадки.

### **Глава 3. СОЗДАНИЕ БАЗЫ ДАННЫХ О ЗЕМЛЯХ, ЗАГРЯЗНЕННЫХ ТОКСИЧНЫМИ ЭЛЕМЕНТАМИ И БАЗЫ ДАННЫХ, СОДЕРЖАЩЕЙ ИНФОРМАЦИЮ О ВИДАХ РАСТЕНИЙ, СПОСОБНЫХ К ВОССТАНОВЛЕНИЮ ПОЧВ**

Концептуальной основой базы данных является обеспечение быстрого доступа к хранящимся данным для решения вопросов утилизации и очистки, с целью снижения негативных последствий токсических элементов; быстрый доступ к хранящимся данным о уровне загрязнения ТЭ почвы и данные о растениях местной флоры, способных к восстановлению почвы, загрязненной органическими и неорганическими загрязнителями [231, 232].

Основные принципы построения базы данных:

информация накапливается и хранится в единой базе данных, в которую поступают сведения из всех возможных источников;

передача данных конечным пользователям осуществляется из единой базы данных;

в качестве интерфейса конечного пользователя выступает автоматизированная информационная система;

инструментарии для ввода, преобразования и анализа данных проводится непосредственно в единой базе данных с сохранением в ней результатов работ;

ключевым объектом управления является обмен сведениями для передачи поступившей информации органам по управлению земельными ресурсами, кадастровому учету, оценке, регистрации и налоговым органам при принятии решений и прогнозировании экологической безопасности техногенно-загрязненных земель.

*Выбор платформы для создания базы данных*

Язык программирования: набор программа Enterprise Architect, SQL Server Management Studio, используя язык С# в Visual Studio 2013. Алгоритм базы данных представлен на рисунке 3.1.

Актуальность выбора программного средства Enterprise Architect связана с тем, что данная программа является CASE-инструментом для проектирования и конструирования программного обеспечения. Microsoft SQL Server основана на технологии .NET, поэтому для установки данной версии на компьютер необходимо установить Microsoft.NET Framework. Выбор СУБД SQL связан с тем, что она поддерживает очень большой список платформ и обрабатывает малые массивы данных (до 100 000 записей) быстрее, чем ORACLE. Использование CASE средства, как eaUML в программе Enterprise Architect позволяет создать схему связи между основной и вспомогательной таблицей. Для реализации клиентской части системы была использована интегрированная среда разработки программного обеспечения Microsoft Visual Studio и язык С#. Этот выбор обусловлен тем, что, имея большие функциональные возможности, клиентское приложение остается легко расширяемым и поддерживаемым.

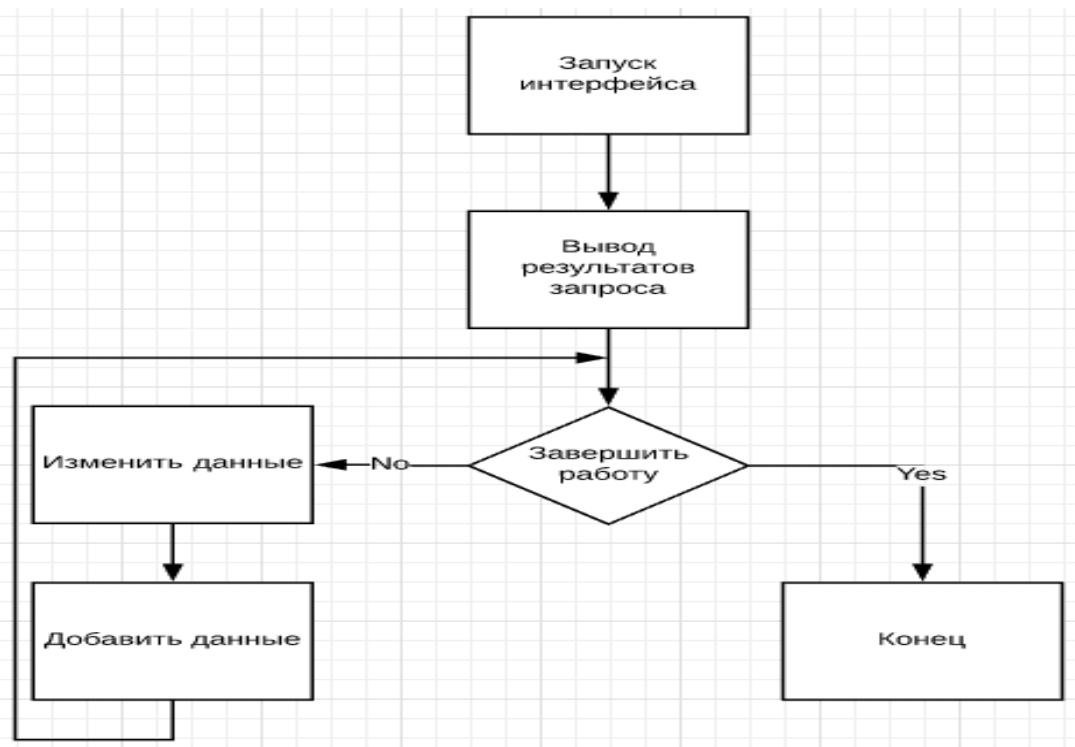


Рисунок 3.1 – Алгоритм базы данных

Выбранный инструмент разработки является признанным лидером в области написания защищенных и многофункциональных объектно-ориентированных приложений. При этом такой инструментарий существенно сокращается время реализации проекта. Программа Visual Studio позволяет обеспечить ввод базы данных в таблицу, их хранение и сопровождение, а также возможность получить из совокупности этой информации нужные данные. Основное взаимодействие пользователей с приложениями осуществляется через базовый элемент графического интерфейса Microsoft Windows.

Для создания графических интерфейсов простой и удобной платформой является Window Forms. Windows Forms – интерфейс программирования приложений, отвечающий за графический интерфейс пользователя и являющийся частью Microsoft.NET Framework. Данный интерфейс упрощает доступ к элементам интерфейса Microsoft Windows и классы, реализующие программное приложение для Windows Forms не зависят от языка разработки.

*Функциональные требования проектируется на языке UML – Enterprise Architect.* Сервер отвечает на запросы клиентов и поддерживает возможности добавления и удаления файлов администратором. Приложение функционирует устойчиво и контролирует входную информацию.

*Основные технические характеристики:*

Для нормального функционирования приложения необходимо использовать компьютеры со следующими техническими характеристиками:

- процессор Intel Pentium Dual-Core 2700 МГц;
- дисковый массив – 300 Гб;

- оперативная память – 2 ГБ;
- сетевая карта – 100 Мбит/с;
- монитор, поддерживающий разрешение экрана 1024x768;
- устанавливается Net Framework (версия 4, 5).

*Функциональные возможности:* база данных позволяет получить доступную информацию. Представленную базу данных можно включить в единую автоматизированную систему управления.

### **3.1 Создание базы данных о землях, загрязненных токсичными элементами, с учетом географического месторасположения территории, количество токсичных элементов и концентрации их в почве**

#### *Описательная информация*

Единой компьютерной информационной БД для решения рекультивационных мероприятий в стране нет. Основная причина нет учета уровня загрязнения антропогенно-нарушенных экосистем и их отходов. В связи с этим, база данных создана на основе сведений, поступивших из различных научных источников о 76 единицах территорий, загрязненных устаревшими пестицидами, количестве идентифицированных и не идентифицированных видов запрещенных пестицидов, количестве контейнеров из-под пестицидов и уровне загрязнения почвы СОЗ-пестицидами, тяжелыми металлами вокруг каждого недействующего склада, на примере 10 районов Алматинской области (таблица 3.1).

Функциональные возможности включают:

- месторасположение складов устаревших пестицидов по районам области;
- количество складов устаревших пестицидов в поселках;
- количество устаревших пестицидов для каждого склада;
- количество идентифицированных пестицидов для каждого склада;
- количество складов пестицидов по районам с уровнем загрязнения почвы в пределах ПДК (предельно-допустимые концентрации);
- количество складов пестицидов по районам с уровнем загрязнения почвы выше ПДК;
- количество складов пестицидов по регионам с уровнем загрязнения почвы ниже ПДК (без загрязнения).

Приложение содержит четыре оконные формы:

- меню навигации о расположении бывших складов хранения устаревших пестицидов по районам области;
- окно, содержащее полную информацию об устаревших пестицидах и уровне загрязнения почвы хлорорганическими пестицидами по районам области;
- окно «о приложении»;
- окно «об авторе».

Таблица 3.1.1 – Описательная информация о количестве пестицидов на территориях недействующих хранилищ пестицидов и уровне загрязнения почвы СОЗ-пестицидами в 10 районах Алматинской области

Районы	Кол-во обследованных поселков, шт	Кол-во недействующих складов, шт	Кол-во устаревших пестицидов, кг
Карасайский	6	6	600
Талгарский	12	14	21650
Джамбульский	4	5	13245
Енбекши-Казахский	10	18	10095
Илийский	3	3	17550
Уйгурский	6	6	3103
Балхашский	3	6	500
Кербулакский	3	7	0
Ескильдинский	2	10	1700
Коксуйский	1	1	0
<b>Всего</b>	<b>50</b>	<b>76</b>	<b>68 443</b>

При запуске приложения клиентом, открывается окно меню навигации о загрязненных территориях. При выборе региона, где расположен склад ядохимикатов, в меню навигации открывается окно информации, где содержится полная информация о количестве токсических веществ и уровне загрязнения почвы.

При проектировании БД с помощью CASE-средств Enterprise Architect первоначально составили модель предметной области, затем диаграмму вариантов использования БД о загрязненных территориях, которая представлена на рисунке 3.1.1. Администратор может добавлять новые данные о загрязненных участках в БД. Измененное приложение обновляется исключительно Администратором. Клиент устанавливает или обновляет измененное администратором приложение. Открыв установленное или обновленное приложение Клиент выбирает нужную информацию. При выборе района Клиент просматривает БД. Диаграммы вариантов использования описывают взаимодействие между пользователем и системой для наглядного представления диаграммы вариантов использования. Известно, что каждый вариант использования – это потенциальное требование к системе и представляет собой последовательность действий (транзакций), выполняемых системой в ответ на событие, инициируемое некоторым внешним объектом (действующим лицом).

*Диаграмма кооперации.* На диаграмме кооперации количество объектов показаны в виде пиктограмм. Стрелки обозначают сообщения и обмен, который осуществляется в рамках данного варианта использования. Их временная последовательность указывается путем нумерации сообщений. Диаграмма последова

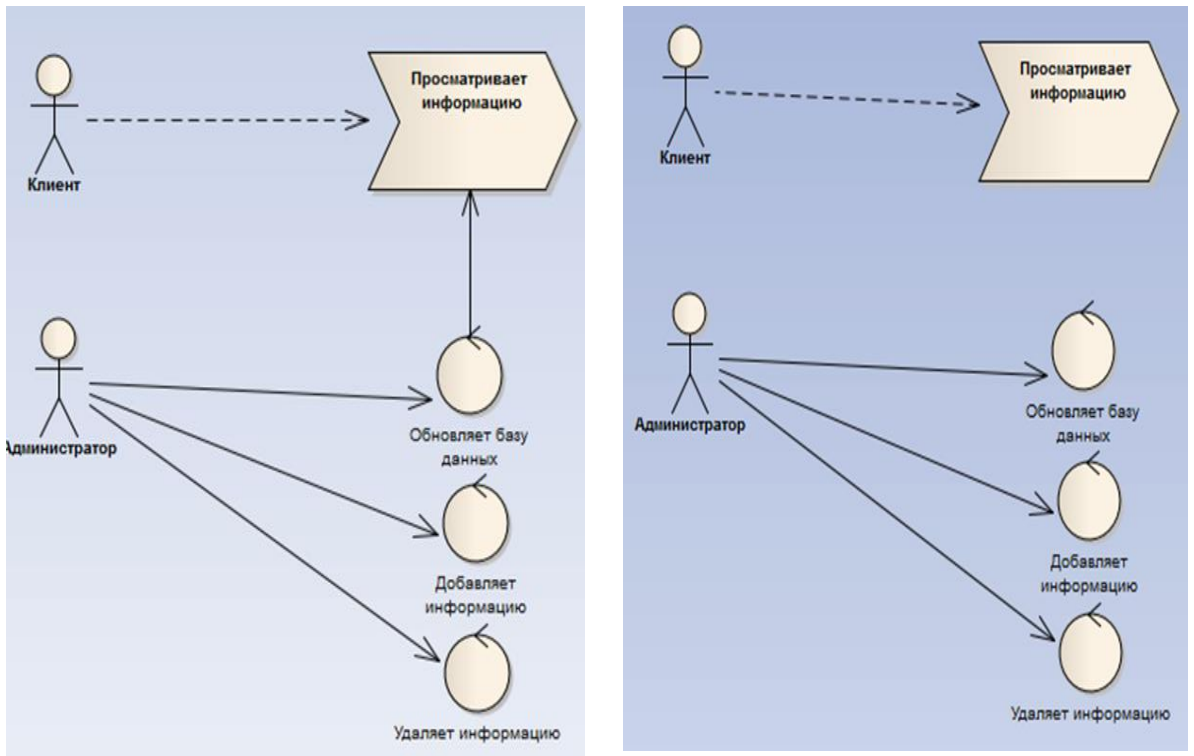


Рисунок 3.1.1 – Модель предметной области и диаграмма вариантов использования

тельностью описывает взаимодействие между объектами, размещая их в хронологическом порядке. Диаграмма кооперации для Администратора представлена на рисунке 3.1.2, Клиента – на рисунке 3.1.3.

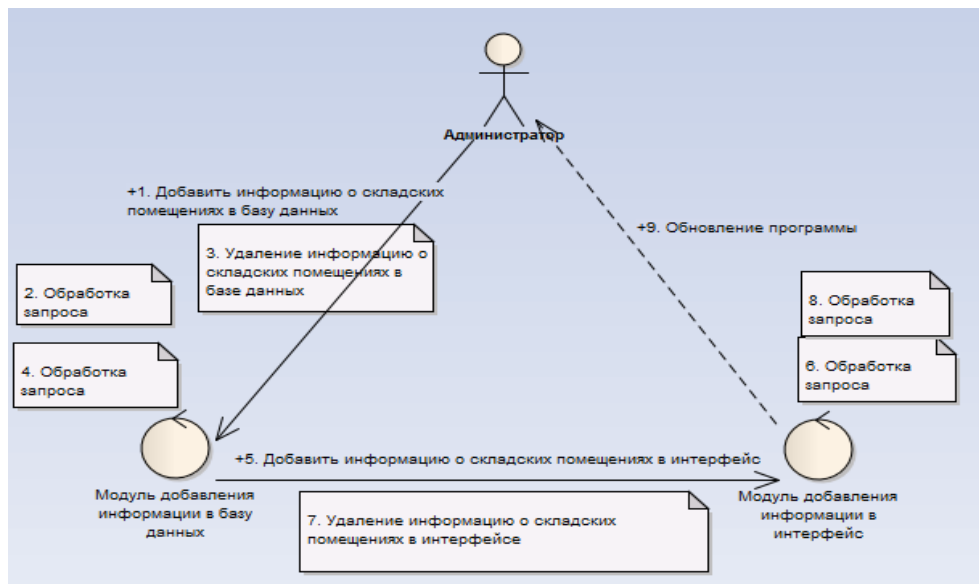


Рисунок 3.1.2 – Диаграмма коопераций для Администратора



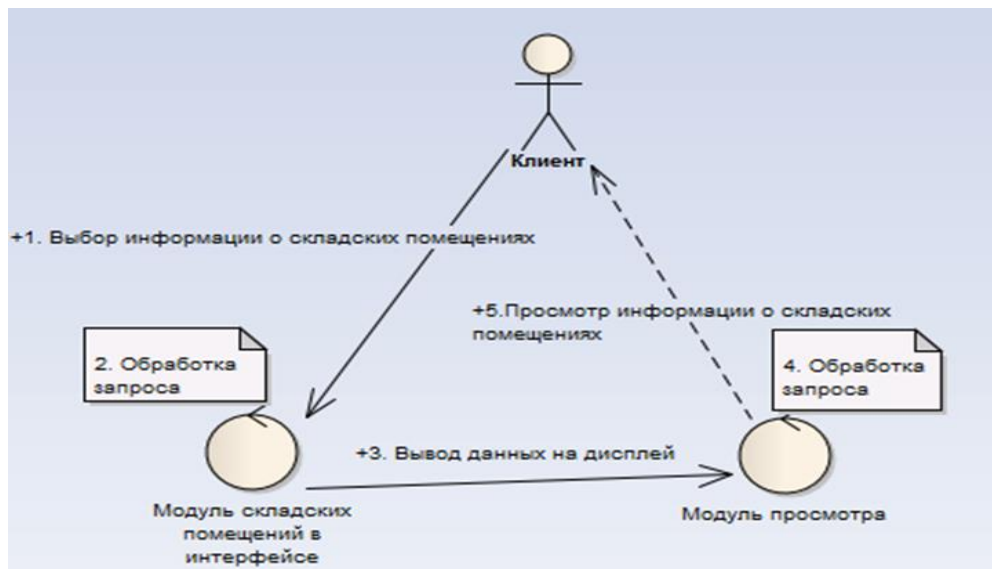


Рисунок 3.1.2 – Диаграмма коопераций для Клиента

Диаграмма последовательностей описывает экземпляр взаимодействий между объектами, размещая их в хронологическом порядке. На диаграмме последовательностей представлен процесс последовательного выполнения действий Администратора (рисунок 3.1.3) и процесс последовательного выполнения действий Клиента (рисунок 3.1.4).

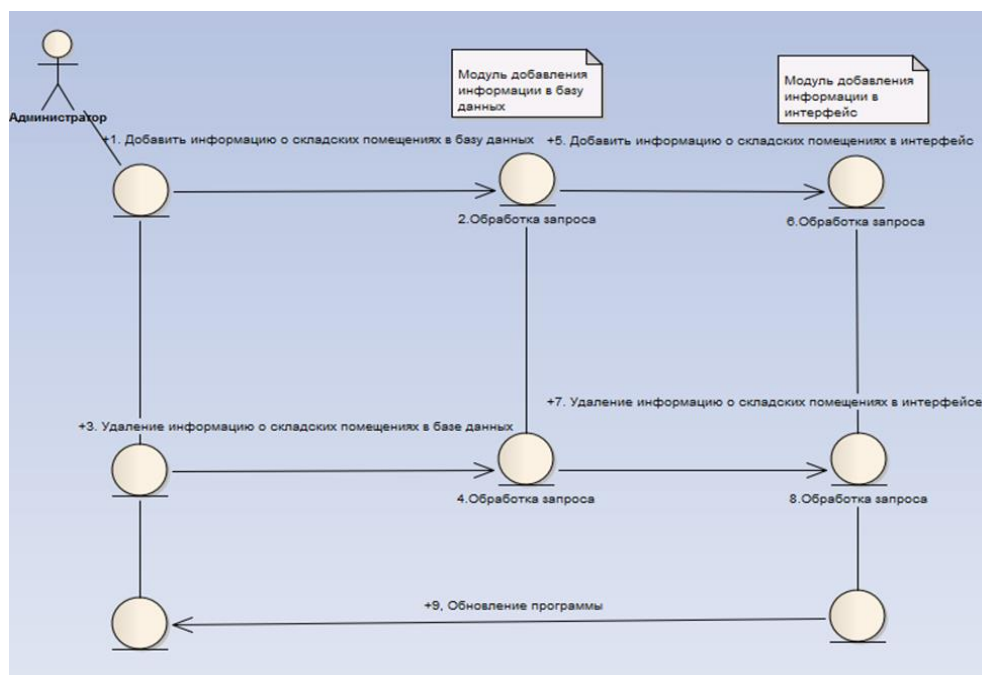


Рисунок 3.1.3 – Диаграмма последовательностей действий Администратора



Рисунок 3.1.4 – Диаграмма последовательностей действий Клиента

В среде Visual Studio 2013 была сформирована диаграмма классов, где отображаются атрибуты классов, операции классов и ограничения, которые накладываются на связи между объектами. Диаграммы классов используются для изучения структур и взаимосвязей приложений и выделения значимых классов. Любые изменения, внесенные в исходном коде, будут отражаться на диаграмме классов, а любое изменение диаграммы классов отражается в коде. На рисунке 3.1.5 изображена типичная диаграмма классов для БД, а на рисунке 3.1.6 диаграмма классов - для интерфейса на языке C#.

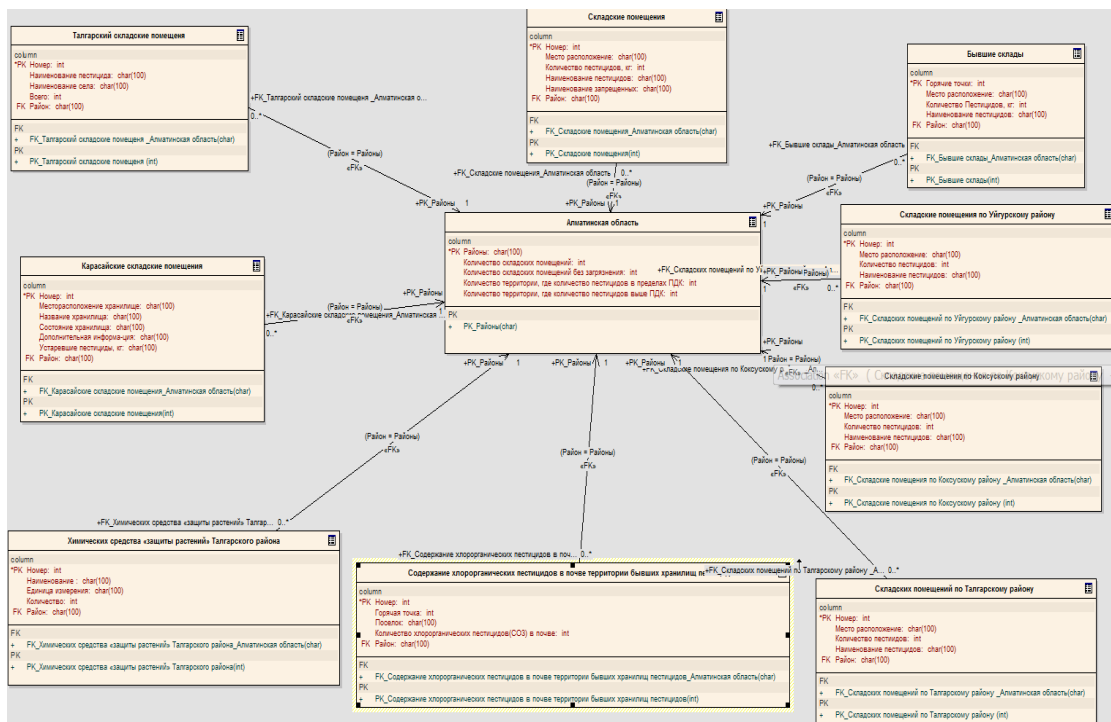


Рисунок 3.1.5– Диаграмма классов для базы данных

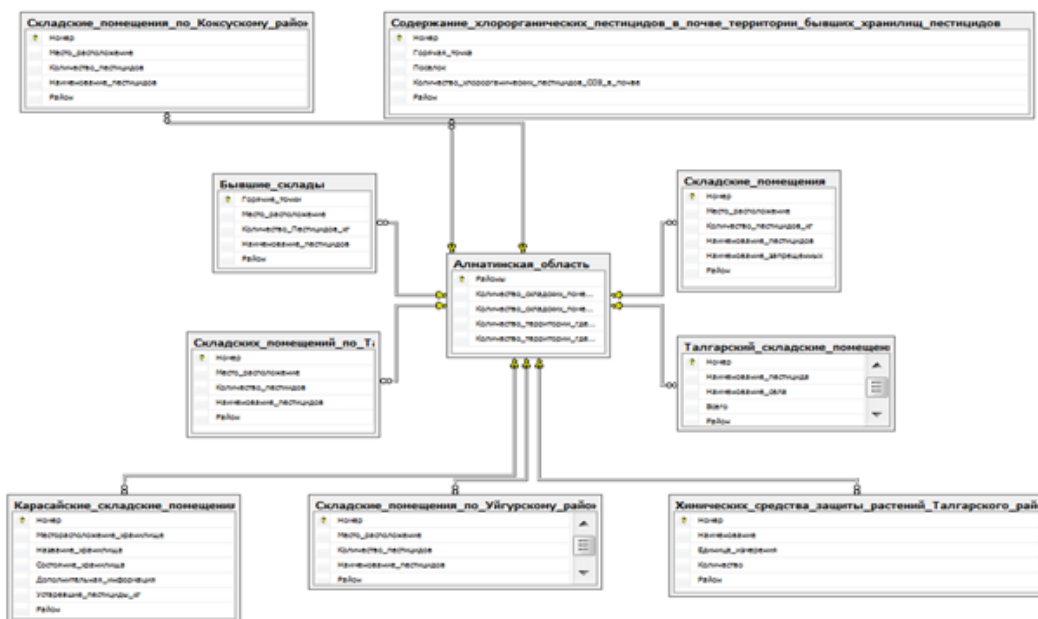
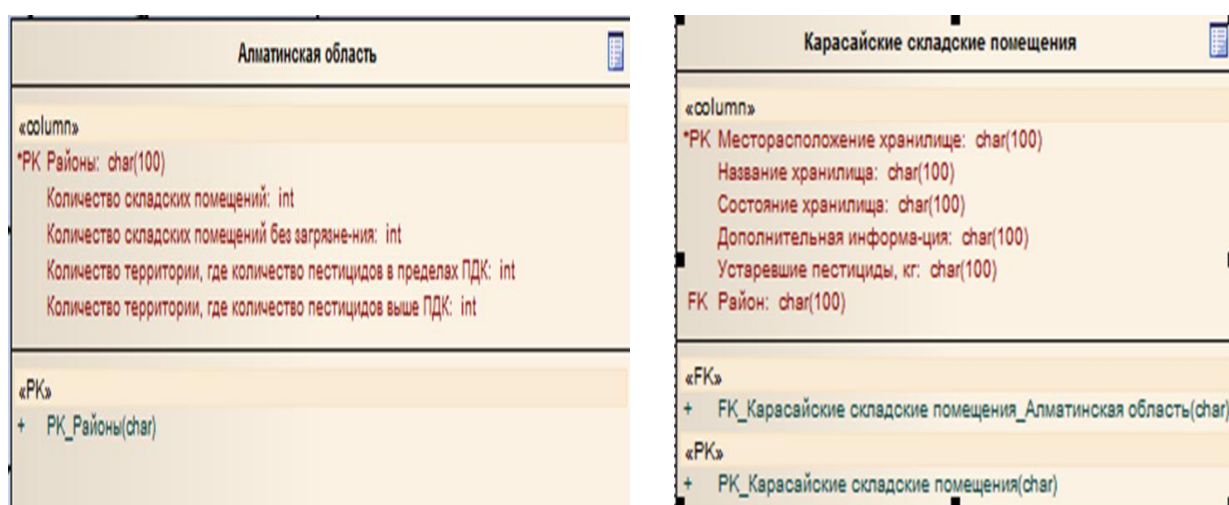


Рисунок 3.1.6 – Диаграмма классов для интерфейса на языке С#

### Разработка приложения

БД о загрязненных пестицидами территориях в разных районах области создавалась в три этапа.

1-ый этап – создание eaUML в программе Enterprise Architect. Применяя CASE средства, такие как eaUML, в программе Enterprise Architect была создана схема, которая содержится в таблице данных. В данной схеме представлена связь между основной и вспомогательной таблицей. Основная таблица содержит БД о районах Алматинской области, а вспомогательные таблицы – данные о загрязненных участках по районам области (рисунок 3.1.7).



Основная таблица

Вспомогательная таблица

Рисунок 3.1.7 – Пример отражения БД в основной и вспомогательной таблицах

В основной таблице в качестве исходных данных использовали код региона, код количества складских помещений, а для вспомогательной таблицы – код региона, код месторасположения склада, код названия склада, код состояния склада, код количества устаревших пестицидов в складе, код уровня загрязнения почвы пестицидами вокруг склада. После создания схемы в Enterprise Architect связи между основной таблицей и вспомогательными таблицами была произведена генерация кода для SQL Server Management Studio, чтобы настроить связанный сервер для использования учетных данных (рисунок 3.1.8).

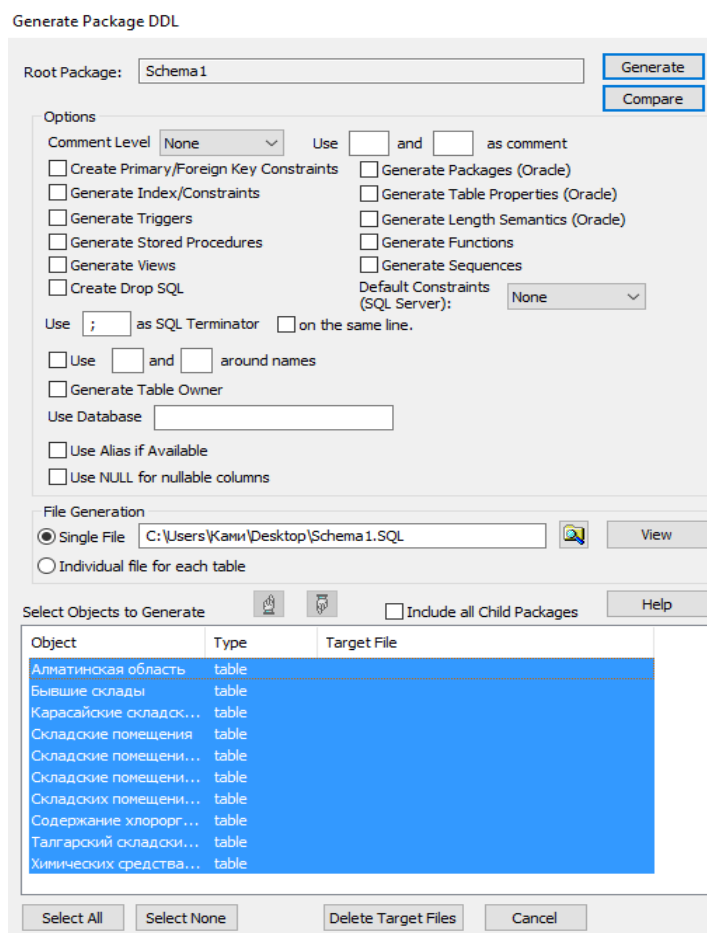


Рисунок 3.1.8– Пример отражения БД при генерации кода в SQL Server Management Studio

2-й этап – создание БД, используя SQL Server Management Studio (рисунки 3.1.9 и 3.1.10).

3-ий этап – создание интерфейса. Используя язык C# в Visual Studio, создали программу в Windows Forms Application. В результате соединили данную программу с программой SQL Server и получили конечную программу (рисунок 3.1.11).

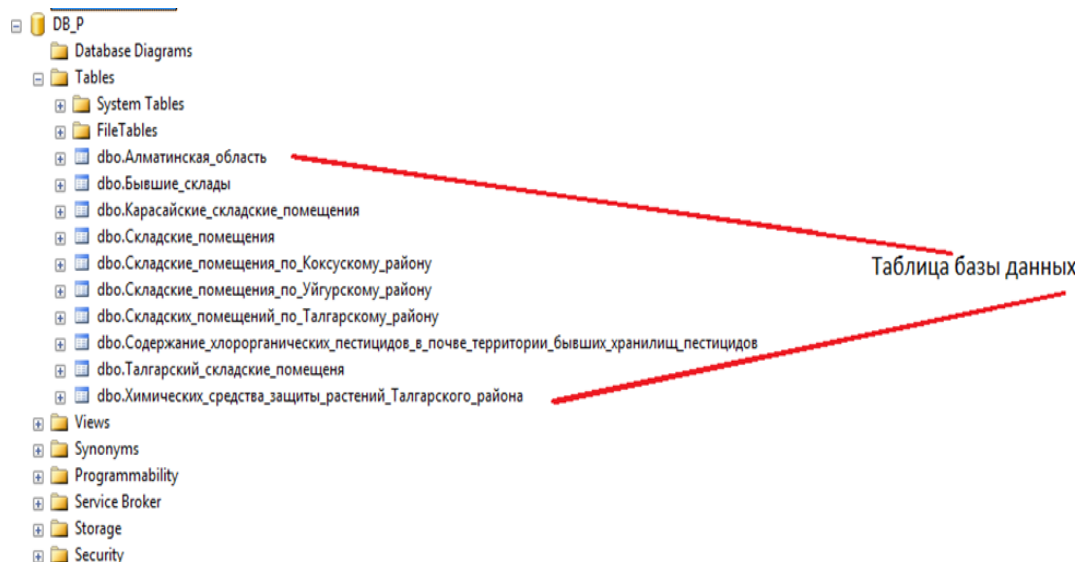


Рисунок 3.1.9 – Пример отражения БД SQL Server Management Studio после генерации кода из Enterprise Architect

Номер	Место_расположение	Количество_п...	Наименование_пестицидов	Район
1	Ассинский с/о ,ПК «Кызыл Шарын» ...	450	Хлорофос, 80% техн. ...	Талгарский ...
2	Ассинский с/о ,ПК «Кызыл Шарын» ...	300	Смесь серы с другими пестицидам...	Талгарский ...
3	Ассинский с/о ,ПК «Кызыл Шарын» ...	500	Неизвестные пестициды	Талгарский ...
4	Октябрьский с/о, СПК «Малы-бай» ...	250	Смесь серы с другими пестицидам...	Талгарский ...
5	Октябрьский с/о, СПК «Малы-бай» ...	300	Неизвестные пестициды	Талгарский ...
6	Массакский с/о, с. Массак	200	Смесь серы с другими пестицидам...	Талгарский ...
7	Каратурукский с/о, Саттарова	350	Неизвестные пестициды	Талгарский ...
8	п. Алмалы, ПК «Коктобе»	1200	Дендробацил-лин	Талгарский ...
9	п. Алмалы, ПК «Коктобе»	720	Тур	Талгарский ...
10	п. Алмалы, ПК «Коктобе»	300	Хлорат магния	Талгарский ...
11	п. Алмалы, ПК «Коктобе»	2160	ТХАН	Талгарский ...
12	п. Алмалы, ПК «Коктобе»	1000	Метальдегид	Талгарский ...
13	С/о Тескенсу, п. Тескенсу	900	Фосулен	Талгарский ...
14	с. Балтабай, ТОО «Балтабай маис»	560	ТХАН	Талгарский ...
15	п. Космос, ТОО АПК «Адал»	484	ТХАН	Талгарский ...
16	Тургень, ЧНФ «Тургень»	800	Меитальдегид	Талгарский ...
17	Ассинский с/о, ПК «Кызыл Шарын» ...	2000	Каунтер	Талгарский ...
18	Каратурукский с/о, п. Саттаров	800	Каунтер	Талгарский ...
19	Массакский с/о, п. Массак	1000	Каунтер	Талгарский ...
20	Тескесуский с/о, п.Тескенсу	2000	Каунтер	Талгарский ...
21	Октябрьский с/о, СПК «Малыбай»	280	Хлорофос, 80% техн. ...	Талгарский ...
22	Каратурукский с/о, П.Саттаров	360	Хлорофос, 80% техн. ...	Талгарский ...

Номер	Местораспол...	Название_хра...	Состояние_хр...	Дополнитель...	Устаревшие_...	Район
1	Ергелинский с...	Бывший хим. с...	Разрушенное ...	Находится в ча...	Симазин - 150 ...	Карасайский ...
2	Ергелинский с...	Бывший хим. с...	Разрушенное ...	Находится в ча...	Смесь 200 ...	Карасайский ...
3	Первомайский...	Бывший хими...	Разрушенное ...	NULL	Энтобактерин ...	Карасайский ...
4	Первомайский...	Бывший хими...	...	Находится в ча...	Севин - 230 ...	Карасайский ...
5	Жандосовский...	Бывший хими...	Разрушенное ...	Находится в ча...	Смесь- 300 ...	Карасайский ...
6	Шамалганский...	Бывший хими...	Разрушенное ...	Находится в ча...	Смесь 120 ...	Карасайский ...
7	Ельтайский с/...	Бывший хими...	Разрушенное ...	Находится в ча...	Смесь - 200 ...	Карасайский ...

Рисунок 3.1.10 – Пример внесения БД, на примере складских помещений Талгарского района в программу SQL Server

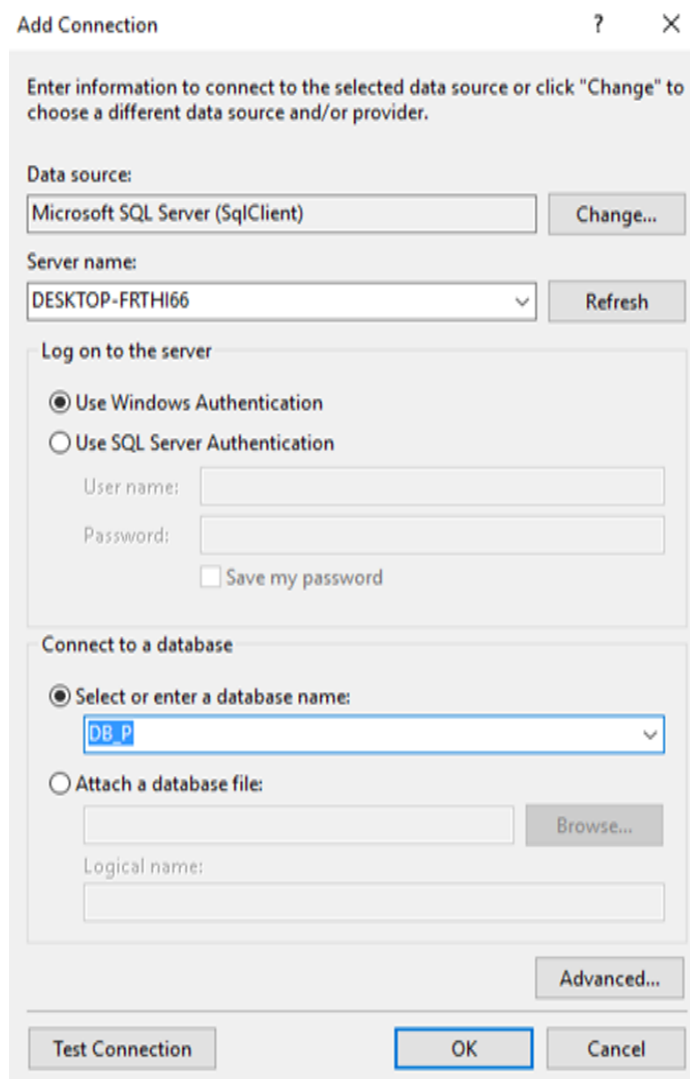


Рисунок 3.1.11 – Пример соединения БД с программой Windows Forms Application

*Тестирование приложения.* Запуск и отладку приложения производили в Visual Studio, а точнее на входящем в состав комплекта средств разработки Visual Studio, а также на реальном компьютере с операционной системой Windows 10. При запуске программы появляется окно программы, в котором можно выбрать нужный район. При нажатии на меню «Файл» появляются окна: «О программе и выход» и «Информация об авторах и версия программы» (рисунок 3.1.12).

Таким образом, созданная программа в среде Enterprise Architect, SQL Server Management Studio с использованием языка C# в Visual Studio, позволяет не только осуществлять сбор и хранение данных, но и получать доступную информацию для решения вопросов охраны окружающей среды.

*Преимущество данной программы:*



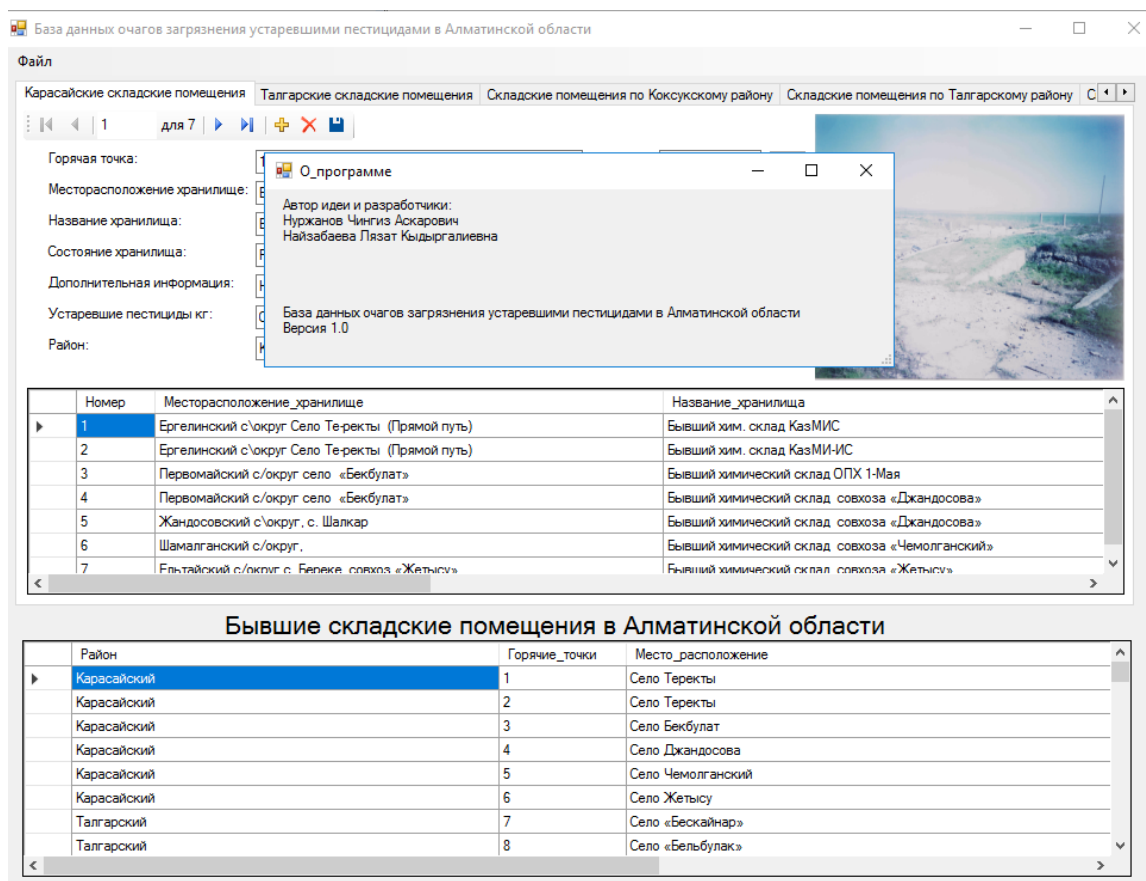


Рисунок 3.1.12 – Окно о программе

возможность учета количества и видов, устаревших и запрещенных пестицидов на территориях недействующих бывших хранилищ химикатов для их утилизации;

возможность оценки уровня риска для экологии и здоровья населения в каждом конкретном регионе области;

база данных может быть использована для решения рекультивационных мероприятий;

данные могут быть использованы при выполнении обязательств Республики Казахстан по Стокгольмской конвенции о СОЗ.

базу данных можно возобновлять новыми данными и детализировать исходные данные об объектах.

Таким образом, созданная база данных содержит данные хроматографического анализа почвы вокруг 76 складов, расположенных в Алматинской области. Химический анализ показал, что почва содержит опасные пестициды ДДТ, ДДД, ДДЭ, концентрации которых превышают казахстанские значения предельно-допустимых концентрации в 65, 19 и 21 раз, особенно на территории склада, расположенных в Талгарском районе поселке Кызылкайрат и посёлке Алдабергена Ескельдинском районе – в 95, 19 и 27 раз. Опасные метаболиты ГХЦГ (альфа, бета и гамма ГХЦГ) в пробах почвы также превышали казахстанские значения предельно-допустимых концентрации в 12, 45 и 4,4 раза соответственно в

поселке Кызылкайрат Талгарского района. Предельно допустимая концентрация в почве Казахстана для отдельных ДДТ и ГХЦГ – 100 мг/кг<sup>1</sup> [26].

Таким образом, территория бывших складов пестицидов представляет собой очаг загрязнения для окружающей среды, человека. При попадании загрязняющих веществ в организм через проглатывание почвы, кожный контакт и вдыхания они представляют канцерогенную опасность для здоровья местного населения [233]. Поэтому мы статистически подсчитали теоретический канцерогенный риск опасных токсичных пестицидов, выявленных в почве вокруг 76 территории бывших складов пестицидов, основываясь на методологию оценки Агентство по охране окружающей среды США и Агентство регистрации токсичных веществ и заболеваний [234].

Теоретический риск развития рака был основан на методологии оценки рисков, предложенный US EPA [235, 236]. Оценку риска рака при проглатывании, кожном контакте и вдыхании учитывали у младенцев в возрасте от 0 до 2 лет, ребенка – от 2 до 16 лет и взрослых – от 16 до 70 лет. Использовали качественное ранжирование оценок риска рака от очень низкого до очень высокого согласно квалификации Агентство регистрации токсичных веществ и заболеваний [237-239]. Качественное ранжирование оценок риска рака согласно Агентству регистрации токсичных веществ и заболеваний: очень низкое ( $\leq 10^{-6}$ ); низкое ( $10^{-6} < \text{величина} \leq 10^{-4}$ ); среднее ( $10^{-4} < \text{величина} \leq 10^{-3}$ ); высокое ( $10^{-3} < \text{величина} \leq 10^{-1}$ ); очень высокое (величина  $\geq 10^{-1}$ ).

Теоретический риск оценивали, используя следующие уравнения:

$$CR_{ingest} = \frac{C_{soil} * IngR * EF * ED}{BW * AT} * CD * SF_{oral} \quad (3.1)$$

где  $CR_{ingest}$  – риск развития рака при случайном проглатывании почвы,  $C_{soil}$  – концентрация загрязняющего вещества в почве (мг кг<sup>1</sup>),  $IngR$  – уровень проглатывания зараженной почвы, мг день<sup>1</sup> (взрослые: 100 мг день<sup>1</sup>, ребенок: 200 мг день<sup>1</sup>),  $EF$  – частота воздействия (дни в год<sup>1</sup>),  $ED$  – продолжительность воздействия (год: взрослые: 70; дети: 14),  $CF$  – коэффициент трансформации (1·10<sup>6</sup> кг мг<sup>1</sup>),  $SF$  – коэффициент наклона полости рта (мг кг<sup>1</sup> d<sup>1</sup>) 1),  $BW$  – масса тела (кг: взрослые: 70 кг; дети: 40 кг) и  $AT$  – усредненное время (дни: 25 500).

$$CR_{dermal} = \frac{C_{soil} * SA * AF_{soil} * ABS * EF * ED}{BW * AT} * CD * SF_{oral} * GIABS \quad (3.2)$$

где  $CR_{dermal}$  – риск развития рака при кожном контакте с почвой,  $SA$  – площадь поверхности кожи, соприкасающейся с почвой (см<sup>2</sup>; взрослые: 3300 см<sup>2</sup>; дети: 2800 см<sup>2</sup>),  $AF_{soil}$  – коэффициент сцепления кожи с почвой (мг см<sup>2</sup>: взрослые: 0,07; ребенок: 0,2),  $ABS$  – коэффициент кожной абсорбции (специфический для



химических веществ: 0,2) и GIABS – фактор желудочно-кишечной абсорбции (безразмерный: 1).

$$CR_{inhale} = \frac{C_{soil} * InhR * AF_{inh} * EF * ED}{PEF * BW * AT} * IUR \quad (3.3)$$

где  $CR_{inhale}$  – риск развития рака при вдыхании почвы,  $InhR$  – скорость вдыхания ( $m^3 \text{дни}^{-1}$ : взрослые: 20; ребенок: 13 лет),  $AF_{inh}$  – коэффициент абсорбции для легких (безразмерный),  $IUR$  – ингаляционная единица риска ( $mg \text{ m}^{-3}$ ),  $PEF$ : вдыхание загрязняющих веществ ( $1,36 \cdot 10^9 \text{ m}^3 \text{ кг}^{-1}$ ) [233].

Данные расчета суммарного риск развития рака при случайном проглатывании почвы, кожном контакте и вдыхании метаболитов ДДТ и изомеров ГХЦГ представлены в Приложении А, таблице 3.2.

Риск развития рака в основном возникает при проглатывании, в незначительной степени при вдыхании и кожном контакте. В случае ДДТ риск развития рака превышает пороговое значение  $10^6$  в 8 пробах почвы из территории бывшего склада, расположенных Карасайском, Талгарском, Енбекшиказахском и Ескельдинском районах, а в одной пробе из села Кызыл Кайрат Талгарского района выше  $10^5$ . В случае изомеров ГХГ риск развития рака во всех случаях низкий, значения ниже  $10^6$ .

Таким образом, наиболее значительный риск рака может быть связан с проглатыванием почвы, что, скорее всего, является результатом проглатывания мелких частиц пыли, которые широко распространены не только на загрязненных территориях, но и в домашнем хозяйстве. Многие исследователи сообщают, что частицы пыли могут быть потенциальным источником воздействия токсикантов на человека, особенно детей [240, 241]. Риск для здоровья детей от токсикантов в домашней пыли может быть в 100 раз больше, чем у взрослых [240]. Особое внимание нужно уделять детям (Pica-kids), они поглощают до 10 г почвы в день [240].

### **3.2 Создание базы данных, содержащей информацию о древесных и травянистых видах растений, способных к восстановлению почв.**

Создание базы данных, содержащей информацию о древесных и травянистых видах растений, способных к восстановлению почв может служить методологической базой при создании казахстанской единой базой о растениях фиторемедиантах (очистителях) и разработке национального плана действий по минимизации экологического риска на все территории Республики.

*Описательная информация.* Информация, представленная в таблице 3.2.1, о 250 различных видах растений казахстанской флоры, способных к очистке почв, загрязненных ТЭ, пестицидами разного класса, нефтеуглеводородами, радионуклидами, взрывчатыми веществами и хлористыми растворителями была собрана из научных литературных данных [242-245].

Таблица 3.2.1 – Описательная информация о количестве видов казахстанской флоры, способных к очистке почв/воды, загрязненных токсичными веществами для БД

Загрязнители	Количество видов, шт			
	Древесные	Многолетние	Однолетние	Технические
Токсические микроэлементы	6	31	10	10
Нефтяные углеводороды	9	18	5	1
Пестициды	5	10	5	8
СОЗ-пестициды	3	7	11	4
Радионуклиды	8	26	3	13
Взрывчатые вещества	0	3	2	3
Хлорированные растворители	4	3	1	2
Аридные зоны	19	0	0	0
Растения с большой биомассой	9	0	0	0
Биоэнергетические виды	2	4	3	2
<b>Всего:</b>	<b>65</b>	<b>102</b>	<b>40</b>	<b>43</b>

*Функциональные возможности включают:*

- русское название вида;
- латинское название вида;
- семейство вида;
- тип вида;
- ботаническое описание вида;
- ареал распространения;
- отношение к загрязнителю;
- литература;
- виды растений – сверхнакопителей мышьяка, цинка, кадмия, никеля и селена;
- виды растений – накопителей мышьяка, цинка, кадмия, никеля и других металлов;
- виды растений – накопителей в корневой системе металлов;
- виды растений – накопителей пестицидов;
- виды растений – накопителей СОЗ-пестицидов;
- виды растений – накопителей нефтеуглеводородов;
- виды растений – накопителей радионуклидов;
- виды растений – накопителей хлористых растворителей;
- виды растений – накопителей взрывчатых веществ;
- виды растений – засухоустойчивых видов для засушливых регионов;

виды растений – биоэнергетических видов с высокой урожайностью биомассы и с высокой скоростью эвапотранспирации;

окно «О приложении»;

окно «Об авторе».

При проектировании БД с помощью CASE-средств Enterprise Architect, также как в предыдущем разделе, составили модель предметной области, диаграмму вариантов использования БД о растениях. Администратор добавляет информацию о растениях в БД, также информацию в интерфейс программного кода приложения. Измененное приложение Администратор обновляет. Клиент устанавливает или обновляет измененное Администратором приложение. Открыв установленное или обновленное приложение, Клиент выбирает нужную информацию, просматривает базу выбранного им вида растений (рисунок 3.2.1).

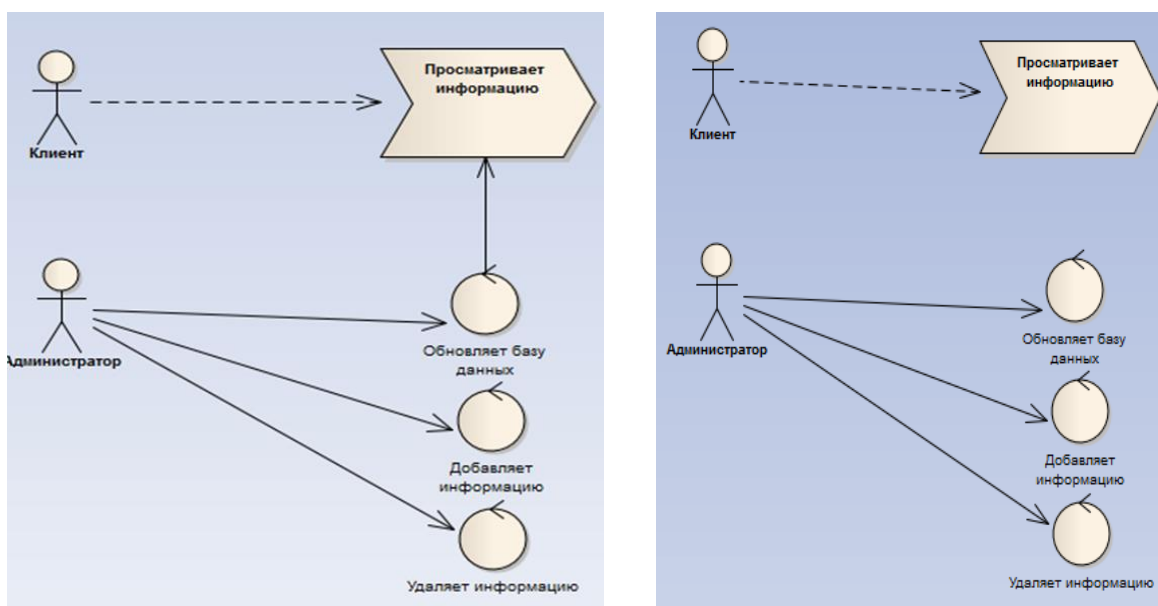


Рисунок 3.2.1 – Модель предметной области и диаграмма вариантов использования БД о растениях

На диаграмме кооперации для Администратора и Клиента стрелки обозначают сообщения и обмен, который осуществляется в рамках данного варианта использования, их временная последовательность – путем нумерации сообщений. Диаграмма коопераций для Администратора и Клиента представлена на рисунке 3.2.2, процесс последовательного выполнения действий Администратора и Клиента – рисунок 3.2.3; диаграмма классов для БД и интерфейса на языке C# – рисунки 3.2.4 и 3.2.5, соответственно.

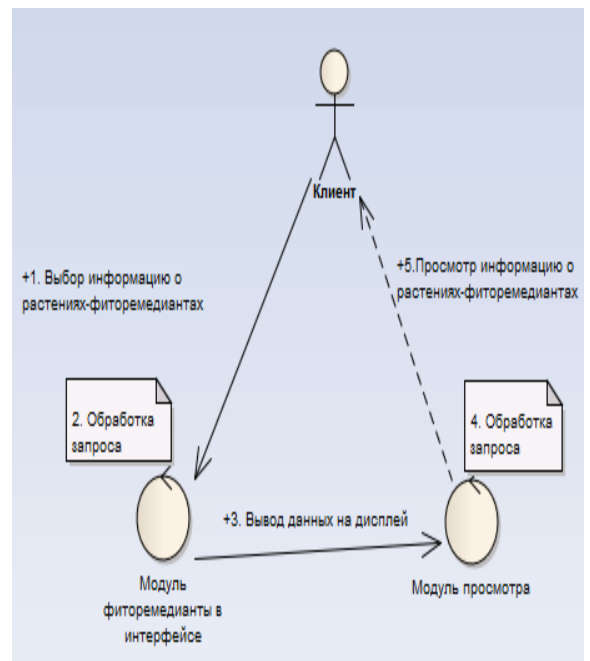
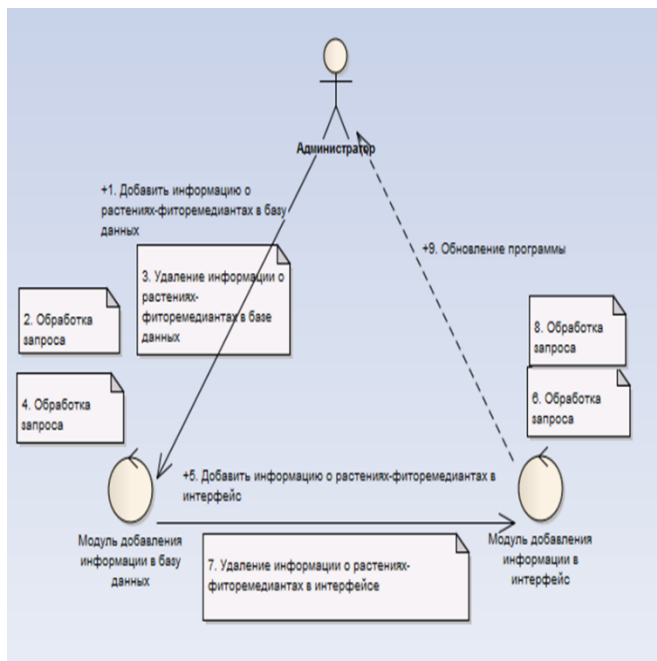


Рисунок 3.2.2 – Диаграмма коопераций для Администратора и Клиента

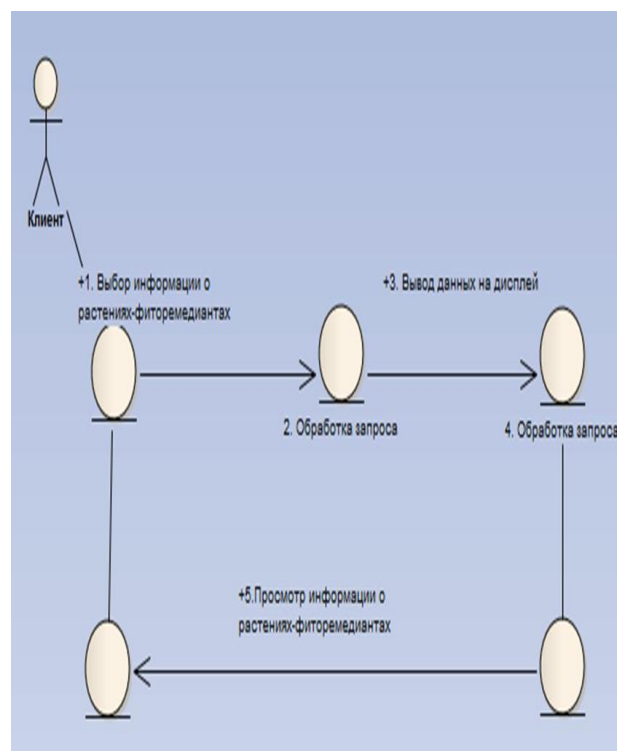
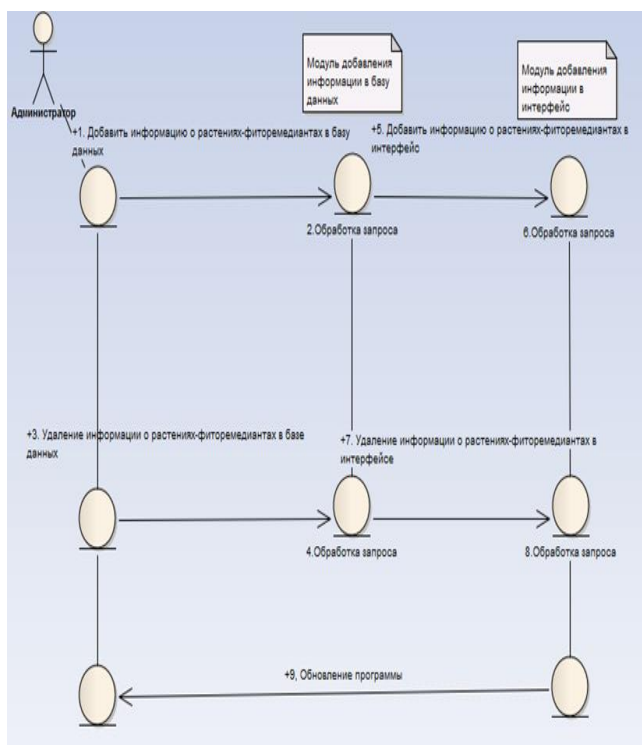


Рисунок 3.2.3 – Диаграмма последовательности действий Администратора и Клиента







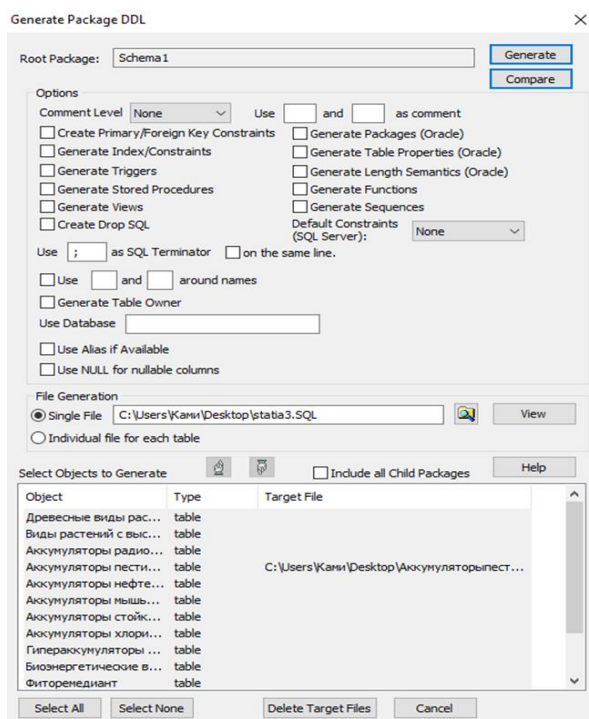


Рисунок 3.2.7 – Пример отражения БД при генерации кода в SQL Server Management Studio

2-й этап – создание БД о растениях сверхнакопителях мышьяка, цинка, кадмия, никеля и селена; растениях накопителях мышьяка, цинка, кадмия, никеля и других металлов; растениях накопителях в корневой системе металлов; растениях накопителях пестицидов; растениях накопителях CO<sub>2</sub>-пестицидов; растениях накопителях нефтеуглеводородов; растениях накопителях радионуклидов; растениях накопителях хлористых растворителей; растениях накопителях взрывчатых веществ и накопителях ТЭ с высокой урожайностью биомассы, с использованием SQL Server Management Studio (рисунки 3.2.8 и 3.2.9).

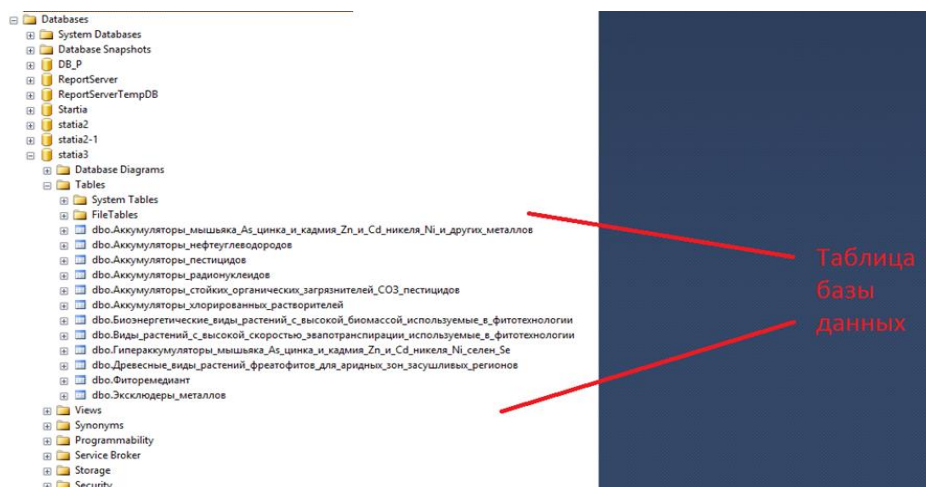


Рисунок 3.2.8 – Пример отражения БД SQL Server Management Studio после генерации кода из Enterprise Architect

Номер	Латинское на...	Русское назе...	Семейство	Тип	Происхожден...	Загрязнители	Литература	Фиторемедиант
1	Morus rubra	Шелковица кр...	Тутовые	Дерево	Северная Аме...	Полихлорбиф...	Olson et al, 200...	Аккумуляторы...
2	Pinus nigra	Сосна черная	Сосновые	Дерево	Европа	Полихлорбиф...	Leigh et al, 200...	Аккумуляторы...
3	Salix caprea	Ива козья	Ивовые	Дерево	Европа Азия	Полихлорбиф...	Leigh et al, 200...	Аккумуляторы...
4	Saxex aquatica	Водоросли	Зигнемовые	Водоросли	Северная Аме...	Полихлорбиф...	Smith et al, 200...	Аккумуляторы...
5	Chrysanthemu...	Нивяник обык...	Астровые	Многолетнее т...	Европа	Полихлорбиф...	Ficko et al, 201...	Аккумуляторы...
6	Festuca arundi...	Овсяница трос...	Злаки	Многолетнее т...	Европа	Полихлорбиф...	Pavlikova et al,...	Аккумуляторы...
7	Lolium multifo...	Плевел много...	Злаки	Многолетнее т...	Европа	ДДЕ	White, 2000	Аккумуляторы...
8	Mentha spicata...	Мята колосист...	Яснотковые	Многолетнее т...	Европа	Полихлорбиф...	Gilbert and Cro...	Аккумуляторы...
9	Rumex crispus	Щавель курча...	Гречишное	Многолетнее т...	Европа	Полихлорбиф...	Ficko et al, 201...	Аккумуляторы...
10	Solidago canad...	Золотарник ка...	Астровые	Многолетнее т...	Северная Аме...	Полихлорбиф...	Ficko et al, 201...	Аккумуляторы...
11	Trifolium repen...	Клевер ползуч...	Мотыльковые	Многолетнее т...	Европа	Полихлорбиф...	McCutcheon a...	Аккумуляторы...
12	Vicia cracca	Горошек мыш...	Бобовые	Однолетнее тр...	Европа	Полихлорбиф...	Ficko et al, 201...	Аккумуляторы...
13	Cucumis sativu...	Огурец посевн...	Тыквенные	Однолетнее тр...	Индия	Диоксины Фур...	Hulster et al, 1...	Аккумуляторы...
14	Cucurbita pepo...	Кабачки	Тыквенные	Однолетнее тр...	Северная и Це...	ДДЕ, ДДТ, ДДД...	Lunney et al, 2...	Аккумуляторы...
15	Cucurbita pepo...	Кабачок Черн...	Тыквенные	Однолетнее тр...	Северная и Це...	ДДЕ, ДДТ, ДДД...	Lunney et al, 2...	Аккумуляторы...
16	Cucurbita pepo...	Кабачки Днам...	Тыквенные	Однолетнее тр...	Северная и Це...	ДДЕ, ДДТ, ДДД...	Lunney et al, 2...	Аккумуляторы...
17	Cucurbita pepo...	Гибрид	Тыквенные	Однолетнее тр...	Северная и Це...	ДДЕ, ДДТ, ДДД...	Lunney et al, 2...	Аккумуляторы...
18	Cucurbita pepo...	Тыква	Тыквенные	Однолетнее тр...	Северная и Це...	ДДТ, Полихл...	Kelsey et al, 20...	Аккумуляторы...
19	Daucus carota	Морковь дика...	Зонтичные	Однолетнее тр...	Северная Аме...	Полихлорбиф...	Ficko et al, 201...	Аккумуляторы...
20	Glycine max	Соя	Бобовые	Однолетнее тр...	Азия	Полихлорбиф...	McCutcheon a...	Аккумуляторы...
21	Lagenaria sicer...	Лагенария об...	Тыквенные	Однолетнее тр...	Азия, Африка	Гептахлор Гепт...	Campbell et al...	Аккумуляторы...
22	Medicago sativ...	Люцерна	Бобовые	Однолетнее тр...	Средиземном...	ДДЕ	White, 2000	Аккумуляторы...
23	Polygonum per...	Горчак почечу...	Гречишное	Однолетнее тр...	Европа	Полихлорбиф...	Ficko et al, 201...	Аккумуляторы...
24	Sesamum indic...	Кунжут	Педалиевые	Однолетнее тр...	Африка	Линдан	Abhilash and Si...	Аккумуляторы...
25	Solanum torvum...	Паслен торву...	Пасленовые	Однолетнее тр...	Северная и Це...	Линдан	Abhilash et al...	Аккумуляторы...
26	Artemisia annu...	Польнь обык...	Астровые	Однолетнее тр...	Южная и Юго...	Хлорорганиче...	Нуржанова, Ка...	Аккумуляторы...
27	Ambrosia arte...	Амброзия пол...	Астровые	Однолетнее тр...	Северная Аме...	Хлорорганиче...	Нуржанова, Ка...	Аккумуляторы...
28	Amarantus retr...	Ширница запр...	Амарантовые	Однолетнее тр...	Мексико	Хлорорганиче...	Нуржанова, Ка...	Аккумуляторы...
29	Xathium strum...	Дурнишник об...	Астровые	Однолетнее тр...	Европа, Средн...	Хлорорганиче...	Нуржанова, Ка...	Аккумуляторы...
30	Helianthus ann...	Подсолнечник...	Астровые	Однолетнее тр...	Северная и Ю...	Хлорорганиче...	Нуржанова, Ка...	Аккумуляторы...
*	NULL	NULL	NULL	NULL	NULL	NULL	NULL	NULL

Рисунок 3.2.9 – Пример внесения БД на примере списка растений-аккумуляторов пестицидов в программу SQL Server

3-ий этап – создание интерфейса. Для создания интерфейса, используя язык C# в Visual Studio, создали программу в Windows Forms Application (рисунок 3.2.10).

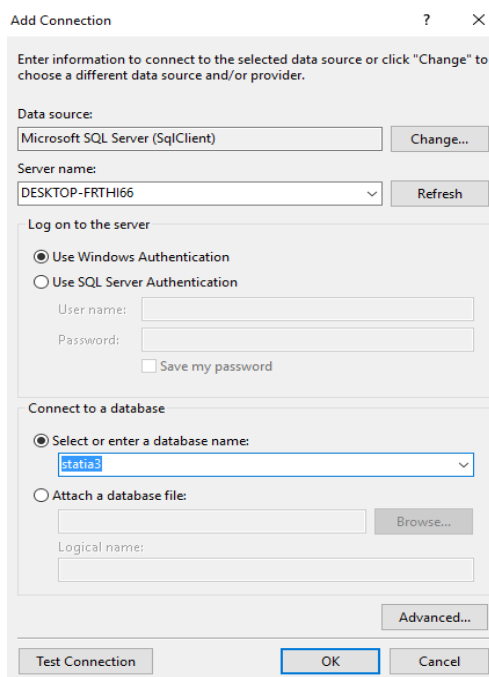


Рисунок 3.2.10 – Пример соединения БД с программой Windows Forms Application

Далее программу Windows Forms Application объединили с программой SQL Server получили конечную программу (рисунок 3.2.11).



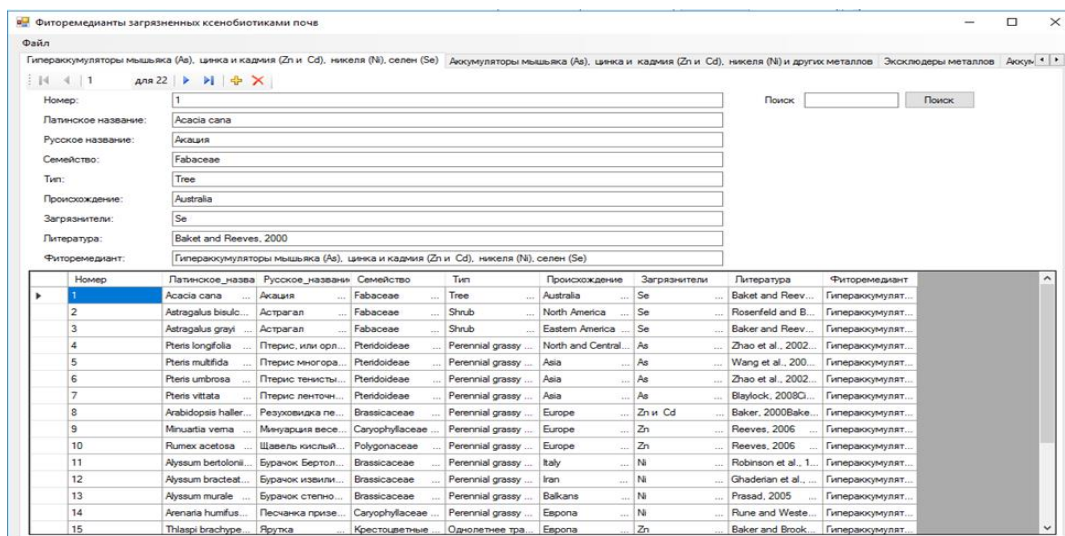


Рисунок 3.2.11 – Пример отражения программы «Фиторемедианты загрязненных ксенобиотиками почв»

*Тестирование программы.* Запуск и отладка приложения производятся на Visual Studio 2013, входящей в состав комплект средств разработки Visual Studio, а также на реальном компьютере с операционной системой Windows 10. При тестировании приложения необходимо запустить программу, после запуска появляется окно программы, в котором можно выбрать нужный район. При нажатии на меню Файл появляются окна «Окно программы», «О программе и выход», и информация «Об авторах и версии программы» (рисунок 3.2.12).

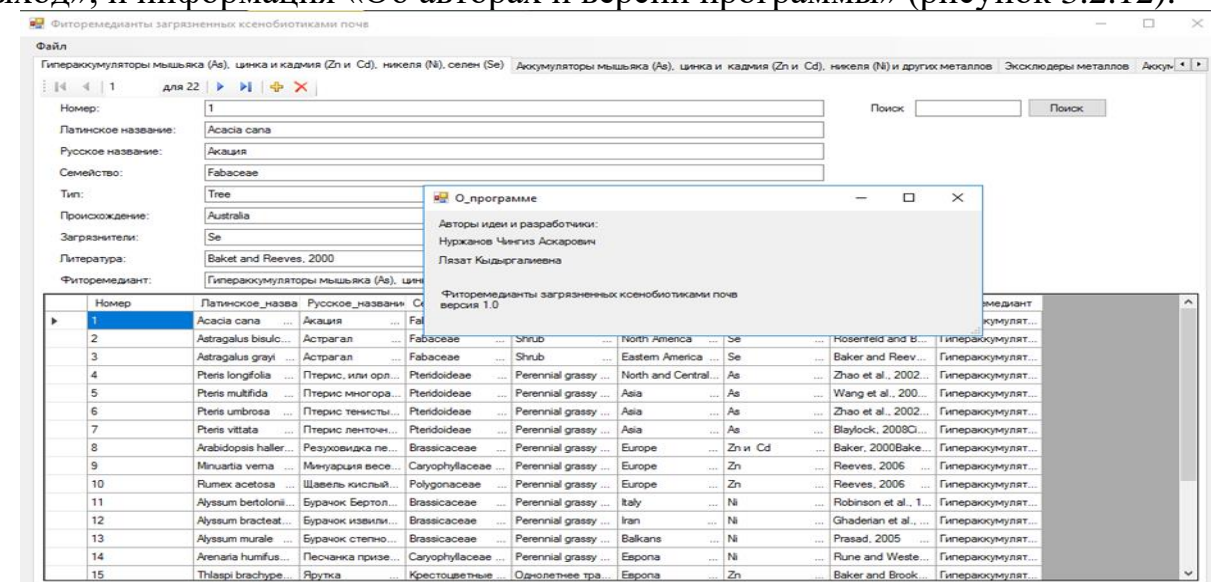


Рисунок 3.2.12 – Окно о программе

Таким образом, созданная БД содержит список растений казахстанской флоры, которые могут быть использованы для рекультивации загрязненных земель/воды органическими и неорганическими токсическими веществами

#### *Преимущество БД:*

- представлена полная информация о растениях локальной флоры, обладающих способностью рекультивировать загрязненную токсическими веществами почву/воду;
- сокращает этапы и время работы при разработке технологии рекультивации техногенно-загрязненных земель/воды с помощью растений;
- может быть использована специалистами, занимающимися садово-парковым искусством и ландшафтной архитектурой;
- может быть использована при проектировании озеленения населенных пунктов с учетом эстетических и экологических качеств;
- может служить методологической основой при создании казахстанской единой базы при разработке национального плана действий по минимизации экологического риска на всей территории Республики;
- базу данных можно возобновлять новыми данными и детализировать исходные данные об объектах

### **3.3 Разработка программного приложения для управления специализированным корпоративным хранилищем данных**

Цель данного раздела – создание корпоративного хранилища базы данных для экологического мониторинга, а именно разработка комплекса технических средств, программно-методического обеспечения для обработки и отображения экологической обстановки [246, 247].

В основу концепции экологического мониторинга нами был предложен интегрированный подход – создание систем, выполняющих весь комплекс необходимых операций, включая:

- экспериментальные показатели;
- сбор, передачу, накопление и обработку измерительных данных;
- анализ экологической ситуации;
- поддержка принятия решений по управлению экологической обстановкой;
- распределение результатов мониторинга между пользователями.

#### *Постановка задачи.*

Первоначальной задачей работ при оценке уровня загрязнения окружающей среды токсичными веществами является расчет суммарных показателей состояния основных компонентов окружающей водной, воздушной и почвенной сред, полученных в результате химического анализа в лабораторных условиях.

Суммарный показатель загрязнения в среде подсчитывают по следующей формуле:

$$Z_{c1} = K_c - (n - 1) \quad (3.4.1)$$

где  $K_c$  коэффициент концентрации  $i$ -го химического элемента;  $n$  – число учитываемых элементов с  $K_c > 1$ .

Коэффициент концентрации  $i$ -го химического элемента оценивают по формуле:

$$K_c = \frac{C_i}{C_{\phi i}} \quad (3.4.2)$$

где  $C_i$  – фактическое содержание  $i$ -го химического элемента в почвах и грунтах, мг/кг;  $C_{\phi i}$  – фоновое содержание  $i$ -го химического элемента (контроль) в почвах, мг/кг.

В зависимости от величины суммарного показателя токсичных элементов в среде, их классифицируют на:

*допустимую* – жизнедеятельность экосистемы с незначительными изменениями;

*опасную* – нарушение жизнедеятельности экосистемы с возрастающим числом обратимых изменений;

*критическую* – отрицательное изменение состояния и структуры экосистемы;

*катастрофическую* – выпадение отдельных звеньев экосистемы либо полное их разрушение.

Нормирование выбросов предприятиями загрязняющих веществ должно производиться на принципах:

минимизации ущерба, наносимого окружающей среде, в сочетании с одновременным обеспечением бесперебойного функционирования предприятия;

рассмотрения аспектов возможного влияния выбросов и сбросов загрязняющих веществ во взаимодействии на окружающую среду (ОС);

использования работ по нормированию выбросов и сбросов в качестве инструмента формирования на предприятии бережного отношения к ОС.

Уровень загрязнения среды с помощью ТЭ, концентрации которых превышают предельно допустимую концентрацию (ПДК) определяют:

$$d_{i\Pi} = \frac{C_{i\Pi}^{cp}}{ПДК_{i\Pi}} \quad (3.4.3)$$

$$d_{iA} = \frac{C_{iA}^{cp}}{ПДК_{iA}} \quad (3.4.4)$$

где  $d_{i\Pi}$ ,  $d_{iA}$  – уровень загрязнения  $i$ -ым ЗВ, соответственно, почв и воздуха;  $ПДК_{i\Pi}$ ,  $ПДК_{iA}$  – предельно допустимая концентрация  $i$ -го ЗВ, соответственно, в почвах (мг/кг) и воздухе (мг/м<sup>3</sup>).

После определения уровней загрязнения ЗВ компонентов окружающей среды рассчитывают превышение их (уровней) над ПДК, по формулам:

$$\Delta d_{i\Pi} = d_{i\Pi} - 1 \quad (3.4.5)$$

$$\Delta d_{iA} = d_{iA} - 1 \quad (3.4.6)$$

где  $\Delta d_{iП}$ ,  $\Delta d_{ia}$  - превышение уровня загрязнения  $i$ -ым ЗВ предельно допустимой концентрации того же вещества, соответственно, в почве и воздухе.

Вычисление суммарного уровня загрязнения компонентов окружающей среды с учетом коэффициентов изоэффективности осуществляют по формулам:

$$d_{П} = 1 + \sum_{i=1}^n \alpha_i d_{iП} \quad (3.4.7)$$

$$d_a = 1 + \sum_{i=1}^n \alpha_i d_{ia} \quad (3.4.8)$$

где  $\alpha_i$  – коэффициент изоэффективности для  $i$ -го ЗВ, равный: для 1-го класса опасности 1,0; для 2-го класса опасности 0,5; для 3-го класса опасности 0,3; для четвертого класса опасности 0,25;  $n$  – число определяемых ЗВ.

*Расчет объемов образования отходов производства.* В общем случае при нормировании в качестве исходной величины принимается количество отходов производства, предусмотренное проектной документацией для конкретного предприятия, при несовпадении реальной производительности предприятия с проектной мощностью объемы образования отходов должны корректироваться по формуле:

$$M_{обр} = M_{пр} \frac{П_{ф}}{П_{пр}} K_{конс} \quad (3.4.9)$$

где,  $M_{обр}$  – объем образования отходов, т/год;  $M_{пр}$  – проектный объем образования отходов т/год;  $П_{ф}$  – реальная производительность предприятия, т/год;  $П_{пр}$  – проектная производительность предприятия, т/год;  $K_{конс}$  – коэффициент консервации отходов производства.

#### *Структура информационной системы*

Эффективное хранение и использование различных типов данных определяет правила поведения пространственно взаимосвязанных географических объектов и классов объектов, а также правила топологии [47].

Полная диаграмма функциональной иерархии экологической системы представлена на рисунке 3.4.1.



Рисунок 3.3.1 – Диаграмма функциональной иерархии экологической системы

Одним из первых шагов управления данными – создание организованного хранилища данных для быстрого доступа необходимой информации. При этом информация должна быть получена с минимальными затратами денежных средств, трудовых ресурсов и с максимальными показателями надежности, точности и объективности.

*Программная реализация системы, основные функции клиентской и серверной частей.*

Разработку информационной системы по клиент-серверной технологии разделили на два этапа: создание серверной части, т. е. базы данных и создание клиентской части. Разработка серверной части заключалась в создании таблиц реляционных систем управления базами данных, классов и методов объектных СУБД. Разработка клиента – создание пользовательского интерфейса к базе данных. БД содержит объекты, в которых хранится первичная и дополнительная служебная информация, необходимая для поддержания нормальной работы других подсистем, а также набор программных модулей бизнес-логики, отвечающий за выполнение необходимых операций над этой информацией. Пользователи напрямую с данной подсистемой не могут взаимодействовать. Информационная система осуществляет анализ в режиме реального времени, прогнозирует процессы и формирует оптимальные варианты решения для улучшения экологической ситуации.

Для проектирования информационной системы использовали систему Erwin исходя из:

создания наглядной модели БД, что позволяет оптимизировать структуру БД и добиться ее полного соответствия требованиям и задачам организации;

инструмент может на основе выявленных информационных потребностей пользователей описать объекты реляционной БД и взаимосвязи между ними, после чего автоматически сгенерировать структуру базы данных;

с помощью ERWin можно провести процедуру реинжиниринга (обратного проектирования), которая автоматически создает метаинформацию о схеме и объектах любого источника данных. Впоследствии эта метаинформация может использоваться при определении процессов определения метаслоя данных, необходимого для генерации произвольных запросов и отчетов к базам данных.

Разработанная информационная система предназначена для решения следующих задач:

1) Расчет, формирование платежей и отчетности выбросов в атмосферу для стационарных и передвижных источников загрязнения:

расчет ежегодных запрашиваемых лимитов выбросов загрязняющих веществ в атмосферу согласно производственным планам;

расчет ежеквартальных данных по выбросам;

формирование платежей за загрязнение атмосферы на основе рассчитанных данных;

формирование статистической отчетности по выбросам за любой период и для любого уровня организационно-структурной иерархии подразделений компании;

2) Информация об образовании и хранении отходов, формирование платежей и отчетности:

ввод информации об образовании и утилизации отходов;

расчет платежей за хранение (размещение) отходов;

формирование отчетности об образовании, утилизации и хранении отходов за любой период и для любого уровня организационно-структурной иерархии подразделений компании;

3) Рекультивация почв:

ввод информации об участках, подвергшиеся рекультивации;

формирование отчетности по рекультивации почв за любой период и для любого уровня организационно-структурной иерархии подразделений компании.

Для наглядного проектирования и представления структуры базы данных использовали концептуальную модель «Сущность – Связь» [248].

По своим функциональным возможностям информационная система включает три подсистемы: подсистему учета выбросов в атмосферу; подсистему учета образования и хранения отходов; подсистему учета рекультивации почв. Диаграмма базы данных экологической системы состоит из 68 таблиц. Подсистемы учета рекультивации почв представлена на рисунке 3.4.2, а на рисунке 3.4.3 – подсистемы учета образования и хранения отходов.

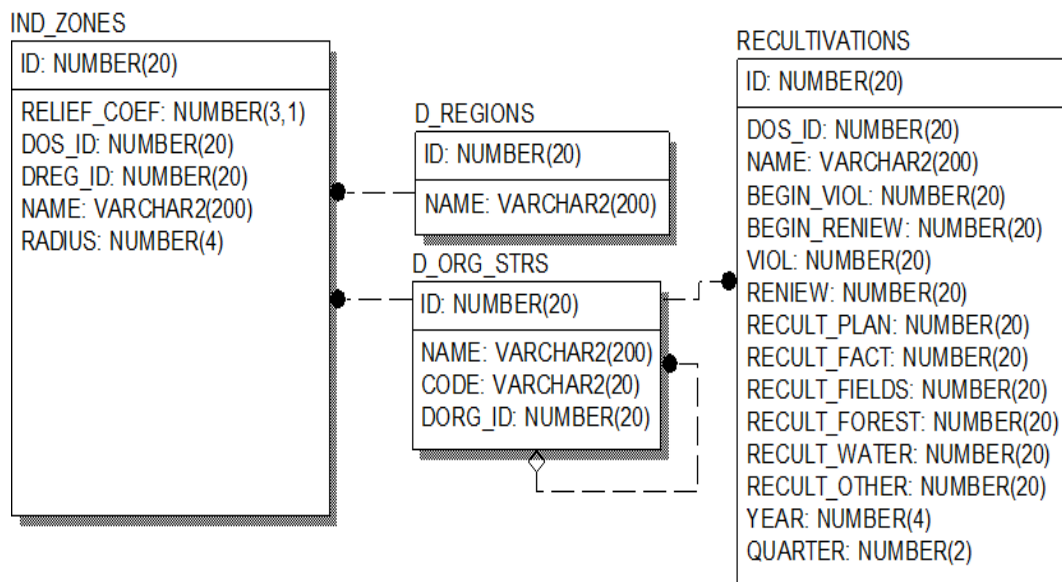


Рисунок 3.4.2 – Диаграмма подсистемы учета рекультивации почв

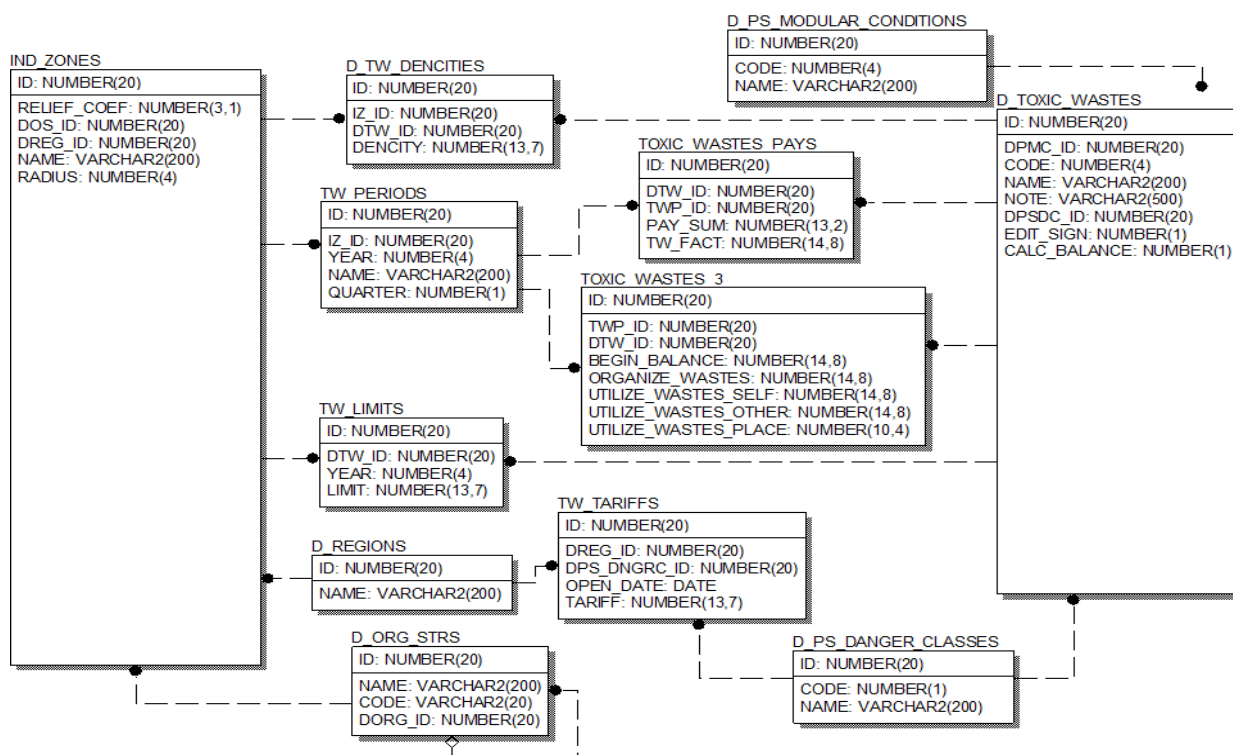


Рисунок 3.3.3 – Диаграмма подсистемы учета образования и хранения ОТХОДОВ

Техническая архитектура интегрированной экологической системы представлена на рисунке 3.4.4

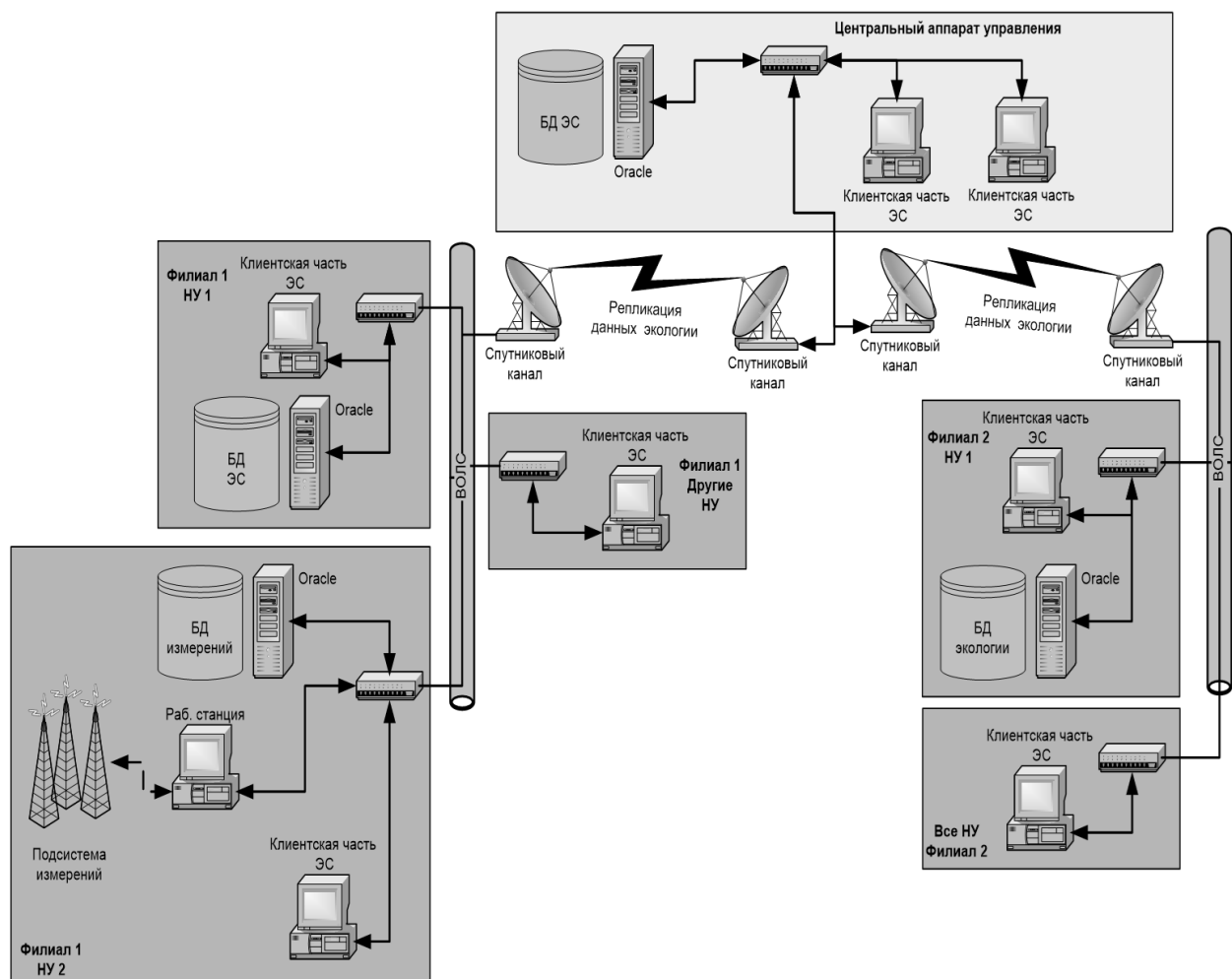


Рисунок 3.3.4 – Техническая архитектура интегрированной экологической системы

Таким образом, разработанная структура базы данных содержит информацию о географическом местоположении территории бывших хранилищ пестицидов, о количестве и идентификации, уровне загрязнения почв устаревшими пестицидами, а также информацию о древесных и травянистых видах растений, способных к очистке почвы/воды различными токсическими элементами. Эта информация позволяет найти оптимальные способы решения проблем очистки и утилизации загрязненной территории, главное, сделать правильный выбор стратегии восстановления загрязненных почв, кроме того, она может быть включена в автоматизированную систему ведения государственного земельного кадастра и автоматизированную систему государственного учета объектов недвижимости, ответственных за проведение рекультивации нарушенных земель, путем электронного обмена сведениями для передачи поступившей информации органам по управлению земельными ресурсами, кадастровому учету, оценке, регистрации и налоговым органам. Информационная система части экологического мониторинга отвечает за прием, накопление, обработку и анализ, а также распределение данных.



### 3.4 Выводы по разделу, постановка задачи

Обобщая результаты данного раздела, следует заметить, что на сегодняшний день в стране нет программ, предназначенных для оценки последствий воздействий отходов на окружающую среду и БД, предназначенных для рекультивации почв, загрязненных различными ТЭ.

База данных создана на основе сведений, поступивших из различных научных источников о 76 единицах территорий, загрязненных устаревшими пестицидами и 250 различных видах растений казахстанской флоры, способных к очистке почв. Язык программирования: набор программа Enterprise Architect, SQL Server Management Studio, используя язык С# в Visual Studio 2013. БД созданы в три этапа, 1-ый этап – создание eaUML в программе Enterprise Architect; 2-й этап – создание базы данных, используя SQL Server Management Studio; 3-ий этап – создание интерфейса, используя язык С# в Visual Studio.

Из-за отсутствия информационной системы для оперативного и долгосрочного мониторинга и прогнозирования состояния окружающей среды, а также предотвращения ее загрязнения в результате негативных воздействий техногенных отходов нами рассмотрен принцип создания информационной системы на базе данных о загрязненных пестицидами территориях и растениях, способных к восстановлению техногенных ландшафтов. Применен подход «Сущность – Связь».

Получены два авторских свидетельства (Приложение Б)

Созданная база данных о загрязненных пестицидами территориях содержит данные хроматографического анализа почвы вокруг 76 складов, расположенных в Алматинской области. Присутствие и возможное использование запрещенных устаревших хлорорганических пестицидов, таких как ДДТ и продукты его разложения ДДД и ДДЭ, а также изомеры ГХЦГ представляют собой источники, откуда эти вещества «доставляются». При расчете теоретического риска развития рака, выявлено, что в 8 образцах канцерогенный риск воздействия ДДТ в течение всей жизни превышал  $10^{-6}$  и в одном участке – риск превысил  $10^{-5}$ . В остальных случаях было установлено, что риск около  $10^{-7}$ . В случае изомеров ГХГ риск развития рака во всех случаях был ниже  $10^{-6}$ . Согласно Агентству регистрации токсичных веществ и заболеваний риска рака, значение выше  $10^{-6}$  представляет канцерогенную опасность.

## Глава 4 Математическая модель поглощения ТЭ растением и перемещение их из почвы в растение в зависимости от условий среды

Для выбора оптимальной стратегии рекультивации загрязненных токсичными элементами почвы с помощью растений разрабатываются различные математические модели и компьютерные программы. Отмечают, что этих исследований недостаточно. Отмечают, что недостаточность модели связана с тем, что поглощение токсических элементов растением и миграции их в почве и грунтовой воде – сложный процесс, в котором участвуют различные физические, биологические и химические факторы [249-251].

### *Общая постановка задачи*

Цель работы состояла в том, чтобы статистически оценить связь и взаимодействие трех факторов (почва, растение и почва-растение) на поглощение ТЭ растением в зависимости от типа почвы: светло-каштановой и песчаной (солончаки) почвы.

В данном разделе приведены результаты создания математической модели поглощения ТЭ растением из загрязненной почвы с применением метода пошаговой регрессии. Для оценки взаимодействия трех факторов (почва, растение и почва-растение) на поглощение ТЭ растением в зависимости от типа почвы использовали дисперсионный многофакторный анализ ANOVA [252-254].

### *Математическая модель*

Исходя из полученных экспериментальных данных, была построена математическая модель, в основу которой легло дифференциальное уравнение. Математическое описание процессов сводиться к нахождению и исследованию функциональной зависимости, т. е. функции отклика:

$$y = f(x_1, x_2, \dots, x_n) \quad (4.1)$$

Предварительно провели нормировку каждой переменной на диапазон разброса ее значений, чтобы исключить влияние на результаты анализа величин измерения. Нормировку данных провели по следующей формуле:

$$Y = \frac{X_i - \min_i}{\max_i - \min_i} \quad (4.2)$$

где  $X_i$  – исходные данные вектора признака  $i$ -го ( $1 \leq i \leq 12$ );  $Y$  – безразмерные значения от 0 до 1;  $\max$ ,  $\min$  – максимальные и минимальные значения признака.

Среднее значение наблюдаемых данных рассчитывали по формуле:

$$\bar{y} = \frac{1}{n} \sum_{i=1}^n y_i \quad (4.3)$$

где  $n$  – число наблюдений,  $y_i$  – наблюдаемые данные

Сумму квадратов остатков оценивали по формуле:

$$SS_{res} = \sum_i (y_i - f_i)^2 \quad (4.4)$$

где  $f_i$  – предсказанные данные

Сумму квадратов подсчитывали:

$$SS_{tot} = \sum_i (y_i - \bar{y})^2 \quad (4.5)$$

Значение коэффициента детерминации рассчитывали по формуле:

$$R^2 = 1 - \frac{SS_{res}}{SS_{tot}} \quad (4.6)$$

Значение коэффициента Фишера определяли по формуле:

$$F_{\text{факт}} = \frac{\frac{R^2}{k-1}}{\frac{1-R^2}{n-k}} \quad (4.7)$$

где  $k$  – количество параметров в модели,  $n$  – число наблюдений

Результаты нормированных экспериментальных накопления ТЭ в растений представлены на рисунке А 4.1 и А 4.2.

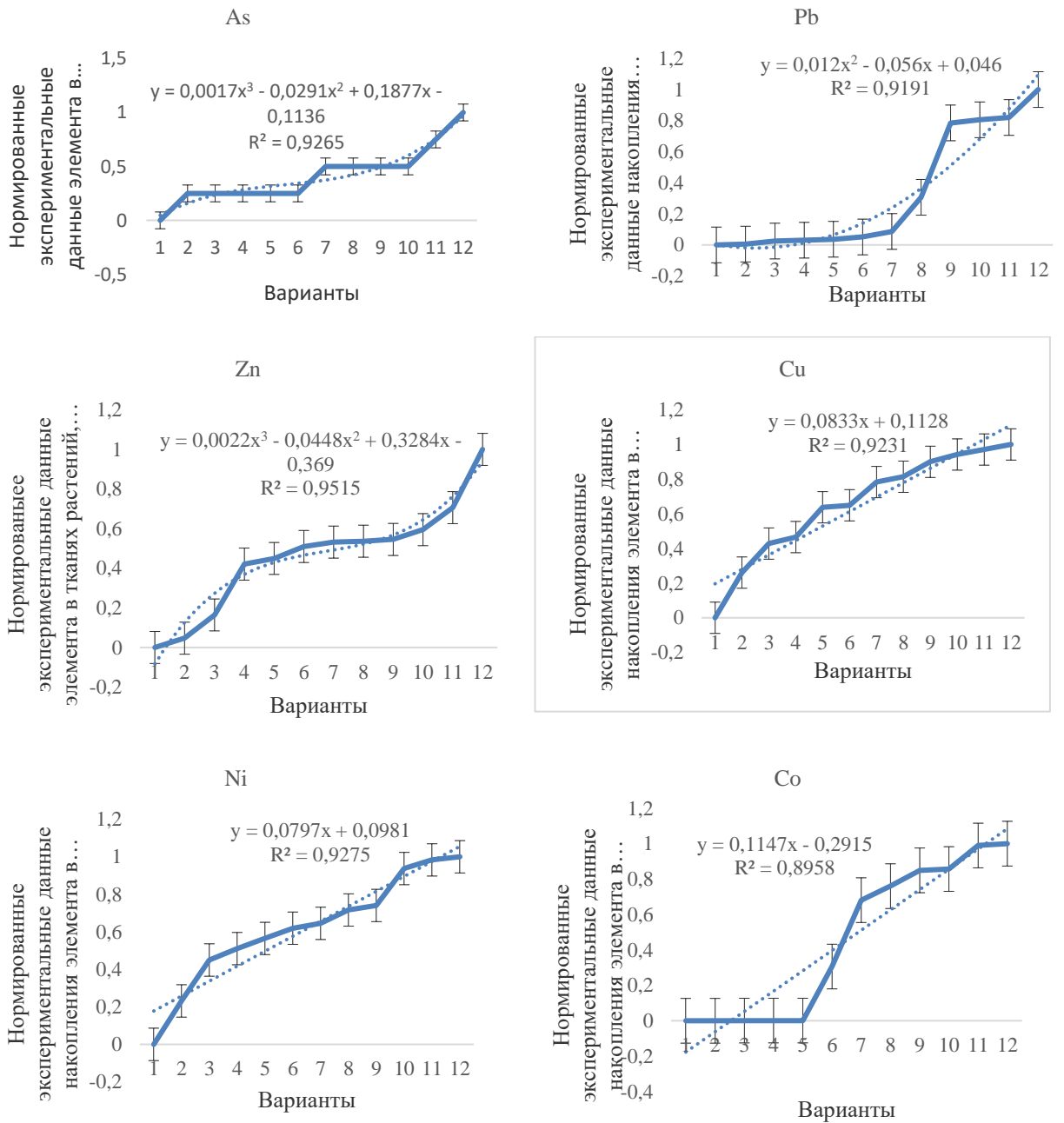


Рисунок А4.1 – Нормированные экспериментальные данные накопления ТЭ 1 и 2 класса опасности растением

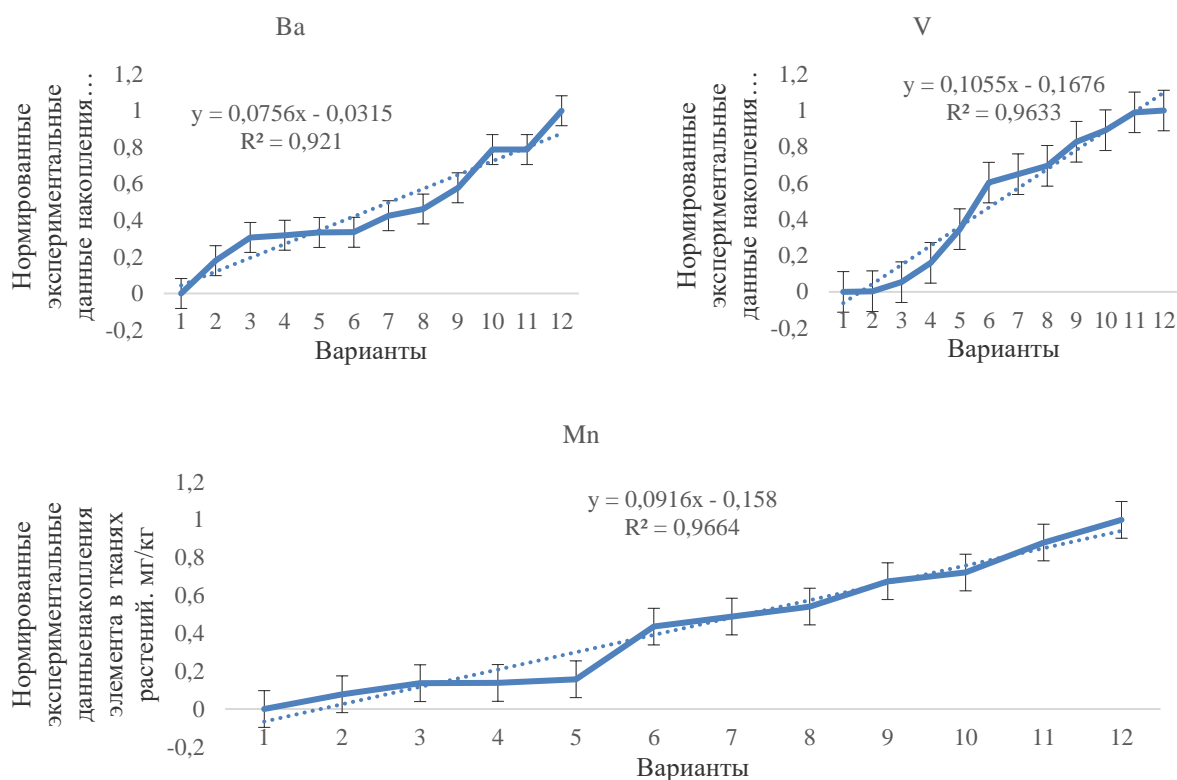


Рисунок 4.2. – Нормированные экспериментальные данные накопления ТЭ 3 класса опасности растением

На основании расчета экспериментальных результатов получили степенное уравнение регрессии поглощения ТЭ растением с учетом величины их достоверности. Значения  $R^2 \geq 1$  свидетельствует о статистической достоверности результатов исследования (таблица 4.1).

Таблица 4.1 – Степенное уравнение регрессии, коэффициент детерминации и значение критерии Фишера

Металлы	Степенное уравнение	$R^2$	$F$
As	$y = 0,0017x^3 - 0,0291x^2 - 0,1817x - 0,1136$	0,9265	49,92
Pb	$y = 0,012x^2 - 0,56x + 0,46$	0,9191	30,64
Zn	$y = 0,0022x^3 - 0,0448x^2 - 0,3284x - 0,369$	0,9515	21,26
Ni	$y = 0,0797x + 0,0981$	0,9275	137,65
Cu	$y = 0,0833x + 0,1128$	0,9231	120,02
Cr	$y = 0,0103x^2 - 0,543x + 0,0798$	0,9669	78,13
Ba	$y = 0,0756x + 0,0315$	0,9210	126,5
Sr	$y = 0,0015x^3 - 0,0276x^2 - 0,2153x - 0,0671$	0,8981	15,45
V	$y = 0,0916x + 0,158$	0,9633	272,57
Mn	$y = 0,0911x + 0,148$	0,9664	72,29

По результатам дисперсионного анализа установлено, что фактор «почва\*растение» оказывает прямое взаимодействие на накопление ТЭ в зависимости от типа почвы (таблица 4.2, таблица 4.3).

Например, растение накапливает ТЭ из каштановой почвы в основном Zn, а из песчаной почвы – As, Pb, Cr ( $p < 0,05-0,001$ ). Статистически значимое различие накопления растением ТЭ из почвы в зависимости типа почвы свидетельствует о том, что на песчаной загрязненной почве спектр поглощения выше, чем у растений, произрастающих на светло-каштановой загрязненной почве. Растение поглощает 10 ТЭ (As, Pb, Cr, Co, Ni, Cu, Mn, V, Sr, U),  $p < 0,05-0,001$ , а при произрастании на светло-каштановой загрязненной почве 8 ТЭ (As, Pb, Cr, Co, Ni, Cu, V, U),  $p < 0,05-0,001$ . Данные результаты свидетельствуют о том, что поглощение ТЭ растением зависит от типа почвы.

#### *ТЭ из почвы в органы растений в зависимости от условий среды*

Для выявления факторов среды, которые влияют на механизм поглощения растением ТЭ разрабатываются различные математические модели, основанные на уравнение баланса массы (закон сохранения массы), которое описывают перенос загрязняющих веществ из почвы в корень и затем надземную часть растительного организма, что позволяет понять причинно-следственные связи между наблюдаемыми параметрами [250,255-257].

#### *Общая постановка задачи*

Цель исследования: разработать математическую модель миграции ТЭ из почвы в органы растений в зависимости от кислотности, влажности почвы, температуры среды, и типа почвы с помощью множественного регрессионного анализа, который позволяет найти наиболее точную и достоверную аппроксимацию стохастической зависимости между откликом  $Y$  и несколькими факторами.

Для регрессионного анализа в качестве результативного показателя ( $y$ ) использовали КБП (коэффициент биологического поглощения) ТЭ (As, Pb, Cd, Zn, Co, Ni, Cu, Cr, Ba, Sr, V, Mn) растением, валовое содержание ТЭ в почве и органах растений (As, Cd, Zn, Co, Ni, Cu, Cr, Ba, Sr, Mn), а в качестве переменных факторов ( $x$ ) – кислотность почвы (рН 8,13-8,94), влажность почвы (влажность почвы 8-11%) и температуру воздуха (21-30°C).

Коэффициент регрессии рассчитывали по следующей формуле:

$$Y = b_0 + b_1X_1 + \dots + b_kX_k \quad (4.8)$$

где  $X_1, \dots, X_k$  – факторы,  $b_0, \dots, b_k$  – коэффициенты регрессии.

На основании расчета получили уравнение регрессии поглощения ТЭ растением из загрязненной почвы, с учетом величины достоверности аппроксимации.

Таблица 4.2 – Влияние типа почвы на поглощение ТЭ растением, при произрастании на каштановой загрязненной почве (почва из территории горно-обогатительного комбината г. Текели)

Факторы	F уровень										
	As	Pb	Zn	Co	Cr	Ni	Cu	Mn	V	Sr	U
Почва	1,10	2,79	3,08	0,47	2,97	3,43	4,05	1,26	<b>47,21***</b>	<b>72,05***</b>	1,57
Растение	<b>74,83***</b>	<b>8,9 *</b>	3,51	<b>57,08***</b>	<b>22,41**</b>	<b>180,47***</b>	<b>65,89***</b>	2,46	<b>76,17***</b>	3,41	<b>168,71***</b>
Почва*Растение	1,10	2,02	<b>7,82*</b>	0,47	2,80	3,43	1,59	1,09	<b>47,21***</b>	3,40	0,59

Примечание: вероятность различия между почвенным и растительным вариантами оценивается по t-критерию Стьюдента: значительные отличия \*  $p < 0,05$ , \*\*  $p < 0,01$ , \*\*\*  $p < 0,001$ , и незначительные экспериментальные значения ( $p > 0,05$ ) не отмечены. Каждое измерение было выполнено в двух повторностях ( $n = 2$ ).

Таблица 4.3 – Влияние типа почвы на поглощение ТЭ растением, при произрастании на песчаной загрязненной почве (почва из территории бывшего военного гарнизона Балхашский район)

Факторы	F уровень										
	As	Pb	Zn	Co	Cr	Ni	Cu	Mn	V	Sr	U
Почва	<b>7,00*</b>	<b>19,94**</b>	2,53	0,29	<b>201,11***</b>	0,48	0,92	2,41	6,69*	1,75	0,00
Растение	<b>38,75***</b>	<b>307,62***</b>	3,75	<b>305,77***</b>	<b>122,74***</b>	<b>596,46***</b>	<b>7,45*</b>	<b>34,13***</b>	<b>1032,08***</b>	<b>9,73*</b>	<b>71,53***</b>
Почва*Растение	<b>6,05*</b>	<b>10,49*</b>	2,24	0,17	<b>122,74***</b>	4,60	2,85	4,01	5,86*	0,68	0,64

Примечание: вероятность различия между почвенным и растительным вариантами оценивается по t-критерию Стьюдента: значительные отличия \*  $p < 0,05$ , \*\*  $p < 0,01$ , \*\*\*  $p < 0,001$ , и незначительные экспериментальные значения ( $p > 0,05$ ) не отмечены. Каждое измерение было выполнено в двух повторностях ( $n = 2$ ).

Цель метода сводилась к установлению коэффициентов, которые отражают степень и направление влияния фактора на результативный признак. Чем выше коэффициент по модулю, тем больше вклад фактора; положительные коэффициенты говорят об усилении поглощения растением токсических элементов под влиянием данного фактора, отрицательные – об ослаблении. Качество полученных регрессионных моделей оценивали с использованием коэффициента детерминации ( $R^2$ ). Для проверки адекватности модели использовали экспериментальные результаты. Предварительно провели корреляционный анализ между КБП и условиями среды. Анализ данных показал корреляционную зависимость миграции ТЭ в системе «почва – растение» от условий среды. Коэффициент корреляции был выше значения 0.85.

#### *Математическая модель*

Результаты расчета коэффициента регрессии показали, что связь КБП с кислотностью и влажностью почвы положительная, а с температурой – отрицательная (таблица 4.3).

Таблица 4.2– Результаты расчета коэффициента регрессии

Металлы	Параметры уравнений регрессии			
	$b_0$	$b_1$	$b_2$	$b_3$
As	-1,85	0,26	0,01	-0,01
Pb	-1,27	0,13	0,01	0,02
Zn	-2,59	0,25	0,01	0,04
Co	-1,44	0,22	0,05	-0,02
Ni	-1,64	0,25	0,03	-0,01
Cu	-0,27	0,01	0,01	0,03
Cr	-1,40	0,39	0,60	-0,58
Ba	-2,42	0,44	0,07	-0,03
Sr	-0,58	0,16	0,08	0,01
V	-2,91	0,38	0,01	-0,01
Mn	-2,41	0,33	0,01	-0,01

Примечание:  $b_0$  – постоянная,  $b_1$  – коэффициент, характеризующий влияние кислотности почвы на коэффициент биологического поглощения;  $b_2$  – коэффициент, характеризующий влияние осадков на коэффициент биологического поглощения;  $b_3$  – коэффициент, характеризующий влияние температуры на коэффициент биологического поглощения.

Полученные результаты подтверждают теоретические данные, что моделируемые процессы по своей сути нестабильные, поскольку они зависят от климатических условий окружающей среды, главным образом осадков, а, следовательно, влажности почвы, также в кислой почве поглощение тяжелых металлов выше, чем в щелочной среде [258].

Математическая модель, в основу которой легло дифференциальное уравнение, описывающее зависимость степени поглощения токсических элементов растением от условий среды свидетельствует о том, что поглощение растением одних металлов (As, Zn, V, Mn) зависит от кислотности почвы, других



(Ni, Ba, Sr) – от влажности почвы, а третьих (Pb, Cu) – от температуры (Приложение А таблица А.4.1, А.4.2, А 4.3).

Таблица 4.6 – Уравнения линейной зависимости КБП элементов растением в зависимости от кислотности почвы

Металлы	Уравнения	Дисперсионный анализ		
		$R^2$	Нормир. $R^2$	F
As	$Y = 0.311752 - 2.42074 \cdot X_1$	0,92	0,91	139,46
Pb	$Y = 0.633107 - 4.93565 \cdot X_1$	0,86	0,85	62,81
Zn	$Y = 1.115822 - 8.63704 \cdot X_1$	0,91	0,91	114,07
Co	$Y = 0.638783 - 5.30143 \cdot X_1$	0,75	0,73	30,76
Ni	$Y = 0.483032 - 3.82095 \cdot X_1$	0,88	0,86	74,10
Cu	$Y = 0.71569 - 5.23147 \cdot X_1$	0,88	0,87	75,29
Cr	$Y = 9.374263 - 6.76103 \cdot X_1$	0,81	0,79	44,86
Ba	$Y = 0.934533 - 7.14674 \cdot X_1$	0,86	0,84	62,48
Sr	$Y = 1.28142 - 9.74377 \cdot X_1$	0,85	0,83	58,30
V	$Y = 0.470819 - 3.87945 \cdot X_1$	0,92	0,92	129,27
Mn	$Y = 0.372201 - 2.82201 \cdot X_1$	0,93	0,93	149,04

Таблица 4.7 – Уравнения линейной зависимости КБП элементов растением в зависимости от влажности почвы

Металлы	Уравнения	Дисперсионный анализ		
		$R^2$	Нормир. $R^2$	F
As	$Y = 0.022314 + 0.125441 \cdot X_1$	0,92	0,92	131,51
Pb	$Y = 0.045223 + 0.235631 \cdot X_1$	0,85	0,83	59,29
Zn	$Y = 0.078748 + 0.482233 \cdot X_1$	0,89	0,87	81,07
Co	$Y = 0.048342 - 0.09831 \cdot X_1$	0,84	0,82	52,58
Ni	$Y = 0.03569 + 0.118146 \cdot X_1$	0,93	0,92	144,14
Cu	$Y = 0.050906 + 0.615522 \cdot X_1$	0,86	0,85	65,87
Cr	$Y = 0.685049 + 8.877057 \cdot X_1$	0,84	0,83	56,20
Ba	$Y = 0.069917 + 0.469688 \cdot X_1$	0,93	0,93	151,27
Sr	$Y = 0.095663 + 0.700877 \cdot X_1$	0,92	0,91	123,01
V	$Y = 0.03364 - 0.0338 \cdot X_1$	0,92	0,91	116,88
Mn	$Y = 0.026047 + 0.221064 \cdot X_1$	0,89	0,88	82,77

Таблица 4.8 – Уравнения линейной зависимости КБП элементов растением в зависимости от температуры

Металлы	Уравнения	Дисперсионный анализ		
		$R^2$	Нормир. $R^2$	F
As	$Y = 0.018316 - 0.24654 \cdot X_1$	0,71	0,62	25,25
Pb	$Y = 0.042494 - 0.66248 \cdot X_1$	0,86	0,85	63,72
Zn	$Y = 0.018316 - 1.13674 \cdot X_1$	0,84	0,84	189,32
Co	$Y = 0.037565 - 0.84744 \cdot X_1$	0,58	0,53	13,83
Ni	$Y = 0.028709 - 0.46109 \cdot X_1$	0,69	0,66	22,49
Cu	$Y = 0.050153 - 0.4577 \cdot X_1$	0,86	0,86	269,07
Cr	$Y = 0.0511853 - 1.18921 \cdot X_1$	0,54	0,49	11,84
Ba	$Y = 0.05477 - 0.62557 \cdot X_1$	0,65	0,62	19,28
Sr	$Y = 0.080972 - 0.95962 \cdot X_1$	0,75	0,73	31,33
V	$Y = 0.028007 - 0.60518 \cdot X_1$	0,73	0,70	27,10
Mn	$Y = 0.022555 - 0.24469 \cdot X_1$	0,76	0,74	32,62

При расчете коэффициента детерминации между КБП элементов и кислотностью среды отмечено что  $R^2$  варьирует в пределах от 75 до 93%, между КБП и влажностью среды – 84-93%, а между КБП и температурой среды – 54 до 86%, а 7-14% изменчивости коэффициента детерминации зависят от действия других факторов среды. При этом поглощение элементов As, Zn, V, Mn из загрязненной почвы растением в зависимости от кислотности почвы выше ( $R^2$ -квадрат в пределах 91-93%), чем ионов Pb, Cd, Co, Ni, Cu, Cr, Ba, Sr ( $R^2$ -квадрат в пределах 75-88%).

Анализируя полученное уравнение, следует отметить, что показатели кислотности и влажности оказывают наибольшее влияние на поглощение ТЭ растением из загрязненной среды. Значения  $R^2$  близки к значению 1, что свидетельствует о статистической достоверности полученных результатов. Следует заметить, что нормированный R-квадрат незначительно отличается от коэффициента детерминации, что доказывает хорошее качество модели. Высокие значения критерия Фишера подтверждает статистическую надежность уравнений в целом, что доказывает возможность использования разработанной модели для прогнозирования интенсивности накопления элементов из загрязненной почвы растением в зависимости от условий среды.

На основании расчета уравнения смоделировали график КБП элементов растением, и сравнили их с экспериментальными данными. Результаты свидетельствует о соответствии КБП между исходными данными и регрессионной моделью, погрешность не более 5% в среднем, что свидетельствует о достаточной адекватности предложенной модели с фактическими данными и возможности практического применения для прогнозирования поглощения токсических элементов растением в зависимости от условий среды (рисунок 4.3).

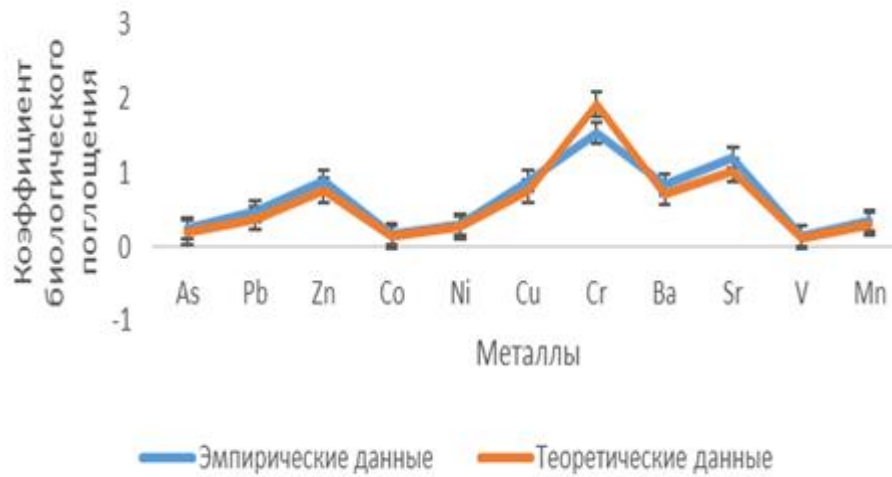


Рисунок 4.3 – Исходные данные и регрессионная модель КБП тяжелых металлов растением

Таким образом, получены математические модели биологического поглощения ТЭ растением в зависимости от факторов среды. Высокие значения критерия Фишера подтверждают статистическую надежность уравнений в целом и предполагают о возможности применения линейного дискриминанта для решения практических задач в области экологии.

#### 4.1 Выводы по разделу, постановка задачи

Поглощение ТЭ растением и миграции их в почве и грунтовой воде – сложный процесс, в котором участвуют различные физические, биологические и химические факторы. Трудности из-за многочисленных факторов возникают при моделировании очистки почвы/воды с помощью растений. В связи с этим, используют один параметр – коэффициент биологического поглощения (КБП). КБП является основным параметром оценки степени накопления элементов растением, и рассчитывается, как отношение содержания элемента в вегетативных органах (корень, стебель, листья) к его валовому содержанию в почве.

Цель работы – статистически оценить связь и взаимодействие трех факторов (почва, растение и почва-растение) на поглощение токсичных элементов растением в зависимости от типа почвы: светло-каштановой и песчаной (солончаки) почвы и разработать математическую модель миграции тяжелых металлов в системе «почва – растение» в зависимости от условий среды (кислотности, влажности почвы, температуры среды) с помощью множественного регрессионного анализа.

При проведении регрессионного анализа в качестве результативного показателя (у) использовали КБП (коэффициент биологического поглощения) тяжелых металлов (As, Pb, Cd, Zn, Co, Ni, Cu, Cr, Ba, Sr, V, Mn) растением, валовое содержание тяжелых металлов (As, Cd, Zn, Co, Ni, Cu, Cr, Ba, Sr, Mn) в почве и растении, а в качестве переменных факторов (х) – кислотность почвы (рН 8,13-8,94), влажность почвы (влажность почвы 8-11%) и температуру воздуха (21-30°C). Получены математические модели биологического поглощения ТЭ растением в зависимости от факторов среды. На основании расчета уравнения смоделировали график КБП элементов растением, и сравнили их с экспериментальными данными. Результаты свидетельствуют о соответствии КБП между исходными данными и регрессионной моделью, погрешность не более 5% в среднем, что свидетельствует о достаточной адекватности предложенной модели с фактическими данными и возможности практического применения для прогнозирования поглощения токсических элементов растением в зависимости от условий среды. Переход ТЭ из почвы в органы растений зависит от факторов среды: элементы, такие как As, Zn, V и Mn зависят от кислотности почвы, Ni, Ba, Sr – влажности почвы, а Pb, Cu – температуры. Статистически оценили связь и взаимодействие трех факторов (почва, растение и почва-растение) на поглощение ТЭ растением в зависимости от типа почвы: светло-каштановой и песчаной (солончаки) почвы. По результатам дисперсионного анализа установлено, что фактор «почва\*растение» оказывает прямое взаимодействие на накопление ТЭ в зависимости от типа почвы: на песчаной загрязненной почве спектр поглощения элементов выше, чем у растений, произрастающих на светло-каштановой загрязненной почве.

## ГЛАВА 5. РАЗРАБОТКА ИНФОРМАЦИОННОЙ СИСТЕМЫ ДЛЯ ПРОГНОЗИРОВАНИЯ И ПРИНЯТИЯ РЕШЕНИЙ В ПРОЦЕССЕ ОЧИСТКИ ПОЧВЫ ОТ ТОКСИЧНЫХ ЭЛЕМЕНТОВ

Актуальность создания единой информационной системы для формирования проектов по проведению фиторекультивационных мероприятий связана с тем, что многие районные и сельские населенные пункты находятся в зоне воздействия накопившихся отходов, которые представляют серьезную экологическую опасность для окружающей среды и человека во многих регионах Республики. Концептуальной основой, созданной БД для информационной системы, было обеспечение быстрого доступа к хранящимся данным для решения вопросов утилизации и очистки, с целью снижения негативных последствий ТЭ; быстрый доступ к хранящимся данным о растениях, способных к очистке почв/воды, загрязненных ТЭ, пестицидами, нефтеуглеводородами, радионуклидами и хлористыми растворителями [231,232]. Исследований по разработке информационной системы для формирования проектов по проведению фиторекультивационных мероприятий незначительны. Основное направление – прогноз распространения и оценка риска для окружающей среды. Для прогноза влияния пространственного распределения, загрязнения и риска для здоровья человека создана карта на основе ковариатной базы данных. Они были протестированы с помощью машинного обучения (Cubist, Random Forest, Support Vector Machine и Linear Model). Модели Cubist и Support Vector Machine показали наилучшие характеристики для пространственного прогнозирования содержания Pb и Zn в почве с более низкими значениями ошибки, что позволило авторам выявить источники загрязнения, которые представляют опасность для здоровья человека [257]. Высокая производительность модели Cubist была также подтверждена Y. Peng et al. [259] в прогнозировании и пространственной локализации содержания потенциально токсичных металлов As, Cr, Ni, Pb, Cu и Zn и D. Kidd [260] на содержание углерода на территориях, находящихся под влиянием горнодобывающей и промышленной деятельности в Катаре. Далее ансамблевые модели сравнили с экспериментальными данными для предсказания содержания мышьяка в аграрном поле. Авторы считают, что для более точной оценки As и других ТЭ необходимо рекомендовать ансамблевые модели. С помощью ансамблевых моделей обнаружили как низкие, так и высокие концентрации As в почве [261]. Последние годы российскими учеными разрабатываются информационно-аналитические системы по выбору технологий влияния и использования сельскохозяйственных угодий для принятия решений по улучшению состояния деградированных агроландшафтов. Создана информационно-аналитическая веб-система «Технологии восстановления плодородия почв и рекультивации деградированных агроландшафтов». Главное меню веб-системы состоит из семи основных пунктов, названия которых полностью соответствуют назначению: «Почвы», «Тип деградации», «Воздействие», «Вид загрязнения», «Объект», «Категория», «Рекомендуемые технологии и подэтапы» [248]. Вопросы, связанные с разработкой и применением научно-обоснованных подходов по выбору мелиоративных

мероприятий, обеспечивающих повышение плодородия деградированных земель при увеличивающейся техногенной нагрузке, остаются открытыми. В связи с этим, целью данного раздела – создание информационной системы для прогнозирования процесса очистки 1 га почвы с глубины залегания 0-20 см, с помощью растений от ТЭ на основе БД, моделей продуктивности растений, модели фитотоксичности почвы, модели поглощения ТЭ растением и содержания их в почве.

В основу концепции информационной системы нами был предложен интегрированный подход – создание систем, выполняющих весь комплекс необходимых операций, включая:

сбор, передачу, накопление и обработку измерительных данных;

математическую модель фитотоксичности почвы;

математическую модели поглощения ТЭ растением и их миграцию в системе «почва–растение» в зависимости от условий среды;

анализ экологической ситуации;

поддержку принятия решений по прогнозированию длительности периода очистки почв, загрязненных ТЭ;

длительность периода очистки почвы с помощью растений.

*Постановка задачи.* Первоначальной задачей работ при оценке уровня загрязнения окружающей среды токсичными веществами является расчет суммарных показателей содержания их в почвенной среде и органах растений, полученных в результате химического анализа в лабораторных условиях. На втором этапе была проведена оценка фитотоксичности для оценки состояния биосистемы, с помощью регрессионного анализ; поглощения ТЭ растением и их переход из почвы в органы растений в зависимости от условий среды, с применением метода пошаговой регрессии.

*Создание математической модели фитотоксичности почвы*

Основным этапом при прогнозировании фитотоксичности почвы, как показателя загрязненности почвы, для оценки состояния биосистемы является создание модели, учитывающей наличие ТЭ и их влияния рост и развитие растений [262,263].

*Общая постановка задачи.* Целью нашей работы явилось построение регрессионных уравнений для описания воздействия загрязненной ТЭ почвы в разной концентрации (суммарный уровень загрязнения почвы – 227,4; 165,4; 20,0 мг кг<sup>-1</sup>) на растение.

МНК – статистическая процедура для нахождения наилучшего соответствия набору точек данных. Метод работает путем минимизации суммы смещений или остатков точек от построенной кривой. Регрессия МНК используется для прогнозирования поведения зависимых переменных. Сущность МНК заключается в выборе в качестве «меры близости» суммы квадратов отклонений левых и правых частей. Параметры в математической модели были оптимизированы путем минимизации суммы квадратов остаточного отклонения точек данных от модели. Отношения считались значимыми при  $p \leq 0,05$ . Ка-

чество полученных регрессионных моделей оценивали с использованием коэффициента детерминации ( $R^2$ ) с помощью дисперсионного анализа. На основании анализа экспериментальных данных рассчитали базовую форму (структуру) модели фитотоксичности почв для растений, произрастающих на незагрязненной и загрязненной ТЭ почве в зависимости от их концентрации [264, 265].

*Математическая модель.* Экспериментальные данные по кинетике роста были близки к логистической зависимости, поэтому сумму квадратов отклонений рассчитывали по модели значений  $\hat{y}_t$  от фактических  $y_t$  уровней временного ряда наименьшей:

$$\sum_{t=1}^n (y_t - \hat{y}_t)^2 = \sum_{t=1}^n e_t^2 \rightarrow \min \quad (5.1)$$

$$e_t = y_t - \hat{y}_t \quad (5.2)$$

где  $e_t$  остаток – расхождение фактических и модельных значений показателя  $Y$ .

При оценке параметров полиномиальных кривых использовали следующую формулу:

$$y_t = a_0 + a_1 t + a_2 t^2 + \dots + a_q t^q \quad (5.3)$$

Далее в результате минимизации получили систему нормальных уравнений. Система состояла из  $(q + 1)$  линейных уравнений, содержащих в качестве известных величин  $\sum y_t, \sum y_t t, \sum y_t t^2, \sum y_t t^q$  и  $(q + 1)$  неизвестных коэффициентов  $a_j$ .

Решение этой системы относительно  $a_0, a_1, a_2, \dots, a_q$  позволил нам вычислить искомые оценки параметров.

$$\sum_{t=1}^n y_t t = a_0 \sum_{t=1}^n t + a_1 \sum_{t=1}^n t^2 + a_2 \sum_{t=1}^n t^3 \dots + a_q \sum_{t=1}^n t^{q+1} \quad (5.4)$$

$$\sum_{t=1}^n y_t = a_0 n + a_1 \sum_{t=1}^n t + a_2 \sum_{t=1}^n t^2 + \dots + a_q \sum_{t=1}^n t^q \quad (5.5)$$

$$\sum_{t=1}^n y_t t = a_0 \sum_{t=1}^n t + a_1 \sum_{t=1}^n t^2 + a_2 \sum_{t=1}^n t^3 \dots + a_q \sum_{t=1}^n t^{q+1} \quad (2.6)$$

$$\sum_{t=1}^n y_t t^{q-1} = a_0 \sum_{t=1}^n t^{q-1} + a_1 \sum_{t=1}^n t^q + a_2 \sum_{t=1}^n t^{q+1} \dots + a_q \sum_{t=1}^n t^{2q-1} \quad (5.7)$$

$$\sum_{t=1}^n y_t t^q = a_0 \sum_{t=1}^n t^q + a_1 \sum_{t=1}^n t^{q+1} + a_2 \sum_{t=1}^n t^{q+2} \dots + a_q \sum_{t=1}^n t^{2q} \quad (5.8)$$

На основании расчета получили эмпирическое уравнение регрессии кинетики роста растений на загрязненной и незагрязненной почвах в зависимости от концентраций ТЭ с учетом величины достоверности аппроксимации (Таблица 5.1).

Таблица 5.1 – Алгебраическое степенное уравнение кинетики роста растений на незагрязненной и загрязненной почвах в зависимости от концентраций ТЭ

Уровень загрязнения почвы, мг кг <sup>-1</sup>	Степенное уравнение	R <sup>2</sup>
Незагрязненная почва	$y = -0,0074x^2 + 1,9942x + 27,358$	0,9853
Zc 165,4 ± 17,1	$y = -0,0042x^2 + 1,429x + 30,985$	0,996
Zc 20,0 ± 3,1	$y = -0,0045x^2 + 1,2436x + 28,795$	0,9582

На основании коэффициента детерминации сделали заключение о тесной взаимосвязи роста растений и концентрации ТЭ в почве (98%).

Для моделирования теоретической кривой кинетики роста растений на загрязненной ТЭ (в разных концентрациях) и незагрязненной (контрольной) почве, использовали логистическое уравнение математической модели одиночной популяции, описанной в статье Е.Горбунова и Е.Колпак [266], так как экспериментальные данные по кинетике роста были близки к логистической кривой.

$$\frac{dl}{dt} = \mu l \left(1 - \frac{l}{l_{max}}\right) \quad (5.9)$$

где  $t$  – время (дни),  $l$  – текущая высота растения (см),  $l_{max}$  – теоретическая максимальная высота (см), которую может достигнуть растение по окончании роста,  $\mu$  – константа (удельная скорость роста, размерность – 1/день).

Решением данного уравнения была функция ( $l_0$  – начальная высота растения):

$$l(t) = l_{max} \frac{l_0 e^{\mu t}}{l_{max} - l_0 + l_0 e^{\mu t}} \quad (5.10)$$

На основании данных уравнения 2.10 получили эмпирическое уравнение теоретической кривой динамики роста растений на загрязненной ТЭ (и незагрязненной почве (рисунок 2.1).



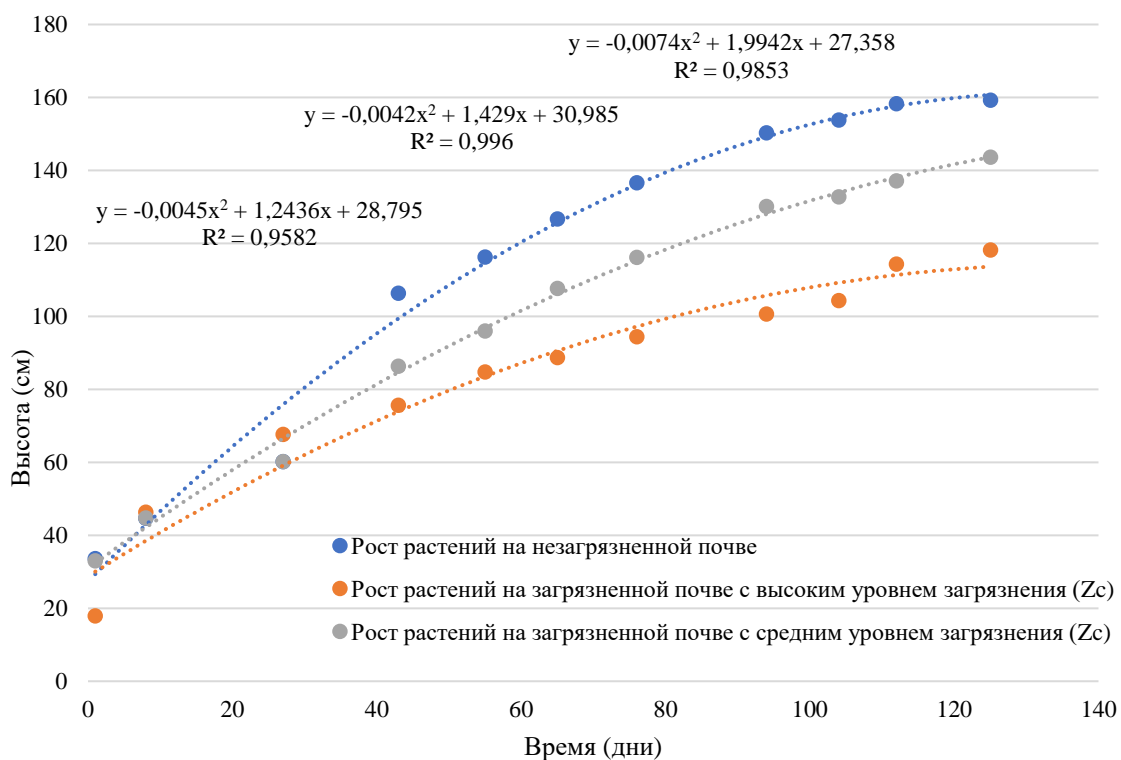


Рисунок 5.1 – Теоретическая кривая кинетики роста растений от времени произрастания, с учетом концентраций ТЭ в почве

При построении теоретической кривой подтвердили, что рост растений зависит от концентрации ТЭ в почве, чем выше концентрация, тем выше ингибирующий эффект. Коэффициенты детерминации, отражающие степень соответствия результатов расчетов экспериментальным данным, при использовании этих данных были близки к значению 1.

Таким образом, установлено, что степень фитотоксичности почвы зависит от концентраций ТЭ в почве. Наибольшее влияние на степень фитотоксичности почвы оказывают высокие концентрации ТЭ (суммарный уровень загрязнения: 281 и 165 мг кг<sup>-1</sup>), которые вызывают либо гибель растений, либо подавляют их рост и развитие. Согласно модели, растение может произрастать на почве с суммарным уровнем загрязнения 20 мг кг<sup>-1</sup>.

*Структура информационной системы* Диаграмма потока информационной системы для очистки почвы от ТЭ представлена на рисунке 5.2

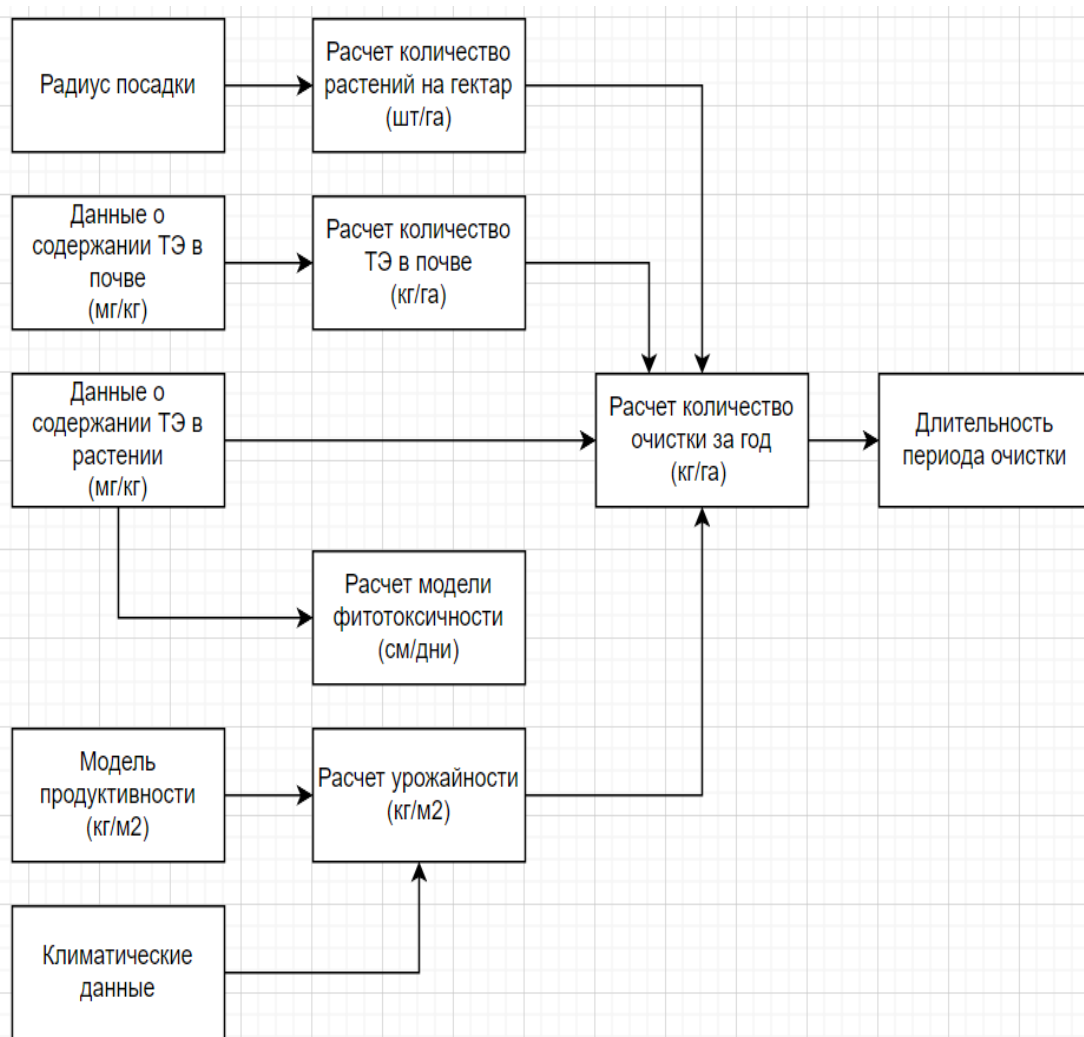


Рисунок 5.2 – Диаграмма потока информационной системы для очистки почвы от ТЭ

При создании информационной системы использовали натренированную модель XGBoost. Модель XGBoost – оптимизированная распределенная библиотека повышения градиента, разработанная для обеспечения высокой эффективности, гибкости и портативности. Она реализует методы машинного обучения в рамках фреймворка Gradient Boosting. Данный метод успешно используется для решения многих задач, в частности для классификации, регрессии, упорядочения. Наиболее эффективен при построении моделей предсказания на структурированных больших данных. Модель обладает более высокой точностью, чем одно дерево решений. Особенности фреймворка модели XGBoost: автоматически предоставлять оценки важности функций из обученной прогностической модели, широкая область применения: может быть использован для решения задач регрессии, классификации, упорядочения и пользовательских задач; совместимость с Windows, Linux и OS X; поддержание ведущих языков программирования: C++, Python, R, Java, Scala и Julia; облачная интеграция поддерживает кластеры AWS, Azure и Yarn, хорошо работает с

Flink, Spark. Отличается высокой скоростью работы, а модели, построенные на его основе, обладают хорошей производительностью. Поэтому он пользуется популярностью при решении задач классификации и регрессии с использованием табличных наборов данных [267].

В качестве прогнозируемых значений использовали экспериментальные данные образцов, созданные модели фитотоксичности и поглощения ТЭ растением, что позволяет получить достоверный прогноз периода очистки почв для конкретного вида. В качестве объекта исследования использовали вид мискантус. Интерес к данному виду обусловлен высокопродуктивностью, высоким содержанием целлюлозы, способностью произрастать на маргинальных неиспользованных землях, способностью к очистке почв, загрязненных ТЭ и нефтепродуктами. Используют в биоэнергетике во многих странах Европы, Восточной Азии и США для получения твердого биотоплива [268].

В наших исследованиях для прогнозирования уровней ТЭ в почве с помощью модели XGBoost использовали 100 проб, отобранных с глубины 0-20 см. Суммарный уровень загрязнения почвы составил 20 мг/кг. Почва была отобрана из территории горно-обогатительного комбината Казцинк (г.Текели). Длительной период интенсивной эксплуатации (с 1942 г.) место добычи цинка, обернулось интенсивным загрязнением земель. Почва загрязнена ТЭ разного класса опасности. Основными загрязнителями, обнаруженные в почве были элементы 1-го и 2-го классов опасности: Pb, Zn, Cu, Ni, Co, Cr) и и металлоид As. Наряду с высокотоксичными элементами почва содержит также умеренно опасные элементы, такие как Mn и Sr.

Графическое изображение накопления ТЭ в почве на площади 1 га представлено на рисунке 5.3

Оптимальное количество растений на 1 га земли зависит от различных факторов, таких как вид растений, площади посадки и нормы посева и др. Используя данные площади и нормы посева, можно оценить количество растений, которые могут быть выращены на данной площади, сохраняя оптимальные условия для роста (Приложение А, Рисунок 5.1). Выявлено, что количество растений на 1 гектаре земли составляет 35367 при радиусе площади посадки 30 см (Приложение А, рисунок 5.2).

Для тренировки модели XGBoost использовали данные продуктивности растений, полученные по отношению к климатическим условиям за 2022 год (Приложение А, рисунок 5.3). Динамика развития мискантуса на площади 1 га загрязненной почвы составила 120 дней, за данный период высота достигала 140 см, продуктивность –1.485 т/га (рисунок 5.3)

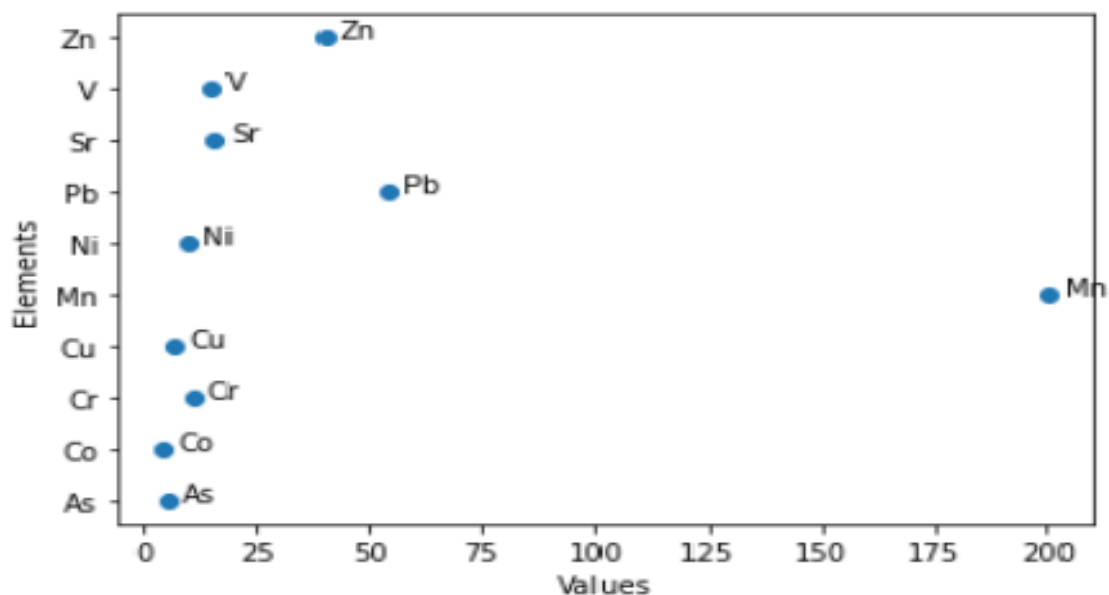
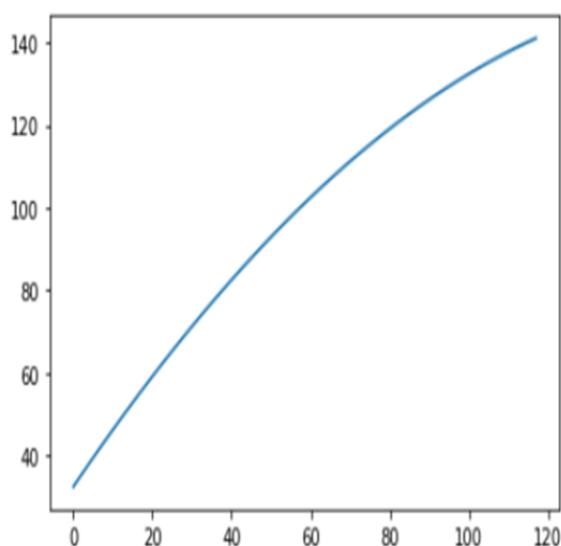
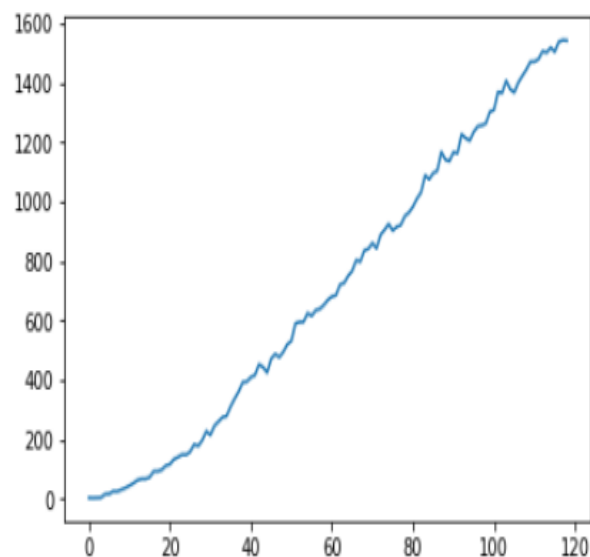


Рисунок 5.3 – Начальный уровень загрязнений почвы ТЭ, кг/га



Модель роста растений, см/дни



Модель урожайности, кг/га

Рисунок 5.4 – Модели роста и продуктивности мискантуса на 1 га загрязненной ТЭ почве

Одним из недостатков очистки почв с помощью растений – длительный процесс. Период полной очистки ТЭ растением из почвы может значительно варьироваться в зависимости от таких факторов, как тип и концентрация токсина, виды растений, и условия окружающей среды. Несмотря на это, использование растений для очистки является экологически безопасным, экономически выгодной технологии, не требует крупных капиталовложений. Поэтому

высаживают многолетние растения для очистки, в частности Мискантус, который может произрастать на одном месте 30 лет. Связь между количеством растений и временем, необходимым для очистки почвы от ТЭ зависит от различных факторов, включая конкретный вид растения, уровень загрязнения почвы и условия окружающей среды. Увеличение количества растений может помочь ускорить процесс ремедиации почвы, увеличивая скорость усвоения и очистки от ТЭ, однако этот эффект может быть ограничен биодоступностью элементов, высокой концентрацией их в почве и условиями почвенной среды. У растений есть различные механизмы для поглощения и детоксикации ТЭ в почве. Они могут фильтровать токсиканты из почвы через свои корни, превращать их в менее вредные формы или хранить их в вакуолях своих клеток. Согласно литературным данным мискантус накапливает 8 высокотоксичных элементов (As, Pb, Zn, Co, Ni, Cr, Cu, V) в корнях и 2 умеренно опасных элементов (Mn, Sr) в стебле и листьях. Мискантус очищает загрязненную опасными ТЭ почву за счет поглощения и накопления их в корневой системе [254]. Поэтому представляло интерес определить: сколько времени потребуется для восстановления ТЭ загрязненной почвы с суммарным уровнем загрязнения 20.0 мкг с помощью мискантуса?

На рисунке 5.5 и 5.6 представлены коды программы информационной системы для получения информации о связи между количеством растений и временем очистки почвы; длительность периода снижения элементов с помощью растений. На рисунке представлено графическое изображение снижения ТЭ в почве из года в год, за счет поглощения и накопления их растением мискантус (рисунок 5.7). Установлено, что длительность периода очистки почвы 1 га почвы с глубины залегания 0-20 см с помощью растения мискантус для разных элементов различна (таблица 5.1). Например, период очистки особо токсичных элементов таких как цинк (Zn), и медь (Cu) составляет 7 лет, стронций (Sr) – 4 года, свинца (Pb) – 12 лет. При этом длительность очистки почвы от мышьяка (As) почти 109 лет, кобальта (Co) и хрома (Cr) – 38 лет, никеля (Ni) – 34 года, а вольфрама (V) – 79 лет.

Таким образом, полученные результаты свидетельствуют о том, что созданная информационная система с помощью модели XGBoost позволяет определить длительность периода очистки почвы от ТЭ.

```
In [54]: metal_concentraision = df_toxic_elemnts_in_soil["Pb"].values * weight # mg in 1-meter square
metal_concentraision = metal_concentraision / (1000 * 1000) #kg in 1-meter square
metal_concentraision
executed in 4ms, finished 18:15:21 2023-04-02

Out[54]: array([0.005934])

In [55]: te_in_1m = metal_concentraision #kg in 1-meter square
te_in_1m
executed in 4ms, finished 18:15:21 2023-04-02

Out[55]: array([0.005934])

In [56]: te_in_20cm = te_in_1m
te_in_20cm
executed in 4ms, finished 18:15:21 2023-04-02

Out[56]: array([0.005934])

In [57]: te_in_hect = te_in_20cm * 10000
te_in_hect
executed in 4ms, finished 18:15:21 2023-04-02

Out[57]: array([59.34])
```

## 11 How plant count related for time nedded for clear soil

```
In [58]: weight_one_plant_on_hectar = y_pred_2022[-1]/plants_per_hectare
weight_one_plant_on_hectar = weight_one_plant_on_hectar * 1000 #gram
weight_one_plant_on_hectar = weight_one_plant_on_hectar //1
weight_one_plant_on_hectar
executed in 5ms, finished 18:15:21 2023-04-02

Out[58]: 42.0

In [59]: #how many years nedded for fully clear a soil
df_metal_concentraision / df_absorb_for_hectar
executed in 9ms, finished 18:15:21 2023-04-02

Out[59]:
```

	As	Co	Cr	Cu	Mn	Ni	Pb	Sr	V	Zn
0	109.275971	38.037661	38.119981	7.20105	23.883469	34.652848	12.365853	4.090107	79.434971	7.42681

```
In [60]: (weight_one_plant_on_hectar / 1000) * 10000 * plants_per_meter
executed in 3ms, finished 18:15:21 2023-04-02

Out[60]: 1485.4461355243564

In [61]: y_pred_2022[-1]
executed in 3ms, finished 18:15:21 2023-04-02

Out[61]: 1507.6465
```

Рисунок 5.5– Скрипшот кода программы информационной системы для получения информации о связи между количеством растений и временем очистки почвы

```
In [45]: age_needed_for_clear_soil = df_metal_concentration / df_absorb_for_hectar
age_needed_for_clear_soil.astype("int64")
```

executed in 10ms, finished 18:15:20 2023-04-02

```
Out[45]:
```

	As	Co	Cr	Cu	Mn	Ni	Pb	Sr	V	Zn
0	109	38	38	7	23	34	12	4	79	7

```
In [46]: period = []
df_period = df_metal_concentration - df_absorb_for_hectar
# create scatter plot
fig, ax = plt.subplots()
ax.scatter(df_period.iloc[0], df_period.columns)

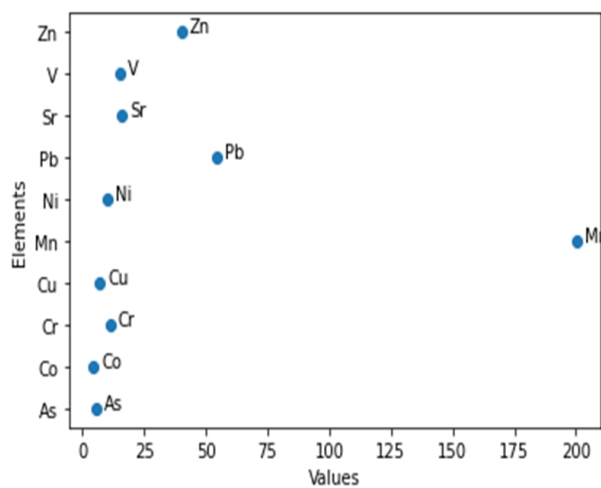
# add labels to the points
for i, (x, y) in enumerate(zip(df_period.iloc[0], df_period.columns)):
    ax.annotate(y, (x, y), xytext=(5, 0), textcoords='offset points')

# set labels for x and y axes
ax.set_xlabel('Values')
ax.set_ylabel('Elements')

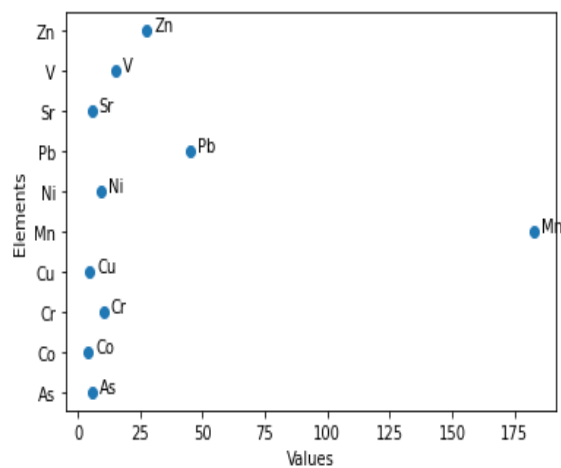
plt.show()
for i in range(9):
    df_period = df_period - df_absorb_for_hectar
    df_period = df_period.clip(lower=0)
    period.append(df_period)
```

executed in 155ms, finished 18:15:20 2023-04-02

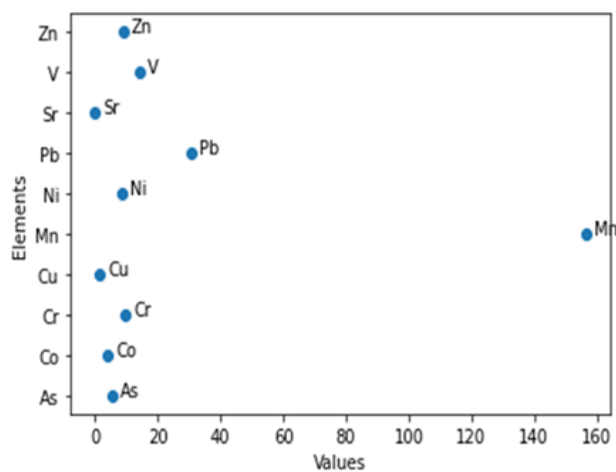
Рисунок 5.6 – Скрипшот кода программы информационной системы для получения информации о длительности периода очистки от ТЭ растением



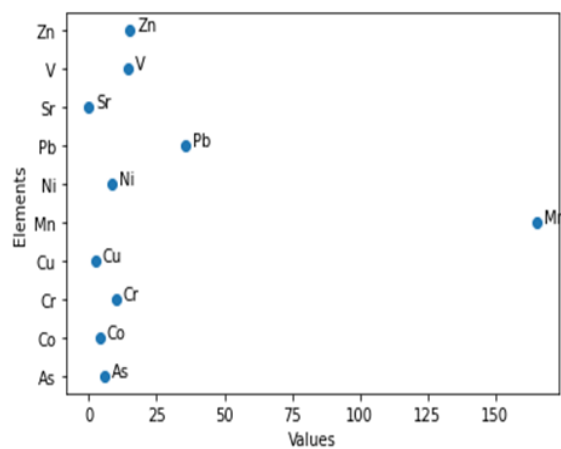
1 год



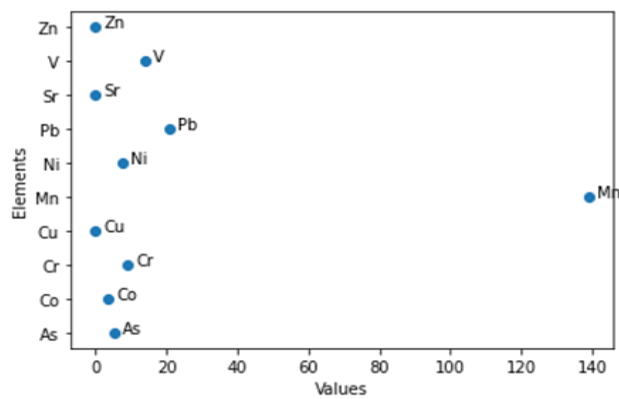
3 года



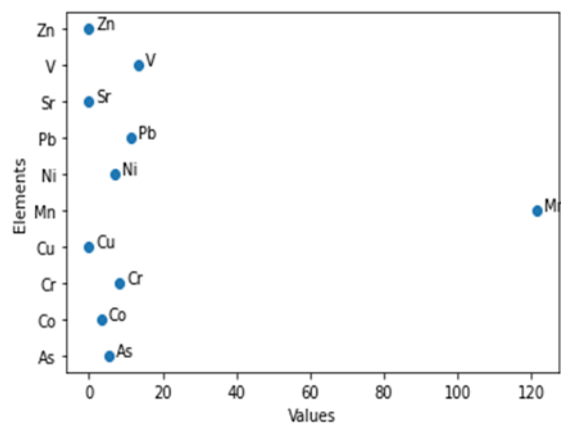
5 лет



7 лет



8 лет



10 лет

Рисунок 5.7– Графическое изображение снижения ТЭ в почве с помощью растений по годам



Таблица 5.1 – Длительность периода очистки 1 га почвы с глубины залегания 0-20 см ТЭ с помощью растения

Токсические элементы	As	Co	Cr	Cu	Mn	Ni	Pb	Sr	V	Zn
Содержание ТЭ в почве, кг/га	5.80	4.64	11.35	8.12	208.98	10.11	59.34	20.97	15.48	46.44
Поглощение растением ТЭ из почвы, кг/га	0.06	0.15	0.37	1.43	11.13	0.37	6.10	6.52	0.24	7.95
Количество ТЭ в почве, после 10 лет очистки, кг/га	5.27	3.46	8.37	0	121.48	7.19	11.35	0	13.53	0
Длительность периода очистки, годы	109	38	38	7	23	34	12	4	79	7
Количество растений на 1 га – 35367 растений.										
Масса урожая одного растения – 42 мг/м <sup>2</sup>										
Масса урожая в год – 1.485 т/га										

## 5.1 Выводы по разделу, постановка задачи

Интерес к фитотехнологии (очистка с помощью растений), как стратегии рекультивации почв, загрязненных ксенобиотиками различного происхождения, растет из года в год в связи с увеличением масштабов загрязнения окружающей среды, вызванного антропогенной деятельностью. Одним из недостатков очистки почв с помощью растений – длительный процесс. Период полной очистки ТЭ растением из почвы может значительно варьироваться в зависимости от таких факторов, как тип и концентрация ТЭ, виды растений, и условия окружающей среды. Несмотря на это, использование растений для очистки является экологически безопасным, экономически выгодной технологией, не требует крупных капиталовложений. Поэтому высаживают многолетние биоэнергетические растения с высокой биомассой, которую можно использовать в различных секторах промышленности.

Цель данной главы – создание информационной системы для прогнозирования процесса очистки 1 га почвы с глубины залегания 0-20 см, с помощью растений от ТЭ на основе БД, моделей продуктивности растений, модели фитотоксичности почвы, модели поглощения ТЭ растением и содержания их в почве.

В качестве прогнозируемых значений использовали экспериментальные данные образцов, созданные модели фитотоксичности и поглощения ТЭ растением, что позволяет получить достоверный прогноз периода очистки почв для конкретного вида. В качестве объекта исследования использовали вид мискантус. Интерес к данному виду обусловлен высокопродуктивностью, высоким содержанием целлюлозы, способностью произрастать на маргинальных землях, способностью к очистке почвы от ТЭ и нефтепродуктами. При моделировании фитотоксичности загрязненной ТЭ почвы получено эмпирическое уравнение регрессии кинетики роста растений, с учетом величины достоверности аппроксимации; построена теоретическая кривая кинетики роста растений от времени произрастания с помощью логистического уравнения. Выявлено, что оптимальным условием произрастания растений является почва с суммарным уровнем загрязнения ТЭ  $20 \text{ мг кг}^{-1}$

В основу концепции информационной системы нами был предложен интегрированный подход – создание систем, выполняющих весь комплекс необходимых операций, включая:

- сбор, передачу, накопление и обработка измерительных данных;
- модели продуктивности;
- математическую модель фитотоксичности почвы;
- математическую модель поглощения ТЭ растением и их переход их почвы в вегетативные органы растений в зависимости от условий среды;
- анализ экологической ситуации;
- поддержку принятия решений по прогнозированию длительности периода очистки почв, загрязненных ТЭ;
- длительность периода очистки почвы с помощью растений.

При создании информационной системы использовали натренированную модель XGBoost, методы машинного обучения, основанный на дереве поиска решений и использующей фреймворк градиантного бустинга. Для тренировки модели XGBoost использовали данные продуктивности растений, в результате разработки моделей по отношению к климатическим условиям за 2022 год проведены расчеты, оптимальное количество растений на 1 га земли, используя данные площади и нормы посева для прогнозирования связи между количеством растений и временем очистки почвы. Получили графическое изображение снижения ТЭ в почве из года в год, за счет их поглощения растением. Установлено, что длительность периода очистки почвы 1 га почвы с глубины залегания 0-20 см с помощью растения мискантус для разных элементов различна: период очистки особо токсичных элементов таких как цинк (Zn), и медь (Cu) составляет 7 лет, стронций (Sr) – 4 года, свинца (Pb) – 12 лет.

Данный подход создания информационной системы может быть использован для принятия решений по восстановлению загрязненной ТЭ почвы с помощью других видов растений.

## ЗАКЛЮЧЕНИЕ

Целью диссертации была разработка интеллектуальной информационной системы для обработки данных о почвах, загрязненных токсичными элементами для прогнозирования и принятия решений об очистке земель Республики Казахстан. Поставленные задачи были решены следующим образом:

1 Адаптирован «Многорядный эвристический метод самоорганизации», предсказывающий биомассу растений в зависимости от окружающих условий. Наибольшее влияние на этот процесс оказывают три фактора: испарение влаги почвы, фотосинтетическая активная радиация и осадки.

1.1 Усовершенствована модель MiscanCalc на основе MiscanMod для прогнозирования биомассы растений на загрязненной почве с учетом климатических данных путем расчета коэффициента разности между загрязненной и чистой почвой.

2. Созданы два организованных хранилища данных: о древесных и травянистых видах растений, способствующие к восстановлению почвы; о количестве устаревших пестицидов и их концентрации в почве.

3. Разработана информационная система для прогнозирования и принятия решений в процессе очистки почвы от ТЭ. Определена длительность периода очистки, затраченного для каждого токсичного элемента.

4 Применены интеллектуальные методы машинного обучения и определен наилучший из них для прогнозирования концентраций загрязнения почвы. По результатам исследования XGB Regressor имеет самые низкие метрики  $R^2 = 0.998$ ;  $MSE = 421.19$ ;  $MAE = 15.03$ ;  $MAPE = 0.065$ .

4.1 Исследованы ансамблевые методы машинного обучения с целью прогнозирования производительности растений на основе климатических данных. Процесс анализа осуществлен через инструментальное программное обеспечение JupyterLab в среде Anaconda.

4.2 Проведена оценка моделей регрессии на урожайность в зависимости от климатических условий. С помощью подхода SHAP отобрано 13 информативных признаков, ответственных за продуктивность, высокая степень влияния оказывает признаки `datetime_1` (месяцы) и `datetime_2` (дни).

4.3 Установлена длительность периода очистки почвы 1 га почвы с глубины залегания 0-20 см с помощью растения мискантус для разных элементов. Для особо токсичных элементов: свинец – 12 лет, цинк – 7 лет.

**Дальнейшие доработки.** Разработка и осуществление специальных программ по оценке влияния ТЭ на окружающую среду и человека является актуальной и важной проблемой для разработки мероприятий по снижению с помощью реабилитации. Остаются до конца нерешенными проблемы, связанные с состоянием окружающей среды, вопросы, связанные с деградацией земель, дефицита водных ресурсов, высоким уровнем загрязненности воздуха в городах и утилизации отходов. До сих пор отсутствует компьютерная информационная БД учета загрязнения окружающей среды, нет модели, описывающей поведение ксенобиотиков в почвенно-грунтовой толще и их переход из почвы

в органы растений (корень, стебель, листья, плоды), которая могла бы стать основой разработки эффективной технологии восстановления загрязненных земель. В связи с этим, возможности для будущей работы достаточно велики, например, создание математических моделей произрастания сельскохозяйственных и лесных культур на деградированных загрязненных землях, с целью рационального использования и их восстановления.

## СПИСОК ИСПОЛЬЗОВАННЫХ ИСТОЧНИКОВ

1 Кобегенова Х. Н., Шакенова Т. К. Деградация свойств почвы в результате воздействия природных и антропогенных факторов на территории Республики Казахстан // Вестник Приамурского государственного университета им. Шолом-Алейхема. – 2017. – № 3(28). – С. 32-38

2 Vaubekova A., Akindykova A., Mamirova A., Dumat C., Jurjanz S. Evaluation of environmental contamination by toxic trace elements in Kazakhstan based on reviews of available scientific data // Environ. Sci. Pollut. Res. – 2021. – № 28. – P. 43315–43328

3 Жумаксанова К.М., Есымханова З.Е., Есенжигитова Р.Г., Кайдарова А.Т. Современное состояние цифровизации сельского хозяйства: проблемы и пути решения // Central Asian Economic Review. – 2019. – Том. 5. – С. 144-155.

4 Hampton S.E., Parker J.N. Collaboration and productivity in scientific synthesis // BioScience. – 2011. – Vol. 61. – P. 900-910.

5 Bolbakov R.G. Big data in information sciences // Educational Resources and Technology. – 2017. – Vol. 1, № 18. – P. 30-35.

6 Nurzhanov C., Pidlisnyuk V., Naizabayeva L., Satymbekov M. Research and trends in computer science and educational technology during 2016-2020: Results of a content analysis // World Journal on Educational Technology: Current Issues. – 2021. – Vol. 13, № 1. – P. 115-128.

7 Honga S., Lee S.P., Park W., Jeon M., Chang H. An analysis of security systems for electronic information for establishing secure internet of things environments: Focusing on research trends in the security field in South Korea // Future Generation Computer Systems. – 2018. – Vol. 82. – P. 769-782.

8 Rodhain F., Fallery B. ICT and Ecology: In favour of research based on the Responsibility principle // AIS Electronic Library (AISeL). Mediterranean Conference of Information Systems. – 2009. – P. 173-186.

9 Eryomin A.L. Information Ecology – a Viewpoint // Int. J. Environ. Sci., Sections A & B. – 1998. – Vol. 3, № 4. – P. 241–253.

10 Patel S., Sayyed I.U. Impact of information technology in agriculture sector // JFAV. – 2014. – Vol. 4, № 2. – P. 17-22.

11 Zhu H. Research on Agricultural Ecological Factors Information Technology based on Internet // Advances in Computer Science Research. 3<sup>rd</sup> International Workshop on Materials Engineering and Computer Sciences (IWMECS 2018). – 2018. – Vol. 78. – P. 19-22.

12 Dara A., Baumann M., Kuemmerle T., Pflugmachera D., Rabe A., Griffiths P., Hölzel N., Kamp J., Freitag M., Hosterta P. Mapping the timing of cropland abandonment and recultivation in northern Kazakhstan using annual Landsat time series // Remote Sensing of Environment. – 2018. – Vol. 213. – P. 49-60.

13 Pazur R., Prishchepov A.V., Myachina K., Verburg P.H., Levykin S., Ponkina E.V., Kazachkov G., Yakovlev I., Akhmetov R., Rogova N., Bürgi M. Restoring steppe landscapes: patterns, drivers and implications in Russia's steppes // Landscape Ecology. – 2021. – Vol. 36. – P. 407-425.

- 14 Michener W.K., Jones M.B. Ecoinformatics: Supporting ecology as a data-intensive science // *Trends in Ecology and Evolution*. – 2012. – Vol. 27. – P. 85-92.
- 15 Schnase J.L., Duffy D.Q., Tamkin G.S., Nadeau D., Thompson J.H., Grieg C.M., McInerney M.A., Webster W.P. MERRA analytic services: Meeting the big data challenges of climate science through cloud-enabled climate analytics-as-a-service // *Computers, Environment, Urban Systems*. – 2017. – Vol. 61. – P. 198-211.
- 16 Olson R.L., Sequeira R.A. An emergent computational approach to the study of ecosystem dynamics // *Ecol. Model.* – 1995. – Vol. 79. – P. 95-120.
- 17 Farley S.S., Dawson A., Goring S.J., Williams J.W. Situating Ecology as a Big-Data Science: Current Advances, Challenges, and Solutions // *BioScience*. – 2018. – Vol. 68, № 8. – P. 563-576.
- 18 Переплетчиков В.И., Носков В.Ю. Разработка информационной системы сбора и анализа данных метеорологических сервисов на основе технологий BIG DATA // *Теплотехника и информатика в образовании, науке и производстве: сборник докладов VII Всероссийской научно-практической конференции студентов, аспирантов и молодых учёных (ТИМ' 2018) с международным участием (Екатеринбург, 17-18 мая 2018 г.)*. – Екатеринбург: ООО АМК «День РА». – 2018. – С. 487-493.
- 19 Athanasiadis N., Mitkas P. A methodology for developing environmental information systems with software agents // *In Advanced Agent-Based Environmental Management Systems, Whitestein Series in Software Agent Technologies and Autonomic Computing*. – 2009. – Birkhauser, Springer. – P. 119-138.
- 20 Athanasiadis N. Towards a virtual enterprise architecture for the environmental sector // *Agent and Web Service Technologies in Virtual Enterprises, Information Science Reference*. – 2007. – P. 256-266.
- 21 Avci C., Tekinerdogan B., Athanasiadis I.N. Software architectures for big data: a systematic literature review // *Big Data Analytics*. – 2020. – Vol. 5, № 1. – P. 2-53.
- 22 Santos A.F.C., Teles I.P., Siqueira O.M.P., de Oliveira A.A. Big Data: A Systematic Review // *Information Technology - New Generations*. – 2017. – Vol. 558. – P. 501-506.
- 23 Majid B., Khalid A., Donald W., Joel B. Examining plant uptake and translocation of emerging contaminants using machine learning: Implications to food security // *Science of the Total Environment*. – 2020. – Vol. 698. – P. 133999.
- 24 Мустафаев Ж.С., Рябцев А.Д., Ибатуллин С.Р., Козыкеева А.Т. Модель природы и моделирование природного процесса // *Тараз*. – 2009. – 190 с.
- 25 Водопьянов В.В. Математическое моделирование динамики роста грибных пропагул в почве, загрязненной нефтью // *Вестник УГАТУ*. – 2007. – Т. 9, № 7(25). – С. 79-82.
- 26 Экологический кодекс Республики Казахстан. 2007, с изменениями и дополнениями по состоянию 27.12.21 г. [http://online.zakon.kz/Document/?doc\\_id=30085593](http://online.zakon.kz/Document/?doc_id=30085593).
- 27 Dedrick J., Green I.S. Concepts and issues for information systems research // *CAIS*. – 2010. – Vol. 27, № 1. – P. 173-184.

- 28 Fodor V.E., Pájer J. Application of Environmental Information Systems in Environmental Impact Assessment (in Hungary) // *Acta Silv. Lign. Hung.* – 2017. – Vol. 13, № 1. – P. 55-67.
- 29 Ekundayo E.O. Environmental monitoring // *Intex.* – 2011. – 528 p.
- 30 Воронич С.С., Роева Н.Н., Зайцев Д.А., Пахов Д.Е., Хлопаев А.Г., Шадская Ю.С., Баранов А.Н. Методология и принципы организации системы экологического мониторинга объектов окружающей среды // *Экологические системы и приборы.* – 2016. - № 4. - С. 3-13.
- 31 Кузнецов С.Д. Основы баз данных // М.: Интернет-университет информационных технологий, 2-е изд. – 2007. – 484 с.
- 32 Peregon A., Maksyutov S., Kosykh N., Mironycheva-Tokareva N., Tamura N., Inoue G. Application of the Multi-Scale Remote Sensing and GIS to Mapping Net Primary Production in West Siberian Wetlands // *Phyton (Austria) Special issue: "APGC 2004".* – 2005. – Vol. 45, № 4. – P. 543-550.
- 33 Kivinen E., Pakarinen P. Geographical distribution of peat resources and major peatland complex types in the world // *Ann. Acad. Scient. Fenn. Ser. A.* – 1981. – Vol. 132. – P. 28-45.
- 34 Astel A.M., Chepanova I., Simeonov V. Soil contamination interpretation by the use of monitoring data analysis // *Water Air Soil Pollut.* – 2011. – Vol. 216, № 1-4. – P. 375-390.
- 35 Белобров В.П., Рожков В.А., Столбовой В.С. База данных о структурах почвенного покрова для их классификации // *Почвоведение.* – 1993. – Т. 7. – С. 83-88.
- 36 Жумадилова М.Б. Возможности использования информационных технологий в создании модели мониторинга // *Аналита и научное планирование.* – 2014. – № 1. – С. 335-344.
- 37 Gupta S.B., Mittal A. Introduction to Database Management System // *Laxmi Publications.* – 2009. – 247 p.
- 38 Марков А.С., Лисовский К.Ю. Базы данных. Введение в теорию и методологию // М.: Изд-во Финансы и статистика. – 2006. – 512 с.
- 39 Drewek K. Data Warehousing Similarities and Differences of Inmon and Kimball. URL: <http://www.beye-network.com/view/743.2011>.
- 40 Кузнецов М., Симдянов И. MySQL 5 // СПб: БХВ-Петербург. – 2010. – 1024 с.
- 41 Dedic N., Clare S. An Evaluation of the Challenges of Multilingualism in Data Warehouse Development // *International Conference on Enterprise Information Systems, 25–28 April 2016, Rome, Italy. Proceedings of the 18<sup>th</sup> International Conference on Enterprise Information Systems (ICEIS 2016).* SciTePress. – 2016. – P. 196-206.
- 42 Львов В.К. Создание систем поддержки принятия решений на основе хранилищ данных // *СУБД.* – 1997. – № 3. – С. 30-40.
- 43 Электронная региональная база данных «Отходы» Приморье / *Экологические новости.* 15.02.2008



- 44 База данных о непригодных пестицидах и территориях, ими загрязненных. Регистрационное свидетельство № 1870800557 от 13.03.2008.
- 45 Щукина В.Н. Тюменский государственный архитектурно-строительный университет // Тюмень. – 2016.
- 46 Иващук О.А., Землякова А.В., Щербинина Н.В., Петров С.П., Буряк Ж.А., Федоров В.И., Богат Д.В. Электронная модель схемы обращения с отходами // Информатика. – 2016 – Т. 2, № 223(37). – С. 162-166.
- 47 ArcGis Desktop Basic 9.3 <http://webhelp.esri.com/arcgisdesktop/9.3>. – 2008.
- 48 Butts-Wilmsmeyer C.J., Rapp S., Guthrie B. The technological advancements that enabled the age of big data in the environmental sciences: A history and future directions // *Current Opinion in Environmental Science & Health*. – 2020. – Vol. 18. – P. 63-69.
- 49 Sharafati A., Haghbin M., Dixon B., Kuma V. Application of Soft Computing Models for Simulating Nitrate Contamination in Groundwater: Comprehensive Review, Assessment and Future Opportunities // *Archives of Computational Methods in Engineering*. – 2020.
- 50 Jacobs A. The pathologies of big data // *Communications of the ACM*. – 2009. – Vol. 52, № 8. – P. 36-44.
- 51 Balakayeva G.T., Darkenbayev D.K. Modeling the processing of a large amount of data // *Вестник КазНУ. Серия "Математика, механика, информатика"*. – 2018. – Т. 97, № 1. – С. 120-126.
- 52 Liu Y., Huang J., Lu N. Research on Environmental Monitoring System Based on Microservices and Data Mining // *E3S Web of Conferences*. – 2020. – Vol. 143. – P. 2031.
- 53 Manyika J., Chui M., Brown B., Bughin J., Dobbs R., Roxburgh C., Byers A.H. Big data: The next frontier for innovation, competition, and productivity. – 2011. McKinsey Company. [www.mckinsey.com/business-functions/digital-mckinsey/our-insights/big-data-the-next-frontier-for-innovation](http://www.mckinsey.com/business-functions/digital-mckinsey/our-insights/big-data-the-next-frontier-for-innovation).
- 54 Ward J.S., Barker A. Undefined by data: A survey of big data definitions. – 2013. - arXiv, Cornell University Library. <https://arxiv.org/abs/1309.5821>.
- 55 Boettiger C., Chamberlain S., Hart E., Ram K. Building software, building community: Lessons from the rOpenSci project // *Journal of Open Research Software*. – 2015. – Vol. 3, № 1. – P. e8.
- 56 Chen M., Mao S., Liu Y. Big data: A survey // *Mobile Networks and Applications*. – 2014. – Vol. 19. – P. 171-209.
- 57 Chang W.L. NIST Big Data // *Interoperability Framework*. – 2015. – Vol. 1, Definitions. National Institute of Standards and Technology, US Department of Commerce.
- 58 The Fourth Paradigm: Data-Intensive Scientific Discovery. – 2009. <http://research.microsoft.com/enus/collaborationad/fourthparadigm>.
- 59 Dietze M.C. Ecological Forecasting // Princeton University Press. – 2017. – 288 p.

60 Ostrowski D.A. Big Data: A Strategic Perspective // International J. of Semantic Computing. – 2014. – Vol. 8, № 3. – P. 319-333.

61 Morris B.D., White E.P. The ecoData retriever: improving access to existing ecological data // PLoS ONE. – 2013. – Vol. 8. – P. 65848.

62 Duncan C., Lutes A.F, Nathan C., Benson B., Keifer M.B., Caratt J.F., Streetman S.A. FFI: a software tool for ecological monitoring // International J. of Wildland Fire. – 2009. –Vol. 18. – P. 310-314.

63 Магеррамов З.Т., Абдуллаев В.Г., Магеррамова А.З. BIG DATA: проблемы, методы анализа, алгоритмы // Информационные технологии. – 2017. – № 3. – С. 42-52.

64 Urazmatov T.Q., Nurmetova B.B., Kuzibayev X.Sh. Analysis of big data processing technologies // Conf. Series: Materials Science and Engineering. – 2020. – Vol. 862. – P. 042006.

65 The role of big data for ICT monitoring and for development // Measuring the Information Society Report. – 2014. – Ch. 5. – P. 173-205.

66 Гринин А.С., Орехов Н.А., Новиков В.Н. Математическое моделирование в экологии: Учебное пособие // М.:ЮНИТИ-Дана. – 2003. – 269 с.

67 Камышова Г.Н., Корсак В.В., Фалькович А.С., Холуденева О.Ю. Математическое моделирование в компонентах природы (интерактивный курс): Учебно-практическое пособие // Саратов: ФГБОУ ВПО СГАУ. – 2012. – 161 с.

68 Булдакова М.В. Методика стохастического факторного анализа финансовых результатов на примере ООО «Диана К» // Экономика и менеджмент инновационных технологий. – 2013. - № 5. [Электронный ресурс].

69 Чумаченко С.И. Имитационное моделирование многовидовых разновозрастных лесных насаждений // Диссертация доктора биологических наук. Мытищи. – 2006. – 287 с.

70 Романов Е.М. Моделирование роста искусственных энергетических лесов // Научный журнал КубГАУ. – 2012. – № 82(8). – С. 1-14.

71 Брюс П., Брюс Э., Гедек П. Практическая статистика для специалистов Data Science. 2-е издание – 2021.

72 Эсбенсен К. Анализ многомерных данных // Черноголовка: ИПХФ РАН. – 2005. – 140 с.

73 Щелкалин В.Н. Системный подход к синтезу математических моделей прогнозирования взаимосвязанных нестационарных временных рядов // Восточно-Европейский журнал передовых технологий. – 2015. – № 2(4). – С. 21-35.

74 Городецкий А.Е., Тарасова И.И. Нечеткое математическое моделирование плохо формализуемых процессов и систем // СПб.: Изд-во Политехи. – 2010. – 336 с.

75 Понятский В.М., Велешки С.И., Жирнова А.В. Использование метода группового учета аргументов для выбора структуры модели динамического объекта // Известия ТулГУ. Технические науки. – 2013. – № 2. – С. 255-267.

76 Ahmadi M.H., Ahmadi M.A., Mehdi Mehrpooya M., Rosen M.A. Using GMDH Neural Networks to Model the Power and Torque of a Stirling Engine // Sustainability. – 2015. – Vol. 7. – P. 2243-2255.

77 Mottaghitab M. Genetic Algorithms and Group Method of Data Handling-Type Neural Networks Applications in Poultry Science // Open access peer-reviewed chapter. – 2012. – P. 219-234.

78 Кошур В.Д. Вычислительные технологии на основе мягких вычислений // Вычислительные технологии. – 2005. – Т. 10. – С. 78-83.

79 Chang J.R., Hung C.T. Development of Pavement Maintenance and Rehabilitation Decision Model by Group Method of Data Handling (GMDH) // International Workshop on Computing in Civil Engineering. – 2007.

80 Круглов В.В., Борисов В.В. Искусственные нейронные сети. Теория и практика // М.:Горячая линия-Телеком. – 2002. – 300 с.

81 Zhang G.P. Time Series Forecasting Using a Hybrid ARIMA and Neural Network Model // Neurocomputing. – 2003. – Vol. 50. – P. 159-175.

82 Garg V. Inductive Group Method of Data Handling Neural Network Approach to Model Basin Sediment Yield // Journal of Hydrologic Engineering. – 2011. – Vol. 20, № 6.

83 Moussa Elbisy M., Osra F. Application of Group Method of Data Handling Type Neural Network for Significant Wave Height Prediction // American Journal of Neural Networks and Applications. – 2019. – Vol. 5, № 2. – P. 51-57.

84 Mehra R.K. Group method of data handling (GMDH): Review and experience - 1977 IEEE Conference on Decision and Control including the 16<sup>th</sup> Symposium on Adaptive Processes and A Special Symposium on Fuzzy Set Theory and Applications. 7-9 December 1977. New Orleans, LA, USA. – 1977. – P. 29-34.

85 Onwubolu G.C. Design of hybrid differential evolution and group method of data handling networks for modeling and prediction // Information Sciences. – 2008. – Vol. 178, № 18. – P. 3616–3634.

86 Samsudin R., Saad P., Shabri A. A hybrid GMDH and least squares support vector machines in time series forecasting // Neural Network World. – 2011. – Vol. 21, № 3. – P. 251–268.

87 Shen C., Asante-Okyere S., Ziggah Y.Y., Wang L., Zhu X. Identification Based on Wavelet Analysis and Dimensionality Reduction as Well Log Data Pre-Processing Techniques // Energies. – 2019. – Vol. 12, № 8. – P. 1-16.

88 Щелкалин В.Н. Гибридные модели и методы прогнозирования временных рядов на основе методов «Гусеница»-SSA и Бокса-Дженкинса // Восточно-Европейский журнал передовых технологий. – 2014. – Т. 5, № 4(71). – С. 43–62.

89 Braga A.L.S., Arias-Garcia J., Llanos C., Dorn M., Foltran A., Coelho L.S. Hardware implementation of GMDH-type artificial neural networks and its use to predict approximate three-dimensional structures of proteins // Reconfigurable Communication-centric Systems-on-Chip (ReCoSoC) – 7<sup>th</sup> International Workshop. – 2012. – P. 1-8.

90 Pereira I.M., Bueno E.I. Variable identification in-group method of data handling methodology // International Nuclear Atlantic Conference - INAC 2011 Belo Horizonte, MG, Brazil, October 24-28, 2011.

91 Farlow S.J. The GMDH algorithm of Ivakhnenko // The American Statistician. – 1981. – Vol. 35, № 4. – P. 210–215.

92 Ivakhnenko A.G., Ivakhnenko G.A. Problems of Further Development of the Group Method of Data Handling Algorithms. Part I // Mathematical theory of pattern recognition. Pattern Recognition and Image Analysis. – 2000. – Vol. 10, № 2. – P. 187–194.

93 Лархер В. Экология растений // М.:Мир. – 1978. – 184 с.

94 Дроздов С.Н., Холопцева Е.С., Попов Э.Г., Курец В.К. Системный подход и моделирование в эколого-физиологических исследованиях // Труды Карельского научного центра РАН. – 2010. – Т. 2. – С. 17-21.

95 Куликова Н.А., Лебедева Г.Ф. Гербициды и экологические аспекты их применения // М.:Книжный магазин ЛИБРОКОМ. – 2010. – 152 с.

96 Viman V., Mihaly Cozmuta L., Mihaly Cozmuta A., Vatca Gh., Varga C., Oprea G. Mathematical Modelling of Pollutants' Dispersion in Soil I. Case Study of Copper Dispersion // Am J Environ Sci. – 2005. – Vol. 1, № 2. – P. 126-129.

97 Stagnitti F., Parlange J-Y., Steenhuis T.S., Barry D.A., Li L., Lockington D.A., Sander G.C. Mathematical equations of the spread of pollution in soils // Hydrolog systems model – 2006. – Vol. II.

98 Афифи А., Эйзен С. Статистический анализ: Подход с использованием ЭВМ // М.:Мир. – 1982. – 388 с.

99 Nimah M.N., Hanks R.J. Model for estimation of soil water, plant and atmospheric interrelations: I. Description and sensitivity // Soil Sci. Amer. Proc. – 1973. – Vol. 37. – P. 522-527.

100 Simunek J., Sejna M., Saito H., Sakai M., van Genuchten M.T. The HYDRUS-1D Package for Simulating the Movement of Water, Heat, and Multiple Solutes in Variably Saturated Media. – V. 4.16. Riverside (CA): Department of Environmental Sciences, University of California. – 2013.

101 Hutson J., Wagenet R. An overview of LEACHM: a process-based model of water and solute movement, transformations, plant uptake and chemical reactions in the unsaturated zone // SSSA special publication. – 2014.

102 Тесленко Р.В. Обоснование технологических схем восстановления качества грунтовых вод для охраны земель от техногенных загрязнений // Автореферат на соискание кандидата технических наук. – 2001. – 23 с.

103 Аксенов С.В. Новосельцев В.Б. Организация и использование нейронных сетей (методы и технологии) // Омск:Изд-во НТЛ. – 2006. – 128 с.

104 Ефремов И.В., Кузьмин О.Н., Колобова Е.А., Дудоров В.Е. Моделирование антропогенной нагрузки на природно-технические системы // Материалы международной научной конференции «Экологические проблемы природных и антропогенных территорий», Чебоксары. – 2011. – С. 167-168.

105 Ефремов И.В. Моделирование почвенно-растительных систем // М.:ЛКИ. – 2008. – 152 с.

106 Мустафаев Ж.С., Рябцев А.Д., Козыкеева А.Т. Современный подход к мелиорации земель в Казахстане // [eessa-water.net/file/mustafaev\\_ryabtsev\\_kozykееva.pdf](http://eessa-water.net/file/mustafaev_ryabtsev_kozykееva.pdf).

107 Finzi A.C., Abramoff R.Z., Spiller K.S., Brzostek E.R., Darby B.A., Kramer M.A., Phillips R.P. Rhizosphere processes are quantitatively important components of terrestrial carbon and nutrient cycles. // *Glob Chang Biol.* – 2015. – Vol. 21. – P. 2082-2094.

108 York L.M., Carminati A., Mooney S.J., Ritz K., Bennett M.J. The holistic rhizosphere: integrating zones, processes, and semantics in the soil influenced by roots. // *J Exp Bot.* – 2016. – Vol. 67. – P. 3629-3643.

109 Vereecken H., Schnepf A., Hopmans J.W. Modeling soil processes: review, key challenges, and new perspectives // *Vadose Zone J.* – 2016. – Vol. 15.

110 Matamala R., Stover D.B. Introduction to a virtual special issue: modeling the hidden half - the root of our problem // *New Phytol.* – 2013. – Vol. 200. – P. 939-942.

111 Tinker P.B., Nye P.H. *Solute Movement in the Rhizosphere* // Oxford University Press. – 2000. – 444 p.

112 Sterckeman T., Perriguet J., Caël M., Schwartz C., Morel J.L. Plant and Soil // – 2004. – Vol. 262. – P. 289-302.

113 Lehto N.J., Davison W., Zhang H., Tych W. Theoretical comparison of how soil processes affect uptake of metals by diffusive gradients in thin films and plants // *J. Environ. Qual.* – 2006. – Vol. 35, № 5. – P. 1903-1913.

114 Barber S.A., Cushman J.H. Nitrogen uptake model for agronomic crops. In: *Modelling wastewater renovation-land treatment* // Wiley-Interscience. – 1981. – 382 p.

115 Fang Y., Yabusaki S., Ahkam A.H., Chen X., Scheibe T.D. An efficient three-dimensional rhizosphere modeling capability to study the effect of root system architecture on soil water and reactive transport // *J. Plant and Soil.* – 2019. – Vol. 441. – P. 33-48.

116 Lopez-Bucio J., Cruz-Ramirez A., Herrera-Estrella L. The role of nutrient availability in regulating root architecture // *Curr Opin Plant Biol.* – 2003. – Vol. 6. – P. 280-287.

117 Hinsinger P., Bravin M.N., Devau N., Gerard F., Le Cadre E., Jaillard B. Soil-root-microbe interactions in the rhizosphere - a key to understanding and predicting nutrient bioavailability to plants // *J. Soil Sci. Plant Nutr.* – 2008. – Vol. 8. – P. 39-47.

118 Chung J.B., Zasoski R.J. Ammonium-potassium and ammonium-calcium exchange equilibria in bulk and rhizosphere soil // *Soil Sci Soc Am J.* – 1994. – Vol. 58. – P. 1368-1375.

119 Mullins G.L., Sommers, L.E. Cadmium and zinc influx characteristics by intact corn (*Zea mays* L) seedlings. // *Plant Soil J.* – 1986. – Vol. 96. – P. 153-164.

120 Tracy J.C., Marino M.A. Flow of water in a root-soil environment // *J. Irrig. and Drain. Eng. ASCE.* – 1989. – Vol. 115, № 6. – P. 954-972.

121 Gérard F., Blitz-Frayret C., Hinsinger P., Pagès L. Modelling the interactions between root system architecture, root functions and reactive transport processes in soil // *Plant Soil J.* – 2017. – Vol. 413. – P. 161-180.

122 Christoffersen B.O., Gloor M., Fauset S., et al. Linking hydraulic traits to tropical forest function in a size-structured and trait-driven model (TFS v.1-hydro) // *Geosci Model Dev. J.* – 2016. – Vol. 9. – P. 4227-4255.

123 Garcia-Tejera O., Lopez-Bernal A., Testi L., Villalobos F.J. A soil-plant-atmosphere continuum (SPAC) model for simulating tree transpiration with a soil multi-compartment solution // *Plant Soil J.* – 2017. – Vol. 412. – P. 215- 233.

124 Manoli G., Huang C.W., Bonetti S., Domec J.C., Marani M., Katul G. Competition for light and water in a coupled soil-plant system // *Adv Water Resour J.* – 2017. – Vol. 108. – P. 216-230

125 Amenu G.G., Kumar P. A model for hydraulic redistribution incorporating coupled soil-root moisture transport // *Hydrol Earth Syst Sci J.* – 2008. – Vol. 12. – P. 55-74.

126 Dunbabin V.M., Postma J.A., Schnepf A., et al. Modelling root-soil interactions using three-dimensional models of root growth, architecture and function // *Plant Soil J.* – 2013. – Vol. 372. – P. 93-124.

127 Tournier P.H., Hecht F., Comte M. Finite element model of soil water and nutrient transport with root uptake: explicit geometry and unstructured adaptive meshing // *Transp Porous Media J.* – 2015. – Vol. 106. – P. 487-504.

128 Nietfeld H., Prenzel J. Modeling the reactive ion dynamics in the rhizosphere of tree roots growing in acid soils. I. Rhizospheric distribution patterns and root uptake of M-b cations as affected by root-induced pH and Al dynamics // *Ecol Model. J.* – 2015. – Vol. 307. – P. 48-65.

129 Хомяков Д.М. Имитационное моделирование влияния абиотических факторов на гео и агроэкосистемы для экологической экспертизы и управления продуктивностью земледелия // Автореферат диссертации доктора технических наук. М.:ИПУ РАН. – 1995. – 42 с.

130 Trapp S., McFarlane J.C. Plant Contamination: Modeling and Simulation of Organic Chemical Processes // Boca Raton USA: Lewis Publishers. – 1995. – P. 107-152.

131 Boersma L., McFarlane C., McCoy E.L. Model of coupled transport of water and solutes in plants // Spec. Report 818. Agric. Exp. Sta., Oregon State University, Corvallis, OR. – 1988. – 109 p.

132 Ouyang Y. Phytoremediation: modelling plant uptake and contaminant transport in the plant – soil – atmosphere // *J Hydrol.* – 2002. – Vol. 266. – P. 66–82.

133 Mathur S. Modeling Phytoremediation of Soils // *Pract. Period. Hazard. Toxic Radioact. Waste Manage.* – 2004. – Vol. 8. – P. 286-297.

134 Ouyang Y., Shinde D., Ma L. Q. Simulation of phytoremediation of a TNT-contaminated soil using the CTSPAC Model // *J. Environ. Qual.* – 2005. – Vol. 34. – P. 1490-1496.

135 Canales-Pastrana R.R., Paredes M. Phytoremediation dynamic model as an assessment tool in the Environmental Management // OJAppS. – 2013. – Vol. 3. – P. 208-217.

136 Sung K., Munster C.L., Corapcioglu M.Y., Drew M.C., Spark S., Rhykerd R. Phytoremediation and modeling of contaminated soil using eastern gama grass and annual ryegrass // Water Air Soil Pol. – 2004. – Vol. 159. – P. 175-195.

137 Moteszarezhadeh B., Navabzadeh M., Liyaghat A.M. Modeling Phytoremediation of Cadmium Contaminated Soil with Sunflower (*Helianthus annuus*) Under Salinity Stress // Int. J. Environ. Res. – 2016. – Vol. 10, № 1. – P. 109-118.

138 Lugli F., Mahler C.F. Phytoremediation of Metals: A Numerical Analysis // Inter. J Phytoremiat. – 2015. – Vol. 17, № 3. – P. 242-248.

139 Sanni E.S, Emeter E.M. Mathematical Modelling of in situ-Bioremediation of Crude Oil Polluted Soil // Science and Engineering Applications. – 2016. – Vol. 1, № 4. – P. 27-32.

140 Yelebe Z.R., Samuel R.J., Yelebe B.Z. Kinetic Model Development for Bioremediation of Petroleum Contaminated Soil Using Palm Bunch and Wood Ash // Inter. J. Engineering Science Invention. – 2015. – Vol. 4, № 5. – P. 40-47.

141 Li K.Y., Xu T., Cawley W.A. Colapret J.A, Bonner J.S., Ernest A., Verramachaneni P.B. Field test and mathematical modeling of bioremediation of an oil-contaminated soil. Part 2: Mathematical modeling // Waste Management. – 1994. – Vol. 14, № 7. – P. 571-579.

142 Korostina O.A., Kuyukina M.S., Ivshina I.B., Kostina L.V., Osipenko M.A., Nyashin Y.I. Modelling of processes of filtration solutions of salts of heavy metals in soil and biosurfactant // Rus J Biomech. – 2009. – Vol. 13, № 3(45). – P. 32-43.

143 Korostina O.A., Kuyukina M.S., Ivshina I.B., Kostina L.V., Osipenko M.A., Nyashin Yu.I. The use of software for solving problems related to the purification of soil from oil and heavy metals through the biosurfactant // Rus J Biomech. – 2011. – Vol. 15, № 3(53). – P. 86-92.

144 Kuyukina M.S., Ivshina I.B., Osipenko M.A., Nyashin Y.I., Korostina O.A. Model of oilflushing contaminated soil-ground under the influence of *Rhodococcus*-biosurfactant // Rus J Biomech. – 2006. – Vol. 10, № 1(31). – P. 59-67.

145 Чеботарева Э.В. Математические модели изменения концентрации нефти в загрязненных почвах под действием сорбентов и микроорганизмов // Вестник ТГГПУ. – 2011. – № 4(26). – С. 48-50.

146 Benito M., Chen-Charpentier H., Kojouharov V. Mathematical modeling of bioremediation of trichloroethylene in aquifers // Computers & Mathematics with Applications. – 2008. – Vol. 56, № 3. – P. 645-656.

147 Dickinson N.M., Baker A.J.M., Doronila A., Laidlaw S., Reeves R.D. Phytoremediation of inorganics: realism and synergies // Inter. J. Phytoremediation. – 2009. – Vol. 11, № 2. – P. 97-114.

148 Григорьев А.А., Руденко О.В., Сова Ю.А., Бородихин А.С. Применение нейронных сетей для решения задач прогнозирования поглощения тяже-

лых металлов растениями // *Материалы 3 международной научно-практической конференции, посвященной 20-летию юбилею ГНУ КНИИХП Россельхозакадемии «Инновационные пищевые технологии в области хранения и переработки сельскохозяйственного сырья» (23-24 мая 2013 г.)* – Краснодар: КНИИХПСП. – 2013. – С. 259-263.

149 Behrendt H., Bruggemann R., Morgenstern M. Numerical and analytical model of pesticide root uptake model comparison and sensitivities // *Chemosphere*. – 1995. – Vol. 30, №10. – P. 1905-1920.

150 Wagenet R.J., Rao P.S.C. Modelling pesticide fate in soils. In ‘Pesticides in the soil environment processes, impacts and modelling’ // *Soil Science Society of America: Madison, WI*. – 1990. – P. 351-399.

151 Paterson K.G., Schnoor J.L. Fate of Alachlor and Atrazine in Riparian Zone Field Site // *Water Environ Res*. – 1992. – Vol. 64. – P. 274-283.

152 Narayanan M., Davis L.C., Erickson L.E. Fate of volatile chlorinated organic compounds in a laboratory chamber with alfalfa plants // *Environ. Sci. Technol*. – 1995. – Vol. 29, № 9. – P. 2437-2444.

153 Kamalu C.I.O., Okere P.C., Egbufor U.C., Nwandikom G.I., Obijiaku J.C., Asomugha C.C. Modeling and Optimization of Phytoremediation Kinetics of Metals in Soil by A Plant Hyperaccumulator // *American Journal of Engineering Research*. – 2017. – Vol. 6, № 11. – P. 196-207.

154 Thoma G.J., Lam T.B., Wolf D.C. A Mathematical Model of Phytoremediation for Petroleum-Contaminated Soil: Model Development // *International Journal of Phytoremediation*. – 2003. – Vol. 5, № 1. – P. 41-55.

155 Santharam G., Sastry P.S., Thathachar M.A.L. Continuous action set learning automata for stochastic optimization // *J. Franklin Inst*. – 1994. – Vol. 331 – P. 607-628.

156 Nair S.S., Kang S., Zhang X., Miguez F.E., Izaurralde C., Post W.M., Dietze M.C., Lyng L.R., Wullschleider S.D. Bioenergy crop models: descriptions, data requirements, and future challenges // *GCB Bioenergy*. – 2012. – Vol. 4. – P. 620-633.

157 Vego G., Kucar-Dragicevic S., Koprivanac N. Application of multi-criteria decision-making on strategic municipal solid waste management in Dalmatia, Croatia // *Waste Manage*. – 2008. – Vol. 28. – P. 2192–2201.

158 Peng Y., Wang G., Wang H. User preferences-based software defect detection algorithms selection using MCDM // *Inform Sciences*. – 2012. – Vol. 191. – P. 3-13.

159 de Almeida A.T., Alencar M.H., Garcez T.V., Ferreira R.J.P. A systematic literature review of multicriteria and multi-objective models applied in risk management // *J Manag Maths*. – 2017. – Vol. 28, № 2. – P. 153-184.

160 Saaty R.W. *Decision Making in Complex Environment: The Analytic Hierarchy Process (AHP) for Decision Making and the Analytic Network Process (ANP) for Decision Making with Dependence and Feedback; Super Decisions: Pittsburgh, PA. USA*. – 2003.



- 161 Cobuloglu H.I., Büyüktaktın E.A. A Two-Stage Stochastic Mixed-Integer Programming Approach to the Competition of Biofuel and Food Production // *Compute Ind Eng.* – 2017. – Vol. 107. – P. 251-263.
- 162 Wang Y., Lee H., Lin K. Fuzzy TOPSIS for multi-criteria decision-making // *Inter J Math.* – 2003. – Vol. 3. – P. 367-379.
- 163 Saelee S., Pawewan B., Tongpool R., Takada J., Manusboonpurmpool K. Biomass type selection for boiler using TOPSIS Multi-Criteria model // *Inter J Environ Develop.* – 2014. – Vol. 5, № 2. – P. 181-186.
- 164 Castillo C.P., Lavalle C., Baranzelli C., Mubareka S. Modelling the spatial allocation of second-generation feedstock (lignocellulosic crops) in Europe // *Inter J Geog Inform Sci.* – 2015. – Vol. 29, № 10. – P. 1807-1825.
- 165 Thomas A., Bong A.J., Hiscock K.M. A multicriteria based review of models that predict environmental impacts of land use-change for perennial energy crops on water, carbon and nitrogen cycling // *GCB Bioenergy.* – 2012. – Vol. 5, № 3. – P. 227–242.
- 166 Demetriades-Shah T.H., Fuchs M., Kanemasu E.T., Flitcroft I. A note of caution concerning the relationship between cumulated intercepted solar radiation and crop growth // *Agr Forest Meteorol.* – 1992. – Vol. 58. – P. 193–207.
- 167 Jing Q., Conijn S.J.G., Jongschaap R.E.E., Bindraban P.S. Modeling the productivity of energy crops in different agro-ecological environments // *Biomass and Bioenergy.* – 2012. – Vol. 46. – P. 618–633.
- 168 Rui J., Tong-tong W., Jin S., Sheng G., Wei Z., Ya-jun Y., Shao-lin C., Ryusuke H. Modeling the biomass of energy crops: Descriptions, strengths and prospective // *J Integr Agric.* – 2017. – Vol. 16, № 6. – P. 1197–1210.
- 169 Tian S., Youssef M.A., Chescheir G.M., Skaggs R.W., Cacho J., Nettles J. Development and preliminary evaluation of an integrated field scale model for perennial bioenergy grass ecosystems in lowland areas // *Environ Modell Softw.* – 2016. – Vol. 84. – P. 226-239.
- 170 Lesur C., Jeuffroy M.-H., Makowski D., Riche A.B., Shield I., Yates N., Fritz M., Formowitz B., Grunert M., Jorgensen U., Laerke P.E. Modeling long-term yield trends of *Miscanthus × giganteus* using experimental data from across Europe // *Field Crops Res.* – 2013. – Vol. 149. – P. 252-260.
- 171 Williams J.R., Jones C.A., Kiniry J.R., Spanel D.A. The EPIC crop growth model // *Trans ASAE.* – 1989. – Vol. 32. – P. 497-511.
- 172 Kiniry J.R., Sanderson M.A., Williams J.R., Tischler C.R., Hussey M.A., Ocumpaugh W.R., Read J.C., Van Esbroeck G., Reed R.L. Simulating Alamo Switchgrass with the ALMANAC Model // *Agron J.* – 1996. – Vol. 80. – P. 602-606.
- 173 Clifton-Brown J.C., Neilson B., Lewandowski I., Jones M.B. The modelled productivity of *Miscanthus × giganteus* (Greef et Deu) in Ireland // *Ind Crop Prod.* – 2000. – Vol. 12. – P. 97-109.
- 174 Hastings A., Clifton-Brown J., Wattenbach M., Mitchell C.P., Smith P. The development of MISCANFOR, a new *Miscanthus* crop growth model: towards

more robust yield predictions under different climatic and soil conditions // GCB Bioenergy. – 2009. – Vol. 1. – P. 154–170.

175 Humphries S.W., Long S.P. WIMOVAC: A software package for modelling the dynamics of plant leaf and canopy photosynthesis // CABIOS. – 1995. – Vol. 11. – P. 361-371.

176 Sampson D.A., Ceulemans R. SECRETS: simulated carbon fluxes from a mixed coniferous/deciduous Belgian forest // In: Forest Ecosystem Modeling, Upscaling and Remote Sensing (eds Ceulemans R., Veroustraete F., Gond V., Van Rensbergen J.B.H.F.). – The Hague: SPB Academic Publishing. – 2000. – P. 95-108.

177 Di Vittorio A.V., Anderson R.S., White J.D., Miller N.L., Running S.W. Development and optimization of an Agro-BGC ecosystem model for C<sub>4</sub> perennial grasses // Ecol Modell. – 2010. – Vol. 221. – P. 2038-2053.

178 Nyakudya I.W., Stroosnijder L. Effect of rooting depth plant density and planting date on maize (*Zea mays* L.) yield and water use efficiency in semi-arid Zimbabwe: Modelling with AquaCrop // Agr Water Manag. – 2014. – Vol. 146. – P. 280-296.

179 Lee J., Pedroso G., Linquist B. A., Putnam D., Kessel C., Six J. Simulating switchgrass biomass production across ecoregions using the DAYCENT model // GCB Bioenergy. – 2012. – Vol. 4. – P. 521-533.

180 Nsanganwimana F., Pourrut B., Mench M., Douay F. Suitability of *Miscanthus* species for managing inorganic and organic contaminated land and restoring ecosystem services. A review // J Environ Manag. – 2014. – Vol. 143. – P. 123-134.

181 Pidlisnyuk B., Erickson L., Kharchenko S., Stefanovska T. Sustainable Land Management: Growing *Miscanthus* in Soils Contaminated with Heavy Metals // J Environ Prot, Special Issue in Environmental Remediation. – 2014. – Vol. 5. – P. 723-730.

182 Techer D., Martinez-Chois C., Laval-Gilly P., Bennisroune A., Formanek P., D’Innocenzo M., Fall J. Assessment of *Miscanthus × giganteus* for rhizoremediation of long-term PAH contaminated soils // Appl. Soil. Ecol. – 2012. – Vol. 62. – P. 42-63.

183 Полуэктов Р.А. Динамические модели агроэкосистем // Л.: Гидрометеоиздат. – 1991. – 312 с.

184 Шмидт Ю.Д., Куликов В.Е. Моделирование урожайности сельскохозяйственных культур // Вестник ТГЭУ. – 2006. – Т. 1. – С. 73-84.

185 Баденко В.Л., Топаж А.Г., Якушев В.В., Миршель В., Нендель К. Имитационная модель агроэкосистемы как инструмент теоретических исследований // Сель биол. – 2017. – Т. 52, № 3. – С. 437-445.

186 Тасекеев М.С., Тажбенов К.А. Пути развития биотопливной энергии Казахстана // Алматы: НИЦ НТИ. – 2011. – 168 с.

187 Riaman R., Sukono S., Supian S., Ismail N. Analysing the decision making for agricultural risk assessment: An application of extreme value theory // Decis. Sci. Lett. – 2022. – Vol. 10. – P. 351–360.

188 Subhadra M., Debahuti M. Gour S., Hari. Applications of Machine Learning Techniques in Agricultural Crop Production: A Review Paper. // Indian J. Sci. Technol. – 2016. – Vol. 9, Is. 38. – P. 1-14.

189 Taylor J.R. Modeling the Potential Productivity of Urban Agriculture and Its Impacts on Soil Quality Through Experimental Research on Scale-Appropriate Systems. // Front. Sustain. Food Syst., Sec. Agroecology and Ecosystem Services 4. – July 2020.

190 Taylor J. R., Lovell S. T., Wortman S. E., Chan M. Ecosystem services and tradeoffs in the home food gardens of African American, Chinese-origin and Mexican-origin households in Chicago, IL. // Renew. Agric. Food Syst. – 2017. – Vol. 32. – P. 69–86.

191 Riaman R., Sukono S., Supian S., Ismail N. Mathematical Modeling for Estimating the Risk of Rice Farmers' Losses Due to Weather Changes. // Computation. –2022. – Vol. 10, Is. 8. – P.140-150.

192 Clifton–Brown J.C., Neilson B., Lewandowski I., Jones M. The modelled productivity of *Miscanthus* × *giganteus* (GREEF et DEU) in Ireland. // Industrial Crops and Products. – 2000. –Vol. 12, Is.2. – P. 97–109.

193 Nurzhanov Ch., Pidlisnyuk V., Nurzhanova A., Naisabayeva L. Mathematical model for growing *Miscanthus* × *giganteus* at the soil contaminated by xenobiotics // The 15<sup>th</sup> Inter Phytotechnol conf., 1-5 October 2018. University of Novi Sad, Serbia. – 2018. – P. 157.

194 Найзабаева Л., Нуржанов Ч., Пидлиснюк В., Сатымбеков М. Верификация модели MISCANMOD для оценки продуктивности биоэнергетического вида мискантус в зависимости от климатических условий Казахстана (Алматинская область) // Материалы Круглого стола "Цифровизация промышленности - основа четвертой промышленной революции». 20.04.2018. Алматы.

195 Найзабаева Л., Нуржанов Ч., Пидлиснюк В., Сатымбеков М. Компьютерное моделирование продуктивности биоэнергетических видов растений (обзор) // Вестник КАЗНИТУ. – 2018. – № 5(129). – С. 34-40.

196 Nurzhanov Ch., Satymbekov M., Kubasheva A., Utenova B., Amanbayeva Zh. MiscanCalc: a model for predicting the productivity of *Miscanthus* × *giganteus* in the ecologically polluted regions of Kazakhstan // Материалы Междунар. Симп. "Устойчивая энергетика и энергомашиностроение- 2021: SUSE-2021". Казань, 18–20 февраля 2021 года Изд-во: Казанский государственный энергетический университет (Казань). – С.111-118.

197 Thornthwaite C.W. Instructions and tables for computing potential evapotranspiration and the water balance // Laboratory Climatology: Centerton, NJ, USA. – 1957. – P. 85-198.

198 Тооминг Х.Г. Солнечная радиация и формирование урожая // Л.:Гидрометеиздат. – 1977. – 200 с.

199 Fortin J.G., Anctil F., Parent L., Bolinder M.A. Site-specific early season potato yield forecast by neural network in Eastern Canada // Prec. Agric. – 2011. – Vol. 12. – P. 905-923.

200 Ruß G. Data mining of agricultural yield data: a comparison of regression models // In: Perner P. (Ed.), *Advances in Data Mining. Applications and Theoretical Aspects: 9th Industrial Conference, ICDM 2009. Leipzig, Germany, July 20-22, 2009. Proceedings.* Springer Berlin Heidelberg, Berlin, Heidelberg. – P. 24-37.

201 Себер Дж. Линейный регрессионный анализ // М.:Мир. – 1980.– 456 с.

202 Белов В.В., Чистякова В.И. Моделирование и прогнозирование бизнес-процессов с помощью алгоритмов самоорганизации формальных описаний // *Бизнес-Информатика.* – 2008. – №4(6). – С. 37-48.

203 Айвазян С.А. Классификация и снижение размерности // М.:Финансы и статистика. – 1989. – 607 с.

204 Ильин В.А., Садовничий В.А., Сендов Бл.Х. Математический анализ // Начальный курс. М.:МГУ. – 1985. – 662 с.

205 Фихтенгольц Г.М. Основы математического анализа. Том. 1 // М.:Наука. – 1968. – 440 с.

206 Калиткин Н.Н. Численные методы. – М.:Наука. – 1978. – 512 с.

207 Бахвалов Н.С. Численные методы. – М.:Наука. – 1975. –Т. 1 – 632 с.

208 Тюкачев Н.А., Рыбак К.С., Михайлова Е.Е. Программирование в Delphi для начинающих // Санкт-Петербург: БХВ-Петербург. – 2007. – 672 с.

209 Бакнелл Дж.М. Фундаментальные алгоритмы и структуры данных в Delphi. – Санкт-Петербург: ООО «ДиаСофтЮП». – 2003. – 560 с.

210 Mazakov T., Dzhomartova Sh., Nurzhanov Ch., Mazakova A., Zhakyp B. Construction of a mathematical model of the influence of various environmental factors on the plant biomass on the contaminated soil by toxic elements // *Advance technologies and computer science ICT.* – 2020. – №1. – P. 28-37.

211 Мазаков Т.Ж., Джомартова Ш.А., Мазакова А.Т., Нуржанов Ч.А., Жакып Б.М., Амирханов Б. С., Шорманов Т.С. Свидетельство о государственной регистрации прав на объект авторского права № 9968 от 19 мая 2020 года «Многорядный Эвристический Метод Самоорганизации для построения регрессионных уравнений» (программа для ЭВМ), Министерство юстиции, 2020, Астана, Казахстан.

212 Wallach D., Mearns L.O., Ruane A.C., Rotter R.P., Asseng S. Lessons from climate modeling on the design and use of ensembles for crop modeling // *Climatic Change.* – 2016.

214 Kundu S.G., Ghosh A., Kundu A., Girish G.P. Aml-ai enabled ensemble model for predicting agricultural yield // *Cogent Food & Agriculture.* – 2022. – Vol. 8, № 1. – P. 2085717.

214 Kung H-Y., Kuo T-H., Chen C-H., Tsai P-Y. Accuracy analysis mechanism for agriculture data using the Ensemble Neural Network Method // *Sustainability.* – 2016. – Vol. 8. – P. 735.

215 Chlingaryan A., Sukkarieh S., Whelan B. Machine learning approaches for crop yield prediction and nitrogen status estimation in precision agriculture: A review // *Computers and Electronics in Agriculture.* – 2018. – Vol. 151. – P. 61-69.

- 216 Subhadra M., Debahuti M., Gour Hari S. Applications of Machine Learning Techniques in Agricultural Crop Production: A Review // Paper. Indian J. Sci. Technol. – 2016. – Vol. 9. – 45-62.
- 217 Tsai C.W., Lai C.F., Chao H.C., Vasilako A.V. BIG DATA analytics a survey // J. BIG DATA. – 2015. – Vol. 2, № 21. – P. 22-32.
- 218 <https://datastart.ru/blog/read/chto-takoe-scikit-learn>
- 219 Guyon I., Elisseeff A. An Introduction to Variable and Feature Selection // J. of Machine Learning Research. – 2003. – Vol. 3. – P. 1157–1182.
- 220 Братон С.Л., Куц Дж.Н. Анализ данных в науке и технике // перевод с англ. А.А. Слинкина. – М.: ДМК Пресс. – 2021 . – 574 с.
- 221 <https://dunyaoguz.github.io/my-blog/dataframemapper.html>
- 222 [https://scikit-learn.org/stable/modules/grid\\_search.html](https://scikit-learn.org/stable/modules/grid_search.html)
- 223 [https://runebook.dev/ru/docs/scikit\\_learn/modules/feature\\_selection](https://runebook.dev/ru/docs/scikit_learn/modules/feature_selection)
- 224 <https://scikit-learn.org/stable/modules/generated/sklearn>
- 225 <https://habr.com/ru/post/428213/>
- 226 <https://www.tensorflow.org/>
- 227 <https://keras.io/>
- 228 Поспелов Г.С. Искусственный интеллект - основа новой информационной технологии // М.: Наука. – 1988. - 280 с
- 229 Погоньшев В.А., Погоньшева Д.А., Ториков В.Е. Нейронные сети в цифровом сельском хозяйстве.
- 230 Брейман Л. Сложенные регрессии // Машинное обучение. – 1996. – № 24. – С. 49–64.
- 231 Найзабаева Л., Нуржанов Ч.А., Мукимбеков М. Создание базы данных о загрязненных территориях устаревшими пестицидами в Казахстане (на примере Алматинской области) // Мат Межд. науч конф ИИВТ МОН РК «Современные проблемы информатики и вычислительных технологии» 28-29 июня 2016 – С. 192-198.
- 232 Нуржанов Ч.А., Найзабаева Л., Мукимбеков М. Создание информационной базы данных о фиторемедиантах загрязненных ксенобиотиками почв // Мат Межд конф «Информатика и прикладная математика», посв. 25 летию Независимости Казахстана и 25 летию ИИВТ МОН РК. 21-24 сентября 2016. – С. 331-338.
- 233 Mostafalou S, Abdollahi M. Pesticides: an update of human exposure and toxicity // Archives of Toxicology. – 2017. – Vol. 91, Is.2. – P. 549–599.
- 234 Sailaukhanuly Y, Nurzhanov Ch., Nurzhanova A., Carlsen L. Evaluation of the potential cancer risk of obsolete organochlorine pesticides in abandoned storehouses throughout the Almaty oblast, Kazakhstan // Int.J. Human and ecological risk assessment. – 2022. – Vol.28, Is.10. – P.1213-1227.
- 235 USEPA. 2011. Exposure factors handbook. EPA/600/R-090/052F. Environmental Protection Agency, Washington; [accessed 2021 Nov 10]. <https://www.nrc.gov/docs/ML1400/ML14007A666.pdf>.

236 USEPA. 2009. Risk assessment guidance for superfund. Vol. I: human health evaluation manual (Part F, Supplemental guidance for inhalation risk assessment), EPA-540-R-070-002, Environmental Protection Agency, Washington; [accessed 2021 Nov 10]. [https://nepis.epa.gov/ Exe/ZyNET.exe/P1002UOM.TXT?ZyActionD=ZyDocument&Client=EPA&Index=2006+Thru+2010&Docs=&Query=Time=&EndTime=&SearchMethod=1&TocRestrict=n&Toc=&TocEntry=&QField=&QFieldYear=&QField-Month=&QFieldDay=&IntQFieldOp=0&ExtQFieldOp=0&XmlQuery=&File=D%3A%5Czyfiles%5CIndex%20Data%5C06thru-10%5CTxt%5C00000006%5CP1002UOM](https://nepis.epa.gov/Exe/ZyNET.exe/P1002UOM.TXT?ZyActionD=ZyDocument&Client=EPA&Index=2006+Thru+2010&Docs=&Query=Time=&EndTime=&SearchMethod=1&TocRestrict=n&Toc=&TocEntry=&QField=&QFieldYear=&QField-Month=&QFieldDay=&IntQFieldOp=0&ExtQFieldOp=0&XmlQuery=&File=D%3A%5Czyfiles%5CIndex%20Data%5C06thru-10%5CTxt%5C00000006%5CP1002UOM).

237 ATSDR. 2005a. Public health assessment guidance manual (Update); [accessed 2021 Nov 10]. [https://www.atsdr.cdc.gov/hac/phamanual/pdfs/phagm\\_final1-27-05.pdf](https://www.atsdr.cdc.gov/hac/phamanual/pdfs/phagm_final1-27-05.pdf).

238 ATSDR. 2005b. Toxicological profile for alpha-, beta-, gamma-, and delta-hexachlorocyclohexane; [accessed 2021 Nov 10]. <https://www.atsdr.cdc.gov/toxprofiles/tp43.pdf>.

239 ATSDR. 2019. Toxicological profile for DDT, DDE, and DDD; [accessed 2021 Nov 10]. <http://www.atsdr.cdc.gov/toxprofiles/tp35.pdf>.

240 Roberts J.W., Wallace L.A., Camann D.E., Dickey P., Gilber S.G., Lewis R.G., Takaro T.K.. Monitoring and reducing exposure of infants to pollutants in house dust // *Reviews of Environmental Contamination and Toxicology*. – 2009. – Vol. 201. – P.1-39.

241 Strbac S., Stojic N., Pucarevic M., Bajic B. Organochlorine pesticides in the Tisza river (Serbia): distribution and risk assessment // *Zbornik Matice srpske za prirodne nauke*. – 2019. – Is. 136. – P.113-122.

242 Kennen R., Kirkwood N. Phyto: Principles and resources for site remediation and landscape design // Bell and Bain Ltd, Glasgow: New York. – 2015. – 346 p.

243 Панченко Л.В., Муратова А.Ю., Дубровская Е.В., Голубев С.Н., Березуцкий М.А., Турковская О.В. Атлас растений фиторемедиантов // Саратов: Научная книга. – 2015. – 560 с.

244 Байтулин И.О., Рубаник В.Г., Рахимбаев И.Р., Ситникова А.С. Зеленая «фабрика» чистого воздуха // *Алма-Ата:Наука КазССР*. – 1979. – 112 с.

245 Абдулина С.А. Список сосудистых растений Казахстана // *Алматы: Стека*. – 1999. – 187 с.

246 Naizabayeva L., Nurzhanov Ch., Orazbekov J., Tleuberdiyeva G. Confirmation for Corporate environmental information system data storage development // *Open Computer Sci*. – 2017. – Vol. 7. – P. 29-35.

247 Найзабаева Л., Нуржанов Ч.А., Султангазы Ж.Б. Моделирование информационной системы мониторинга атмосферного загрязнения // *Труды Межд науч-прак конф «Математические методы и информационные технологии макроэкономического анализа и экономической политики»*, посвященной празднованию 80-летнего юбилея академика НАН РК А.А.Ашимова, 11-12 апреля 2017, Алматы. – С.468-473.

248 Коломийцев Н.В., Матвеев А.В., Корженевский Б.И. Информационно-аналитическая веб-система «Технологии восстановления плодородия почв и рекультивации деградированных агроландшафтов». // Природоустройство. – 2019. – № 3. – P. 13-20.

249 Kamalu C.I.O., Okere P.C., Egbufor U.C., Nwandikom G.I., Obijiaku J.C., Asomugha C.C. Modeling and Optimization of Phytoremediation Kinetics of Metals in Soil by A Plant Hyperaccumulator // Amer. J. Engineering Research. – 2017. – Vol.6, Is.11. – P.196-207.

250 Thoma G.J., Lam T.B. Wolf D.C. A Mathematical Model of Phytoremediation for Petroleum-Contaminated Soil: Model Development // International Journal of Phytoremediation. – 2003. – Vol.5. No. 1. – P. 41–55.

251 Sung K., Munster C.L., Corapcioglu M.Y., Drew M.C., Spark S., Rhykerd R. Phytoremediation, and modeling of contaminated soil using eastern gamma grass and annual ryegrass // Water Air Soil Pol. – 2004. – Vol.159. – P. 175 -195.

252 Найзабаева Л.К. Нуржанов Ч.А., Мукумбеков М.Ж. Математическое моделирование в экологической биотехнологии // ДАН Республики Казахстан. – 2016. – № 1. – С.70-82.

253 Нуржанов Ч., Найзабаева Л., Пидлиснюк В., Сатымбеков М. Математическая модель изменчивости коэффициента биологического поглощения тяжелых металлов растением в зависимости от факторов окружающей среды // Вестник КАЗНИТУ – 2018. – № 5 (129). – С.83-87.

254 Nurzhanova A., Pidlisnyuk V, Abit K., Nurzhanov Ch., Kenessov B., Stefanovska T., Erickson L. Comparative assessment of using *Miscanthus × giganteus* for remediation of soils contaminated by heavy metals: a case of military and mining sites // Environmental Science and Pollution Research. – 2019. – Vol. 26. – P.13320-13333.

255 Undeman E., Czub G., McLachlan M.S. Addressing temporal variability when modeling bioaccumulation in plants // Environmental Science and Technology. – 2009. – Vol. 43. – P. 3751–3756.

256 Paterson S., Mackay D., McFarlane C. A model of organic-chemical uptake by plants from soil and the atmosphere // Environ Sci Tech. – 1994. – Vol. 28. – P. 2259-2266.

257 Verma P., George K.V., Singh H.V., Singh K., Juwarkar A., Singh R.N. Modeling rhizofiltration: heavy-metal uptake by plant roots // Environ model asses – 2006. – Vol. 11. – P. 387-394.

258 Diwan H., Ahmad A., Iqbal M. Uptake-related parameters as indices of phytoremediation potential // Biol. – 2010. – Vol. 65, № 6. – P. 1004–1011.

259 Peng Y., Kheir R.B., Adhikari K., Malinowski R., Greve M.B., Knadel M., Greve M.H. Digital mapping of toxic metals in qatari soils using remote sensing and ancillary data. // Remote Sens. . – 2016. – Vol. 8, Is. 12. – P. 1003.

260 Kidd D., Webb M., Malone B., Minasny B., McBratney A. Eighty-metre resolution 3D soil-attribute maps for Tasmania, Australia. // Soil Res. . – 2015. – Vol. 58, Is.8. – P. 932–955.

261 Kobina J., Biney M., Vašát R., Blöcher J. R., Borůvka L., Němeček K. Using an ensemble model coupled with portable X-ray fluorescence and visible near-infrared spectroscopy to explore the viability of mapping and estimating arsenic in an agricultural soil. // *Sci.Total Environ.* – 2022. – Vol. 818. – P. 151805

262 Ma Y., Zhu G., Li J. New model describing copper dose–toxicity to tomato and bok choy growth in a wide range of soils // *Int J Environ Res Public Health.* – 2019. – Vol. 16, № 2. – P. 264-274.

263 Gharbi F., Rejeb S., Ghorbal M.H., Morel J.L. Plant response to copper toxicity as affected by plant species and soil type // *J. Plant Nutr.* – 2005. – Vol. 28. – P. 379-392.

264 Nurzhanov Ch., Amirkhanova G., Naizabayeva L., Pidlisnyuk V., Satymbekov M. Mathematical Model of Kinetics Growth Plants on Contaminated Soil by Heavy Metals // II International Conference on Modern Problems of Computer Science and Computer Technology. National Academy of Sciences of the CS MES RK, September 27-30, 2017, Almaty, Kazakhstan. – 2017. – С. 232-237.

265 Нуржанов Ч., Найзабаева Л. Модель кинетики роста нового фиторемедианта *Miscanthus × giganteus* (мискантус гигантский) на загрязненной тяжелыми металлами почве // *Вестник КАЗНИТУ* – 2018. – № 6(130). – С. 83-87.

266 Горбунова Е.А., Колпак Е.П. Математические модели одиночной популяции // *Вестник Санкт-Петербург универ.* – 2012. – Т. 10, № 4. – С. 18-30.

267 <https://biconsult.ru/products/uchebnyy-kurs-po-data-science-ml-ai>

268 Pidlisnyuk V, Erickson L, Kharchenko S, Stefanovska T. Sustainable land management: growing miscanthus in soils contaminated with heavy metals. // *J Environ Prot.* – 2014. – Vol. 5. – P. 723–730.



## ПРИЛОЖЕНИЕ А

Таблица А 2.2.1 – Параметры программы Самоорганизации

Biomass	Pe	PAR	RAINFALL	Temperature
1	2	3	4	5
1,050=	2,408	8,280	0,250	15,500
2,532=	3,515	7,314	7,110	17,000
3,535=	2,152	3,542	5,080	14,000
7,215=	2,785	9,660	4,060	14,000
12,295=	4,519	11,684	3,050	14,500
17,181=	4,047	10,028	3,050	13,500
23,268=	4,807	11,178	5,100	13,500
25,129=	2,200	3,174	2,896	18,500
29,027=	2,360	6,210	2,896	18,000
35,456=	2,951	9,384	3,050	6,500
42,407=	4,214	9,476	0, 510	9,000
51,585=	4,659	12,512	0,000	11,500
60,808=	4,350	12,558	0,000	14,500
64,115=	1,764	4,462	8,890	17,000
66,279=	0,000	2,852	3,200	18,500
69,517=	1,850	3,956	3,200	16,500
77,822=	3,216	9,522	3,200	18,500
87,465=	3,256	10,350	1,600	12,500
94,551=	2,969	7,113	1,000	11,000
105,603=	4,043	10,902	1,000	10,000
112,070=	2,234	4,232	1,000	13,000
127,766=	4,245	10,258	4,060	16,500
138,685=	4,219	7,084	4,060	16,500
142,788=	2,311	2,622	7,110	17,000
145,272=	3,285	1,564	5,080	17,500
155,950=	2,674	6,624	5,100	18,500
159,336=	0,000	2,070	0,000	18,000
174,318=	3,084	9,016	0,000	16,500
187,872=	2,641	8,050	0,000	15,000
210,551=	3,435	13,340	0,000	16,500
222,359=	2,304	6,900	0,000	18,500
239,359=	2,973=	9,660	6,100 0	21,000
255,286=	3,060	9,338	2,030	23,000
269,071=	0,000	7,820	0,000	24,000
279,836=	2,490	6,026	0,000	23,500
300,642=	3,332	11,500	0,000	22,000
324,859=	3,446	13,248	0,000	14,500
349,307=	3,813	13,340	0,000	23,000
372,355=	4,034	13,846	0,000	24,000
392,449=	4,294	13,524	0,000	24,000
400,566=	0,000	6,256	0,000	24,500
413,066=	3,421	9,614	0,000	25,500
423,983=	2,785	9,522	0,000	26,000
437,536=	3,868	13,294	29,970	25,500

Продолжение таблицы 3.2.1				
451,542=	4,295	13,708	38,100	26,000
465,906=	3,805	14,030	38,100	27,500
472,561=	0,000	6,408	8,080	24,500
476,522=	0,000	2,806	5,080	19,000
497,295=	3,712	12,512	2,030	20,000
516,555=	0,000	12,144	7,870	19,500
520,999=	0,000	2,300	0,250	18,500
537,163=	3,913	8,326	7,870	18,500
557,790=	3,666	10,580	7,870	18,500
573,731=	3,495	8,142	0,000	17,500
591,136=	3,263	9,430	20,070	17,500
597,608=	0,000	3,496	20,070	18,500
601,620=	0,000	2,024	0,000	22,000
618,498=	2,805	8,464	0,000	23,000
631,162=	0,000	6,302	4,060	19,000
650,187=	3,602	9,384	4,060	22,500
664,913=	3,709	7,222	3,050	24,000
677,765=	0,000	6,256	7,110	25,500
696,267=	2,799	8,924	7,110	24,000
713,981=	0,000	8,464	2,030	22,500
730,965=	3,306	8,050	0,000	21,500
750,598=	0,000	9,246	0,000	19,500
762,383=	0,000	5,520	0,000	21,000
786,154=	0,000	11,086	0,000	20,500
791,604=	0,000	2,530	0,000	21,500
806,135=	0,000	6,716	0,000	22,500
817,431=	1,844	5,198	0,000	23,500
833,897=	2,999	7,544	2,030	24,500
868,316=	3,822	12,374	1,905	23,500
881,330=	1,454	3,313	2,030	25,500
892,253=	2,311	4,922	12,950	27,00
895,507=	0,000	1,460	13,970	24,500
899,107=	0,000	1,610	13,970	26,500
915,543=	2,705	7,314	3,050	24,000
931,540=	3,003	7,084	1,020	24,500
956,250=	3,399	10,902	0,000	24,500
967,862=	1,999	5,106	1,020	25,500
994,146=	3,904	11,546	1,020	22,500
1019,129=	4,650	10,902	9,910	21,000
1027,209=	2,037	3,542	2,030	23,000
1049,209=	3,169	9,568	7,870	26,000
1077,030=	3,891	12,098	5,080	28,000
1098,759=	0,000	9,430	6,100	27,000
1111,295=	1,861	3,496	0,000	20,000
1119,385=	0,000	2,852	0,000	22,500
1125,995=	3,312	10,120	0,000	18,200
1149,477=	3,606	10,304	0,000	19,000
1173,406=	2,790	7,314	0,000	21,000
1190,402=	0,000	4,784	0,510	22,000

Продолжение таблицы 3.2.1				
1201,525	0,000	4,784	0,510	22,500
1207,519	0,000	2,576	0,000	22,000
1214,374=	0,000	2,944	0,000	22,500
1214,374=	2,014	6,624	0,000	23,000
1249,112	0,000	8,280	0,000	22,500
1264,565=	2,014	6,624	0,000	22,500

Таблица А 2.3.1 – Параметры машинного обучения

Данные	Описание
Дата и время	Период
Max T <sup>0</sup> C	Max T <sup>0</sup> C за период
Min T <sup>0</sup> C	Min T <sup>0</sup> C за период
T <sup>0</sup> C	T <sup>0</sup> C
Max снижение T <sup>0</sup> C	Максимальная снижение T <sup>0</sup> C, за счет ветра и жары
Min снижение T <sup>0</sup> C	Минимальное снижение T <sup>0</sup> C, за счет ветра и жары
Снижение T <sup>0</sup> C	Среднее значения снижения T <sup>0</sup> C, за счет ветра и жары
Роса	Роса
Влажность	Влажность
Ветер	Ветер
Индекс жары	Индекс жары
Осадки	Типы осадков
Атмосферные осадки	Количество осадков (дождь, снег)
Защита от осадков	Период выпадения осадков.
Вероятность осадков	Вероятность осадков (дождь, снег)
Снег	Количество выпавших осадков снега
Глубина снега	Глубина снега
Скорость ветра	Средняя скорость ветра в течение 2 мин.
Порыв ветра	Скорость ветра в период порыв ветра
Направление ветра	Среднее значение направления ветра в течение 2 мин
Давление над уровнем моря	млбар
Видимость	Расстояние видимости
Облачность	Процент облачности в небе
Относительная влажность	Относительная влажность
Давление над уровнем моря	Давление над уровнем моря
Солнечная радиация	Солнечная радиация
Солнечная энергия	Солнечная энергия
Тип погоды	Информация о погоде метеослужбы
Условия	Информация о погоде других служб.
Восход	Время восхода солнца
Закат	Время заката солнца
Фаза луны	Фаза луны
Адрес	Местонахождение
Широта	Широта местонахождения
Долгота	Долгота местонахождения
Содействующие метеостанции	Метеостанция
Дата и время 1	Месяц
Дата и время 2	Часы

Таблица А.2.3.2 – Регрессионные модели без тюнинга и без отбора признаков с учетом показателей качества моделей ( $R^2$ )

Модели	$R^2$
Linear Regression	0,9903
Lasso	0,9926
Ridge	0,9676
Elastic Net	0,3672
SGD Regressor	0,9850
Random Forest Regressor	0,9912
Decision Tree Regressor	0,9913
Gradient Boosting Regressor	0,9944
XGB Regressor	0,9923
Примечание: $R^2$ - коэффициент детерминации.	

Таблица А 2.3.3 – Регрессионные модели без тюнинга и без отбора признаков с учетом показателей качества моделей (MSE, MAE, MAPE и  $R^2$ )

Модели	MAE	MAPE	$R^2$	MSE
Linear Regression	33,49	1,66	0,992	1820,71
Lasso	33,68	1,69	0,993	1783,7
Ridge	38,79	1,34	0,991	2215,87
Elastic Net	344,22	8,71	0,390	157013
SGD Regressor	34,98	1,62	0,992	1921,89
Decision Tree Regressor	23,37	0,06	0,995	1153,43
Gradient Boosting Regressor	17,41	0,069	0,9977	585,95
Random Forest Regressor	18,3	0,084	0,9975	623,7
XGB Regressor	17,6	0,064	0,9975	637,25
Stacking Regressor (without Linear Regression, Elastic Net)	17,47	0,250	0,9978	545,62
Stacking Regressor	17,81	0,234	0,997	564,67
Voting Regressor (without Linear Regression, Elastic Net)	21,81	0,6612	0,9971	732,79
Voting Regressor	45,13	0,320	0,9884	2978
Keros	28,94	0,24	0,994	1433,71
H <sub>2</sub> OAuto ML	4,47	0,081	0,9998	39,04
Примечание: $R^2$ – коэффициент детерминации; MSE – средняя квадратная ошибка; MAE – средняя абсолютная ошибка; MAPE – средняя абсолютная ошибка в процентах.				

Таблица А 2.3.4 – Регрессионные модели с тюнингом и без отбора признаков с учетом показателей качества достоверности ( $R^2$ )

Модели	$R^2$
Linear Regression	0,9903
Lasso	0,9916
Ridge	0,9905
Elastic Net	0,9908
SGD Regressor	0,9907
Random Forest Regressor	0,9910
Decision Tree Regressor	0,9903
Gradient Boosting Regressor	0,9902
XGB Regressor	0,9949
Примечание: $R^2$ – коэффициент детерминации.	

Таблица А 2.3.5 – Регрессионные модели с тюнингом и без отбора признаков с учетом показателей качества достоверности (MSE, MAE, MAPE и R<sup>2</sup>)

Модели	MAE	MAPE	R <sup>2</sup>	MSE
Linear Regression	33,4	1,66	0,992	1820,7
Lasso	32,58	1,7	0,993	1731,97
Ridge	33,74	1,64	0,9929	1808,03
Elastic Net	33,69	1,63	0,993	1800,92
SGD Regressor	34,98	1,62	0,992	1921,89
Decision Tree Regressor	21,63	0,013	0,996	844,019
Gradient Boosting Regressor	15,05	0,065	0,998	438,35
Random Forest Regressor	18,27	0,0865	0,997	611,12
XGB Regressor	15,3	0,065	0,998	471,19
Stacking Regressor (without Linear Regression, Elastic Net)	15,57	0,2	0,9982	451,71
StackingRegressor	15,25	0,17	0,998	434,47
Voting Regressor (without Linear Regression, Elastic Net)	20,43	0,68	0,997	634
Voting Regressor	20	0,63	0,997	624
Keros	28,94	0,24	0,994	1433,71
H <sub>2</sub> OAuto ML	2,91	-	0,9999	15,3
Примечание: R <sup>2</sup> – коэффициент детерминации; MSE – средняя квадратная ошибка; MAE – средняя абсолютная ошибка; MAPE – средняя абсолютная ошибка в процентах.				

Таблица А 2.3.6 – Регрессионные модели для отбора признака [1] без тюнинга с учетом показателей качества достоверности ( $R^2$ )

Модели	$R^2$
Linear Regression	0,9918
Lasso	0,9926
Ridge	0,9785
Elastic Net	0,3783
SGD Regressor	0,9907
Random Forest Regressor	0,9923
Decision Tree Regressor	0,9910
Gradient Boosting Regressor	0,9928
XGB Regressor	0,9921
Примечание: $R^2$ – коэффициент детерминации.	

Таблица А 2.3.7 – Регрессионные модели для отбора признака [2] без тюнинга с учетом показателей качества достоверности ( $R^2$ )

Модели	$R_2$
Linear Regression	0,9922
Lasso	0,9932
Ridge	0,974
Elastic Net	0,360
SGD Regressor	0,9885
Random Forest Regressor	0,9916
Decision Tree Regressor	0,9909
Gradient Boosting Regressor	0,9926
XGB Regressor	0,9922
Примечание: $R^2$ – коэффициент детерминации.	



Таблица А 2.3.8 – Регрессионные модели для отбора признака [1] без тюнинга с учетом показателей качества достоверности (MSE, MAE, MAPE и R<sup>2</sup>)

Модели	MAE	MAPE	R <sup>2</sup>	MSE
Linear Regression	32,12	1,68	0,993	1666,85
Lasso	33,79	1,7	0,993	1786,67
Ridge	37,97	1,41	0,99	2096,56
Elastic Net	344,10	8,7	0,39	156646,41
SGD Regressor	33,27	1,68	0,993	1743,2
Decision Tree Regressor	22,86	0,049	0,995	1089,18
Gradient Boosting Regressor	17,89	0,08	0,9976	597,79
Random Forest Regressor	18,48	0,08	0,997	632,44
XGB Regressor	18,22	0,06	0,997	682,02
Stacking Regressor (without Linear Regression, Elastic Net)	17,74	0,26	0,9978	554,17
Stacking Regressor	17,74	0,26	0,997	554,17
Voting Regressor (without Linear Regression, ElasticNet)	21,85	0,68	0,997	713,00
Voting Regressor	21,85	0,68	0,997	713,70
Keros	20,79	0,12	0,997	709,19
H <sub>2</sub> OAuto ML	6,630	0,085	0,9995	86,35

Примечание: R<sup>2</sup> – коэффициент детерминации; MSE – средняя квадратная ошибка; MAE – средняя абсолютная ошибка; MAPE – средняя абсолютная ошибка в процентах.

Таблица А 2.3.9 – Регрессионные модели для отбора признака [2] без тюнинга с учетом показателей качества достоверности (MSE, MAE, MAPE и R<sup>2</sup>)

Модели	MAE	MAPE	R <sup>2</sup>	MSE
Linear Regression	32,32	1,68	0,993	1690,85
Lasso	33,69	1,69	0,993	1776,69
Ridge	36,78	1,42	0,992	2033,92
Elastic Net	343,70	8,68	0,39	156419,244
SGD Regressor	33	1,67	0,993	1746,2
Decision Tree Regressor	24,04	0,068	0,995	1196,68
Gradient Boosting Regressor	17,28	0,06	0,9977	576,86
Random Forest Regressor	17,95	0,07	0,997	601,39
XGB Regressor	17,84	0,069	0,997	642,92
Stacking Regressor (without Linear Regression, Elastic Net)	17,19	0,25	0,9979	532,05
Stacking Regressor	17,19	0,25	0,9979	532,05
Voting Regressor (without Linear Regression, Elastic Net)	21,57	0,68	0,997	715
Voting Regressor	21,57	0,68	0,997	715,85
Keros	21,63	0,12	0,996	774,78
H <sub>2</sub> OAutoML	4,38	0,079	0,9998	35,63
Примечание: R <sup>2</sup> – коэффициент детерминации; MSE – средняя квадратная ошибка; MAE – средняя абсолютная ошибка; MAPE – средняя абсолютная ошибка в процентах.				

Таблица А 2.3.10 – Регрессионные модели для отбора признака [1] с тюнингом с учетом показателей качества достоверности (R<sup>2</sup>)

Модели	R <sup>2</sup>
Linear Regression	0,991
Lasso	0,992
Ridge	0,9918
Elastic Net	0,9918
SGD Regressor	0,9881
Random Forest Regressor	0,9917
Decision Tree Regressor	0,990
Gradient Boosting Regressor	0,986
XGB Regressor	0,994
Примечание: R <sup>2</sup> - коэффициент детерминации.	

Таблица А 2.3.11 – Регрессионные модели для выбора признака [2] с тюнингом с учетом показателей качества достоверности ( $R^2$ )

Модели	$R^2$
Linear Regression	0,9922
Lasso	0,9926
Ridge	0,9922
Elastic Net	0,9922
SGD Regressor	0,9857
Random Forest Regressor	0,9916
Decision Tree Regressor	0,9900
Gradient Boosting Regressor	0,9866
XGB Regressor	0,9940
Примечание: $R^2$ – коэффициент детерминации.	

Таблица А 2.13 – Регрессионные модели для отбора признака [1] с тюнингом с учетом показателей качества достоверности (MSE, MAE, MAPE и  $R^2$ )

Модели	MAE	MAPE	$R^2$	MSE
Linear Regression	32,12	1.68	0.9935	1666.77
Lasso	32	1.68	0.9935	1656.77
Ridge	32,12	1.68	0.9935	1665.83
Elastic Net	32,1	1.68	0.9935	1664.11
SGD Regressor	33,27	1.68	0.993	1743.2
Decision Tree Regressor	20,43	0.062	0.997	769.43
Gradient Boosting Regressor	15,6	0.058	0.998	464.41
Random Forest Regressor	18,2	0.07	0.9975	621.26
XGB Regressor	15,89	0.061	0.998	514.03
Stacking Regressor (without Linear Regression, ElasticNet)	16,61	0.22	0.9981	479.855
Stacking Regressor	16,61	0.22	0.9981	479.49
Voting Regressor (without Linear Regression, ElasticNet)	20,59	0.71	0.997	640.88
Voting Regressor	22,74	0.93	0.996	783.31
Keros	20,79	0.12	0.997	709.19
H <sub>2</sub> OAuto ML	6,630	0.085	0.9995	86.35
H <sub>2</sub> O Gradient Boosting Estimator	18,6	0.16	0.9975	629.37
Примечание: $R^2$ – коэффициент детерминации; MSE – средняя квадратная ошибка; MAE – средняя абсолютная ошибка; MAPE – средняя абсолютная ошибка в процентах.				

Таблица А 2.14 – Регрессионные модели для отбора признака [2] с тюнингом с учетом показателей качества достоверности (MSE, MAE, MAPE и R<sup>2</sup>)

Модели	MAE	MAPE	R <sup>2</sup>	MSE
Linear Regression	32,32	1,68	0,9934	1690,5
Lasso	32,31	1,69	0,9935	1672,18
Ridge	32,32	1,68	0,9934	1690,51
Elastic Net	32,32	1,68	0,9935	1690,55
SGD Regressor	33,02	1,67	0,9934	1746,45
Decision Tree Regressor	21,5	0,067	0,996	844,98
Gradient Boosting Regressor	14,8	0,066	0,998	419,15
Random Forest Regressor	17,91	0,078	0,9976	595,96
XGB Regressor	15,03	0,066	0,998	421,19
Stacking Regressor (without Linear Regression, ElasticNet)	15,56	0,26	0,9981	425,061
StackingRegressor	15,58	0,268	0,9983	425,89
Voting Regressor (without Linear Regression, Elastic Net)	19,25	0,55	0,997	590,80
Voting Regressor	19,48	0,67	0,997	594,80
Keros	21,63	0,12	0,996	774,78
H <sub>2</sub> OAutoML	4,38	0,079	0,9998	35,630
H <sub>2</sub> OGradientBoostingEstimator	18,1	0,13	0,9976	607,47
Примечание: R <sup>2</sup> - коэффициент детерминации; MSE - средняя квадратная ошибка; MAE - средняя абсолютная ошибка; MAPE - средняя абсолютная ошибка в процентах.				

Таблица А 2.3.14 – Анализ эффективности и точности регрессионных моделей с помощью метрики ошибок

Регрессионные методы	RMSE	MAE	R <sup>2</sup>
Linear Regression	5208,87	37,21	0,97
ElasticNet	5216,01	37,36	0,97
Linear SVR	134768,39	100,67	0,36
LassoLars	1467,48	31,44	0,99
SGD Regressor	4,25	16957371182189,20	-2,0
Support Vector Regression (SVR)	272393,64	434,55	-0,28
ARDR Regression	1374,22	27,80	0,99
Bayesian Ridge	5565,35	37,84	0,97
Gradient Boosting Regressor	220,04	19,3	0,99
Extra Trees Regressor	70,0	6,83	0,99
Ada Boost Regressor	867,67	11,3	0,99
Bagging Regressor	450,0	17,0	0,99
Passive Aggressive Regressor	20852,16	67,77	0,90
Random Forest Regressor	180,8	11,3	0,99
Примечание: RMSE – ошибка корневого квадрата; MSE – средняя квадратная ошибка; R <sup>2</sup> – коэффициент детерминации.			

Приложение А Таблица 3.2 – Значения суммарного канцерогенного риска для различных путей воздействия пестицидов из загрязненной почвы

№	Старые склады пестицидов, поселки	Суммарный канцерогенный риск					
		ДДТ	ДДД	ДДЭ	Альфа ГХЦГ	Бета ГХЦГ	Гамма-ГХЦГ
<b>Карасайский район</b>							
1	Теректы	3.6E - 07	0.0E + 00	1.3E - 07	4.7E - 10	3.1E - 10	4.1E - 11
2	Теректы	3.3E - 07	0.0E + 00	6.3E - 08	1.7E - 10	1.7E - 10	1.2E - 11
3	Бекболат	3.2E - 06	0.0E + 00	1.6E - 06	1.4E - 11	1.2E - 10	2.9E - 12
4	Шалкар	1.1E - 07	9.4E - 09	3.8E - 07	1.4E - 11	5E - 12	1.4E - 10
5	Чемолган	1.6E - 06	2.2E - 07	4.9E - 07	2E - 10	1.2E - 10	1.9E - 11
6	Береке	1.4E - 06	1.1E - 07	6.8E - 07	8.1E - 11	9.6E - 10	3.4E - 11
<b>Талгарский район</b>							
1	Бескайнар	2.9E - 08	5.6E - 07	9.1E - 07	3.5E - 10	1.4E - 10	4.7E - 12
2	Бельбулак	2.9E - 08	0.0E + 00	2.7E - 08	1E - 10	1.1E - 10	1.2E - 11
3	Талгар	0.0E + 00	0.0E + 00	0.0E + 00	0	0	0
4	Кайрат	3.0E - 06	6.5E - 07	2.7E - 06	1.2E - 09	2.7E - 10	1.6E - 11
5	Кызылкайрат	1.3E - 05	2.5E - 06	4.0E - 06	8.4E - 09	8.8E - 10	5.3E - 09
6	Кызылкайрат	3.9E - 07	2.8E - 07	5.9E - 08	1.7E - 09	1.2E - 10	2.3E - 09
<b>Енбекши-казахский район</b>							
1	Космос	5.9E - 07	7.8E - 08	5.3E - 07	7.5E - 09	1.2E - 09	1.4E - 10
2	Космос	5.2E - 07	8.5E - 08	7.0E - 07	5.3E - 09	1.2E - 09	9.4E - 11
3	Акчи	0.0E + 00	0.0E + 00	0.0E + 00	0	6E - 11	0
4	Акчи	0.0E + 00	0.0E + 00	0.0E + 00	0	6.57E - 11	0
5	Балтабай	6.4E - 07	0.0E + 00	7.2E - 07	8.8E - 11	2.3E - 10	3.5E - 12
6	Балтабай	4.0E - 07	9.4E - 09	5.4E - 07	3.7E - 10	2.3E - 10	7E - 12
7	Балтабай	7.4E - 08	0.0E + 00	8.2E - 08	0	9.8E - 11	0
8	Саймасай	1.5E - 06	1.9E - 07	7.4E - 07	2.8E - 09	2.2E - 11	4.8E - 10
9	Саймасай	1.4E - 06	1.5E - 07	7.1E - 07	3.2E - 09	2.4E - 11	4.4E - 10
<b>Илийский район</b>							
1	Жанадаур	2.6E - 07	6.7E - 09	7.4E - 08	0	0	1.2E - 12
2	Карой	0.0E + 00	0.0E + 00	0.0E + 00	1.2E - 10	8.4E - 11	4.1E - 12
<b>Балхашский район</b>							
1	Береке	0.0E + 00	0.0E + 00	0.0E + 00	0	5.6E - 11	0
2	Бакбакты	0.0E + 00	0.0E + 00	0.0E + 00	0	6.6E - 11	0
3	Бакбакты	0.0E + 00	0.0E + 00	0.0E + 00	0	5E - 11	9.3E - 12
4	Бакбакты	0.0E + 00	0.0E + 00	0.0E + 00	1.2E - 09	4.1E - 10	1E - 10
5	Бакбакты	0.0E + 00	0.0E + 00	0.0E + 00	0	6.4E - 11	0
<b>Уйгурский район</b>							
1	Большой Аксу	0.0E + 00	0.0E + 00	0.0E + 00	7.4E - 11	8E - 12	4E - 11
2	Ават	0.0E + 00	0.0E + 00	0.0E + 00	4.7E - 11	0	1.2E - 12
3	Актам	0.0E + 00	0.0E + 00	0.0E + 00	0	2E - 12	0
4	Кыргызсай	0.0E + 00	0.0E + 00	0.0E + 00	0	2.6E - 11	0
5	Сумбе	0.0E + 00	0.0E + 00	0.0E + 00	0	3.2E - 11	0
<b>Ескельдинский район</b>							
1	Алдабергенова	0.0E + 00	0.0E + 00	3.8E - 08	2E - 11	1.3E - 10	1.2E - 12
2	Алдабергенова	4.8E - 08	1.5E - 08	5.7E - 08	0	1.6E - 10	0
3	Алдабергенова	3.7E - 06	3.7E - 06	5.5E - 06	3.4E - 09	1.7E - 10	2E - 09
4	Алдабергенова	0.0E + 00	0.0E + 00	0.0E + 00	0	9.2E - 11	

Продолжение таблицы 3.2							
5	Бактыбай	0.0E + 00	0.0E + 00	0.0E + 00	0	4E -11	0
6	Бактыбай	0.0E + 00	0.0E + 00	0.0E + 00	0	5.4E - 11	0
7	Бактыбай	3.6E- 08	2.3E - 08	5.0E - 08	0	5.4E -11	0
8	Бактыбай	0.0E + 00	0.0E + 00	0.0E + 00	0	3.8E -11	0
Жамбульский район							
1	Каракестек	3.1E -07	8.5E - 08	1.7E - 08	3.4E - 10	1.2E - 11	3.1E -10
2	Шиын	3.8E - 09	0.0E + 00	3.8E - 09	1.3E -10	1.9E -10	2.1E -11
3	Мынбаево	0.0E + 00	0.0E + 00	0.0E + 00	1.7E - 11	4E - 11	3.5E -12
4	Умбетали	9.5E - 09	0.0E + 00	0.0E + 00	1.4E -11	0	2.4E - 11

## 8 Using radius find optimum for plant count

```
# Define the depth in meters
depth = 0.2 # meters

# Define the area in square meters
area = 1 # square meter

# https://www.sciencedirect.com/science/article/pii/S2095311918619306
# Define the bulk density in kilograms per cubic meter
bulk_density = 1.29*100 # kilograms per cubic meter

# Calculate the volume of soil in cubic meters
volume = area * depth

# Calculate the weight of the soil in kilograms
weight = volume * bulk_density # kilograms per cubic meter

# Print the result
print("The weight of soil in a 1-meter square with a depth of 20 cm is:", weight, "kilograms")
```

executed in 4ms, finished 18:15:20 2023-04-02

The weight of soil in a 1-meter square with a depth of 20 cm is: 25.8 kilograms

```
25.8 * 5 # 1-meter square
```

executed in 4ms, finished 18:15:20 2023-04-02

129.0

Рисунок А 5.1 – Скрипшот кода программы информационной системы для расчета оптимальных условия для роста растений на загрязненной ТЭ почве



## 9 Period of full clearance

```
In [34]: import math

# Calculate the area of the circle with a radius of 50 cm
radius = 30 # cm plant with radius 30 cm
radius = radius/100 # m
area = (radius * radius) * math.pi #m^2

plants_per_hectare = 10000/area # plants in hectare

executed in 3ms, finished 18:15:20 2023-04-02
```

```
In [35]: area

executed in 3ms, finished 18:15:20 2023-04-02
```

```
Out[35]: 0.2827433388230814
```

```
In [36]: plants_per_hectare

executed in 4ms, finished 18:15:20 2023-04-02
```

```
Out[36]: 35367.765131532295
```

```
In [37]: # How much plant can be planted in meter square
plants_per_meter = plants_per_hectare / 10000
plants_per_meter

executed in 4ms, finished 18:15:20 2023-04-02
```

```
Out[37]: 3.5367765131532294
```

```
In [38]: absorbsion_for_each_element

executed in 9ms, finished 18:15:20 2023-04-02
```

```
Out[38]:
```

	As	Co	Cr	Cu	Mn	Ni	Pb	Sr	V	Zn
0	1.502	3.452	8.42	31.91	247.4	8.252	135.68	145	5.51	176.8

```
In [39]: plants_per_hectare

executed in 4ms, finished 18:15:20 2023-04-02
```

```
Out[39]: 35367.765131532295
```

```
In [40]: df_absorb_for_hectar = pd.DataFrame()
for col in absorbsion_for_each_element.columns:
    absorb = absorbsion_for_each_element[col].values * plants_per_hectare #mg in kg for the hectar
    absorb_for_hectar = absorb / (1000 * 1000) #kg in hectar
    df_absorb_for_hectar[col] = absorb_for_hectar

executed in 7ms, finished 18:15:20 2023-04-02
```

```
In [41]: df_toxic_elemnts_in_soil

executed in 8ms, finished 18:15:20 2023-04-02
```

```
Out[41]:
```

	As	Co	Cr	Cu	Mn	Ni	Pb	Sr	V	Zn
0	22.5	18	44	31.5	810	39.2	230	81.3	60	180

Рисунок А 5.2 – Скрипшот кода программы информационной системы для получения информации о количестве растений на 1 га, загрязненной ТЭ почве

df_weather_2022_transform[feature]													
executed in 16ms, finished 23:00:57 2023-04-01													
	temp	dew	humidity	precip	windgust	windspeed	winddir	sealevelpressure	cloudcover	visibility	moonphase	datetime_1	datetime_2
0	0.467391	0.521739	0.409556	0.000000	0.052632	0.000000	0.160985	0.425439	0.631868	0.282051	0.000000	0.0	0.000000
1	0.521739	0.586957	0.383959	0.000000	0.045680	0.000000	0.191063	0.232456	0.654945	0.256410	0.051546	0.0	0.033333
2	0.472826	0.644928	0.457338	0.000000	0.571003	0.272727	0.654540	0.254386	0.726374	0.717949	0.082474	0.0	0.066667
3	0.413043	0.876812	0.744027	0.016846	0.145978	0.181818	0.723288	0.434211	0.969231	0.410256	0.113402	0.0	0.100000
4	0.233696	0.855072	0.948805	0.298199	0.248262	0.181818	0.503581	0.688596	1.000000	0.256410	0.144330	0.0	0.133333
...	...	...	...	...	...	...	...	...	...	...	...	...	...
114	0.342391	0.043478	0.203072	0.000000	0.249255	0.090909	0.177886	0.631579	0.154945	0.230769	0.896907	1.0	0.733333
115	0.434783	0.014493	0.124573	0.000000	0.035750	0.000000	0.698654	0.657895	0.518681	0.461538	0.927835	1.0	0.766667
116	0.434783	0.304348	0.254266	0.008932	0.035750	0.095960	0.183901	0.692982	0.856044	0.282051	0.958763	1.0	0.800000
117	0.331522	0.413043	0.424915	0.032703	0.642502	0.234848	0.592094	0.763158	0.825275	0.435897	0.989691	1.0	0.833333
118	0.407609	0.217391	0.266212	0.009101	0.074479	0.090909	0.099398	0.771930	0.315385	0.282051	0.000000	1.0	0.866667

119 rows x 13 columns

Рисунок А 5.3 – Климатические данные за 2022 год

## ПРИЛОЖЕНИЕ Б

ҚАЗАҚСТАН РЕСПУБЛИКАСЫ  РЕСПУБЛИКА КАЗАХСТАН

**АВТОРЛЫҚ ҚҰҚЫҚПЕН ҚОРҒАЛАТЫН ОБЪЕКТІЛЕРГЕ ҚҰҚЫҚТАРДЫҢ  
МЕМЛЕКЕТТІК ТІЗІЛІМГЕ МӘЛІМЕТТЕРДІ ЕНГІЗУ ТУРАЛЫ**

**КУӘЛІК**

2018 жылғы « 13 » желтоқсан № 976

Автордың (лардың) жөні , аты, әкесінің аты (егер ол жеке басын куәландыратын құжатта көрсетілсе):  
НАЙЗАБАЕВА ЛЯЗАТ ; НУРЖАНОВ ЧИНГИЗ АСКАРОВИЧ

Авторлық құқық объектісі: деректер базасы

Объектінің атауы: База данных о растениях фиторемедиантах

Объектіні жасаған күні: 25.06.2017

**СВИДЕТЕЛЬСТВО**

**О ВНЕСЕНИИ СВЕДЕНИЙ В ГОСУДАРСТВЕННЫЙ РЕЕСТР  
ПРАВ НА ОБЪЕКТЫ, ОХРАНЯЕМЫЕ АВТОРСКИМ ПРАВОМ**

№ 976 от « 13 » декабря 2018 года

Фамилия, имя, отчество, (если оно указано в документе, удостоверяющем личность) автора (ов):  
НАЙЗАБАЕВА ЛЯЗАТ ; НУРЖАНОВ ЧИНГИЗ АСКАРОВИЧ

Вид объекта авторского права: база данных

Название объекта: База данных о растениях фиторемедиантах

Дата создания объекта: 25.06.2017





Құжат түпнұсқалығын <http://www.kazpatent.kz/ru> сайтының  
"Авторлық құқық" бөлімінде тексеруге болады <https://copyright.kazpatent.kz>

Подлинность документа возможно проверить на сайте [kazpatent.kz](http://kazpatent.kz)  
в разделе «Авторское право» <https://copyright.kazpatent.kz>

Подписано ЭЦП Оспанов Е. К.



ҚАЗАҚСТАН РЕСПУБЛИКАСЫ



РЕСПУБЛИКА КАЗАХСТАН

АВТОРЛЫҚ ҚҰҚЫҚПЕН ҚОРҒАЛАТЫН ОБЪЕКТІЛЕРГЕ ҚҰҚЫҚТАРДЫҢ  
МЕМЛЕКЕТТІК ТІЗІЛІМГЕ МӘЛІМЕТТЕРДІ ЕНГІЗУ ТУРАЛЫ

КУӘЛІК

2018 жылғы «13» желтоқсан № 971

Автордың (лардың) жөні, аты, әкесінің аты (егер ол жеке басын куәландыратын құжатта көрсетілсе):

НАЙЗАБАЕВА ЛЯЗАТ, НУРЖАНОВ ЧИНГИЗ АСКАРОВИЧ

Авторлық құқық объектісі: деректер базасы

Объектінің атауы: База данных о загрязненных пестицидами территориях

Объектіні жасаған күні: 25.06.2016

СВИДЕТЕЛЬСТВО

О ВНЕСЕНИИ СВЕДЕНИЙ В ГОСУДАРСТВЕННЫЙ РЕЕСТР  
ПРАВ НА ОБЪЕКТЫ, ОХРАНЯЕМЫЕ АВТОРСКИМ ПРАВОМ

№ 971 от «13» декабря 2018 года

Фамилия, имя, отчество, (если оно указано в документе, удостоверяющем личность) автора (ов):

НАЙЗАБАЕВА ЛЯЗАТ, НУРЖАНОВ ЧИНГИЗ АСКАРОВИЧ

Вид объекта авторского права: база данных

Название объекта: База данных о загрязненных пестицидами территориях

Дата создания объекта: 25.06.2016



Құжат түпнұсқасын <http://www.kazpatent.kz/ru> сайтының  
"Авторлық құқық" бөлімінде тексеруге болады. <https://copyright.kazpatent.kz>

Подлинность документа возможно проверить на сайте [kazpatent.kz](http://kazpatent.kz)  
в разделе «Авторское право» <https://copyright.kazpatent.kz>

Подписано ЭЦП

Оспанов Е. К.



ҚАЗАҚСТАН РЕСПУБЛИКАСЫ



РЕСПУБЛИКА КАЗАХСТАН

## СВИДЕТЕЛЬСТВО

О ВНЕСЕНИИ СВЕДЕНИЙ В ГОСУДАРСТВЕННЫЙ РЕЕСТР  
ПРАВ НА ОБЪЕКТЫ, ОХРАНЯЕМЫЕ АВТОРСКИМ ПРАВОМ

№ 9968 от «19» мая 2020 года

Фамилия, имя, отчество, (если оно указано в документе, удостоверяющем личность) автора (ов):

МАЗАКОВ ТАЛГАТ ЖАКУПОВИЧ, ДЖОМАРТОВА ШОЛПАН АБТРАЗАКОВНА, МАЗАКОВА  
ӨНГЕРІМ ТАЛГАТҚЫЗЫ, НУРЖАНОВ ЧИНГИЗ АСКАРОВИЧ, ЖАҚЫП БОТАҒӨЗ МАРАТҚЫЗЫ,  
АМІРХАНОВ БАҒЫРЖАН САЙДАШЕВИЧ, ШОРМАНОВ ТИМУР СЕРИКХАНОВИЧ

Вид объекта авторского права: программа для ЭВМ

Название объекта: Многорядный Эвристический Метод Самоорганизации для построения регрессионных  
уравнений

Дата создания объекта: 15.05.2020



Служба государственной регистрации интеллектуальной собственности  
"Авторские права" Белгінде тіркелген болсаңыз: <http://copyright.kazpatent.kz>

Подлинность документа возможно проверить на сайте [kazpatent.kz](http://kazpatent.kz)  
в разделе «Авторские права»: <http://copyright.kazpatent.kz>

Подписано ЭЦП

Куантыров Е.С.

## ПРИЛОЖЕНИЕ В Акты внедрения

HUMAN HEALTH  
INSTITUTE



HUMAN HEALTH  
INSTITUTE

110000 Казахстан Республикасы, Астана қаласы, Желтоқсан көшесі, 33/1, бөліме 102  
010000 Республика Казахстан, город Астана, ул. Желтоқсана, 33/1, офис 102  
Тел./Факс +7(7172)7797489, моб. +7 777 778 78 66, e-mail: hhkaz@gmail.com  
BHO58549012387, свидетельство о гос. регистрации №7379-1992-00

Шығыс/Бел. №139  
Күні/Дата 08.11.2018г.

КазНУ им аль-Фараби,  
г.Алматы  
докторанту Нуржанову Ч.А.  
Научному руководителю  
д.т.м. Найзабаевой Л.К.

*о внедрении «Базы данных о загрязненных пестицидами территориях»*

Становится общезвестным фактом, что особо опасные пестициды (ООП), применявшиеся ранее в странах с развитой агропромышленностью, в том числе и Казахстане, представляют до сих пор угрозу для здоровья человека и окружающей среды. Это связано с отсутствием в настоящее время достоверной исторической информации о ассортименте ООП, возможных их остатках на складах, длительности хранения, географии и особенностей применения пестицидов в регионах страны. До сих пор в Республике не проведена полная инвентаризация таких мест размещения складов сельхозхимии, захоронения пестицидов, нет региональных карт.

В этой связи разработанная в рамках подготовки докторской диссертации «База данных о загрязненных пестицидами территориях» по 10 регионам Алматинской области, учитывающая количество устаревших пестицидов, идентификацию устаревших пестицидов, количество контейнеров, уровень загрязнения почв хлорорганическими пестицидами представляет большой интерес для дальнейшего управления подобными объектами в целом по Казахстану. Внедрение нашей разработки в практическом аспекте, **несомненно, актуально.** Это позволит в решении, прежде всего, проблем с

бесхозяйными отходами и объектами, оставшимися с преданных лет, которые до сих пор представляют угрозу природной среде и здоровью населения. Наличие такой информационной базы данных по всей территории Республики позволит привлечению инвестиций в разработку новых инновационных технологий по утилизации оставшихся ООП, зачистке (рекультивации) загрязненных территорий, создаст определенные основы по управлению здоровьем населения загрязненных территорий.

Председатель  
ОО Human Health Institute



Капасов А.К.



## АКТ

### о внедрении материалов диссертационного исследования Нуржанова Чингиза Аскарловича «Проектирование информационной системы для прогнозирования и принятия решений в процессе очистки почвы, содержащей токсичные элементы»

Комиссия: председатель - Дюсенов О.К., члены: Шажиров В., Резаев П. составили настоящий акт о том, что результаты диссертационного исследования Нуржанова Ч.А. представленного на соискание ученой степени доктора философии, могут быть рекомендованы для использования в деятельности ТОО «KazEcoSolutions» в виде:

– практического использования информационной системы на базе данных о загрязненных токсичными элементами территорий и растений, способных к восстановлению техногенных ландшафтов, что позволит оперативно в режиме реального времени формировать оптимальные варианты рекультивационных решений для улучшения экологической ситуации (глава 4 'Разработка комплексного подхода к созданию информационной системы на базе данных о загрязненных пестицидами территориях и растениях, способных к восстановлению техногенных ландшафтов);

– практических рекомендаций при выборе растений из базы данных с учетом экологии почвы для рекультивационных мероприятий загрязненных земель (нефтью, тяжелыми металлами и пестицидами), что позволит сократить время и сроки выполнения работ (раздел 4.2. 'Создание базы данных о растениях, способных к восстановлению загрязненных токсическими элементами земель/воды);

– практических рекомендаций при формировании выбора растений из базы данных для ландшафтной архитектуры (раздел 4.2. 'Создание базы данных о растениях, способных к восстановлению загрязненных токсическими элементами земель/воды');

- результаты диссертационной работы могут быть использованы при формировании карты загрязнения для минимизации экологического риска на загрязненных территориях.

Председатель комиссии

Члены комиссии:



Дюсенов О.К.

Шажиров В

Резаев П.