Министерство образования и науки Российской Федерации
Российский государственный геологоразведочный университет имени Серго Орджоникидзе

На правах рукописи

РУКАВИЦЫН ВАДИМ ВЯЧЕСЛАВОВИЧ

ОПРЕДЕЛЕНИЕ УСТОЙЧИВОСТИ ГЕОЛОГИЧЕСКОЙ СРЕДЫ С ПРИМЕНЕНИЕМ МЕТОДОВ МАШИННОГО ОБУЧЕНИЯ (НА ПРИМЕРЕ Г. МОСКВЫ)

Специальность 25.00.08 – Инженерная геология, мерзлотоведение и грунтоведение

Диссертация
на соискание ученой степени
кандидата геолого-минералогических наук

Научный руководитель:
доктор геолого-минералогических наук,
профессор Экзарьян В. Н.

Москва – 2018

СОДЕРЖАНИЕ

ВВЕДЕНИЕ 3

1. СОВРЕМЕННОЕ СОСТОЯНИЕ ПРОБЛЕМЫ ОПРЕДЕЛЕНИЯ УСТОЙЧИВОСТИ ГЕОЛОГИЧЕСКОЙ СРЕДЫ 10

1.1. Понятие устойчивости геологической среды 10

1.2. Основные факторы, определяющие устойчивость геологической среды 16

1.3. Определение критериев оценки устойчивости геологической среды 31

1.4. Обоснование предельно допустимых норм критериев оценки устойчивости геологической среды 39

2. ОБОСНОВАНИЕ ПРИМЕНЕНИЯ МАШИННОГО ОБУЧЕНИЯ ДЛЯ ОПТИМИЗАЦИИ АНАЛИЗА УСТОЙЧИВОСТИ ГЕОЛОГИЧЕСКОЙ СРЕДЫ 51

2.1. Выбор подхода к оптимизации анализа устойчивости геологической среды 51

2.1.1. Применимость машинного обучения для решения задачи определения устойчивости геологической среды 57

2.1.2. Применимость других методов моделирования для решения задачи определения устойчивости геологической среды 62

2.2. Принцип определения устойчивости геологической среды при помощи машинного обучения 67

2.3. Алгоритмы машинного обучения 80

3. ОПЫТ РЕШЕНИЯ ИНЖЕНЕРНО-ГЕОЛОГИЧЕСКИХ ЗАДАЧ МЕТОДАМИ МАШИННОГО ОБУЧЕНИЯ 89

4. АНАЛИЗ ИНЖЕНЕРНО-ГЕОЛОГИЧЕСКИХ УСЛОВИЙ ИЗУЧАЕМОЙ ТЕРРИТОРИИ ДЛЯ ИНТЕРПРЕТАЦИИ РЕЗУЛЬТАТОВ МОДЕЛИРОВАНИЯ 106

5. МЕТОДИКА ОЦЕНКИ УСТОЙЧИВОСТИ ГЕОЛОГИЧЕСКОЙ СРЕДЫ НА ОСНОВЕ МАШИННОГО ОБУЧЕНИЯ 127

5.1. Методика создания модели устойчивости геологической среды 131

5.2 Методика тестирования модели устойчивости геологической среды 136

6. МОДЕЛИРОВАНИЕ УСТОЙЧИВОСТИ ГЕОЛОГИЧЕСКОЙ СРЕДЫ ПРИ ПОМОЩИ МАШИННОГО ОБУЧЕНИЯ НА ПРИМЕРЕ РЕАЛЬНОЙ ИНЖЕНЕРНО-ГЕОЛОГИЧЕСКОЙ ИНФОРМАЦИИ 140

6.1. Создание модели геологической среды на основе машинного обучения 142

6.2. Оценка результатов моделирования устойчивости геологической среды по выбранным критериям 158

6.3. Применение методики и развитие исследования 168

ЗАКЛЮЧЕНИЕ 170

ЛИТЕРАТУРА 175

Введение

Актуальность работы

Вопрос определения устойчивости геологической среды является важным, актуальным и довольно сложным в решении. Моделирование устойчивости геологической среды чаще всего базируется на использовании различных модификаций экспертных методов, а применение математического моделирования крайне затруднено из-за сложности формализации исходных данных.

В России и за рубежом проводится широкий комплекс исследований по разработке и внедрению моделей машинного обучения для автоматизации систем принятия решений и других важных задач. Машинное обучение – это научная дисциплина, которая занимается проектированием и разработкой алгоритмов, позволяющих компьютерам изменить своё поведение на базе некой информации, например, на базе информационных сенсоров или базы данных [Mitchell T. 1997]. Одним из основных направлений исследований машинного обучения является способность учиться автоматически решать сложные проблемы и принимать разумные решения, основанные на содержательной формализации и обработке фактического материала

Для решения инженерно-геологических задач машинное обучение применяется не часто и, в основном, решаемые задачи довольно просты. Однако опыт использования данного метода в других отраслях показывает актуальность и перспективность этого направления для решения инженерно-геологических задач.

Заимствуя логику принятия решений, описанную в работах Бондарика Г.К., Осипова В.И., Трофимова В.Т., Пендина В.В., Шестопалова В.М., Азроянца Э.А, Зайонца И.О. и других исследователей, метод машинного обучения позволил формализовать информацию о геологической среде и оптимизировать определение оценки её устойчивости.

Цель работы

Разработка методики оптимизации решения задачи определения устойчивости геологической среды при помощи машинного обучения.

Задачи исследований

В соответствии с целью научных исследований были поставлены следующие задачи:

- Обоснование применения машинного обучения для оптимизации анализа устойчивости геологической среды
- Разработка методики определения устойчивости геологической среды при помощи машинного обучения;
- Выбор критериев, характеризующих устойчивость геологической среды;
- Создание представительной базы данных, характеризующей устойчивость геологической среды посредствам выбранных критериев.
- Обработка исходных данных, для дальнейшей их обработки при помощи машинного обучения;
- Определение устойчивости геологической среды г.Москвы при помощи машинного обучения и сравнение результатов с данными многолетних исследований анализа;
- Интерпретация результатов и обоснование применимости машинного обучения для определения устойчивости геологической среды на предпроектной стадии строительства.

Фактический материал, методы исследования

Основным объектом исследования устойчивости геологической среды послужили результаты инженерных изысканий г. Москвы. В их основе лежали карты, которые содержали в себе как некую базовую информацию, так и результаты экспертной оценки территории.

В ходе выполнения работы автором был проведен анализ литературных источников, описывающих возможности применения машинного обучения для решения инженерно-геологических задач и способы определения устойчивости геологической среды. В результате этого анализа была разработана методика определения устойчивости геологической среды при помощи машинного обучения с целью оптимизации выполнения данного вида исследований для получения возможности его более широкого применения.

На основе разработанной методики были собраны исходные данные, которые впоследствии были обработаны, специальным образом проанализированы и адаптированы для дальнейшего использования в машинном обучении. Данные включали в себя как параметры геологической среды, так и результаты экспертного моделирования ее устойчивости.

С помощью программного продукта WEKA была построена компьютерная модель устойчивости геологической среды г. Москвы, копирующая результаты ранее проведенного экспертного анализа, что позволяет частично доверить проведение такой сложной комплексной оценки искусственному интеллекту.

Разработка методики, создание базы исходных данных и моделирование проводилось в Российском Государственном Геологоразведочном Университете им. Серго Орджоникидзе на кафедре экологии и природопользования и в Университете Менделя г. Брно.

Научная новизна работы

1. Разработана уникальная методика определения устойчивости геологической среды на основе методов машинного обучения;
2. Предложенная методика определения устойчивости геологической среды на основе методов машинного обучения была апробирована при решении реальной инженерно-геологической задачи;
3. На основе ряда экспериментов были определены наиболее значимые параметры, характеризующие устойчивость геологической среды;
4. В результате внедрения новой методики создана представительная исходная база данных, адаптированная для работы с машинным обучением и позволяющая решать задачи по определению устойчивости геологической среды в городах центральной России.
5. На основе представленных экспериментальных данных определен программный алгоритм, способный наиболее точно проводить оценку устойчивости геологической среды;
6. При помощи проведенной оценки показана возможность применения машинного обучения для определения устойчивости геологической среды, что позволяет унифицировать решение подобных задач в будущем.

Защищаемые научные положения

1. Определение устойчивости геологической среды, выполненное на предпроектной стадии, позволяет оптимизировать и унифицировать проведение инженерных изысканий.
2. Одним из перспективных методов математического моделирования и прогнозирования устойчивости геологической среды является машинное обучение, позволяющий обрабатывать большие объемы разнородной и сложной по составу инженерно-геологической информации.

3. Модель устойчивости геологической среды г. Москвы, разработанная в данной работе, следует рассматривать в качестве аналога при определении состояния геологической среды урбанизированных территорий, находящихся в аналогичных инженерно-геологических условиях.

Практическая значимость работы

1) Проведенный анализ различных способов применения искусственного интеллекта для решения инженерно-геологических задач и подходов к анализу геологической среды позволил сформировать подход к определению ее устойчивости на основе машинного обучения.

2) Усовершенствованы существующие практические подходы к определению устойчивости геологической среды для целей оценки территории строительства.

3) Созданная представительная исходная база данных, характеризующая устойчивость геологической среды г. Москвы позволяет проводить ее оценку на сходных по условиям территориях в кратчайшие сроки и при наличии минимального набора информации.

4) Разработанная методика использования машинного обучения доказала свою состоятельность при решении реальных задач и результаты этих решений коррелировались с данными многолетних изысканий, что имеет важное значение для прогноза негативных инженерно-геологических процессов при строительстве.

Апробация работы

Основные результаты работы и её отдельные положения докладывались на научных конференциях:

Международная научная конференция студентов, аспирантов и молодых учёных «Ломоносов-2018», Москва 2018; Новые идеи в науках о Земле, Москва 2017; Науки о Земле на современном этапе, Москва 2014; 9th Summer School of Aclied Informatics, Бедрихов, Чешская Республика, 2012; Развитие экологической экономики и образования в мире и в Украине, Киев, Украина,2011; MECON 2011, Острава, Чешская Республика, 2010; ISDMCI 2010, Евпатория, Украина, 2010; MendelNet 2009, Брно, Чешская Республика, 2009. В 2010 году автор получил грант IGA Университета Менделя.

Публикации

По теме диссертационной работы опубликовано 16 печатных работ, в том числе 2 в рецензируемых монографиях и 4 в реферируемых журналах ВАК.

Объем и структура работы

Диссертация состоит из введения, 6 глав, заключения и списка использованных источников из 114 наименований. Работа изложена на 187 страницах, содержит 29 рисунков, 13 таблиц.

Благодарности

Автор выражает глубокую благодарность своему научному руководителю – доктору геолого-минералогических наук, профессору Экзарьяну Владимиру Нишановичу.

Автор признателен всем сотрудникам кафедры экологии и природопользования за ценные советы и помощь в написании настоящей работы.

Автор выражает благодарность всем сотрудникам Университета Менделя г. Брно за помощь, внимание и ценные советы в работе. Автор считает своим долгом выразить благодарность доценту Яну Жижка за содействие в изучении машинного обучения и методов его прикладного применения в инженерной геологии.

1. Современное состояние проблемы определения устойчивости геологической среды

1.1. Понятие устойчивости геологической среды

История цивилизации неразрывно связана с использованием земли - сначала просто как территории обитания, потом в качестве плодородной среды для земледелия. Постепенно глубина, степень освоения и использования земли возрастали и территориально дошли до планетарного масштаба, а по глубине - до самых древних геологических слоев.

С усилением влияния человечества на природную среду, все больше усиливается ответное влияние природной среды на человека и создаваемые им сооружения. Поскольку население планеты неуклонно растет, использование и территориальной, и ресурсной, и земледельческой составляющих природной среды должно становился более рациональным. Геологические условия территории необходимо учитывать не только при добыче полезных ископаемых, но и при освоении новых районов для заселения и для строительства на уже освоенных и нарушенных техногенными процессами участках.

При выборе участка, в частности, для строительства нужно учитывать не только планировочные качества (выровненность поверхности, необходимость подрезок и насыпей и т.д.), но и возможные последствия вмешательства в природную среду на конкретном участке. Реакция природной среды на техногенные воздействия зависит как от вида и степени такого воздействия, так и от свойств самой среды. В связи с этим можно говорить о такой интегральной оценке, как устойчивость геологической среды.

Термин устойчивость широко используется в математике, технике, механике, кибернетике, термодинамике, биологии, но до настоящего времени он является термином общего пользования.

С общефилософской точки зрения «устойчивость» определяется, как единство изменчивости и неизменчивости. Специалисты из других областей знаний, принимая это общее определение, пытаются найти для него более точное, более математическое выражение. Наиболее полно вопрос устойчивости разработан в механике, математике, кибернетике.

С течением времени такой термин, как устойчивость начал использоваться в биологии и термодинамике, ну а в конце 60-х годов XX века - в географических и геологических наук [Rodriguez J. 2009].

А.И. Ляпунов дал первое определение термина «устойчивость». Оно звучало как «способность явления (состояния) или процесса мало изменяться (или совсем не изменяться) под действием причин, стремящихся его изменить». Именно это определение использовалось в механике до конца XIX века. Под устойчивостью тогда подразумевали способность явления (состояния) или процесса мало изменяться (или совсем не изменяться) под действием причин, стремящихся его изменить. Недостаток этого определения в том, что оно рассматривает явление как простое и неделимое, в то время как природные явления характеризуются многокомпонентностью и сложностью функционирования. Кроме того, в нем отсутствует определение того, что называют «малым изменением» [Герасимова А.С. 1994].

По представлению У.Р. Эшби [Эшби У.Р. 1959] в кибернетике устойчивое состояние определяется как такое множество состояний, в которых преобразование «I» не порождает из состояния "А" нового состояния, то есть, вошедшая в это множество точка не может его покинуть. У.Р. Эшби считает, что устойчивость - это слово с большой перегрузкой и неустоявшимся значением. Понятие «устойчивость», по его мнению, является составным. Лишь когда указан каждый его аспект, оно может применяться к конкретным случаям. Простейшим случаем устойчивости,

считает У.Р. Эшби, является равновесие, когда состояние и преобразование связаны между собой так, что преобразование не заставляет состояние изменяться, причем через все значения устойчивости проходит идея инвариантности, т.е. идея о том, что хотя система в целом претерпевает последовательные изменения, некоторые ее свойства (инварианты) сохраняются неизменными. Он предлагает такие слова, как «устойчивость» и «равновесие», употреблять в качестве сокращений, когда явления достаточно просты.

Л.А. Чульский [Чульский Л.А. 1965], определяет устойчивость, как способность сохраняться, выживать, выполнять поставленную задачу и идти к намеченной цели, несмотря на мешавшую этому воздействию силу. Он полагает, что необходимо количественно оценивать, далеко ли находится изучаемая система от границ неустойчивости, т.е. запас устойчивости или меру ее инертности.

Однако понятие устойчивости имеет и более специализированные значения. Если говорить об устойчивости гелогической среды, то начать нужно с глобальной устойчивости. Под глобальной устойчивостью понимается концепция ноосферы и устойчивого развития. Основы учения о ноосфере были намечены В.И. Вернадским, однако в последующем учение получило сильное развитие.

Устойчивое развитие — это процесс экономических и социальных изменений, при котором эксплуатация природных ресурсов, направление инвестиций, ориентация научно-технического развития, развитие личности и институциональные изменения согласованы друг с другом и укрепляют нынешний и будущий потенциал для удовлетворения человеческих потребностей и устремлений. Во многом речь идёт об обеспечении качества жизни людей [Фролова В. И. 2011].

Но для обеспечения такого глобального устойчивого развития, помимо рационального экономических и социальных преобразований,

необходимо сохранение устойчивости природных сред, в том числе устойчивости геологической среды.

Среди существующих определений понятия устойчивости геологической можно выделить следующие:

- возможность сохранения объекта на протяжении некоторого промежутка времени;

- стабильность состояний или качественных характеристик объекта во времени;

- способность возвращаться в исходное состояние после некоторого воздействия;

- способность приспосабливаться к изменившимся условиям, искать новые состояния равновесия;

- способность сохранять наиболее значимые характеристики в допустимом состоянии за счет изменения других, менее значимых;

- способность не блокировать внешние воздействия;

- способность длительное время накапливать вредные воздействия, избегая при этом значительного повреждения системы;

- способность развиваться в том же направлении, несмотря на воздействия;

- способность сохранять свои свойства при внешнем воздействии [Татаринов В.Н. 2000].

В итоге можно выделить три направления интерпретации термина:

- способность сохранять некоторые характеристики;

- способность к изменениям, но в пределах некоторых толерантных пределов;

- способность к возвращению в базовое состояние, несмотря на воздействия.

То же относится и к устойчивости геологической среды. Однако устойчивость геологической среды является информационным и базовым обеспечением устойчивости окружающей среды в целом, поскольку каждая

экосистема развивается, подстраиваясь под сложившиеся инженерно-геологические условия, которые проявляют себя в рельефе, тектонических условиях, наличии и степени активности различных эндогенных и экзогенных геологических процессов.

Моделирование и оценка устойчивости геологической среды должна включать в себя решение трех задач:

1. Определение основных факторов, определяющих стабильность и устойчивость геологической среды;

2. Установление критериев оценки и прогноза устойчивости;

3. Обоснование предельно допустимых норм критериев оценки и прогноза.

При этом для моделирования геологической среды и определения ее устойчивости необходимо рассматривать ее на различных иерархических уровнях. Их можно классифицировать следующим образом:

1.Глобальный, соответствующий масштабу всей Земли или одного материка.

2.Национальный, соответствующий территории одного государства.

3.Региональный, соответствующий территории большого участка одного государства или сопредельных участков нескольких государств.

4.Локальный, соответствующий относительно небольшой территории района крупного предприятия, водного объекта, крупного поселения города.

5.«Точечный» уровень, соответствует конкретным объектам в масштабах одного дома.

В данной работе рассматривать геологическую среду на глобальном уровне не имеет смысла. Здесь больше всего сказываются именно тектонические процессы, и такое моделирование имеет слишком общий характер. Также отсутствуют детальные данные о глобальной устойчивости геологической среды, что затрудняет последующую компьютерную аналитику данных.

Национальный и региональный уровни уже достаточно изучены. Существуют карты районирования РФ и СССР, проведено множество исследований. Однако разнообразие условий геологической среды слишком велико, для того, чтобы начинать моделирование с этих уровней.

Для нужной степени оценки достоверности моделирования и при этом из расчета доступности исходной информации имеет смысл оценить устойчивость геологической среды на локальном уровне. Точечный уровень будет иметь смысл разрабатывать после детального изучения моделирования на локальном уровне и получении исходных данных и масштабах точечных объектов.

В итоге устойчивости геологической будет являться результатом такого моделирования.

Далее разберем особенности решения поставленных задач для максимальной оптимизации процесса моделирования и создания наиболее универсального метода определения устойчивости геологических систем.

1.2. Основные факторы, определяющие устойчивость геологической среды

Устойчивость геологической среды на определенной территорий характеризуется двумя основополагающими группами факторов:

1) Природными, которые характеризуют непосредственно геологическую среду и процессы ее развития,
2) Техногенными, которые определяют воздействия на нее со стороны деятельности людей.

К природным факторам относятся все осбенности инженерно-геологических условий территорий; характеристики геологического строения, гидрогеологические условия, рельеф, и геодинамическая ситуация. Некой общепринятой классификации группы природных факторов, которые бы определяли устойчивость геологической среды, на сегодняшний день нет. Но при этом их можно разделить на группы, соответственно классификации Золотарева Г.С:

1. Формирующие среду факторы. Под ними понимаются ы структурно-тектонический план территории, комплекс горных пород, общие особенности обводненности пород и т.д.

2. Факторы, которые изменяют состояния и свойства среды. К таким относят процессы выветривания, разгрузки, карстовые и суффозионные процессы; проявления современных геодинамических и тектонических процессов, влияющие на распределение естественных напряжений; обводнение массивов пород; трещинообразование [Назарова О.В. 2001].

Техногенные воздействия оцениваются и исследуются по-разному для разных территориальных уровней. Так для регионального уровня оцениваются крупные отраслевые технические системы, со сложной структурой организации, состоящие из разнородных техносферных объектов и осуществляющие разнообразные как по природе (химические, физические

и т.д.) так и способам приложения (поверхностные, объемные, линейные, площадные и т.д.) воздействия на геологическую среду.

Для локального уровня исследований в границах технических систем регионального уровня выделяют крупные отдельные объекты техногенного воздействия на природную среду. Их называют природно-техническими системами (ПТС).

В инженерной геологии понятие устойчивости начало употребляться в отношении оценки склоновых процессов, и этот фактор являлся основополагающим при ее определении. Для этого разработано множество методов. Однако при появлении понятия «геологическая среда» термин устойчивость приобрел более общее значение и стал применяться при прогнозировании и других процессов.

Так Бондарик Г.К. [Бондарик Г.К. 1984] употребляет термин «устойчивость» в контексте способности системы подавлять внешние импульсы и длительное время сохранять свои функции. Методика оценки ресурсов устойчивости основывалась на определении максимально-допустимых значений импульсов воздействия. Однако предварительно необходимо было определить границы допустимых изменений геологической среды. При этом определение устойчивости геологической среды усложняется тем, что необходимо определить не только перечень компонентов инженерно-геологических условий, но и выразить их изменения количественно.

Котлов В.Ф. и Кофф Г.Л. [Котлов В.Ф. 1997] предлагали оценивать устойчивость геологической среды тремя способами:

1) Составление серии карт, отражающих инженерно-геологическую характеристику территории, состоящую из комплекта покомпонентных оценок. При их совмещении определяются все детали воздействия каждого компонента, эти воздействия суммируются, и получается некая интегральная оценка, которая отражается на соответствующей карте.

2) Определение геодинамической устойчивости через оценку развития природных и антропогенных процессов. В этом случае устойчивость определяется, как активность проявления процессов антропогенных изменений геологических явлений.

3) Оценка геологической среды по косвенным показателям.

Ими же с соавторами была составлена серия аналитических оценочно-прогнозных детальных карт масштаба 1:10 000 опасности воздействия геологических процессов для разных стадий проектирования городской застройки, наземной и подземной инфраструктуры города, реконструкции и капитального ремонта [Кофф Г. Л. 1999]. Опасность воздействия геологических процессов на здания и сооружения оценивается с точки зрения устойчивости самих зданий. Модель взаимодействия геологической среды и сооружения для г. Москвы выражена авторами на графе, представленном на рисунке 1.2.1

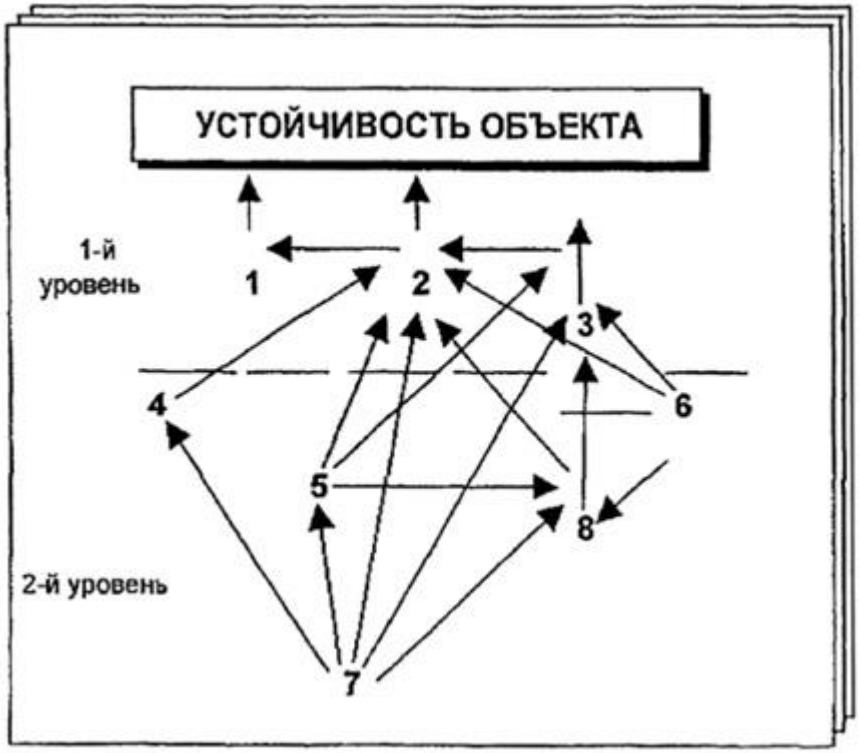


Рисунок 1.2.1. Граф трехмерной модели геологической среды. Базовые элементы модели, отражающие развитие и протекание геологических процессов: 1 — рельеф; 2 — состояние грунтовой толщи; 3 — уровеньподземных вод; 4 — высотное положение кровли каменноугольных отложений; 5 — ее рельефность; 6 — литологический тип пород, залегающих в кровле каменноугольных толщ; 7 — наличие или отсутствие регионального юрского водоупора; 8 — наличие или отсутствие разуплотненных зон [Кофф Г. Л. 1999]

Устойчивость зданий оценивалась с позиций рассмотрения протекания различных процессов: 1) на границе раздела «сооружение — геологическая среда»; 2) в области взаимодействия сооружения с толщей грунтов; 3) в пределах геологической среды, окружающей данную область. Соответственно процессы были подразделены на прямые, воздействущие непосредственно на сооружение (позиции 1, 2, рис. 1.2.1), и косвенные, действующие опосредованно через область взаимодействия (позиция 3) [Кофф Г. Л. 1999].

В этой модели в качестве критериев, определяющих состояние геологической среды, были использованы рельеф, толщи пород и подземные воды. По степени воздействия эти критерии были разделены на два уровня:

- на первом уровне на здания воздействуют прямые процессы, которые определяются: 1) рельефом, 2) состоянием грунтовых толщ в области взаимодействия с сооружением и 3) близостью к дневной поверхности (а значит и к фундаменту) первого от поверхности водоносного горизонта;

- на втором уровне действуют косвенные процессы, которые определяются, применительно к рассматриваемой территории: 1) характером поверхности кровли каменноугольных отложений; 2) распространением регионального юрского водоупора, отделяющего эти отложения от вышележащих четвертичных, относящихся к первому уровню геологической среды; 3) наличием зон разуплотнения, контролирующих тектонические нарушения, а также линий метрополитена [Кофф Г. Л. 1999].

Оценка опасности на 1-м модельном уровне, при учете особенностей рельефа, типа грунтовой толщи и уровня грунтовых вод, показывает, по мнению авторов карт, опасность воздействия геологических процессов, которую необходимо учитывать на стадиях проекта детальной планировки и проекта застройки.

Петренко А.С. совместно с этими авторами на основе указанных выше выводов были составлены карты геологического риска г.Москвы и опасности воздействия геологических процессов на объекты строительства [Петренко А.С. 2001]. Он также при оценке риска и опасности воздействия оценивались не только геологическая устойчивость территории, но и типы зданий, их

фундаменты и другие конструктивные особенности объектов. Пример таких карт представлен на рисунке 1.2.2 – 1.2.5.

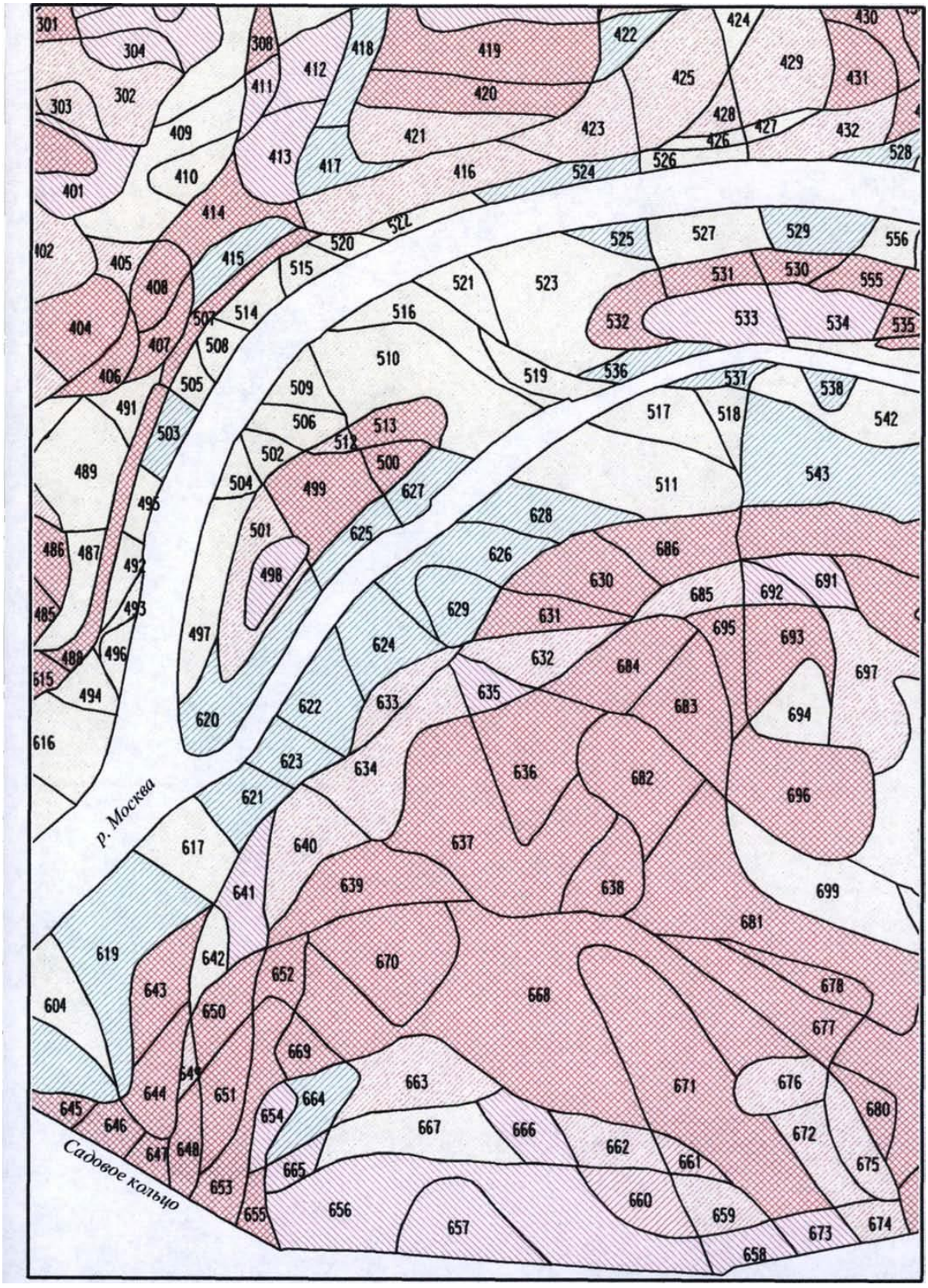


Рисунок 1.2.2. Фрагмент карты опасности воздействия геологических процессов для проектирования наземной и подземной инфраструктуры города (г. Москва, центральная часть) [Петренко А.С. 2001]

Рисунок 1.2.3. Условные обозначения к картам опасности воздействия геологических процессов на территории г. Москвы [Петренко А.С. 2001]

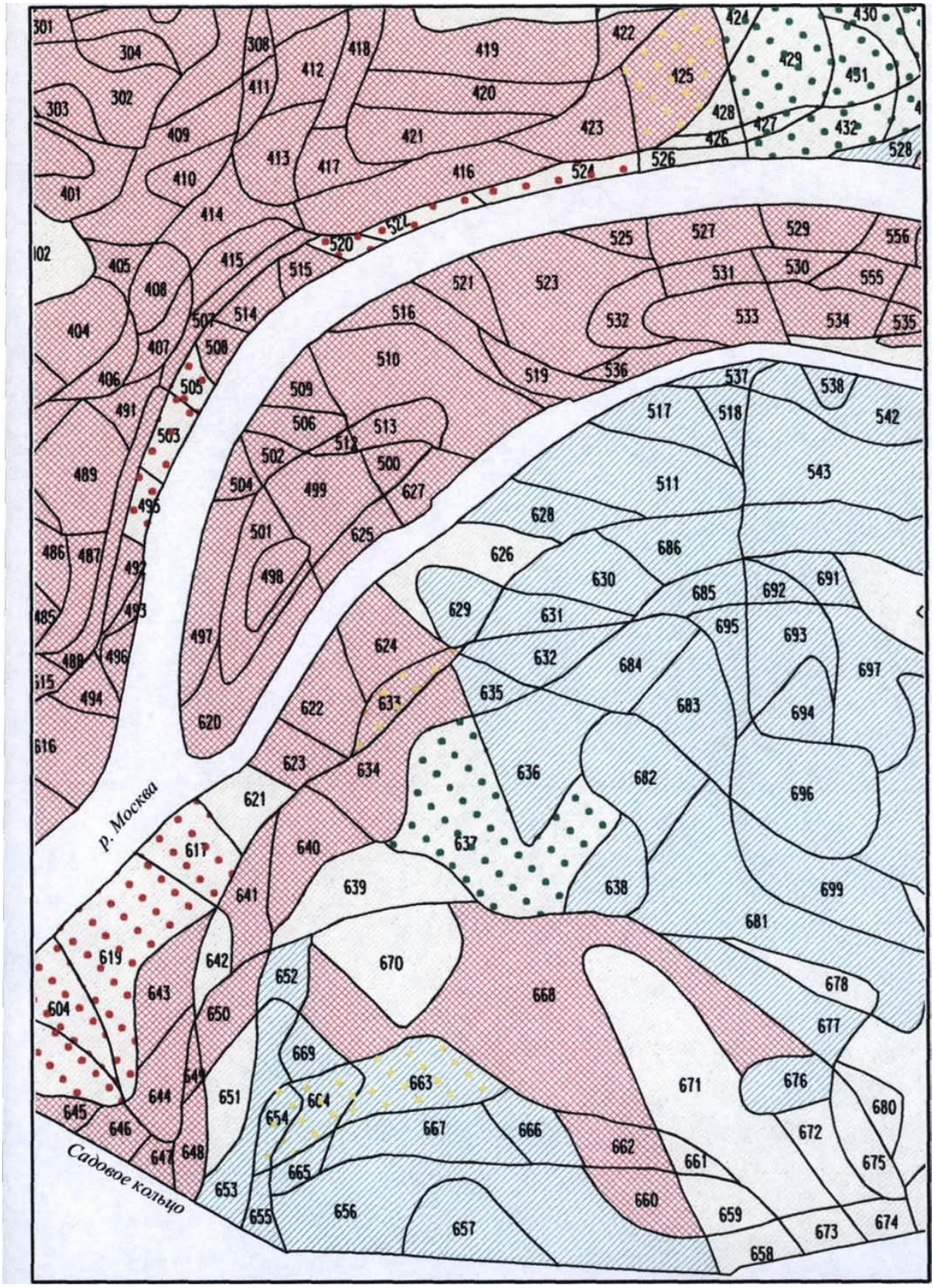


Рисунок 1.2.4. Фрагмент карты геологического риска при наземной застройке микрорайонов (г. Москва, центральная часть) [Петренко А.С. 2001].

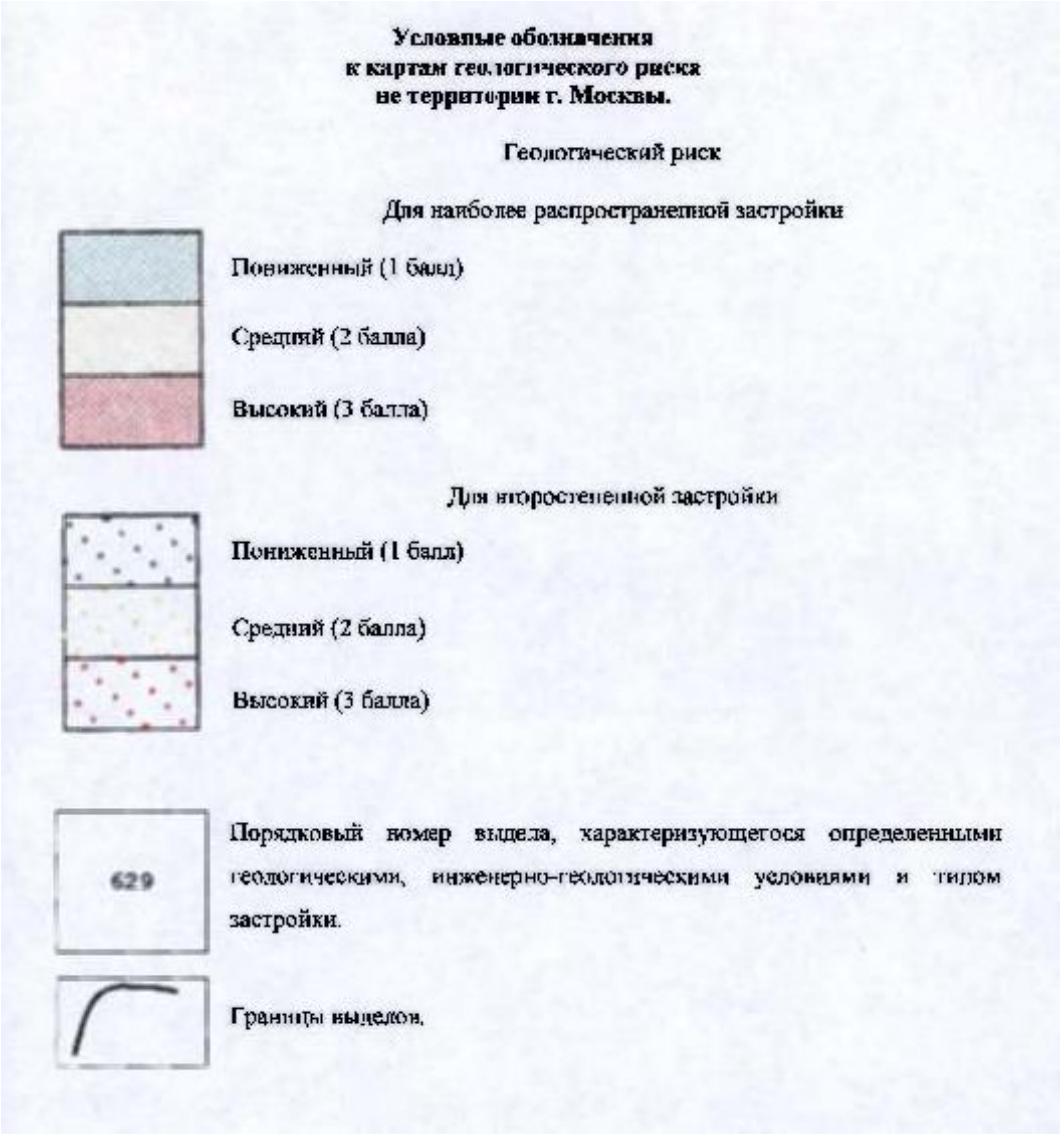


Рисунок 1.2.5. Условные обозначения к картам геологического риска на территории г. Москвы [Петренко А.С. 2001].

В работах Грибановой Л.И. фактором определения устойчивости геологической среды выступают количественные показатели подземной и поверхностной закарстованности. А Дзекцер Е.С. с соавторами [Дзекцер Е.С. 1984] в качестве одного из таких факторов выделяет степень подтопляемости территории.

Инамов А.Н. [Инамов А.Н. 1989] в качестве ведущих факторов устойчивости инженерно-геологических условий выделял: геологическое строение, характер и степень расчлененности рельефа, тип фильтрационного разреза, интенсивность инженерно-геологического воздействия,

дренированность территории, уровень минерализации грунтовых вод, просадочность и засоленность грунтов. Им же с соавтором была составлена Карта изменения геологической среды под воздействием ирригационно-мелиоративного строительства в центральной части Каршинской степи масштаба 1:50 000, представленная на рисунке 1.2.6.

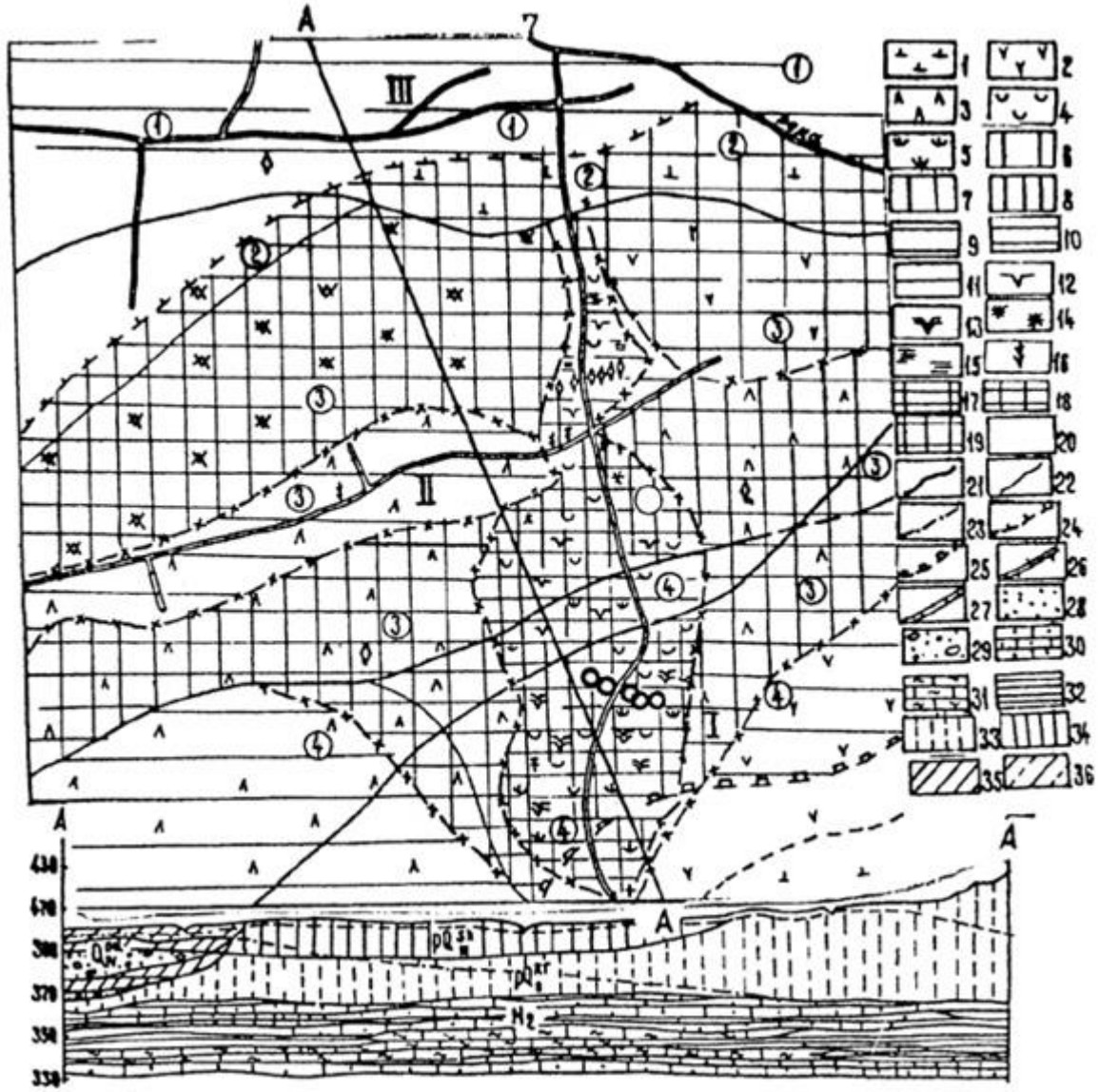


Рисунок 1.2.6. Макет Карты изменения геологической среды под воздействием ирригационно-мелиоративного строительства в центральной части Каршинской степи. Геолого-геоморфологическое строение и гидрогеологические условия. Равнины: I — пролювиальная волнистая сильнорасчлененная pIIkr (лесс, 20–40 м); II —

пролювиальная плоская, pIIsh (лессовидные суглинки и супеси, 5–15 м); III — аллювиальная плоская aIVad (переслаивающиеся суглинки, супеси и пески, 20–30 м). Подъем уровня грунтовых вод, м: 1 — 2–5; 2 — 5–10; 3 — 10–20; 4 — 20–30; 5 — 30–40. Увеличение содержания легкорастворимых солей в грунтах, %: 6 — 0,1–0,3; 7 — 0,3–0,6; 8— более 0,6; естественной влажности, %: 9 — 1,5; 10 — 5–10; 11 — более 10. Инженерно-геологические процессы и явления, просадка: 12 — до 15 см, 13 — более 15 см; 14 — засоление; 15 — подтопление и заболачивание; 16 — размыв и обрушение берегов каналов, коллекторов и дрен. Степень измененности геологической среды: 17 — сильная; 18 — средняя; 19 — слабая; 20 — неизмененная. Границы: 21 — морфогенетических типов рельефа; участков с различной: 22 — глубиной залегания УГВ, 23 — степенью измененности геологической среды; 24 — староорошаемой зоны; 25 — новоорошаемой; 26 — каналы; 27 — коллекторы; 28 — песок; 29 — гравий, галька; 30 — песчаник; 31 — алевролит; 32 — глина; 33 — лессовидные суглинки и супеси; 34 — лесс; 35 — суглинки плотные; 36 — супеси плотные. [Э. В. Мавлянов, А. Н. Инамов, 1986]

Оценка изменений проводилась путем наложения однотипных карт, составленных на 1965 (условно принятое за исходное состояние) и 1985 (за современное состояние) годы. Эти аналитические карты позволили оценить произошедшие за 20 лет изменения глубин залегания грунтовых вод, их минерализации, засоленности грунтов зоны аэрации, региональные изменения инженерно-геологических свойств грунтов. Это дало возможность качественно и количественно отобразить изменения гидрогеологических и инженерно-геологических компонентов геологической среды, а также оценить интенсивность и направленность ее изменений в целом [Э. В. Мавлянов, А. Н. Инамов, 1986].

Знаково, что Шейдеггер А.Е. ввел понятие «риск», который, по его мнению, является мерой оценки устойчивости территории. Под риском он понимал оценку частоты и размеров потенциальных катастроф [Шейдеггер А.Е. 1981].

При этом Бахирева Л.В. [Бахирева Л.В. 1989] использует термин «геологический риск», понимая вероятность проявления и активизации природных и техногенных процессов, которые вызывают негативные изменения технического состояния объектов и зон их влияния, также используя его для оценки устойчивости территории.

Цуриков К.О. [Цуриков К.О. 1991] предлагает использовать коэффициент риска проявления экзогенных геологических процессов под воздействием техногенных факторов при определении устойчивости геологической среды. Каждый вид экзогенных процессов в этом случае выступает в качестве фактора ее оценки.

Таким образом, по его мнению, факторами, определяющими устойчивость геологической среды, являются виды негативного на нее воздействия и ответные реакции среды на эти воздействия. Поэтому для определения устойчивости необходимо определить эти воздействия, которые будут различными при разных видах хозяйственной деятельности.

На основе этих выводов он составил Карту районирования территории Крымской области по условиям реакции геологической среды на воздействие техногенного фактора. На карте показаны коэффициенты риска проявления ЭГП под воздействием техногенного фактора и определяемые ими степени риска, используемые для качественной прогнозной оценки реакции геологической среды на техногенное воздействие (Рис. 1.2.7.). Коэффициенты риска колеблются от 1 (процессами слабой активности поражено менее 10 % площади участка) до 20 (процессами очень сильной активности поражено более 75 % площади участка). Коэффициент риска обозначается на карте

цифровыми индексами от 1 до 20, типы опасных процессов — общепринятыми буквенными индексами, степени риска — штриховкой и крапом.

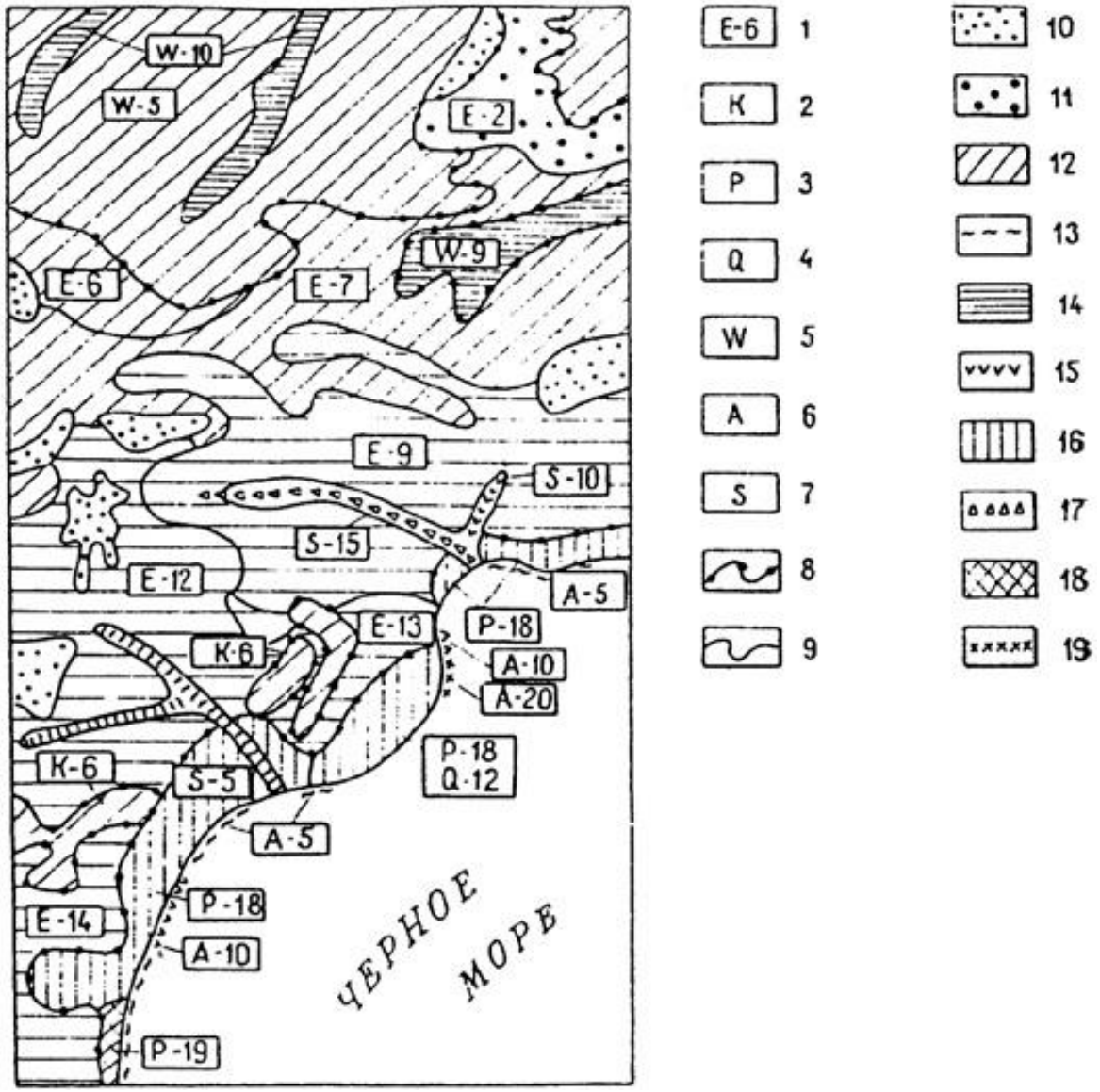


Рисунок 1.2.7. Фрагмент карты районирования по условиям реакции геологической среды на воздействие техногенного фактора [К. О. Цурикову, 1991]. I — индексы участков с различными условиями реакции (цифра — значение коэффициента риска, буква — тип ЭГП (в данном случае эрозия)); индексы ЭГП: 2 — карст, 3 — оползни, 4 — обвально-осыпные процессы; 5 — подтопление изаболачивание, 6 — абразия, 7 — сели; границы: 8 —

районов преимущественного развития различных типов ЭГП, 9 — участков с различными значениями коэффициентов риска проявления ЭГП; степени риска проявления ЭГП: 10 — проявления ЭГП не ожидаются, 11 — слабая; 12, 13 — средняя, 14, 15 — сильная, 16, 17 — очень сильная, 18, 19 — весьма сильная (13, 15, 17, 19 — для линейных районов)

1.3. Определение критериев оценки устойчивости геологической среды

Как было указано выше, общепринятый интегральный критерий оценки устойчивости геологической среды отсутствует. Существующие критерии можно разделить на три группы: количественные, полуколичественные и качественные.

Для определения критериев оценки устойчивости геологической среды на основе анализа геологической среды необходимо понимать, где она располагается. А если учитывать, что рассматривать геологическую среду нельзя без техногенного фактора, то наилучшим способом определения критериев оценки будет выделение показателей деградации или же стабилизации территориально-промышленных комплексов.

Территориально-промышленный комплекс (ТПК) – это природно-техногенный объект, функционирование которого определяется взаимодействием природных условий и хозяйственной нагрузки. ТПК занимает земельное пространство, в пределах которого преимущественно развит определенный вид хозяйственной нагрузки.

По видам и интенсивности воздействия на окружающую природную среду выделяются следующие основные типы ТПК:

1) Градопромышленный комплекс
2) Энергетический комплекс
3) Горнодобывающий комплекс
4) Агропромышленный комплекс
5) Рекреационный комплекс [Экзарьян В.Н. 2009]

В каждом из этих комплексов можно выделить критерии его оценки, с точки зрения устойчивости природной среды. Эти критерии представлены в таблице 2.3.1.

Таблица 1.3.1. Критерии оценки устойчивости природных систем в ТПК [Экзарьян В.Н. 2009]

Территориально-промышленный комплекс	Критерий оценки устойчивости природной среды
Градопромышленный комплекс	Изменение атмосферы Изменение микроклимата Изменение растительного покрова Изменение почвенного покрова Изменение гидрографической сети Изменение рельефа Изменение гидрогеологических условий Изменение свойств горных пород Процессы подтопления и заболачивания городских территорий Изменение экзогенных процессов Изменение здоровья людей
Энергетический комплекс	
Тепловые электростанции	Изменение атмосферы Загрязнение гидросферы Изменение рельефа
Гидроэлектростанции	Затопление территории (изменение рельефа) Изменение гидрогеологических условий Процесс переформирования береговых склонов Изменение температурного режима Активизация сейсмических явлений
Атомные электростанции	Изменение рельефа Изменение температурного режима водоемов Изменение атмосферы

Горнодобывающий комплекс	Изменение рельефа Изменение атмосферы Изменение растительного покрова Изменение почвенного покрова Изменение животного состава Изменение гидрогеологических условий Изменение экзогенных процессов Изменение почвенного покрова Загрязнение горных пород
Агропромышленный комплекс	Изменение растительного покрова Изменение почвенного покрова Изменение гидрографической сети Загрязнение горных пород Изменение экзогенных процессов Изменение гидрогеологических условий
Рекреационный комплекс	Изменение растительного покрова Изменение почвенного покрова Изменение животного состава Изменение гидрографической сети

Моделирование в денной работе производилось на примере градопромышленного комплекса, поэтому в качестве критериев оценки выбраны именно критерии этого ТПК.

Как было показано в предыдущей главе, в качестве критериев оценки устойчивости геологической среды в основном выделяют экзогенные процессы, особенности геологического строения, рельеф, уровень грунтовых вод, наличие поверхностных вод и различные техногенные воздействия.

Но так как моделирование в данной работе проводилось на примере г. Москвы итоговый выбор критериев должен был соответствовать информации об устойчивости геологической среды именно этой территории.

Для этого было рассмотрено детальное многолетнее исследование, проведенное группой специалистов во главе с Осиповым В.И..

Осиповым В.И. и соавторами [Осипов В.И. 1997] при оценке устойчивости геологической среды города Москвы, под которой понималась стабильность функционирования и долговечность жилой промышленной застройки, транспортных магистралей, инженерных сетей, зеленых насаждений и других объектов городской среды, определялся геологический риск территории. Он оценивался на основе учета трех важнейших опасных геологических процессов, развивающихся на территории города: карстово-суффозионном, подтоплении, образовании оползней (Рисунок 1.3.1). Также Осипов В.И. предлагает учитывать повышенную сжимаемость техногенных грунтов, загрязнение геологической среды промышленными и бытовыми отходами, техногенные физические поля, такие как вибрационное, электрическое, тепловое (Рисунок 1.3.2) [Осипов В.И. 1994].

Эти модели устойчивости геологической среды позволили выбрать критерии, наиболее полно и универсально ее описывающие, при этом с большим количеством сопоставимых разноплановых данных по каждому из них.

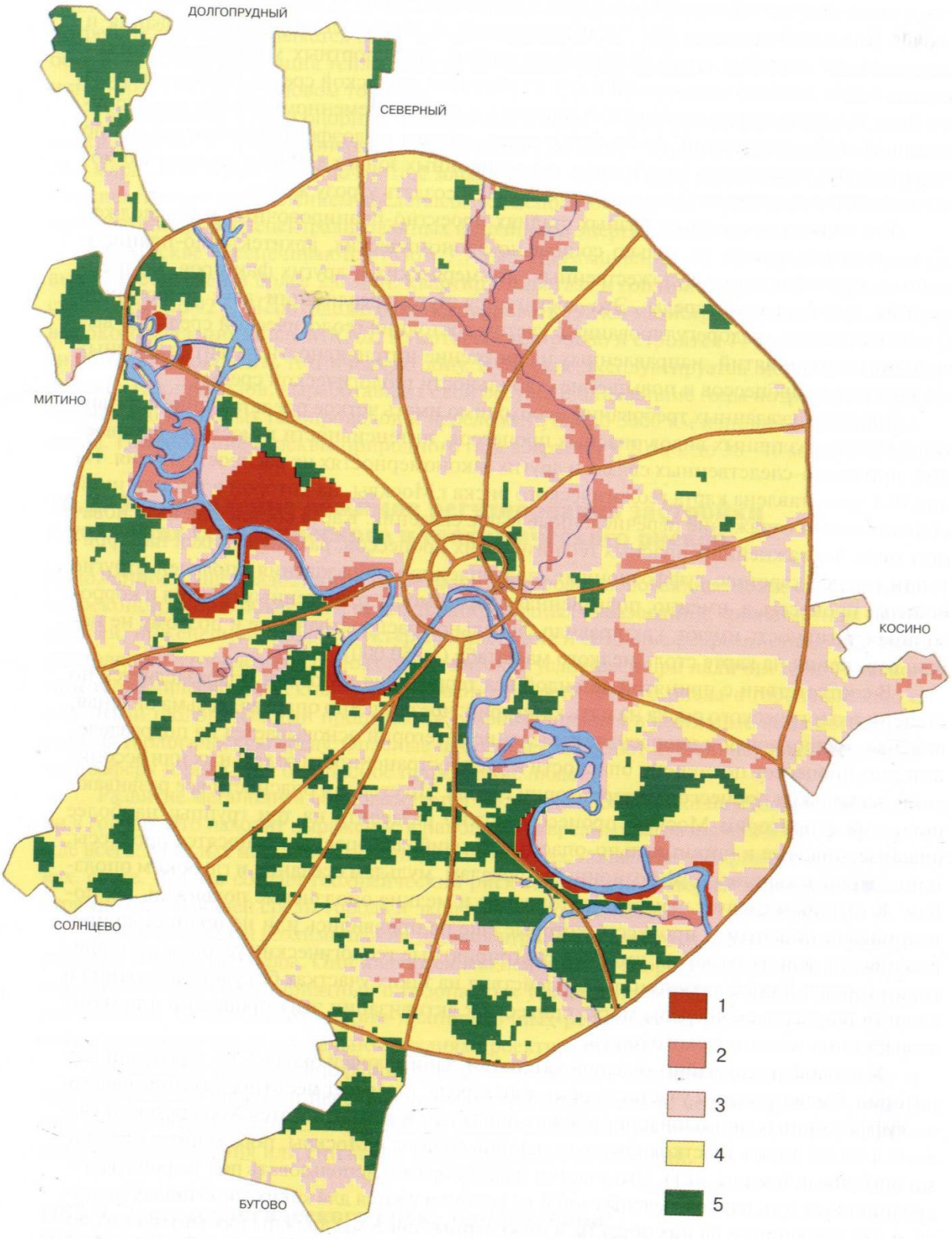


Рисунок 1.3.1. Карта геологического риска территории г. Москвы. Категории геологического риска 1 – чрезвычайно опасная, 2 – весьма опасная, 3 – опасная, 4 – малоопасная, 5 – неопасная. [Осипов В.И. 1997]

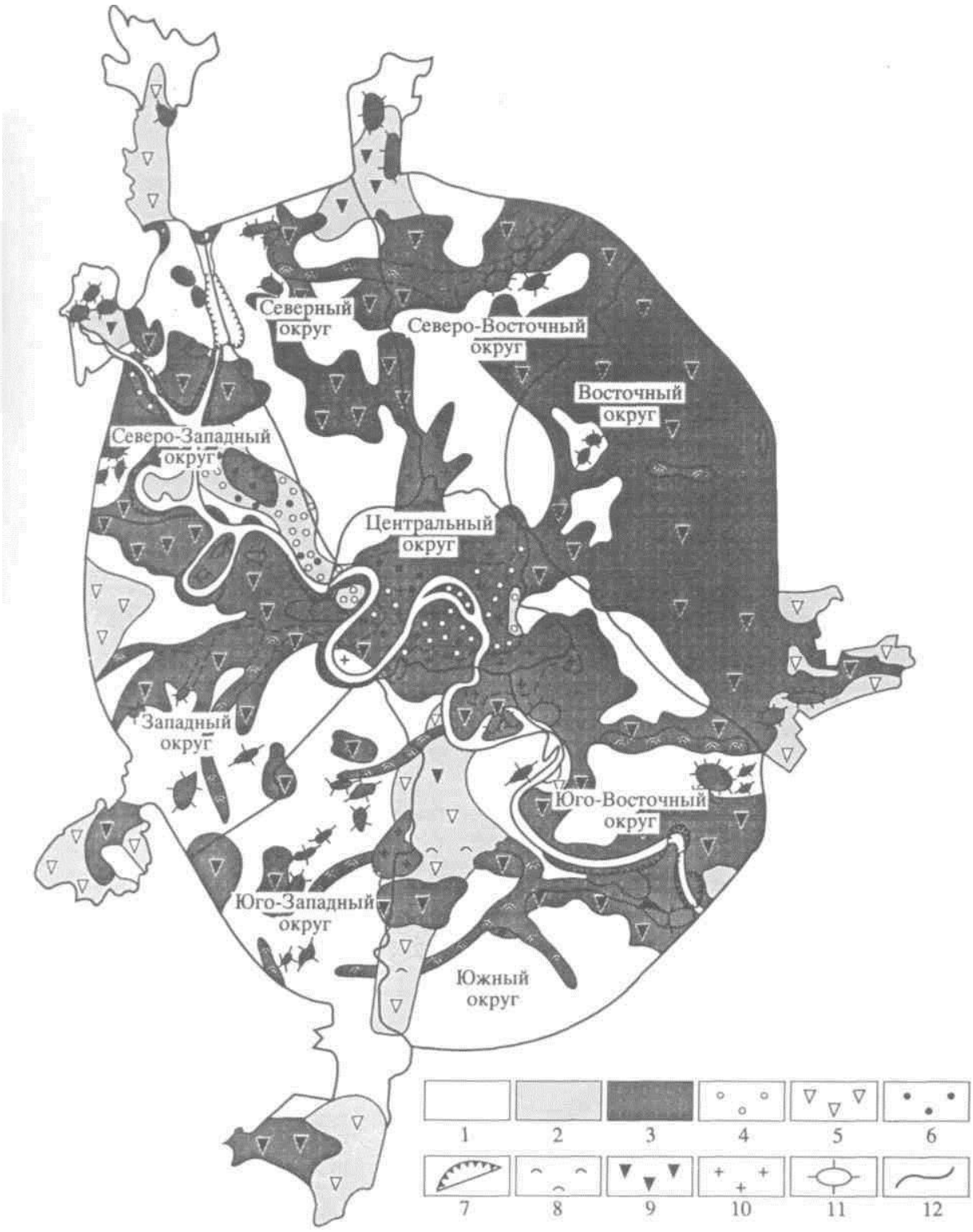


Рисунок 1.3.2. Схема районирования территории Москвы по геологическому риску. Территории: 1 – неопасные (геодинамически стабильные, 2 – потенциально опасные (потенциального риска), 3 –опасные (актуального риска); опасности: 4 – потенциальное развитие карстово-суффозионных провалов, 5 – потенциальное подтопление, 6 – карстово-

суффозионные провалы, 7 – глубокие оползни, 8 – мелкие оползни и оплывины, 9 – подтопление, 10 – повышенная сжимаемость техногенных образований, 11 – газогенерирующие грунты; 12 – границы административных округов. [Осипов В.И. 1994]

Как видно из представленных карт (Рисунки 1.3.1 и 1.3.2), при изменении состава критериев оценки устойчивости геологической среды, границы опасных и безопасных территорий сильно изменяются. Поэтому выбор наиболее значимых критериев является основой точности модели.

В связи с этим поиск критериев оценки был сконцентрирован на их универсальности и

точности отражения реальной ситуации.

Оценка опасных процессов возможна лишь после проведения анализа факторов, которые влияют на их формирование. Поэтому в качестве критериев устойчивости геологической среды было бы правильно рассматривать не сами процессы, а определенные инженерно-геологические условия, которые приводят к формированию вышеуказанных процессов, а также показатели геологической среды, являющиеся индикатором их наличия.

Однако, для получения информации по этим процессам исследователями предварительно были обработаны данные по геологическому строению, гидрогеологии, тектонике, гидрологии, ландшафтно-геоморфологическим условиям и др. Т.е. следует учитывать, что данные по опасным геологическим процессам сами по себе уже являлись результатом обработки огромного числа исходных параметров, описывающих инженерно-геологические условия, т.е. были результатом промежуточного моделирования. Такие данные не могут быть получены на первоначальном этапе. А, как указывалось выше, с точки зрения полезности оценки устойчивости геологической среды, намного лучше было бы ее получить в самом начале исследований.

Поэтому для оценки устойчивости геологической среды при помощи можно выделить следующие:

1. Рельеф, как показатель, который, так или иначе, проявляет все геологические процессы, как экзогенные, так и эндогенные;
2. Разломы, и линиаментные зоны, поскольку они обычно приурочены к тектонически неустойчивым территориям, что приводит к образованию оползней и картово-суффозионных процессов;
3. Уровень грунтовых вод, как показатель подтопления территории;
4. Гидрологическая сеть, как фактор, во многом влияющий на все главные экзогенные процессы;
5. Мощность техногенных образований, как показатель степени техногенного воздействия на территорию и объема слабых грунтов.

Именно эти параметры непосредственно связаны с устойчивостью, являются своеобразными индикаторами всех раннее рассматриваемых инженерно-геологических процессов и доступны на первом этапе исследований и изысканий. Часто они уже имеются в качестве архивной информации ещё на предпроектной стадии, что позволит оценить устойчивость геологической среды в самом начале работ. В случае же создания модели при помощи машинного обучения для решения задач определения устойчивости геологической среды на предпроектной стадии необходимо, чтобы базовые факторы были максимально универсальны, и возможно было использовать данную модель в широком интервале их флуктуаций.

1.4. Обоснование предельно допустимых норм критериев оценки устойчивости геологической среды

В качестве примера для моделирования устойчивости геологической среды был выбран градопромышленный комплекс и были определены критерии для его оценки. Однако ещё необходимо выделить предельно допустимые нормы для указанных критериев. Непосредственно к выбранным критериям нормы практически не разрабатывались, однако, существуют нормы связанные с защитой от опасных экзогенных и эндогенных процессов.

Так наличие разрывных нарушений и линиаментных зон самих по себе не является препятствием для градостроения. Но по СП 14.13330.2014 «Строительство в сейсмических районах» [СП 14.13330.2014] Площадки строительства, в пределах которых отмечены тектонические нарушения, перекрытые чехлом рыхлых отложений мощностью менее 10 м, участки с крутизной склонов более 15°, с оползнями, обвалами, осыпями, карстом, селями, участки, сложенные грунтами III и IV категорий являются неблагоприятными в сейсмическом отношении.

При необходимости строительства зданий и сооружений на таких площадках следует принимать дополнительные меры по укреплению их оснований, усилению конструкций и инженерной защите территории от опасных геологических процессов.

Так же рельеф и гидрологическая сеть не нормируется, но существуют показатели оценки опасных геологических процессов. Согласно СП 116.13330.2012 «границы оползнеопасных территорий устанавливают по данным комплексных инженерных изысканий с использованием расчетов устойчивости склонов и материалов сравнительного инженерно-геологического анализа применительно к особенностям рельефа, геологического строения, гидрогеологических и сейсмических условий, характера растительного покрова и климата» [СП 116.13330.2012]. Расчет опасности оползня в каждом конкретном случае выполняется отдельно.

«По характеру карстовой опасности для строительных объектов следует выделять следующие основные ее виды:

1) Карстоопасность вида А обусловлена повышенной чувствительностью закарстованных территорий к загрязнению геологической среды (в первую очередь, подземных вод) вследствие образования провалов, наличия карстопроявлений (карстовых воронок, мульд оседания, разуплотненных зон в грунтовой толще и т.д.). Карстоопасность вида необходимо учитывать, прежде всего, при проектировании и эксплуатации полигонов захоронения промышленных и бытовых отходов, нефтепроводов, канализационных коллекторов, автомобильных и железных дорог, химических предприятий, атомных электростанций и т.п.

2) Карстоопасность вида В обусловлена вероятностью повреждения строительных объектов или земельных участков поверхностными карстопроявлениями (провалами, локальными и общими оседаниями и др.).

3) Карстоопасность вида С обусловлена возможными осложнениями при строительстве и эксплуатации подземных сооружений и фундаментов глубокого заложения вследствие наличия и развития карстопроявлений (обводненных карстовых полостей и зон повышенной трещиноватости горных пород, локальных разуплотненных зон в грунтовой толще и др.).

4) Карстоопасность вида D обусловлена недопустимыми утечками воды из водоемов, каналов, водоотводных канав и др.» [СП 116.13330.2012]

Что касается подтопления, то согласно недействующему на данный момент СНиП 2.06.15-85 «нормы осушения (глубины понижения грунтовых вод, считая от проектной отметки территории) при проектировании защиты от подтопления принимаются в зависимости от характера застройки защищаемой территории» [СНиП 2.06.15-85] в соответствии с табл. 1.4.1.

Таблица 1.4.1. Нормы осушения грунтовых вод [СНиП 2.06.15.-85]

Характер застройки	Норма осушения, м
1. Территории крупных промышленных зон и комплексов	До 15
2. Территории городских промышленных зон, коммунально-складских зон, центры крупнейших, крупных и больших городов	5
3. Селитебные территории городов и сельских населенных пунктов	2
4. Территории спортивно-оздоровительных объектов и учреждений обслуживания зон отдыха	1
5. Территории зон рекреационного и защитного назначения (зеленые насаждения общего пользования, парки, санитарно-защитные зоны)	1

Соответственно в городе на селитебной территории область считалась подтопленной при уровне грунтовых вод менее 2 м. На текущий момент расчет подтопления и присвоения территории статуса подтопленной осуществляется согласно СНиП 2.02.01-83 [СНиП 2.02.01.-83].

Геологическое строение для определения инженерно-геологического риска регламентируется СНиП 2.02.01-83 [СНиП 2.02.01.-83]. Характеристики грунтов, которые приведены в табл.1.4.2-1.4.4, могут быть использованы для расчетов оснований сооружений.

Таблица 1.4.2. Нормативные значения удельного сцепления c_n, кПа (кгс/см^2), угла внутреннего трения φ_n, град., и модуля деформации E, МПа (кгс/см^2), песчаных грунтов четвертичных отложений [СНиП 2.02.01.-83].

Песчаные грунты	Обозначения характеристик грунтов	Характеристики грунтов при коэффициенте пористости , равном			
		0,45	0,55	0,65	0,75
Гравелистые и крупные	c_n	2 (0,02)	1 (0,01)	-	-
	φ_n	43	40	38	-
	E	50 (500)	40 (400)	30 (300)	-
Средней крупности	c_n	3 (0,03)	2 (0,02)	1 (0,01)	-
	φ_n	40	38	35	-
	E	50 (500)	40 (400)	30 (300)	-
Мелкие	c_n	6 (0,06)	4 (0,04)	2 (0,02)	-
	φ_n	38	36	32	28
	E	48 (480)	38 (380)	28 (280)	18 (180)
Пылеватые	c_n	8 (0,08)	6 (0,06)	4 (0,04)	2 (0,02)
	φ_n	36	34	30	26
	E	39 (390)	28 (280)	18 (180)	11 (110)

Таблица 1.4.3. Нормативные значения удельного сцепления c_n, , кПа (кгс/см^2), угла внутреннего трения φ_n, град., пылевато-глинистых нелессовых грунтов четвертичных отложений [СНиП 2.02.01.-83].

Наименование грунтов и пределы нормативных значений их показателя текучести		Обозначения характеристик грунтов	Характеристики грунтов при коэффициенте пористости e, равном						
			0,45	0,55	0,65	0,75	0,85	0,95	1,05
Супеси	0 $\leq I_L \leq$ 0,25	c_n	21 (0,21)	17 (0,17)	15 (0,15)	13 (0,13)	-	-	-
		φ_n	30	29	27	24	-	-	-
	0,25 $< I_L \leq$ 0,75	c_n	19 (0,19)	15 (0,15)	13 (0,13)	11 (0,11)	9 (0,09)	-	-
		φ_n	28	26	24	21	18	-	-
Суглинки	0 $< I_L \leq$ 0,25	c_n	47 (0,47)	37 (0,37)	31 (0,31)	25 (0,25)	22 (0,22)	19 (0,19)	-
		φ_n	26	25	24	23	22	20	-
	0,25 $< I_L \leq$ 0,5	c_n	39 (0,39)	34 (0,34)	28 (0,28)	23 (0,23)	18 (0,18)	15 (0,15)	-
		φ_n	24	23	22	21	19	17	-
	0,5 $< I_L \leq$ 0,75	c_n	-	-	25 (0,25)	20 (0,20)	16 (0,16)	14 (0,14)	12 (0,12)
		φ_n	-	-	19	18	16	14	12

Глины	0 $< I_L \leq$ 0,25	c_n	-	81 (0,81)	68 (0,68)	54 (0,54)	47 (0,47)	41 (0,41)	36 (0,36)
		φ_n	-	21	20	19	18	16	14
	0,25 $< I_L \leq$ 0,5	c_n	-	-	57 (0,57)	50 (0,50)	43 (0,43)	37 (0,37)	32 (0,32)
		φ_n	-	-	18	17	16	14	11
	0,5 $< I_L \leq$ 0,75	c_n	-	-	45 (0,45)	41 (0,41)	36 (0,36)	33 (0,33)	29 (0,29)
		φ_n	-	-	15	14	12	10	7

Таблица 1.4.4. Нормативные значения модуля деформации пылевато-глинистых нелессовых грунтов [СНиП 2.02.01.-83].

Происхождение и возраст грунтов		Наименование грунтов и пределы нормативных значений их показателя текучести		Модуль деформации грунтов Е, МПа (кгс/см^2), при коэффициенте пористости *e*, равном										
				0,35	0,45	0,55	0,65	0,75	0,85	0,95	1,05	1,2	1,4	1,6
Четвертичные отложения	Аллювиальные Делювиальные Озерные Озерно-аллювиальные	Супеси	$0 \le I_L \le 0,75$	-	32 (320)	24 (240)	16 (160)	10 (100)	7 (70)	-	-	-	-	-
		Суглинки	$0 \le I_L \le 0,25$	-	34 (340)	27 (270)	22 (220)	17 (170)	14 (140)	11 (110)	-	-	-	-
			$0,25 < I_L \le 0,5$	-	32 (320)	25 (250)	19 (190)	14 (140)	11 (110)	8 (80)	-	-	-	-

			$0,5 < I_L \leq 0,75$	-	-	-	17 (170)	12 (120)	8 (80)	6 (60)	5 (50)	-	-	-
		Глины	$0 \leq I_L \leq 0,25$	-	-	28 (280)	24 (240)	21 (210)	18 (180)	15 (150)	12 (120)	-	-	-
			$0,25 < I_L \leq 0,5$	-	-	-	21 (210)	18 (180)	15 (150)	12 (120)	9 (90)	-	-	-
			$0,5 < I_L \leq 0,75$	-	-	-	-	15 (150)	12 (120)	9 (90)	7 (70)	-	-	-
	Флювиогляциальные	Супеси	$0 \leq I_L \leq 0,75$	-	33 (330)	24 (240)	17 (170)	11 (110)	7 (70)	-	-	-	-	-
		Суглинки	$0 \leq I_L \leq 0,25$	-	40 (400)	33 (330)	27 (270)	21 (210)	-	-	-	-	-	-
			$0,25 < I_L \leq$	-	35	28	22	17	14	-	-	-	-	-

			0,5 $0,5 < I_L \leq 0,75$	- -	(350) -	(280) -	(220) 17 (170)	(170) 13 (130)	(140) 10 (100)	 7 (70)	 -	 -	 -	 -
	Моренные	Супеси Суглинки	0,5	75 (750)	55 (550)	45 (450)	-	-	-	-	-	-	-	-
Юрские отложения оксфордского яруса		Глины	$-0,25 \leq I_L \leq 0$ $0 < I_L \leq 0,25$ $0,25 < I_L \leq 0,5$	- - -	- - -	- - -	- - -	- - -	- - -	27 (270) 24 (240) -	25 (250) 22 (220) -	22 (220) 19 (190) 16 (160)	- 15 (150) 12 (120)	- - 10 (100)

Геодинамические характеристики, определяющие деформацию геологической среды с проявлением негативных последствий, можно представить в виде масштаба и интенсивности проявления текущего напряженно-деформированного состояния верхней части литосферы. Сами же эти характеристики определяются масштабами предполагаемого сейсмического эффекта и параметрами критических скоростей деформации. Если исходить из того, что порог разрушения любого твердого тела составляет порядка 0,0001 отн.ед., то предельным (или критическим) уровнем геодинамического воздействия для какого-либо объекта будет величина деформации в 0,00001 отн.ед., которая применяется при осуществлении оценки аномальных техногенных деформаций. Принимая во внимание установленные факты изменения деформационных процессов в разломных зонах в пространстве и времени, критический уровень деформации (0,00001) достигается в течение 15 - 30 лет в локальных зонах.

Такие сроки сопоставимы с минимальными сроками эксплуатации объектов и сооружений особой важности. Нарушение функционирования таких объектов может привести к экологическому бедствию. И именно уровень деформации в 0,0001 отн.ед. и приводит к таким последствиям (Таблица 1.4.5).

Экзогенные геологические процессы могут проявляться вне зависимости от человеческой деятельности. Но при этом техногенные факторы могут усилить или ослабить их проявление. Необоснованное вмешательство людей в естественное развитие экзогенных процессов способно привести к необратимым изменениям природных ландшафтов из-за их катастрофической активизации. Примерами таких проявлений являются сели, оползни, оседание поверхности, карсты, и др.

Таблица 1.4.5. Критерии оценки экологической опасности деформаций и изменения геологической среды [Критерии… 1992].

N N пп	Показатели	Параметры		
		Экологическое бедствие (ст.59)	Чрезвычайная экологическая ситуация (ст.58)	Относительно удовлетворительная ситуация
1.	Аномальные техногенные деформации горного массива (более 0,00001 отн.ед.) и индуцированная сейсмичность, % территории	более 40 - 50	20 - 40	2 - 3
2.	Механические нарушения горного массива при недропользовании, ведущие к загрязнению геологической среды, аномальные деформации горных пород, отн.ед.	0,0001	0,00001	менее 0,000001
3.	Просадки земной поверхности, оползни,	более 30	20 - 30	менее 5

	сели, карсты, обусловленные техногенной нагрузкой, % территории			

На основе всего этого ряда нормативных критериев, возможно оценить устойчивость геологической среды посредством покомпонентного анализа каждого негативного фактора и их последующей суммации. Однако, в случае применения машинного обучения, возможно избежать всего обилия индивидуальных расчетов и сложностей интегральных оценок и произвести анализ устойчивости геологической среды основываясь на результатах уже проведенных подобных анализов и базовых критериях оценки, не требующих дополнительных расчетов.

При этом, в случае наличия достаточно большой базы данных и результатов оценки устойчивости геологической среды, можно довести точность моделирования при помощи машинного обучения до 100% и минимизировать время, затрачиваемое на него. Таким образом проводить подобную интегральную оценку будет возможно при любом новом крупном строительстве без большого объема предварительных исследовательских работ, что снизит инженерно-геологические риски и повысит безопасность построек.

2. Обоснование применения машинного обучения для оптимизации анализа устойчивости геологической среды

2.1. Выбор подхода к оптимизации анализа устойчивости геологической среды

Определение устойчивости геологической среды во многом сводится к моделированию и прогнозированию, в данном случае произведенным при помощи методов машинного обучения.

Под инженерно-геологическим прогнозированием понимается предсказание инженерно-геологических условий [Бондарик Г.К. 1986]. Инженерно-геологическое прогнозирование выполняется методами: индикаторов (геологических, физико-географических, антропогенных); природных (геологических) аналогий (сравнительно-геологический метод - качественный прогноз реакции территории на планируемое техногенное воздействие по комплексу известных инженерно-геологических условий; метод инженерно-геологических аналогий - сравнительный анализ с однотипной по инженерно-геологическим условиям территорией, на которой уже ведётся аналогичная техногенная деятельность); физического (воссоздание геологической системы для прогноза тех процессов, исследование которых в естественных условиях затруднено) и математического (основаны на использовании детерминированных и статистических моделей геологической среды, вероятностных моделей для изучения динамики развития геологического и инженерно-геологического процессов) моделирований. Выбор методов инженерно-геологического прогнозирования связан со стадийностью инженерно-геологических исследований. На начальных стадиях используются сравнительно-геологические методы прогноза, дополняемые отдельными поверочными расчётами, а также вероятностно-статистические методы. На последующих

стадиях, когда выполняется детальное обоснование схем размещения отдельных сооружений и возрастает необходимость в количественных прогнозах, применяются в большом объёме методы моделирования [Трофимов В.Т. 2015].

Однако, в первую очередь, говоря об инженерно-геологическом прогнозе необходимо дать его классификацию. Цель этой классификации состоит в упорядочивании представлений о возможных классах инженерно-геологических прогнозов и в выявлении тех классов, которые представляются интересными, но методы которых пока не разработаны. Эта цель достигается, если для разделения полного множества прогнозов на подмножества выбрать небольшое число признаков-оснований. Ограниченное число делящих признаков обеспечивает простоту классификации и в то же время классификация удовлетворяет поставленной цели.

Первым признаком, по которому следует разделить прогнозы, является характер оценок, получаемых в результате прогнозирования: качественные или количественные. Качественный прогноз, как показывает его название, содержит качественную оценку будущего состояния геологической среды в целом или оценку каких-либо ее свойств (или структуры) в будущем. Оценка предусматривает указание тенденций в развитии геологической среды, результатов ее развития на некоторый более или менее точно установленный момент будущего времени (или без привязки к какому- то моменту времени) и указание вероятности достижения этих результатов. Количественный прогноз состоит в описании структуры и свойств геологической среды в количественных мерах, в получении показателей (или функций), характеризующих ее структуру и свойства в заданные моменты будущего времени или в некоторый конечный момент времени. При этом предполагается, что количественные показатели прогноза отвечают заданной точности и надежности, или указывается их изменение во времени.

Второй признак-основание предусматривает разделение прогнозов по

группам методов прогнозирования. Набор включает четыре группы методов прогнозирования: 1) индикаторов (геологических, физико-географических, антропогенных); 2) инженерно-геологических аналогий; 3) физического и 4) математического моделирований.

Третий важнейший признак, который разделяет полное множество прогнозов, является их отношение ко времени. Набор прогнозов, выделяемый по этому признаку, предусматривает три категории прогнозов: 1) разрабатываемые безотносительно ко времени; 2) срочные, даваемые на конкретный срок,— прогнозы количественные по времени; 3) бессрочные прогнозы на будущее время, срок которых не установлен или устанавливается путем качественных оценок (прогнозы качественные по времени).

Четвертый набор признаков, учитываемый классификацией, предполагает разделение инженерно-геологических прогнозов по их отношению к геологической среде. По этому признаку такие прогнозы разделяются на локальные, разрабатываемые для квазиоднородной геологической среды; региональные, составляемые для неоднородной среды; глобальные, разрабатываемые с целью предсказания изменения ее структуры и свойств — компонентов геологической среды в масштабе планеты.

По охвату свойств геологической среды — ее компонентов, прогнозы целесообразно разделить на общие, предусматривающие предсказание инженерно-геологических условий (структуры и свойств) в целом, и на частные, целью которых является предсказание одного какого-либо свойства геологической среды, одного компонента инженерно-геологических условий. Последовательное введение рассмотренных признаков-оснований позволяет разделить полное множество инженерно-геологических прогнозов на классы и по пяти однородным наборам (Рисунок 2.1.1) [Бондарик Г.К. 1986].

.

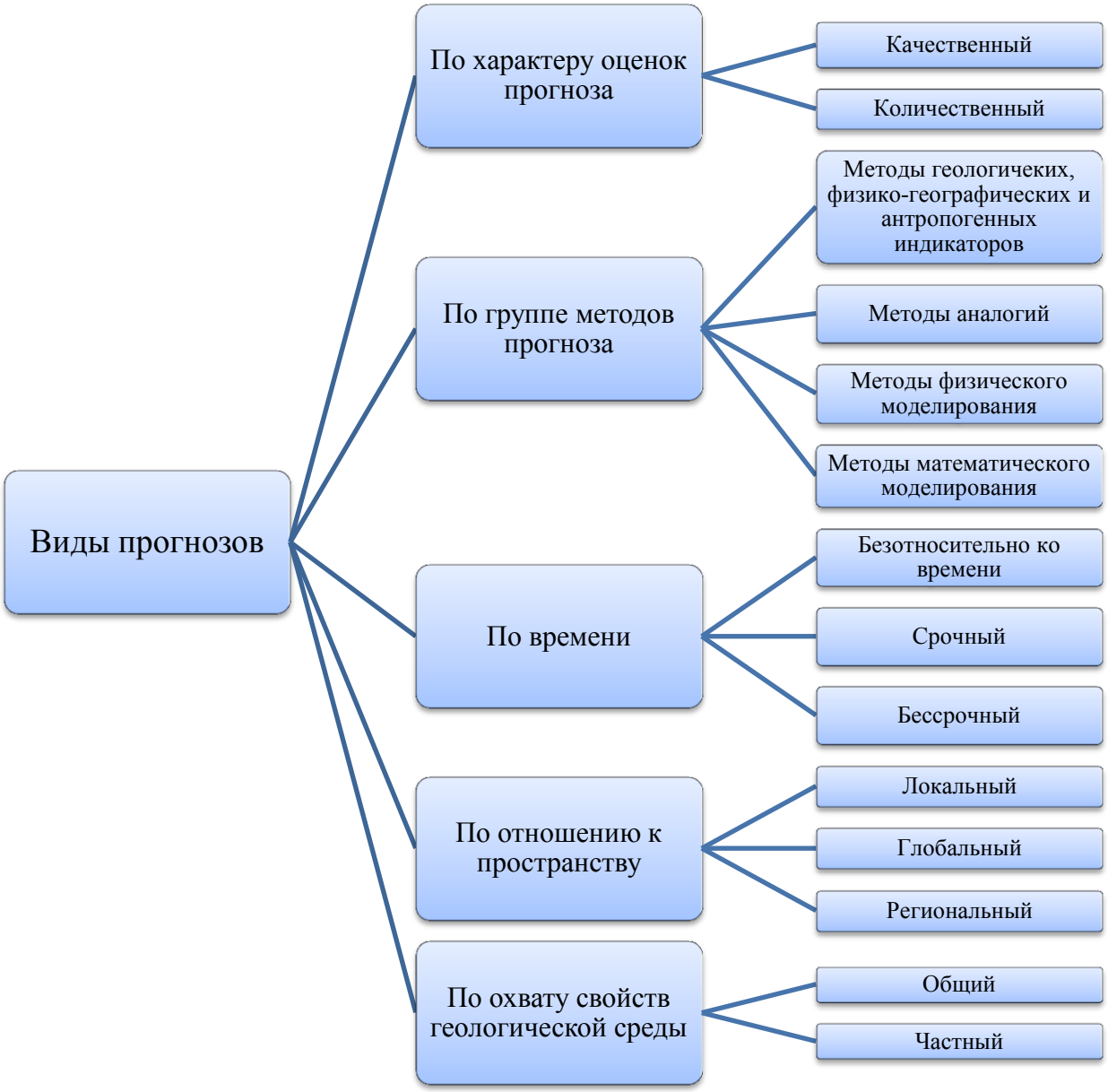


Рисунок 2.1.1. Классификация инженерно-геологических прогнозов [Бондарик Г.К. 1986].

В настоящее время различают два подхода к временным прогнозам природных явлений и процессов.

Первый подход основан на корреляционно-регрессионном принципе, т.е. на основе установленных корреляционных связей между прогнозируемым процессом и факторами, его обусловливающими, делается экстраполяция развития данного процесса с заданной заблаговременностью. При этом подходе необходимо выявить причинно-следственные связи между факторами и прогнозируемыми явлениями.

Второй подход заключается в том, что на основе анализа рядов данного явления при помощи спектрального или гармонического анализа выявляются закономерности развития во времени и при использовании этих закономерностей делается экстраполяция развития этого процесса на заданный срок. Преимущество второго подхода заключается в том, что он не требует выявления причинно-следственных связей между прогнозируемым явлением и факторами, его обусловливающими, что представляет собой обычно сложную задачу, на решение которой затрачивается много усилий и не всегда может увенчаться успехом. При этом принято считать, что в самом ряду проявления того или иного процесса в интегральной форме заложена информация о всех факторах, которые определяют развитие этого процесса. В этом случае важно иметь длину ряда развития процесса, которая охватывала бы время, достаточное для установления основных закономерностей [Шеко А.И. 1984].

При решении задачи определения устойчивости геологической среды речь идет о качественном бессрочном прогнозе различного масштаба общего характера. Так как методика определения позволяет производить анализ территории разного масштаба в зависимости от исходных данных, а сам прогноз определяет степень устойчивости всей геологической среды и тенденцию её развития в целом методом индикаторов.

Однако можно заявить абсолютно точно, что это задача в первую очередь относится к задаче интеллектуального анализа данных.

Интеллектуальный анализ данных — собирательное название, используемое для обозначения совокупности методов обнаружения в данных ранее неизвестных, нетривиальных, практически полезных и доступных интерпретации знаний, необходимых для принятия решений в различных сферах человеческой деятельности [Паклин Н.Б. 2009].

Под данными при этом понимается фактическая информация в формализованном виде, которую можно передавать и обрабатывать в некотором информационном процессе, т.е. в процессе создания, поиска,

получения, сбора, обработки, хранения, накопления, распространения и использования информации.

Информация же при этом представляет собой сведения о каком-либо явлении или процессе. Информация является базовой единицей измерения при анализе данных и присутствует как в самих данных, так и после их обработки для получения знаний, т.е. проверенной информации, позволяющей решать конкретную задачу.

Интеллектуальный анализ данных объединяет различные методы типизации, моделирования, основанные на применении различных способах получения знаний из набора данных.

Также при интеллектуальном анализе важно наглядно представлять результаты вычислений, чтобы даже не специалисты могли пользоваться результатами его обработки.

Изначально интеллектуальный анализ используется при наличии достаточно большой базы данных, содержащей сложно структурированную, разноплановую и часто «скрытую» информацию. Извлечение подобной информации из большого объема данных является сложной задачей. Поэтому необходимо найти или разработать методы получения знаний, которые скрыты в большом количестве исходных данных.

Под скрытыми знаниями понимаются знания:

- неизвестные ранее – т.е. знания, являющимися новыми, а не подтверждающими ранее полученную информацию;

- нетривиальные – т.е. не очевидные при просматривании или при проведении простой статистической обработки;

- полезные – т.е. представляющие собой ценность для ученого или потребителя;

- легко интерпретируемые – т.е. знания, которые легко привести в наглядную для пользователя форму и которые можно легко объяснить в рамках исследуемой задачи.

Все эти требования удовлетворяют задаче определения устойчивости геологической среды.

Помимо метода машинного обучения предварительно необходимо рассмотреть другие возможные методы решения задачи определения устойчивости геологической среды. А так как это в первую очередь задача моделирования, ниже рассмотрены методы моделирования с точки зрения их пригодности для её решения.

2.1.1. Применимость машинного обучения для решения задачи определения устойчивости геологической среды

В данной работе определяется возможность применения передовых методов анализа данных для выявления информации и знаний, скрытых внутри реальных баз данных. Изложенный подход исходит из того, что существуют большие объемы данных, полученные при помощи различных методов измерения. Эти данные могут предоставить интересную и полезную информацию при условии, что они обрабатываются квалифицированным персоналом. Эксперты могут выявить, такие параметры данных, их значения или их комбинации, которые могут иметь отношение к различным конкретным задачам и, при помощи обобщения, такая информация может в дальнейшем использоваться для улучшения и оптимизации конечных результатов различных процессов. Тем не менее, в реальных данных часто возникают препятствия для их легкой обработки из-за больших объемов, недостающих значений, непредставительных параметров типизации, шумов, содержания как цифровых, так и буквенных значениях и т.д. Эксперты могут справиться с такими трудностями, если объемы данных не слишком большие, степень размерности не слишком высока, ситуация не очень осложнена большим количеством малозначимых параметров типизации и имеется достаточно времени для решения данной проблемы. В противном случае машина (компьютер) может помочь разобраться в задаче или указать,

как это сделать. Ожидаемый результат анализа данных должен быть в форме знания, например, набора правил, деревьев решений, или общих типизаторов, которые могут быть использованы в будущем в качестве ценных полезных инструментов.

Очевидно, что такие инструменты должны функционировать «разумно», то есть, так же как их человеческие аналоги. Современная компьютерная наука предоставляет широкий спектр алгоритмов в области искусственного интеллекта. Одна из различных возможностей – это использование набора доступных образцов данных, подготовленных экспертами, чтобы на основе этих обучающих примеров «научить» машину, что и как делать в конкретной ситуации, когда новые неизвестные, но так или иначе подобные ситуации возникнут в будущем. Этот метод называется обучением с учителем. При этом исходные данные обрабатываются и разбиваются на две выборки: обучающую и тестовую.

Детальная схема метода машинного обучения, при решении задачи типизации, представлена на рисунке 2.1.1.1.

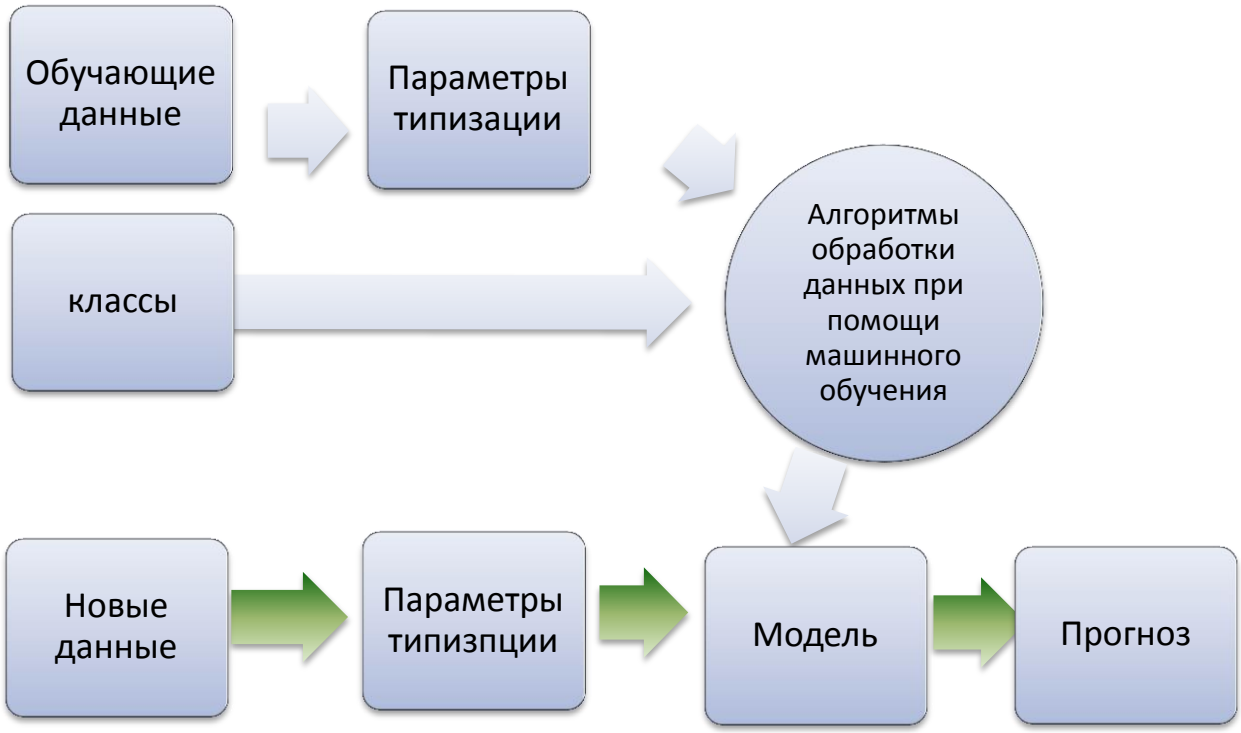


Рисунок 2.1.1.1 Схема машинного обучения [Witten H. 1999].

Обучающая выборка — это выборка, по которой производится оптимизация параметров модели, которые оптимизируются (обучаются) таким образом, чтобы на объектах выборки определялись нужные классы, или близкие к ним.

Если модель зависимости построена по обучающей выборке, то оценка качества этой модели, сделанная по той же выборке, оказывается, как правило, слишком завышенной. Это называется переобучением. На практике оно встречается очень часто. Хорошую эмпирическую оценку качества построенной модели даёт её проверка на независимых данных, которые не использовались для обучения.

Для этого используется тестовая выборка. Если обучающая и тестовая выборки независимы, то оценка, сделанная по тестовой выборке, является достоверной.

При обучении с учителем значение по каждой точке представляет собой пару «объект, класс». Требуется найти функциональную зависимость классов от описаний объектов и найти алгоритм, на входе дающий описание объекта, состоящее из формализованного списка параметров, и выдающий на выходе определение класса объекта с достаточной точностью для каждого объекта. Критерий качества, как правило, определяется как средняя ошибка в определении классов, выданных алгоритмом, по всем объектам выборки. Под учителем понимается либо сама обучающая выборка, либо тот, кто указал на заданных объектах правильные классы.

Для решения рассматриваемой проблемы решалась задача типизации. Она состоит в том, что множество допустимых классов конечно. Их называют метками классов. Класс — это множество всех объектов с данным значением метки. Т.е. в результате типизации каждый объект получает номер или наименование класса, выдаваемые алгоритмом типизации в результате проводимых вычислений по каждому объекту.

После получения при обучении знаний, машина может предложить решение, основанное на определенном сходстве с одним или более обобщенными случаями, известными из прошлых построений модели. Такой подход в обучении машин пытается подражать поведению человека, «учителя» машины.

В настоящее время существуют десятки и сотни различных компьютерных алгоритмов обработки данных в рамках машинного обучения: некоторые из них работают быстро и легко, другие сложные, но гораздо лучше приспособлены для решения определенных задач, некоторые предоставляют понятные знания (например, деревья решений и правила), а некоторые являются черным ящиком (например, искусственные нейронные сети). Из опубликованной и доказанной no-free-lunch (NFL) теоремы следует, что не существует единственного универсального алгоритма поиска и оптимизации данных. Другими словами, конкретные данные могут обрабатываться одними алгоритмами хорошо, а другими хуже или

неэффективно, однако, у различных данных могут быть совершенно противоположные результаты с той же группой алгоритмов. Поиск наиболее подходящего для решения изучаемой задачи алгоритма производится экспериментально и является отдельной задачей данной работы.

Для разных данных, различные алгоритмы могут достичь наивысшей точности типизации или прогноза, которая определяется качеством и количеством обучающей выборки. Оба этих параметра часто ограничены (или, наоборот, обеспечиваются) доступными ресурсами. Так, например, сбор и подготовка необходимых данных требует времени, денег и специалистов. Во многих случаях, однако, эти расходы могут быть позже возвращены путем применения результатов интеллектуального анализа данных, так как некоторые скрытые факты, неочевидные общие параметры типизации, сильные или слабые связи и комбинации значений, могут, таким образом, быть выявлены и использованы для улучшения инженерно-геологической ситуации. Результаты прогнозирования также зависят от соответствующей предварительной обработки данных.

Задачу машинного обучения можно представить на примере поиска гиперплоскостей, которые определяют границы гиперповерхностей. Разделяющая гиперплоскость – это гиперплоскость, которая отделяет группу объектов, имеющих различную классовую принадлежность. Для решения таких задач необходимо производить экспериментальный подбор алгоритмов и их параметров. В случае, если классов больше, чем два, задача ещё больше усложняется и число алгоритмов и подходящих параметров этих алгоритмов, способных с ней справиться соответственно уменьшается.

Машинное обучение является хорошим примером инструмента интеллектуального анализа. Но для его нормального функционирования необходимо предварительно обработать данные. Для этого лучше всего подходят методы статистики. Машинное обучение - очень деликатный инструмент, и он нуждается в точной настройке для создания хорошей модели. При такой высокой чувствительности интеллектуального анализа,

обработка данных должна компенсировать все трудности его настройки. В таком случае, обработка должна выполняться такими методами, которые хорошо сочетаются с методами машинного обучения.

Так как машинное обучение позволяет решать даже самые сложные задачи типизации быстро и без специальной подготовки, в случае уже построенной модели, оно является продуктивным методом решения имеющейся задачи оптимизации определения уровня устойчивости геологической среды. Однако его точную продуктивность можно установить путем проведения экспериментов.

2.1.2. Применимость других методов моделирования для решения задачи определения устойчивости геологической среды

Все методы прогнозирования можно объединить в две группы: логические и формализованные (рисунок 2.1.2.1).

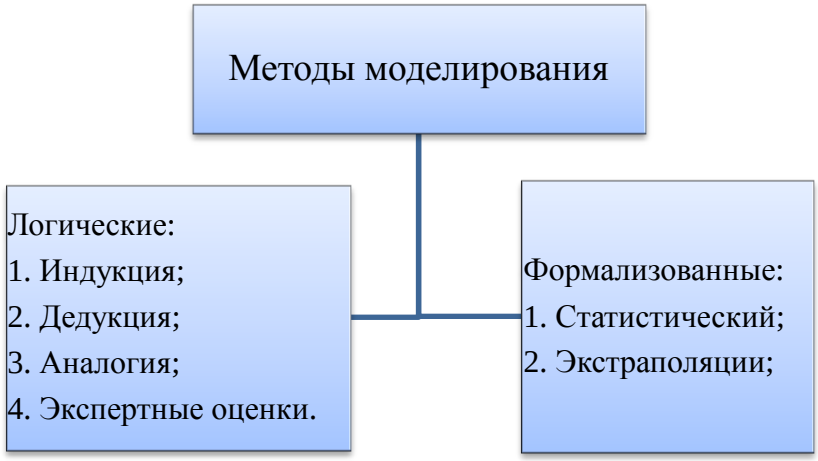


Рисунок 2.1.2.1. Методы моделирования [Степановских А.С. 2001]

В учетом того, что в инженерной геологии, приходится иметь дело, главным образом, со сложными взаимосвязями и зависимостями, то прежде всего используются именно логические методы, так как они способны установить связь между объектами. К таким методам относятся методы дедукции, индукции, аналогии и экспертных оценок.

Методом индукции устанавливают причинные связи предметов и явлений. Индукция – процесс логического вывода на основе перехода от частного положения к общему. Индуктивный метод исследования выявляет сходные и отличительные черты объектов, что позволяет сделать первые попытки обобщения. Однако индуктивные заключения легко ведут к ошибкам, которые обычно возникают из-за большого количества причин развития процесса и из-за непонимания развития процесса вследствие временного фактора или источника воздействия на процесс. В индуктивном исследовании речь идет о следствиях, к которым нужно подыскать причины; их поиск называется объяснением явления, но известное следствие может быть вызвано целым рядом различных причин [Михаленко Ю.П. 1990]. Метод индукции может быть очень продуктивен, однако, процесс взаимосвязи исходных геологических критериев и интегральной устойчивости геологической среды на данное время изучен плохо. Общие представления по этому вопросу появились благодаря также и индуктивному методу, однако одной из задач работы является разработка автоматизированной и доступной для широкого круга специалистов методики прогноза устойчивости геологической среды. Исключительно методом индукции решить такую проблему будет нелегко, поэтому индуктивный метод использован исключительно для поиска наиболее подходящих критериев оценки устойчивости геологической среды.

При **дедуктивном методе** исследование ведется в обратном направлении, от общего к частному, т.е., зная некие общие правила и строя модели на их основе, можно прийти к нужному умозаключению. Дедукция в большей степени определяет стратегию прогноза, но при этом индуктивный и дедуктивный методы сильно взаимосвязаны. Как и в случае с индуктивным методом, отсутствие достаточной изученности рассматриваемой взаимосвязи не позволяет применять метод для прогноза в полной мере. Предварительно необходимо установить точные механизмы формирования геологической среды. При рассмотрении вопроса автоматизации прогноза такое

исследование не представляется возможным, поэтому метод дедукции использовался в данной работе только в части подготовки данных и анализе результатов.

При отсутствии достоверных сведений об объекте исследований, и при этом если такой объект нельзя анализировать матиматически, используется **метод экспертных оценок**, суть которого состоит прогнозе на основе мнений квалифицированных экспертов, которых привлекают для решения определенной проблемы. Выделяют индивидуальную и коллективную экспертизу. При совершении прогноза с использованием метода экспертных оценок специалисты пользуются картографическими, статистическими и другими материалами. При использовании метода экспертных оценок точность прогноза получается достаточно высокой, несмотря на возможные проблемы базы исходных данных, однако метод трудно использовать широко из-за малого числа экспертов, а также медленной работы с большим количеством информации. И даже в случае объединения метода с формализованными методами, широкого применения для решения задачи определения устойчивости геологической среды будет сложно добиться из-за вышеперечисленных причин.

Метод аналогий исходит из того, что один процесс и закономерности его развития в некоторой степени можно использовать для прогноза развития другого сходного процесса. Метод аналогий в основном используют при локальном прогнозировании. Процесс прогнозирования по методу аналогии:

1) Изучить развитие домашней системы.

2) Найти другую систему, которая развивалась подобным образом. Получить данные о ней.

3) Узнать момент времени t, когда чужая система находилась в таком же состоянии, что и домашняя система в настоящее время.

4) Скомпенсировать разницу между домашней и чужой системами, исключая, например, разницу в размере, путем умножения данных чужой системы в соответствии с соотношением размеров.

5) Данные чужой системы, начиная с момента t до настоящего времени, теперь могут быть приняты в качестве прогноза для домашней системы, начиная с настоящего времени.

Метод аналогий очень прост в использовании и позволят работать с любыми данными, но его точность непредсказуема. Не всегда известно, является ли аналог идентичным или только кажется таковым. Поэтому, для повышения точности прогноза, база данных объектов аналогов должна быть очень большой, что очень трудноосуществимо на данный момент.

Формализованные методы подразделяют на статистический, метод экстраполяции, и метод моделирования.

Статистический метод базируется на количественных характеристиках, благодаря которым делается вывод о характере эволюции процесса. К статистическим методам относятся: дескриптивный анализ, корреляционный и регрессионный анализ, факторный анализ, дисперсионный анализ, компонентный анализ, дискриминантный анализ, анализ временных рядов, анализ выживаемости, анализ связей. Статистические методы хороши для обработки информации или в сочетании с другими методами. В качестве типизатора статистический метод может выступать в случае небольшого числа используемых критериев оценки, сбалансированной базы данных и априорного представления об анализируемых данных. Т.е. когда информация заложена уже в самих исходных данных, а не должна быть получена путем каких-либо сложных преобразований обработанных данных. Частично методы статистики используются для решения проблемы определения устойчивости геологической среды в части обработки исходных данных, а также в качестве базы некоторых алгоритмов машинного обучения. Однако опять же сами по себе они не могут предоставить все условия для решения поставленной задачи и даже в сочетании с экспертными методами не дают возможности широкого применения метода.

Метод экстраполяции предполагает перенос текущего характера развития процесса в будущее без особых изменений. Метод может быть полезен для определения будущего состояния, но не решает проблему определения текущего состояния объекта исследования. Может быть использован только после завершения типизации для произведения временных прогнозов различного временного характера, но не определять устойчивость геологической среды.

В общем, моделирование заключается в построение и изучение моделей реально существующих объектов, процессов или явлений с целью получения объяснений этих явлений, а также для предсказания явлений, интересующих исследователя. В принципе машинное обучение можно отнести к части методов моделирования, и правильно построенная модель способна отобразить основные тенденции в развитии геологической среды. Встает вопрос каким способом модель будет простроена и с использованием каких критериев. В том случае, когда механизм не достаточно изучен и формализован для заложения в модель, машинное обучение может быть прекрасной основой для такого комплексного и сложного моделирования, особенно в сочетании со всеми перечисленными методами.

2.2. Принцип определения устойчивости геологической среды при помощи машинного обучения

Так как был сделан вывод, что для решения поставленной задачи быстро и качественно подходит машинное обучение, необходимо описать принцип его работы.

Машинное обучение в своей основе представляет собой анализ влияния группы факторов на некий признак. В нашем случае это анализ влияния взаимосвязанных между собой природных и техногенных факторов на устойчивость геологической среды. В итоге для того, чтобы выбор критериев для ее оценки был наиболее эффективен и применим к использованию в машинном обучении, необходимо понять, какие подходы применялись к многофакторному анализу в инженерной геологии ранее. Но для начала нужно сформулировать саму суть машинного обучения и способ его реализации. Лучше всего эту суть описывает идея интеллектуальной реконструкции, рассмотренная Кисселем М.А.

«Понять концепции прошлого возможно, если воссоздать ход мышления, который привел ученого к определенному утверждению. Интеллектуальная реконструкция начинается с понимания проблемы или целевой задачи, которая породила объективную часть концепции. Далее устанавливается:

— центральное понятие;

— ход реализации принципа;

— соблюдается правило критерия адекватности — удовлетворение требованиям «внутренняя связность» и «внешнее оправдание»;

— анализируется группа эмпирических фактов и т. д.» [Киссель М.А. 1976]

Также был рассмотрен метод функционально-энергетического анализа (ФЭА) в системе инновационных инструментов, позволяющих исследовать взаимоотношения в природе. Это определенный комплекс методов, которые

обеспечивают понимание направления и оценку характеристик природных процессов, объектов и явлений, определяющих взаимодействие природы с человеком для приобретения ими устойчивости от вешних воздействий

По мнению автора для ФЭА обычно свойственны следующие черты:

— «объект, процесс или явление рассматривается как комплекс функций и оценивается по степени его влияния или собственной реакции от других влияний в процессе взаимодействия с окружающей средой в единицах выделения, поглощения или преобразования энергии;

— функции, которые отражают поведение или состояние объекта, процесса или явления, оцениваются с точки зрения пространственно-временных и энергетических характеристик реализации функции. Путем сравнения этих показателей оценивается значение и эффективность функциональных свойств в модели. Так оценивается состояние объектов, процессов и явлений, прогнозируется дальнейшее развитие системы и определяется степень допустимого активного воздействия на человека или окружающую природную среду;

— критерием эффективности функций является оценка их значимости в обеспечении самоорганизации окружающей среды и ее устойчивости к антропогенным воздействиям» [Киссель М.А. 1976]. В нашем случае будет проводиться скорее обратный анализ. А именно анализ влияния состояния геологической среды на наличие геологического;

— «проведение ФЭА требует определенных методов, а также технологий, организованных в программы или планы;

— комплексность решения задач с помощью ФЭА требует системного подхода и применяется там, где решения задач могут быть альтернативными» [Киссель М.А. 1976].

Эти понятия закладывают базу в подходе к решению задачи анализа устойчивости геологической среды по ряду совокупных характеристик.

В.М. Шестопалов в своем исследовании [Шестопалов В.М. 1997] показал пригодность геологических структур, в первую очередь, в

кристаллических породах для надежной изоляции радиоактивных отходов. Как правило, для получения такой предварительной оценки территории широко используются результаты структурно-геологических и петрологических работ. Но в этом случае из-за очень слабой степени изученности, этот подход не обеспечивал надежного результата. По этой причине было решено применить сложную методологию, которая включает в себя разработку двух моделей - структурно-геодинамической и гидрогеологической. Области их исследования охватывали восточную часть Плутона Коростень и его склонов, достигнув Чернобыльской АЭС.

В практике существуют только три прямых метода реконструкции деформации поля, а именно полевые исследования на основе крупномасштабного исследования пласта, петротектонические методы и методы оптической поляризации. К сожалению, все эти методы являются чрезвычайно трудоемкими и обеспечиваются только регулярными полевыми испытаниями, которые трудно проводить в области обнажения горных пород.

Предложенная методика основана на анализе вертикальной и горизонтальной анизотропии гравитации и геомагнитного поля, а также квази 3D пространственно-временной топографической поверхности.

Задачами геосинематического анализа являлись:

1. анализ на основе геодинамических карт, для восстановления схем направления силы и распределения деформаций в процессе формирования основных структур и структурных парагенезисов исследуемой территории;

2. составление схемы на основе определенного напряжения, для прогнозирования реального местоположения ожидаемого тектонического нарушения, зон разрушения и дезинтеграции пород и массово устойчивых блоков.

Были проанализированы более 30 преобразований каждого геофизического поля, в том числе их первые и вторые производные, карты гистограмм и коэффициенты корреляции между полями. Основываясь на этих данных, создавались геодинамические карты, отражающие динамику

развития территории на основании палеореконструкции последовательных этапов активности образования современной литосферы.

Гидрогеологическая модель строилась, начиная с верхней гидродинамической зоны, для которой данные гидравлических параметров были доступны. После взаимной калибровки верхних элементов модели, была разработана более глубокая часть модели с использованием данных о проницаемости структуры кристаллического массива горных пород, полученных из геодинамической модели. Результаты анализа и несоответствие режима гидродинамического моделирования были использованы для уточнения репрезентативности модели по структуре проницаемости пород. Этот повторяющийся процесс последовательной корреляции и уточнение модели может повторяться много раз. В результате реализации этого метода были найдены области (массивы) активного и наоборот очень медленного водообмена, и в системе были выявлены вертикально чередующиеся зоны повышенной фильтрации и слабой проницаемости. На основании этих данных предварительно были выбраны участки, которые оказались перспективными для дальнейшего более детального исследования и последующего захоронения радиоактивных отходов в геологических структурах.

В результате были выбраны два региона, характеризующиеся существованием геодинамических процессов охлаждения, термической усадки и структурной упаковки вещества геологической среды. Анализ неотектонических процессов также показал геодинамическую устойчивость выбранных структур. Эти участки представляются наиболее перспективным для глубокого захоронения радиоактивных отходов [Шестопалов В.М. 1997].

При анализе устойчивости геологической среды при помощи машинного обучения важно выбрать наиболее значимые ее характеристики и сгруппировать их в единый вектор значений. Совокупность этих векторов будет образовывать некое поле значений, заключающее в себе информацию о всех значимых параметрах геологической среды одновременно.

В своей работе Бондарик Г.К. [Бондарик Г.К. 2009] сформулировал теорию геологического поля. В его представлении геологическое взаимодействие представляет собой систему, состоящую из гравитационного, электромагнитного, термодинамического и других взаимодействий, которые являются элементами физических полей. Сумма всех этих полей, которые в свою очередь иерархически связаны друг с другом, представляет собой случайное поле, которое следует считать полем геологического процесса или геологическим полем. Также им были рассмотрены виды энергий, порождающие природные процессы, взаимосвязь этих видов энергий и соответствие уровней организации материи различным взаимодействиям (физическому, химическому, геологическому, биологическому). Это позволяет понять структуру появления природного процесса и энергию, влияющую на его развитие. Соответственно можно сделать вывод, что при анализе геологического поля либо его наиболее иерархически высокой части можно определить направление развития всей геологической среды за счет иерархической взаимосвязи различных ее характеристик.

Комплексная количественная оценка информации при региональных инженерно-геологических изысканиях была рассмотрена Пендиным В.В. [Пендин В.В. 2009]. Им был сделаны следующие выводы:

1. При различных целях проведения исследований следует использовать различные методы комплексной оценки инженерно-геологических условий.

2. При районировании территории с целью выбора оптимальных участков для определенного вида деятельности следует использовать интегральный показатель инженерно-геологических условий.

3. Для разбиения территории на однородные участки в отношении состава и объема дальнейших исследований следует применять меры теории информации при проведении количественной оценки.

4. Для выявления закономерностей пространственной изменчивости территории целесообразно использовать факторный анализ исходных данных.

5. На заключительной стадии целесообразно использовать дискриминантный анализ для подтверждения достоверности моделирования.

Эти выводы и описание каждого из методов позволяет обобщить информацию о возможностях проведения анализа устойчивости геологической среды.

Также интересна работа Азроянца Э.А. [Арзоянц Э.А. 1995]. Он с соавторами провел исследования физических явлений, происходящих в земной коре и на ее поверхности, для оценки возможности наступления катастрофических явлений. Предложенный способ направлен на расширение области прогнозирования появления опасных катастрофических явлений с точки зрения возможности его использования в любой местности земного шара независимо от геологических, географических и климатических условий, так и с точки зрения возможности прогнозирования не только землетрясений, но и других катастрофических явлений, включая техногенные, расширение арсенала технических средств для осуществления прогноза.

Суть метода заключается в том, что проводится непрерывный контроль изменяющегося во времени параметра геофизического поля, определяются значения периода и частоты его колебаний, измеряют амплитуду контролируемого параметра, а вывод о возможности наступления катастрофических явлений делают в случае обнаружения в изменении параметра геофизического поля синусоидального колебательного процесса с периодом от 100 до 1000000, имеющего амплитуду колебаний, статистически достоверно отличающуюся от фонового значения для данной местности. В качестве контролируемого параметра могут быть использованы значения температуры, давления атмосферного воздуха, деформаций земной поверхности, напряженности естественного электромагнитного поля, уровня

естественной радиоактивности, температуры приповерхностных слоев литосферы и гидросферы, силы тяжести, микросейсмической активности, содержание гелия в подземных флюидах, интенсивность излучения электромагнитного поля в диапазоне волн от 7 до 12 мм. Совместный учет величин амплитуды и периодов колебаний контролируемого параметра позволяет увеличить достоверность прогноза и исключить из него влияние различного рода помех.

В свою очередь Бондаренко Я.И. создал модель генетического рисков и риска здоровью населения Днепропетровска на основе анализа геофизической информации. Основными целями были: многофакторное моделирование и корреляция природных и антропогенных изменений в окружающей среде и, соответственно, в здоровья человека; определение показателей, которые определяют риск возникновения разрушения структур на разных уровнях организации и функционирования экосистем города (геофизических и геохимических полей, почвы, гидросферы, атмосферы, биосферы); анализ закономерностей взаимодействия природных, антропогенных и биологических ритмов. На первом этапе формирования многослойной геоинформационной модели была выполнена математическая обработка космических снимков, литогеохимических и геофизических карт с использованием компьютерных систем JEIS и ERDAS. В результате была обнаружена корреляция гравитационных и магнитных полей с частотой заболеваемости онкологическими болезнями у взрослых. Результаты многослойных пространственно-временные корреляции геофизических параметров поля и вариации ритмов заболеваемости населения позволили им создать базу и разработать медико-биологическую и био-геодинамическую классификацию геопатогенных зон. Био-геодинамическая модель служила основой для определения контуров антропогенно-ландшафтной и ландшафтно-геофизической аномалий с точки зрения опасности для здоровья людей в Днепропетровске [Бондаренко Я.И. 2001].

Здесь стоит рассмотреть исследования связанные с обнаружением и влиянием геопатогенных зон и геофизических аномалий на здоровье человека.

Согласно трудам Балтунова В.А. и [Болтунов В.А. 1999] и Гриндина В.И. [Гриндин В.И. 1994]. Физические поля Земли тесно связаны с полями околоземного пространства и образуют общую геофизическую систему, которая управляет всеми процессами, происходящими в земной коре. Эта совокупность геофизических полей активно влияет и может контролировать ландшафтные и экологические процессы, включая их изменение и активизацию. Котловым В.Ф. [Котлов В.Ф. 1997] даже была найдена зависимость между участками с повышенной заболеваемостью раком и локальными тектоническими разломами. Мельников Е.К. [Мельников Е.К. 1994] обнаружил такую взаимосвязь в местах проявления региональных разломов, разрывных активных структур и погребенных палеорусел.

Лобацкой Р.М. было показано, что различные геофизические и геохимические аномалии связаны с эндогенными и экзогенными процессами, которые происходят в разломных зонах. Разломные зоны, прежде всего, определяются гравитационными аномалиями, а измененная напряженность и деформация горных пород как правило связаны с аномалиями естественных магнитных, электрических и импульсных электромагнитных полей в радиочастотном диапазоне. Геопатогенные зоны в таком случае являются наиболее активными при развитии микротрещин в горных породах [Лобацкая Р.М. 1997].

В ходе исследования многих ученых было установлено, что геопатогенные зоны являются особым видом геофизических аномалий. Эти аномалии обычно возникают вследствие наличия трех факторов: активных геологических разломов, пересечения подземных водных потоков и жил, наличия узлов глобальных «энергетических» сеток и их совместное действие [Кречмар Б. 1994, Мельников Е.К. 1997, Осипов В.И. 1994].

Бондаренко Я.И. и Зайонц И.О. разработали метод моделирования и прогнозирования риска заболеваемости населения в зависимости от геофизических условий региона. Данные исследования проводились в Киеве [Бондаренко Я.И. и др. 2001] и Днепропетровске [Бондаренко Я.И. 2000]. Анализ основан на комплексной переработке данных магнитного поля, гравитационного поля и рельефа. Авторы предположили, что эти данные содержат информацию риске для проживания населения. Они сделали некоторый анализ в городах под Днепропетровском и выяснили, что количество опасных зон, определенных их методом, соответствует количеству больных людей в этих городах.

Определить эти зоны можно было в результате специальной математической обработки данных и экспертной оценке результатов. Обработка основывается на расчете вектора напряженности и геодинамических преобразованиях.

Расчет вектора напряжения основан на преобразовании данных геофизических полей в векторное поле. После создания векторного поля создается множество различных карт. Эти карты в дальнейшем использовались для экспертизы. Для создания карты авторы использовали различные методы обработки данных. Использовались: анализ главных компонент, корреляционный анализ, сравнение градиентов геофизических полей и др. Также в расчет вектора напряжения включено создание сглаженной поверхности геофизических полей.

Расчет геодинамических преобразований основан на расчете основных параметров векторного поля. Они включают в себя дивергенцию и ротор. Эти параметры определяют динамику природных полей.

Метод включает в себя следующие стадии:

1) Сбор данных геофизических полей и рельефа земли;

2) Расчет вектора напряжения;

3) Расчет дивергенции и ротора.

4) Экспертный анализ результатов и создания структурной модели территории.

Для создания структурной модели авторы собрали и объединили все карты и проделанные ранее расчеты, и построил новую карту. Эта карта разделила площадь на 12 классов в соответствии уровню воздействия на человека. В итоге информация, полученная из построенной карты, соответствовала реальной ситуации на исследуемой территории [Бондаренко Я.И. и др. 2001].

Теперь, разобравшись в существующих подходах к многофакторному анализу можно перейти к принципу моделирования при помощи машинного обучения.

Описание принципа моделирования при помощи методов машинного обучения приведено на рисунок 2.2.1

Рисунок 2.2.1. Моделирование при помощи методов машинного обучения

Не каждый набор данных может быть обработан автоматически, либо использован без предварительной обработки. Из-за этого, в первую очередь необходимо подготовить данные для обработки. Различные типы программного обеспечения требуют различной предварительной обработки. Это означает, что иногда нужно использовать несколько различных программ для одной обработки данных. В качестве программного обеспечения анализа данных в данной работе использовалась WEKA

[Weka 3]. Это программное обеспечение представляет собой набор алгоритмов машинного обучения для решения задач интеллектуального анализа данных. Алгоритмы могут быть применены либо непосредственно к набору данных, либо создаваться посредством собственного кода Java. WEKA содержит инструменты для предварительной обработки данных, типизации, регрессии, кластеризации, ассоциативных правил, и визуализации. Этот программный продукт также хорошо подходит для разработки новых схем машинного обучения. Этот программный продукт использует текстовый формат документов и может анализировать как слова, так и числа. Однако иногда необходимо сделать несколько предварительных расчетов. Эту задачу решают различные программы статистических расчетов. При предварительной обработке данных для их подготовки к использованию в машинном обучении используются различные текстовые редакторы для приведения данных в нужный формат перед началом расчетов. Основной задачей при этом является превращение некоторых данных в векторы значений критериев и классов, которые было бы возможно определить по этим параметрам. Но это только первый шаг обработки. Затем необходимо проверить данные на предмет отсутствующих значений, шумов и т.д., и решить эти проблемы, если они есть. Это можно сделать с WEKA. Она имеет много различных фильтров данных, которые могут преобразовать данные в правильном направлении.

После этого, при наличии слишком большого числа критериев, выбираются наиболее подходящие. Это делает расчеты более быстрыми и точными. По завершении всех описанных процедур данные являются подготовленными для обработки в WEKA.

После получения вектора значений критериев начинается самая важная часть работы - выбор наиболее значимых алгоритмов для обработки выборки. WEKA содержит около 100 алгоритмов. В данной работе исследовались следующие группы алгоритмов WEKA, выбранные по точности проводимой ими типизации (при предварительной оценке точности

алгоритмов использовался метод перекрестной проверки): *Decision trees, Bayes, Functions, Lazy, Meta-algorithms, Miscellaneous* и *Rules.* Из них, в начале, нужно выбрать несколько наиболее значимых. Кроме того, каждый алгоритм имеет несколько параметров. Изменение этих параметров может увеличить точность на 5-10%. Детальный поиск параметров проводиться только для алгоритмов, показавших наилучшие результаты на первом этапе обучения.

Перекрестная проверка – это метод оценки аналитической модели и её поведения на независимых данных. При оценке модели, имеющиеся в наличии данные разбиваются на k частей. Затем на k−1 частях данных производится обучение модели, а оставшаяся часть данных используется для тестирования. Процедура повторяется k раз; в итоге каждая из k частей данных используется для тестирования. В результате получается оценка эффективности выбранной модели с наиболее равномерным использованием имеющихся данных. В работе использовалось разбиение на 10 частей при перекрестной проверке.

Она помогает предварительно выбрать группу алгоритмов, подходящих для решения поставленной задачи, но из-за опасности переобучения не может служить окончательным критерием точности модели.

Все эти шаги должны быть сделаны для данных, характеризующих предварительно интерпретированных экспертом. Чем больше будет интерпретированных примеров, тем лучше будет результат, показанный компьютером, при обработке неопределенных данных. После обработки всех данных обучающей выборки и нахождения лучшего алгоритма, настаёт время для создания прогноза по данным с неопределенным классом. Но этот набор данных должен содержать точно такие же критериии и иметь точно такой же тип данных. Только в этом случае можно будет использовать WEKA для создания прогноза. Суть заключается в том, что новые наборы данных могут быть обработаны быстро и с хорошей точностью только, если

содержат информацию, заключённую в критериях оценки, и связаны с той же проблемой, что и ранее обработанные данные.

Кроме машинного обучения методами WEKA необходимо использовать некоторые статистические методы и другие программы машинного обучения для сравнения. В качестве такого программного обеспечения использовался See5. В качестве статистического программного обеспечения использовались Excel, Minitab и Mathlab.

Полученные результаты далее анализируется, интерпретируются и из них извлекаются полезные для решения поставленной задачи знания.

Представленные результаты необходимо визуализировать для более точного их понимания и интерпретации. Для перевода числовой информации в графическую использовался программный продукт Surfer11. Он предоставляет хорошие возможности для построения карт изолиний и карт типизации.

2.3. Алгоритмы машинного обучения

В данном разделе представлено описание алгоритмов машинного обучения, использованных для анализа данных. Эффективность алгоритмов оценивалось в рамках экспериментов. В данном разделе приводится описание принципа действия и основных характеристик используемых алгоритмов машинного обучения.

В качестве метода проверки алгоритмов использовалась 10-кратная перекрестная проверка. Перекрестная проверка – это методика оценки того, насколько результаты статистического анализа действительны в независимых наборах данных. Она в основном используется тогда, когда целью является прогноз, и появляется необходимость оценить, насколько точно прогнозная модель будет действовать на практике. Один этап перекрестной проверки включает в себя разбиение выборки данных на дополнительные подмножества, анализ данного подмножества (обучение), а также проверку анализа на других подмножествах (тестирование). Для уменьшения изменчивости, несколько этапов перекрестной проверки выполняются с использованием различных частей выборки, а также результаты проверки усредняются по этапам. 10-кратная перекрестная проверка используется довольно часто. При 10-кратной перекрестной проверке, оригинальный набор данных случайно разбивается на 10 частей. Процесс перекрестной проверки повторяется 10 раз, и каждая из частей используется только один раз в качестве тестовой выборки. Для проведения оценки результаты всех этапов усредняются (или комбинируются иным способом). Преимущество этого метода перед проверкой при помощи тестовой выборкой состоит в том, что все данные используются и для обучения, и для проверки, при этом каждое наблюдение используется для проверки только один раз.

Лучший алгоритм был протестирован на отдельном наборе данных. Таким образом, была оценена его работа при реальном моделировании.

Деревья решений

Деревья решений - метод, применяемый при многоэтапном процессе анализа данных и принятии решений о типизации. Ветви дерева изображают события, которые могут иметь место, а узлы и вершины - момент выбора направления действий. Принятие решений осуществляется на основе логической конструкции «если… то…», путем ответа на вопрос вида «является ли значение переменной меньше значения порога?». При положительном ответе осуществляется переход к правому узлу дерева, при отрицательном – к левому узлу. После этого осуществляется принятие решения уже для выбранного узла.

Обучение деревьев решений представляет собой метод, широко используемый в интеллектуальном анализе данных. Суть которого состоит создании модели, предсказывающей значение искомого параметра на базе некоторых входных переменных. Каждый узел дерева решений соответствует одному такому входному параметру, в то время как листья этого дерева соответствуют значениям искомого параметра. Значения входных параметров определяют путь от корня к листу.

Деревья решений, которые используются при интеллектуальном анализе данных, подразделяются на два типа:

- Деревья анализа типизации, когда происходит разделение группы значений на различные типы в зависимости от их характеристик;

- Деревья регрессионного анализа, когда искомый результат представляет собой вещественное число [Breiman Leo 2001].

В целом можно указать общие достоинства и недостатки метода деревьев решений при интеллектуальном анализе.

Достоинства:

1. Простота понимания и интерпретации.

2. Не требуется сложная подготовка данных. Другие алгоритмы требуют добавления переменных, нормализации данных, устранения пустых значений.

3. Можно использовать модель «белого ящика». В том случае, когда определенная ситуация встречается в модели, её возможно интерпретировать с помощью бинарной логики.

4. Способность работать и с интервальными, и с категориальными переменными. Другие алгоритмы работают только с такими данными, в которых присутствует лишь один из указанных типов переменных.

5.Есть возможность оценить модель при помощи статистических тестов, что позволяет оценить надежность модели.

6.Надежный метод. Метод хорошо себя показывает даже в случае, когда исходные характеристики модели были нарушены.

7. Позволяет обрабатывать большие объемы данных без особой подготовки.

Недостатки:

1. Проблема получения оптимального дерева решений решается не полностью даже в случае решения простых задач. В деревьях решений оптимальное решение может быть принято только в каждом конкретном узле, что не позволяет обеспечить оптимальную работу всего дерева.

2. Возможно создание слишком сложного дерева, которое при этом будет плохо соответствовать решаемой задаче. Такая проблема получила название проблемы «чрезмерной подгонки» [Bramer M. 2007]. Для ее устранения необходимо уменьшать размер дерева решений, удаляя его участки с наименьшим весом.

3. Дерево может быть слишком большим, из-за чего трудно интерпретируемым [Horvath. 2003].

4. При использовании данных, включающих многоуровневые категориальные переменные, наибольший информационный вес дается тем критериям, которые являются наиболее сложными, что не всегда верно [De’ath G. 2000].

В исследовании использовались следующие алгоритмы:

1. BFTree
2. DesStump
3. FT
4. J48
5. LADTree
6. LMT
7. NB Tree
8. RandomForest
9. RandomTree
10. REPTree
11. SimpleCart
12. See5

Алгоритмы на основе теоремы Байеса

Теорема Байеса — одна из главных теорем теории вероятностей. Она дает возможность определить вероятность какого-либо события при условии, что произошло другое статистически взаимосвязанное с ним [Aha D. 1991].

Теорема Байеса имеет вид:

$$P(A|B) = \frac{P(B|A)\,P(A)}{P(B)},$$

где

$P(A)$ — априорная вероятность гипотезы *A*;

$P(A|B)$ — вероятность гипотезы *A* при условии наступления события *B*;

$P(B|A)$ — вероятность наступления события *B*, если гипотеза A истинна;

$P(B)$ — полная вероятность наступления события *B*.

Алгоритм байеса оптимален, то есть обладает минимальной вероятностью ошибок, если известна плотность распределения классов. На практике плотности распределения классов, как правило, не известны. Их приходится оценивать (восстанавливать) по обучающей выборке. В результате байесовский алгоритм перестаёт быть оптимальным, так как восстановить плотность по выборке можно только с некоторой погрешностью. Чем короче выборка, тем выше шансы «подогнать» распределение под конкретные данные и столкнуться с эффектом переобучения.

Байесовский подход к типизации является одним из старейших, но до сих пор сохраняет прочные позиции в теории распознавания. Он лежит в основе многих удачных алгоритмических моделей

В исследовании использовались следующие алгоритмы:

1. BayesNet

2. NaiveBayesSimple
3. NaiveBayes

Функции

Это алгоритмы, которые строят простые и расширенные линейные модели для типизации. Эти методы лучше всего работают с числовыми характеристиками.

В исследовании использовались следующие алгоритмы:

1. Логистическая регрессия
2. Многослойный перцептрон
3. Перцептрон
4. Сеть RBF
5. SimpleLogistic
6. SMO

«Ленивые» алгоритмы

«Ленивое» обучение - это метод обучения, в котором генерализация обучающей выборки откладывается до внесения в систему запроса, в отличие от интенсивного обучения, когда система пытается обобщить данные обучение до получения запроса.

Основным преимуществом, полученным от применения метода ленивого обучения, таким, как определение на основе факта, является то, что целевая функция будет аппроксимирована локально, например, как в алгоритме к-ближайших соседей. Из-за того, что целевая функция аппроксимируется локально для каждого запроса к системе, ленивые системы обучения могут одновременно решать несколько задач и успешно справляться с изменениями в предметной области.

Недостатки ленивого обучения включают в себя потребности в большом объеме памяти для хранения всего набора обучающих данных. В особенности, шумы в обучающей выборке понижают необходимость базового этапа, потому что в ходе обучающего этапа не происходит обобщения. Еще одним недостатком является то, что ленивые методы

обучения, как правило, медленнее работают, хотя это связано с более быстрым этапом обучения.

В исследовании использовался метод k–ближайших соседей.

Метаалгоритмы

Метаалгоритмы представляют собой вычислительные методы, которые оптимизируют проблемы путём многократных попыток улучшить предлагаемое решение в связи с данным показателем качества. Метаалгоритмы несколько раз или даже ни разу не пропускают через себя проблему, которую необходимо оптимизировать и могут при этом найти большое количество возможных решений. Тем не менее, метаалгоритмы не гарантируют, что оптимальное решение будет найдено. Многие метаалгоритмы реализуют ту или иную форму стохастической оптимизации.

В исследовании использовались следующие алгоритмы:

1. AdaBoostM1
2. Классификатор выбора параметров
3. Bagging
4. Классификация через кластеризацию
5. Классификация через регрессию
6. Выбор параметров перекрестной проверкой
7. Dagging
8. Decorate
9. Grading
10. LogitBoost
11. MultiBoostAB
12. Мультиклассовый классификатор
13. Мульти схема
14. Class Balanced ND; Data Near Balanced ND; ND
15. OrdinalClassClassifier
16. Raced Incremental Logit Boost
17. RandomCommittee

18. RandomSubSpace
19. Rotation Forest
20. Stacking
21. StackingC
22. Vote

Правила

В данном разделе описаны алгоритмы, генерирующие правила для проведения типизации.

В исследовании использовались следующие алгоритмы:

1. Соединительное правило
2. DecisionTable
3. DTNB
4. JRip
5. OneR
6. ZeroR

Другие алгоритмы машинного обучения

В данном разделе представлены алгоритмы, которые трудно отнести к какой-либо группе, поскольку они сочетают в себе качества нескольких из них или не имеют много общего с другими алгоритмами.

В исследовании использовались следующие алгоритмы:

1. HyperPipes
2. VFI

Выше были описаны алгоритмы, которые использовались в процессе машинного обучения при определении устойчивости геологической среды при помощи анализа инженерно-геологических данных. Сравнение этих алгоритмов по степени их эффективности показано в последующих главах. На основании базового описания можно сказать, что лучше всего себя проявляют деревья решений, функции и мета алгоритмы, так как изначально настроены на решения задач с несбалансированными данными и неочевидными решениями. «Ленивое» обучение и правила не могут показать адекватные результаты из-за своих особенностей. Алгоритмы, основанные на теореме Байеса, проявляют себя не плохо, но хуже чем функции и деревья решений. Более точно в этом можно удостовериться только исходя из результатов экспериментов.

3. Опыт решения инженерно-геологических задач методами машинного обучения

Существует множество различных методов моделирования. Все они используются в различных ситуациях и имеют право на существование. Метод моделирования при помощи машинного обучения полезен только в тех ситуациях, когда мы имеем дело с очень большим объемом данных, проблемами с базами данных, трудностями в определении класса и при наличии некоторых архивных данных с решенной проблемой типизации. В подобной ситуации машинное обучение является отличным помощником в работе, делает работу быстрее и имеет большие экономические преимущества.

В 1990-х годах, методы машинного обучения начали проникать в природопользование. Сегодня это уже не является экзотикой, поскольку их присутствие можно наблюдать повсеместно в науке о земле. Они активно используются в обработке спутниковых данных, в модели общей циркуляции (МОЦ) для физических моделей, в пост-обработке выходов МОЦ, в прогнозировании погоды и климата, моделировании инженерно-геологических условий, анализе и моделировании геологических данных, гидрогеологическом прогнозированиях, инженерно-геологическом мониторинге состояния и т.д.

Эта работа представляет новый метод инженерно-геологического моделирования, и в этой главе показано, как подобные проблемы были решены с помощью машинного обучения.

Кобринским Б.А. были разработаны механизмы отражения образного мышления и интуиции специалиста слабоструктурированной предметной области. Его работа интересна своим подходом и постановкой задачи, поскольку инженерная геология также является слабоструктурированной предметной областью. Автор в качестве примера предметной области выбрал

медицину, но это не умаляет её ценность для данного исследования, поскольку основной подход к вопросу информационного отображения образного мышления остается неизменным. При обработке информации, которая крайне похожа на логику специалиста, неизбежно присутствуют черты, отражающие личный профессиональный опыт, убеждения, знание, которое трудно формализовтаь, уверенность и другие особенности мышления [Хай Г.А. 1988]. «Специалист, способный на интуитивное решение, как правило, в довольно высокой степени уверен в своем предположении, базирующемся на его знаниях. То же касается и ситуации с возникновением у него образа-решения. Образ - это то, что обычно более или менее четко фокусируется мысленным зрением, но именно его “ядерная” составляющая, тогда как периферия образа выглядит расплывчатой или в форме неясных теней, которые могут являться как составной частью образа, так и быть примыкающими к нему, т.е. сопутствующими (в том числе случайными) проявлениями. Исходя из этого, условие выдвижения гипотезы об “образе” в базе знаний интеллектуальной системы должно сопровождаться указанием о степени уверенности эксперта в выдвигаемой им гипотезе:

- “абсолютно достоверно” (полная определенность),
- “скорее всего” (крайне высокая степень определенности),
- “весьма вероятно” (высокая степень определенности),
- “противоречивые или спорные сведения”, что относится, как правило, к комиссионному решению при проведении консилиума (практически полная неопределенность),
- “маловероятно” (высокая степень определенности),
- “сомнительно, но не исключено” или “крайне маловероятно”, т.е. почти полное отрицание (крайне высокая степень определенности),
- “абсолютно невозможно”, т.е. полное отрицание (полная определенность)» [Корбинский Б.А. 1996].

Это позволяет сформировать классы для проводимой оценки и типизации информации для поиска решений. А также предложенная автором лингвистическая шкала, в отличие от порядковых линейных шкал, задает размытый порядок объектов вследствие пересечения интервалов числовой шкалы.

Кобринским было показано, что образные представления в базе знаний можно представить в виде:

-«комплексно описанных ситуаций-аналогов, включая метафорически представленные (диагностически значимые) интегральные проявления процесса;

- специфических проявлений процесса в форме ассоциирующих признаков;

- визуализированных проявлений процесса (рисунки, фотографии и др.)» [Корбинский Б.А. 1996].

В отношении визуализации результатов было показано, что имеет смысл использование разнообразных элементов визуализации. Это приведет к более эффективному восприятию результатов работы экспертной системы специалистом. Это особенно обосновано для образных представлений, так как из всех возможных способов передачи знаний графический способ является наиболее древним. Можно полагать, что использование образов окажется условием, которое трудно описать словами. «В то же время, визуальное представление выходных параметров требует “поднастройки” экспертной системы на конкретного пользователя, учитывая его индивидуально-опытные представления, характерологические особенности личности и способность к формированию образных представлений или аналитико-синтетической деятельности мозга в зависимости от преобладания деятельности правого или левого полушария мозга» [Корбинский Б.А. 1996].

Подобные подходы закладывают основы для обработки информации инструментами искусственного интеллекта и её предоставления, принимая во внимание её изначальную неструктурность.

Одним из основных инструментов искусственного интеллекта являются искусственные нейронные сети. Они получили широкое применение для решения различных экологических задач.

Группа ученых во главе с Потылицыным Е.Н. применили нейронные сети для проверки приемлемости нейросетевых технологий для оценки риска негативного воздействия на природную среду. Была построена модель влияния вредных выбросов на здоровье населения Красноярска. «В качестве индикаторов состояния окружающей среды использовались концентрации загрязняющих веществ в воздухе по данным центра по мониторингу загрязнения окружающей среды ГУ «Красноярский ЦГСМ-Р», осуществляющего наблюдения за качеством атмосферного воздуха на 8 стационарных постах в г. Красноярске» [Потылицын Е.Н. 2013].

Для индикации здоровья населения были использованы данные санитарно-демографической статистики.

Таким образом, строилась модель состояния окружающей среды, точность которой оценивалась через статистику заболеваний. Подобный подход является очень близким и исследованию, представленному в данной работе. Относительная ошибка модели составила 0,40%, в то время как средняя абсолютная ошибка составляла 0,93%, т.е. сеть обеспечивала соответствие фактических и расчетных значений.

Однако столь хороший результат был получен вследствие небольшого размера исходной базы данных (тренировочной выборки), и, главное, сложности решаемой проблемы типизации, которую можно было решить при помощи статистических методов оценки.

Машинное обучение активно используется в геологии. Так, например, Гай Дешарне, Жан-Филипп Пимон и Дуг Хэтфилд [Buettner Mason 2015] применили их для поиска новых золоторудных месторождений на основе 26 Гб исходных геологоразведочных данных. Авторы обработали исторические данные, составили блоковую модель до двух километров на их основе, а затем составили схему перспективности обнаружения полезных ископаемых.

В полученной модели каждому блоку был присвоен определенный балл, который и указывал на вероятность (перспективность). После этого все полученные данные обрабатывались методом машинного обучения, с применением алгоритма Байса. При этом результат моделирования был представлен в форме виртуальной реальности на очках «Oculus Rift». Это дало возможность спланировать систему буровых скважин, а также их вычислить угол и ориентацию, наиболее подходящие для изучения объекта.

Такую оценку перспективности можно при этом легко обновлять данными, которые будут поступать после начала буровых работ.

Строкова Л.А. применила машинное обучение для прогноза оседания земной поверхности при строительстве городских туннелей а также при строительстве метрополитена [Строкова Л.А. 2009]. Использовались нейронные сети и деревья решений при обучении без учителя. Исходными данными послужили материалы мониторинга за оседанием поверхности при строительстве метрополитена в г. Мюнхен. В качестве параметров моделирования автор выбрала геометрические параметры, характеризующие расположение туннеля в массиве (глубина залегания массива, отношение мощности перекрывающих пород над туннелем к диаметру проходки, отношение расстояния между осями двойных туннелей к диаметру, величина осадки, ширина корыта оседания) и геологические параметры (состав, возраст, условия залегания горных пород от дневной поверхности до подошвы туннеля). В результате было выявлено, что нейронные сети способны достоверное определить результаты просадки в 23 случаях из 24 при имеющейся выборки. Хотя, стоит отметить, что данный пример является лишь иллюстрацией процесса разработки базы знаний по планированию и проектированию сооружений в условиях плотной городской застройки. Для окончательного воплощения требуются более детальные исследования на более большой выборке.

Белозеров Б.В. и соавторы [Белозеров Б.В. 2014] использовал машинное обучение для решения задача типизации данных с месторождения

по типам фациальных обстановок каротажных при использовании спектрального представления геофизических полей. Эта методика с успехом применялась для интерпретации фаций в реально существующем месторождении. При решении поставленной задачи необходимо было разделить множество каротажных данных с месторождения по типам, используя информацию о различных типах обстановок осадконакопления. В обучающей выборке присутствовали данные о проинтерпретированных специалистом каротажных кривых, где для каждого значения была определена фациальная обстановка он относится, при том что присутсвовал хотя бы один выделенный тип обстановки осадконакопления.

Для подготовки типизации были выбраны признаки объектов. Они должны были определены так, чтобы было возможным эффективно определять объекты из разных типов. В итоге в качестве признаков объектов использовалась совокупность коэффициентов $\{a_n, b_n\}$ при разложении функции каротажа в ряд Фурье. Выбор признаков типизации осуществлялся при помощи определения порядка разложения функции, необходимого для правильного описания каротажных данных, в ряд. В качестве алгоритма использовался алгоритм ближайшего соседа («ленивые» алгоритмы). Исходная выборка содержала данные по 183 скважинам. На базе обучающей выборки выделены фациальные зоны при помощи алгоритма машинного обучения. Точность типизации при этом составила в среднем 65%. Стоит отметить, что использование более мощных алгоритмов возможно дало бы более точный результат.

Серкова В.И. применила нейронные сети для решения задач типизации и прогноза геологической интерпретации геофизических данных [Серкова В.И. 2009]. Результаты исследований показали, что искусственная нейронная сеть (многослойный персептрон) обеспечивает корректное решение нефтепоисковых задач с вероятностью более 98% (против 54% при типизации методом линейно-статистического анализа с помощью дискриминантных функций). Обучающая выборка была сформирована на

основе данных ГИС, включающих в себя набор стандартных каротажей: боковой, акустический, гамма-каротаж, нейтронный гамма-каротаж, кавернометрия.

Казнечеев П.Ф. и др. рассмотрел применение методов искусственного интеллекта для повышения эффективности в нефтегазовой и других сырьевых отраслях [Казначеев П.Ф. 2016]. Автор указывает, что применение нейронных сетей при решении геофизических и геотехнических задач позволяет сократить требуемое количество скважин и проводимых тестов для определения характеристик грунтов и свойств коллекторов, проводя к значительной экономии денег и времени. Были примеры удачного применения нейронных сетей для интерпретации данных сейсмических наблюдений, определения литологичекой структуры, границ литологичеких объектов, анализа водонасыщенности и проницаемости. А при использовании машинного обучения для работы на месторождениях, показатели добычи повышаются на 2-10% главным образом за счёт сокращения внеплановых простоев.

Джероски С. использовал символическое машинное обучение для моделирования состояния окружающей среды и воздействия на геологическую среду. Символические методы машинного обучения генерируют знаковые модели из данных. Это отличается от таких методов, как нейронные сети, которые наиболее часто для решения подобных проблем. Модели, построенные при помощи символического машинного обучения, могут быть проверены, использованы и изменены экспертами и имеют потенциал, чтобы стать частью знаний в соответствующей предметной области. В своей работе он использовал машинное обучение для моделирования пригодности среды обитания, а также для мониторинга и защиты окружающей среды (мониторинг окружающей среды, воздействие на окружающую среду химических веществ, прогнозирования биологического разложения соединений). В итоге были сделаны выводы: 1), что символические методы были успешно применены ко многим проблемам в

области экологии, инженерии и управления, 2), что символические методы машинного обучения весьма перспективным для дальнейшего применения в экологическом моделировании [Dzeroski Sas 2001].

Каневский М. с соавторами представил некоторые современные подходы к пространственному анализу данных состояния геологической среды. Решались задачи проблемно-ориентированного пространственного анализа данных и моделирования: валоризации и репрезентативности данных с помощью анализа данных, пространственного прогноза, вероятностных карт и картирования рисков, разработки и применения условного моделирования стохастических моделей. В качестве исходных данных принимались данные с места аварии Чернобыльской АЭС. Многомерный характер картины обусловлен многочисленными факторами воздействия: параметрами источника, погодными условиями (особенно осадками), сухими и влажными осадками, свойствами рельефа (орографии, травяного покрова, типа почв, землепользования и т.д.). Наиболее значительное влияние на долгосрочные загрязнения было произведено радионуклидом цезия ^{137} Cs. Период полураспада этого изотопа составляет около 30 лет. Выбранный регион охватывал 7428 км^2 с 845 населенными пунктами.

Образцы, отражающие пространственную структуру загрязнения, являются предметом исследовательского анализа пространственных данных для определения пространственной непрерывности. Пространственная непрерывность является функцией пространственных процессов, имеющих общую физику процесса.

Поскольку образцы представляют собой лишь один вариант реализации пространственного процесса, автором было предложено использовать статистические методы.

Декластеризацией ячеек была использована для разбиения данных на обучающий и тестовый наборы для обеспечения эффективного машинного обучения. Регион был разделен на прямоугольные ячейки регулярной сетки и одна или несколько точек были выбраны случайным образом из каждой

ячейки. Тестовый набор данных был получен таким способом, а остальные данные сформировали обучающую выборку. Таким образом, тестовый набор представлял региональные данные. Были использованы модели многослойного перцептрона (МП) со следующими параметрами: два входных нейрона, пространственные координаты (X, Y), один скрытый слой и выходной нейрон, описывающий загрязнения ^{137}Cs. Анализ вариограмм полученных разностей для обученной нейронной сети с различным количеством нейронов на скрытом уровне показал, что оптимальные результаты (в смысле моделирования нелинейных тенденции) получали с использованием МП с пятью нейронами в одном скрытом слое. Дальнейшее увеличение количества скрытых нейронов приводит к извлечению более подробных местных особенностей образца, отражается несколько диапазонов корреляции оценок тренда вариограмма. Затем МП используется для пространственного отображения прогноза распределения ^{137}Cs. Прогнозы были выполнены на прямоугольной регулярной сетке с размером ячейки 1 × 1 км. Обученный многослойный персептрон и регрессионный метод опорных векторов (МОВ) смогли извлечь некоторую информацию из данных, описывающих крупномасштабные пространственные корреляции. Остальная информация, мелкомасштабные пространственно структурированные остатки, была проанализирована и смоделирована с использованием условного геостатистического стохастического моделирования. Коэффициенты корреляции между остатками и выборочными значениями ^{137}Cs были равны 0,77 (для МП) и 0,79 (для МОВ). Были разработаны нестационарные модели НС и МОВ для анализа и картирования пространственно распределенных данных. Нелинейные тенденции в данных об окружающей среде были эффективно определены при помощи трехслойного персептрона. Другие виды моделей искусственного интеллекта (в частности местные аппроксиматоры) были использованы с возможными изменениями предлагаемой структурой. Был сделан вывод, что статистический метод требует незначительных финансовых затрат для

решения подобной проблемы, но применение метода требует глубоких знаний эксперта в геостатистическом моделировании. Кроме того, применение метода для решения более сложных проблем провидят к появлению многомерных величин, тогда как алгоритмы машинного обучения способны справиться с многомерной информацией и могут интегрировать различные типы данных. [Kanevski M. 2004].

Краснопольский В.М. с соавторами представил и сформулировал новый тип численной модели, - гибридную численную модель на основе объединения детерминированного моделирования и компонентов машинного обучения. Она применялась для комплексного гидрогогического и метеорологического моделирования. Полученные результаты показывают, что:

1. концептуально возможна разработка гибридной модели с точной эмуляцией нейронной сетью компонентов модели, которая сохраняет целостность и все подробные характеристики оригинальных экологических числовых моделей (ЭЧМ).

2. точные эмуляции нейронной сети являются надежными и очень быстрыми (до 105 раз быстрее, чем оригинальные параметризации), так что значительное ускорение расчетов гибридной модели может быть достигнуто без ущерба для ее точности;

3. статистические компоненты и компоненты машинного обучения могут быть успешно соединены с компонентами детерминированной модели в гибридную модель так, что их взаимодействие может быть эффективно использовано для экологического моделирования без каких-либо негативных последствий для качества моделирования;

4. эта новая комбинация детерминированного и статистического подходов к обучению приводит к новым возможностям в гибридной модели для моделирования и прогнозирования состояния геологической среды. Например, новые более сложные параметризации, которые невозможны при использовании численной модели в их первоначальном виде, станут

вычислительно «доступными» при использовании их точной и вычислительно более эффективной эмуляции машинного обучения в гибридной модели.

Это исследование определяет основные рамки и первые экспериментальные результаты для нового гибридного подхода к моделированию. Развитие эмуляции нейронных сетей, ядра гибридного подхода к моделированию, в значительной степени зависит от способности генерировать представительную обучающую выборку, чтобы избежать использования нейронных сетей для экстраполяции далеко за пределами области покрытия обучающей выборки. Из-за высокой размерности входного домена, которая составляет порядка нескольких сотен или более, достаточно сложно освоить всю область, особенно её «дальние уголки», связанные с редкими событиями, даже когда используются модельные данные для обучения нейронных сетей. Другая связанная с этим проблема состоит в том, что эмуляции нейронных сетей разработаны для изменяющейся во времени геологической среды. Это означает, что при ее моделировании конфигурация домена может изменяться во времени. Для обеих описанных ситуаций, эмуляция нейронной сети может быть вынуждена экстраполироваться за ее способность обобщения, что может привести к ошибкам на выходе нейронной сети и в результате к соответствующей ошибке моделирования гибридной модели. Чтобы решить проблемы такого рода и сделать подход эмуляции нейронной сети приемлемым для моделирования долгосрочного изменения климата и других программ, развиваются два новых метода: соединение параметризации (СП) и динамической регулировки нейронной сети (ДР). Соединение параметризации состоит из трех элементов: оригинальный параметризации, его эмуляции нейронной сети и блока контроля качества. Во время обычного моделирования гибридной модели с СП, эмуляция нейронной сети используется по умолчанию, и она генерирует физические параметры (выходы), которые проходят контроль качества (КК). Если КК принимает выходные параметры, они используются в гибридной

модели. Если КК отвергает выходы эмуляции НС, вместо этого используется оригинальная параметризация для генерации физических параметров, которые используются в ней. Если оригинальная параметризация используется вместо эмуляции НС, ее входы и выходы будут сохранены для дальнейшей корректировки эмуляции НС. После накопления достаточного количества записей, динамической регулировки эмуляции НС производится короткое переобучение с использованием накопленных входящих / выходящих записей. Таким образом, адаптированные эмуляции нейронных сетей становятся динамически приспособленным к изменениям и / или новым событиям / состояниям, возникающим в сложных экологических системах.

Существуют различные возможные модели, которые соответствуют КК. Первая и самая простая модель основана на множестве регулярных физических и статистических тестов, которые используются для проверки согласованности выходов нейронных сетей. Вторая модель основана на обучении дополнительной НС специально для оценки ошибок в выходе эмуляции НС. Если эти ошибки превышают заданный порог, вместо эмуляции НС используется оригинальная параметризация. Третья и наиболее перспективная модель основана на технике проверки домена, предложенная в контексте применения нейронной сети для спутникового дистанционного зондирования. В этом случае контроль качества основан на комбинации прямой и обратной НС [Krasnopolsky V.M. 2006].

Димитр Рута с соавторами [Dymitr Ruta 2009] показал, что, несмотря на недавние успехи и достижения в области искусственного интеллекта и машинного обучения, эта область все ещё содержит в себе множество проблем и уходит от адекватного анализа явлений и процессов, наблюдаемых в реальном мире. Люди остаются непревзойденными по своей эффективности обработки и извлечения информации из неопределенных данных, поступающих в различных формах, в то время как все больше и больше алгоритмов обучения и оптимизации базируют свой аналитический

потенциал на некоторых естественных природных явлений. Превосходство методов обучения нейронных сетей и обучения на основе ядра, появление методов оптимизации на основе частиц, группы и социального поведения – лишь некоторые из многих фактов, свидетельствующих о тенденции к большей эксплуатации моделей и систем, создатели которых были вдохновлены природными явлениями. Работа Дititra продемонстрировала, как простая концепция физического поля может быть принята для создания полноценной основы для создания модели по методике контролируемого и неконтролируемого обучения. Вдохновение для искусственного обучения было обнаружено в механике физического поля на микро и макро уровнях. Используя аналогию между полученными данными и полями заряженных частиц подвергающихся гравитации, электростатических и газовых частиц, было разработано семейство новых алгоритмов, после чего применено к типизации, кластеризации данных и конденсации, в то время как свойства поля были использованы в уникальной визуализации типизации и модели синтеза типизатора. Его работа описывает обширные живописные примеры и визуальные интерпретации представленных методов наряду с некоторыми более известными реальными и искусственными базами данных сравнительного тестирования.

Понятие поля в типизации не новы и на самом деле связаны с методами ядра [Caruna 2006]. Смыслом использования понятия поля является обеспечение того, чтобы каждый образец данных активно участвовал в формировании окончательного решения типизации. Все данные, которые считаются заряженными частицами, каждая из которых является источником центрального поля, влияет на другие образцы. Все характеристики такого поля являются результатами определения потенциала и могут быть произвольно выбраны в зависимости от различных приоритетов. Для целей типизации суть состоит в том, чтобы присвоить метки класса ранее неизвестным образцам на основе класса пространственной топологии, полученной из обучающих данных. Эта цель достижима в рамках полевых

данных, если предположить, что тестовые образцы являются мобильными и вынуждаются полем изменять прикрепленные обучающие данные, делясь своим лейблом. Общее поле, измеренное в конкретной точке исходного пространства, является результатом суперпозиции локальных полей из всех источников. Таким образом, положение обучающих данные однозначно определяет поле во всем исходном пространстве и тем самым определяет изменения тестовых данных во время процесса типизации. Если поле сконструировано таким образом, что все возможные варианты изменения попадают в один из источников, то все пространство вариантов может быть разделено на области, представляющие различные классы. Границы между этими регионами образуют определенную границу выбора класса, который завершает процесс создания модели.

В своей работе он показал, что ряд естественных физических явлений могут быть непосредственно использованы для разработки модели искусственного обучения. С точки зрения обучающей выборки, автор старался показать, что простое понятие потенциального поля может быть легко адаптировано для формирования интересных способов применения типизации, кластеризации и модели конденсации данных и обеспечивает появление новых идей для визуализации, типизации процессов и синтеза модели. Тщательно проанализировав понятия поля, были рассмотрены образцы данных, как частицы, несущие определенный заряд и являются источниками классово-ориентированного поля.

В случае типизации, суперпозиция полей, генерируемых из каждого источника данных, полностью определяет потенциальный ландшафт, который образует тестовые данные, для попадания в потенциальные ямы и разделения класса исходного образца на дне ямы. Электростатическая метафора зависимых от заряда притяжений или отталкиваний поля была принята для создания внутриклассового притяжения и межклассового отталкивания для дальнейшего поощрения отделимости класса и

сглаживания границ класса, также отражается в повышение производительности типизации.

Было показано, что понятие поля также находит свое применение в конденсации данных. Предлагаемые семейства мягких динамических моделей конденсации позволяют сливать данные различных классов, кумулятивно сохранять мягкую/нечеткую границу разделения класса. Новый подход к динамической конденсации дает возможность переместить сложность типизированных данных из исходного пространства в пространство разделения классов, и было показано, что классификация стремиться к увеличению четкости базы данных с увеличением качества и уменьшением объема данных, который может достигать 99%.

Осознание того, что дискриминант функции, используемой в типизации, на самом деле является определенными реакциями поля на данные, показало, что оба типизатора, а также модель и объединитель, можно интуитивно визуализировать так, что он объяснял и давал понимание того, как модель или объединитель определяет границы классов и сглаживает выходы типизации. Такая визуализация особенно важна для модели слияния моделей, обычно рассматриваемой как очень сложный черный ящик, в настоящее время четко визуализируемой и прозрачной [Dymitr Ruta 2009].

Исследования Минаева Ю.Н. и его коллег по применению методов искусственного интеллекта для проведения мониторинга на предприятиях нефтеперерабатывающей промышленности показали, что этот способ применения искусственного интеллекта интересен постоянно пополняющейся исходной базой данных, которую требуется постоянно анализировать для постоянной оценки изменения ситуации. Исследования показали возможность решения задачи оперативного мониторинга, в частности определения источника загрязнения природной среды на основе методов и моделей искусственного интеллекта, ситуационного моделирования.

Решение поставленной задачи выполнялось с учетом того, что она является недостаточно формализованной и содержит: а) параметры, учитываемые в виде нечетких утверждений; б) параметры, задаваемые неточно [Минаев Ю.Н. 1996].

Интересной является работа Метью Дж. Кракнелла и Анны Ридинг [Matthew J. 2014] в которой были сравнены 5 алгоритмов (Байес, Ближайший сосед, Random Forest, машина опорных векторов, нейронные сети) машинного обучения при решении задачи литологической типизации. Исходные данные были взяты с территории порядка 160 км^2 рядом с Австралийским городом Брокен Хил. Необходимо было определить 13 основных литологических разностей. В качестве критериев были использованы координаты, геомагнитное поле, результаты гамма-спектрометрии территории, наличие калия, тория, уран и общего содержания радионуклидов, а также различные значения по полосам спектра. Связан такой набор признаков с составом пород, являющихся по большей части метаморфическими и вулканическими. В результате лучше всего себя проявил алгоритм Random Forest со значением ошибки менее 24%.

Лисенковым А.Б. с коллегами была разработана методика решения инженерно-геологических задач с помощью аппарата теории информации [Лисенков А.Б. 2003]. Ими был выполнен пространственный прогноз загрязнения подземных вод на основе методики обучения информационной модели по ключевым участкам, определены оптимальные критерии для решения ряда эколого-гидрогеологических задач, и приняты проектные решения на основе выполненного моделирования. Моделирование на основе информационного анализа показало, что пространственная информационная прогнозная модель хорошо себя проявляет и позволяет решать прогнозную пространственную задачу в сходных эколого-гидрогеологических условиях с необходимой точностью. В качестве необходимой точности был выбран порог ошибки в 25%. Эти работы показали перспективность использования машинного обучения в экологическом прогнозировании и указывают

основные шаги при построении прогнозной модели при помощи машинного обучения

Выше были описаны исследования, описывающие типы применения искусственного интеллекта для решения различных прикладных задач, связанных с изучением геологической среды. Здесь были представлены не все возможные применения, но эти способы являются наиболее интересными и представительными. Показаны основные темы, самые популярные алгоритмы и методы расчета. Благодаря этому можно понять, что искусственный интеллект является отличным инструментом для инженерно-геологического моделирования и делает процесс моделирования быстрее и проще во многих ситуациях. Были описаны примеры работы с анализом геофизических полей и с моделированием состояния геологической среды при помощи искусственного интеллекта. В результате анализа статей задача свелась к тому, чтобы выбрать лучший из всех описанных подходов машинного обучения и применять эту комбинацию в конкретной задаче инженерно-геологического моделирования.

Как было показано в данной главе, при решении инженерно-геологических задач, чаще всего машинное обучение используется для решения проблем, связанных с обработкой именно геофизических данных.

В данной главе были описаны исследования, которые показали различные способы применения машинного обучения для решения инженерно-геологических задач. В следующих главах показана методика разработанная в результате анализа всех описанных исследований методика определения устойчивости геологической среды при помощи машинного обучения. Как было продемонстрировано, уровень устойчивости геологической среды не определялся с помощью методов машинного обучения и при этом сама проблема является актуальной для исследования. В связи с этим дальнейшие эксперименты имеют под собой научную базу и решают актуальную проблему оптимизации решения данной задачи наиболее подходящим способом.

4. Анализ инженерно-геологических условий изучаемой территории для интерпретации результатов моделирования

В качестве базовых параметров модели, как упоминалось ранее, использовались данные о рельефе, тектонике, гидрологии, гидрогеологии, техногенной нагрузке и геологическом строении.

В работе произведено моделирование геологической среды города Москва, поэтому информация собиралась именно по этой территории. На сегодняшний момент г. Москва изучен довольно хорошо, поэтому вся информация была найдена в свободном доступе.

Далее приведено общее геологическое и тектоническое описание изучаемой области, которое использовалось для уточнения результатов исследований. Геологическое строение московского региона было описано многими авторами, но в данной работе приводится описание, представленное в работе Б.Б. Вагнера, Б.О. Манучарянца а также в материалах геологического атласа Москвы

Территория Московского региона расположена в центральной части Русской (или Восточно-Европейской) платформы. Как и всем платформенным сооружениям, Русской платформе присуще двухъярусное строение. Ее нижний структурный этаж - кристаллический фундамент - сложен древними породами архейской и протерозойской эры, а верхний этаж (платформенный чехол) слагают преимущественно осадочные породы палеозоя, мезозоя и кайнозоя [Вагнер Б.Б. 2003] (Рисунок 4.1).

Породы фундамента Русской платформы представлены различными магматическими и метаморфическими образованиями, включающими гнейсы, амфиболиты, филлиты, различные сланцы и кварциты, прорванные интрузиями гранитов, сиенитов и диоритов. Они залегают на значительной глубине и вскрыты в пределах нашей области только буровыми скважинами. Наименьшие глубины залегания фундамента отмечаются в Московском

регионе к югу от г. Серебряные» [Вагнер Б.Б. 2003]. Глубина их залегания составляет 1000 м. При этом наибольшие глубины наблюдаются к востоку от г. Зарайск и южнее г. Ногинск и составляют 4200 м. Породы фундамента выходят на дневную поверхность только на Балтийском щите к северу от Подмосковья. Среди моренных валунно- галечниковых отложений часто встречаются характерные образцы гнейсов, гранитов, и кварцитов, которые показывают облик горных пород, слагающих кристаллический фундамент Русской платформы [Вагнер Б.Б. 2003].

В верхнепротерозойское время глубокие прогибы закладывались на подвижных участках Русской платформы, один из которых в палеозое превратился в обширную Московскую синеклизу, являющейся главным структурным элементом осадочного чехла Московского региона. В основании слагающей синеклизу осадочной толщи, залегают породы вендского и рифейского отделов верхнего протерозоя. (Рисунок 4.2) «Они представлены песчаниками, конгломератами, алевритами, аргиллитами и глинами с прослоями вулканогенных пород - туфов и туффитов. Органические остатки в них встречаются крайне редко. Это отпечатки медуз и так называемые акритархи (сферические скопления отпечатков одноклеточных и планктонных организмов), которые обнаружены в глинах вендского возраста. Эти отложения образовывались при разрушении горноскладчатых сооружений фундамента и накапливались преимущественно в прогибах, где их мощность достигает иногда 2500 м. На разделяющих эти прогибы выступах мощность рифей-вендских отложений не превышает первых сотен метров, а порой они отсутствуют вовсе» [Вагнер Б.Б. 2003].

В раннепалеозойскую эпоху в течение кембрийского, ордовикского и силурийского периодов Московский регион был под влиянием каледонской складчатости вовлечен в процесс воздымания. В результате этого процесса осадконакопление происходило лишь в незначительной степени. «От этого времени сохранились только маломощные (60-80 м, на севере области до 300

м) пласты морских мелководных и лагунных (а в силуре также континентальных) отложений, вскрытых скважинами и представленных песчаниками, песками, глинами, реже мергелями и доломитами. В нижнедевонское время море окончательно покинуло пределы Московской синеклизы, и осадки этого возраста в Подмосковье не обнаружены. Начиная со среднего девона море вновь покрывает территорию региона, оставив повсеместно мощные пласты осадочных пород морского и лагунного генезиса: известняков, мергелей, песчаников, глин, каменной соли и гипса. Мощность пород среднего и верхнего девона достигает почти километра, но на дневную поверхность они нигде в Подмосковье не выходят» [Вагнер Б.Б. 2003].

Погружение Русской платформы было продолжено и в каменноугольном периоде. В это время шло мощное осадконакопление морских отложений. Отложения каменноугольного периода являются самыми древними из тех, что можно обнаружить на дневной поверхности в столичном регионе. При этом причем их мощность значительно превосходит общую мощность более молодых, кайнозойских и мезозойских [Вагнер Б.Б. 2003].

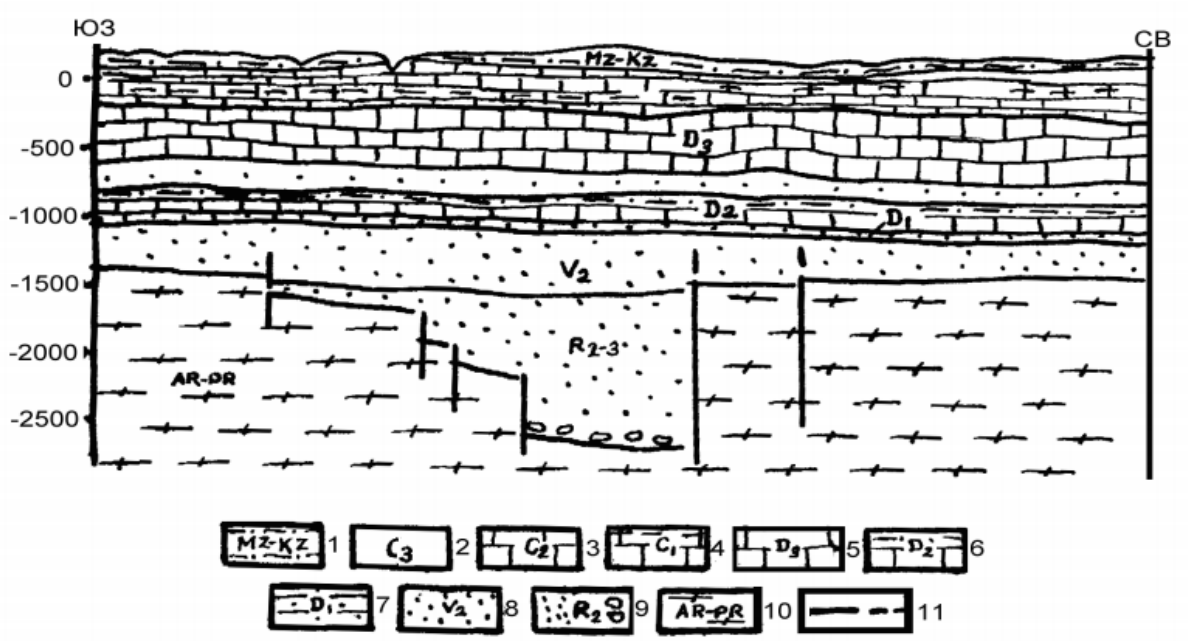


Рисунок 4.1. Геологический профиль через центральную часть Московского региона 1 – мезокайнозойские отложения; 2 – верхнекаменноугольные отложения; 3 – среднекаменноугольные отложения; 4 – нижнекаменноугольные отложения; 5 – верхнедевонские отложения; 6 – среднедевонские отложения; 7 – нижнедевонские отложения; 8 – вендские отложения, 9 – рифейские отложения; 10 – кристаллические породы архей-протерозойского возраста, 11 – глубинные разломы [Вагнер Б.Б. 2003]

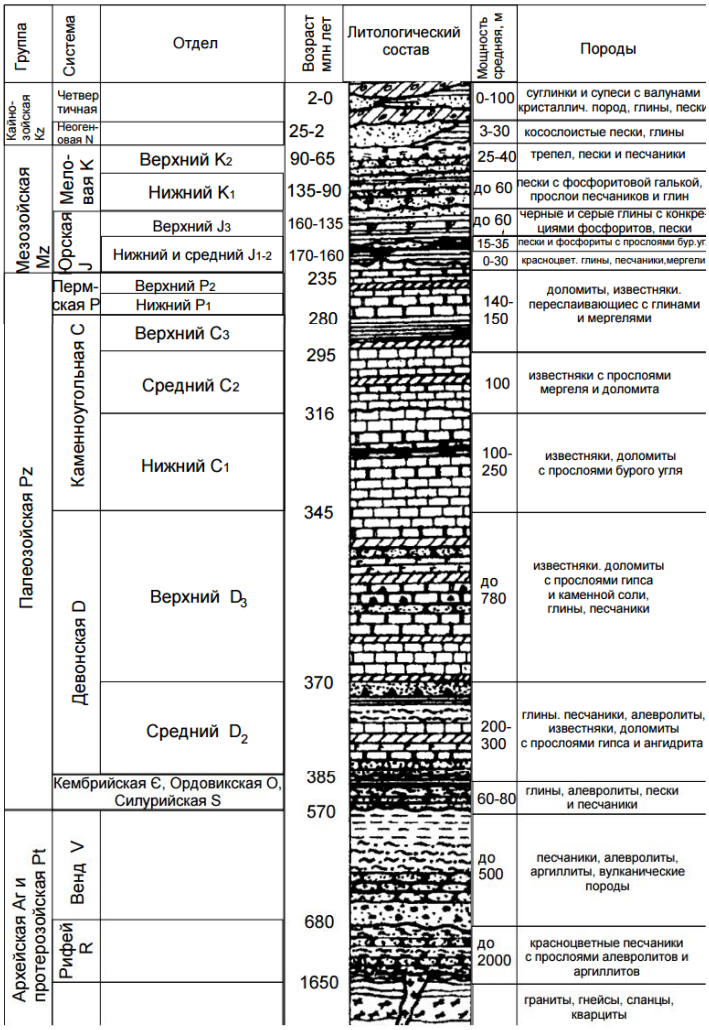


Рисунок 4.2 Сводная стратиграфическая колонка Московской синеклизы [Вагнер Б.Б. 2003]

Каменноугольные отложения распространены повсеместно и принадлежат нижнему, среднему и верхнему отделам. Тектоническая позиция территории г. Москвы в осадочном чехле - южное крыло Московской синеклизы, определяет условия залегания каменноугольных пород: их подошва погружается в северо-восточном направлении.

На большей части территории каменноугольные отложения перекрываются породами мезозоя и лишь на отдельных участках - четвертичными отложениями.

Средне- и верхнекаменноугольные породы вскрывались в бортах долины р. Москвы, на юго-западе г Москвы в районах Дорогомилово и Шелепихи, где в настоящее время они засыпаны.

Поверхность карбона сильно эродирована. На ней прослеживаются несколько крупных доюрских долин, выполненных аллювиальными и ингрессивными морскими отложениями средней юры.

На домезозойскую поверхность выходят отложения среднего и верхнего отделов. В районе Солнцево, в четвертичном врезе вскрываются отложения каширской свиты. Среднекаменноугольные отложения занимают всю южную треть площади г. Москвы, а также выходят в тальвегах доюрских долин (Рисунок 4.3) [Геологический Атлас Москвы].

Мезозойские отложения на территории Москвы представлены юрскими и меловыми. В основу принятой легенды к картам дочетвертичных отложений были положены местные и региональные стратиграфические схемы юрских и меловых отложений. В общей шкале бывшим нижнему и среднему подъярусам волжского яруса соответствует титонский ярус верхней юры, верхнему - нижняя часть берриасского яруса нижнего мела.

Выходы мезозойских отложений на поверхность дочетвертичного рельефа, отраженные на карте, контролируются изогипсами с колебаниями абс. отметок от 80-125м до 242 м, свидетельствующими о совпадении, в основных чертах, погребенного и современного рельефов. На фоне общего погружения палеорельефа к северо-востоку, можно выделить систему погребенных долин, которые в большинстве случаев являются унаследованными современными. Центральной является пра-Москва, которая имеет множество палеопритоков. Наиболее глубокие из них: пра-Яуза, пра-Лихоборка, пра-Городня и пра-Чертановка, вскрывающие весь мезозойский разрез и пра-Сходня, пра-Серебрянка, пра-Чермянка, достигающие кровли среднекелловейских глин, а также другие более мелкие палеодолины (Рисунок 4.4.).

Юрская система на территории г. Москвы представлена байосским, батским и келловейским ярусами среднего отдела, а также оксфордским, кимерижским и титонским ярусами верхнего отдела. Юрские отложения трансгрессивно, с резким угловым и стратиграфическим несогласием

залегают на эродированной поверхности каменноугольных образований [Геологический Атлас Москвы].

Меловая система представлена нижним и верхним отделами.

Нижний отдел представлен берриасским, готеривским, барремским, аптским и альбским ярусами.

Верхнемеловые отложения слагают высокую часть водораздела Теплостанской возвышенности. В их составе выделяются сеноманский, коньякский и сантонский ярусы [Геологический Атлас Москвы].

Четвертичные отложения распространены практически повсеместно, исключение составляют некоторые участки склонов берегов рек. На поверхности дочетвертичных отложений выделяется система погребенных палеодолин. Их дно имеет абсолютные отметки от 80 до 100-120 м, водоразделы - 130-190 м. На Теплостанской возвышенности они достигают 230-243 м.

Мощность четвертичных отложений в Москве колеблется в среднем в диапазоне от 1,5 до 20 м. Однако на западе и северо-западе возрастает до 25-35 м.

На территории г. Москвы выделяются водно-ледниковые образования, а также моренные горизонты. Они относятся к нижнему плейстоцену ликовской толще, сетуньской и донской свитам, среднеплейстоценовой московской свите. Межледниковые отложения представлены окатовской, сукромненской свитами, рославльской серией, лихвинская, микулинская, мончаловская свитами.

В долинах рек выделяются аллювий двух уровней второй (мневниковской) и первой (серебряноборской) террас аллювиально-флювиогляциальные отложения третьей (ходынской) надпойменной террасы, и пойменный аллювий.

На склонах речных долин выделяется коллювий. Помимо пойменного аллювия, в голоценовых отложениях присутствуют болотные, хемогенные и техногенные образования (Рисунок 4.5) [Геологический Атлас Москвы].

Краткий обзор тектонических и неотектонических особенностей строения территории

Тектонические особенности строения рассматриваемой территории определяются ее положением на южном крыле Московской синеклизы, именуемой Тульской моноклиналью.

Породы фундамента расчленены разрывными нарушениями. Выделяется в основном три направления их простирания: субширотное, северо-западное, северо-восточное и близкое к субмеридиональному. Разломы, окаймляющие блоки, часто при этом совпадают с границами архейских и нижнепротерозойских структур. Разнонаправленные тектонические движения блоков относительно друг друга формируют блоково-клавишную структуру поверхности фундамента. В пределах северо-западной части г. Москвы выделяется Московский (Красногорский) горст, на северо-востоке наблюдается выход Мытищинский грабена, а на западе отмечается Рузский выступ. Рельеф поверхности фундамента г. Москвы характерен перепадами ее глубин залегания порядка 1000 м.

Узкие грабенообразные прогибы – авлакогены можно встретить вдоль разрывных нарушений фундамента, образовавшихся в предрифейское время. Эти прогибы в последствии заполнялись осадочно-вулканическим материалом рифейского времени. Так рифейский Подмосковный авлакоген выходи на юге Москвы. Территория города ограничивается двумя зонами разрывных нарушений субширотного распространения: Московской и Раменской. Перепад высот в структуре авлакогена составляет порядка 1000 м (Рисунок 4.6).

Что касается каменноугольных отложений, то большая часть г. Москвы, кроме южных районов, находится в пределах структуры II порядка, с юга ограниченной цепью наклоненных к северу флексур. На западе впадина граничит с чередой флексур и структурных уступов, которые простираются в направлении от Одинцово к Зеленограду. Амплитуды флексур составляют порядка 10-25 м. Структура II порядка осложнена структурами III-го:

Тушинским выступом, Щелковским выступом, Пироговской впадиной и Болшевской впадиной. Юг Москвы располагается в пределах Подольского выступа (структура II порядка), где он прослеживается по Чертановской впадине III-го порядка , которая с юга ограниченна флексурами.

В каменноугольных отложениях выделяется также ряд мелких структур. Их длина составляет от 100 м до первых километров, а ширина порядка 0,5 км (Рисунок 4.7.).

Тектоника юрских отложений во многом схода с тектоникой каменноугольных. Структуры II порядка сохраняют свое положение, лишь немного сдвинув свои границы. Структуры ограниченны флексурами с амплитудами 5-20 м. При этом структуры III-го порядка резко изменили свои очертания (Рисунок 4.9).

Неотектоника изучаемой территории в основном определяется по очертаниям гидрографической сети. Долины рек г. Москвы имеют главным образом три направляния простирания: субширотное, северо-западное - юго-восточное и субмеридиональное. Субширотное направление характерно для юга (Рисунок 4.8.). Северо-западное - северо-восточное направление связано с р. Москвой в черте города. Субмеридиональное простирание заметно на востоке территории.

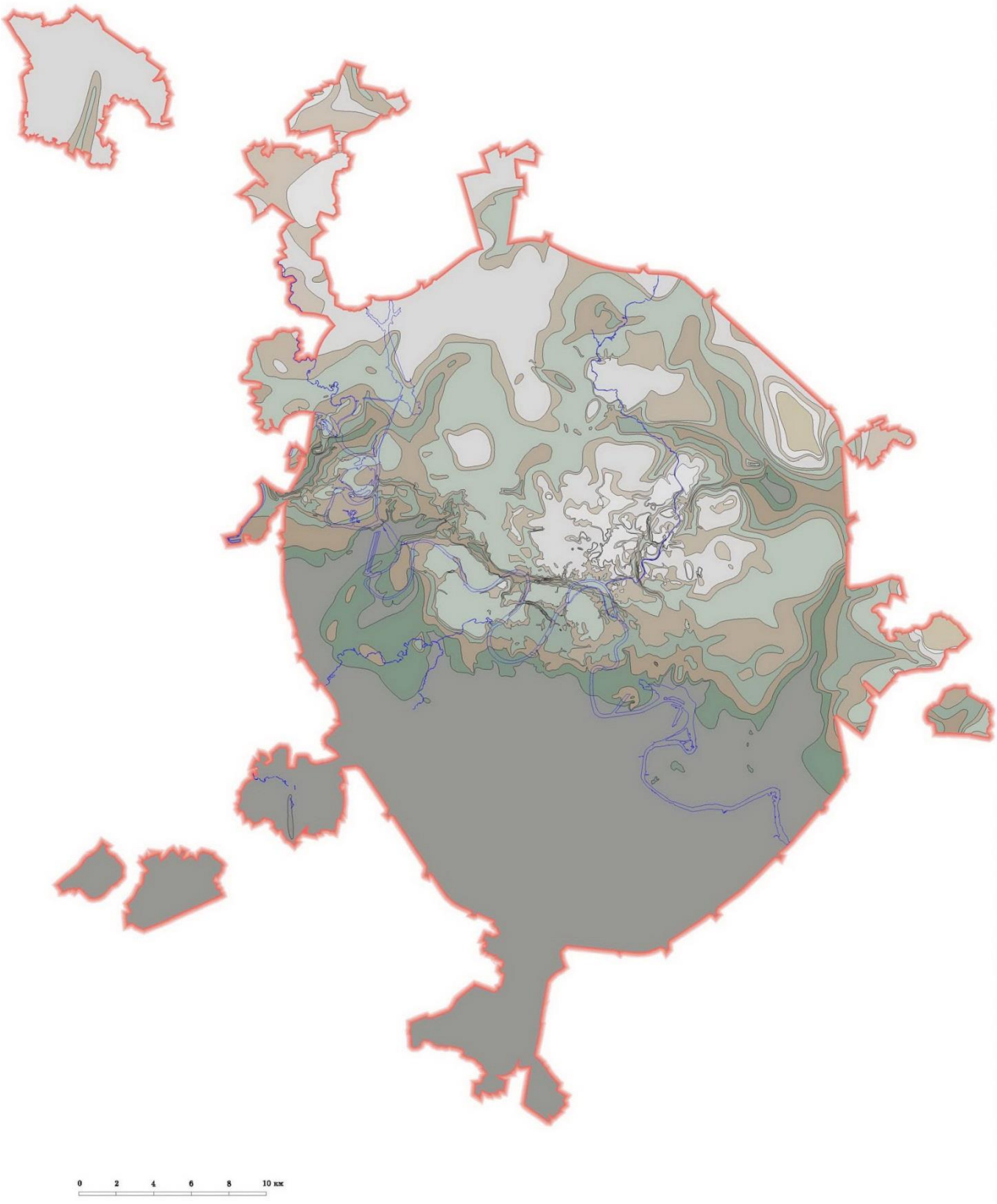
Долины рек Десна, Пахра, Городня и широтный отрезок долины р. Москвы практически параллельны между собой. Это может являться признаком их тектонического происхождения, особенно если сравнить на простирание разломов в этой части. Долины рек Пахра, Десна и Городня с севера и юга ограничивают Подмосковный авлакоген. Долины притоков р. Москвы широтного простирания совпадают при этом с полосой изгибов слоев отложений каменноугольной системы. Отсюда следует, что широтное распространение рек во многом связано с разрывными нарушениями.

При этом связь долин рек, простирающихся меридионально, с разрывными нарушениями не прослеживается.

Линиаментная зона же северо-западного - юго-восточного направления совпадает с простиранием р. Москвы, что также может говорить о тектоническом происхождении.

В блоковом строении современного рельефа выделяется три блока: Южный (Подольский), северо-западный (Рузский) и Северный.

Также на территории г. Москвы выделяются кольцевые структуры. В основном они представлены дуговыми и кольцевыми отрезками долин рек. Такие структуры тесно связаны с фундаментом платформы [Геологический Атлас Москвы].


0 2 4 6 8 10 км

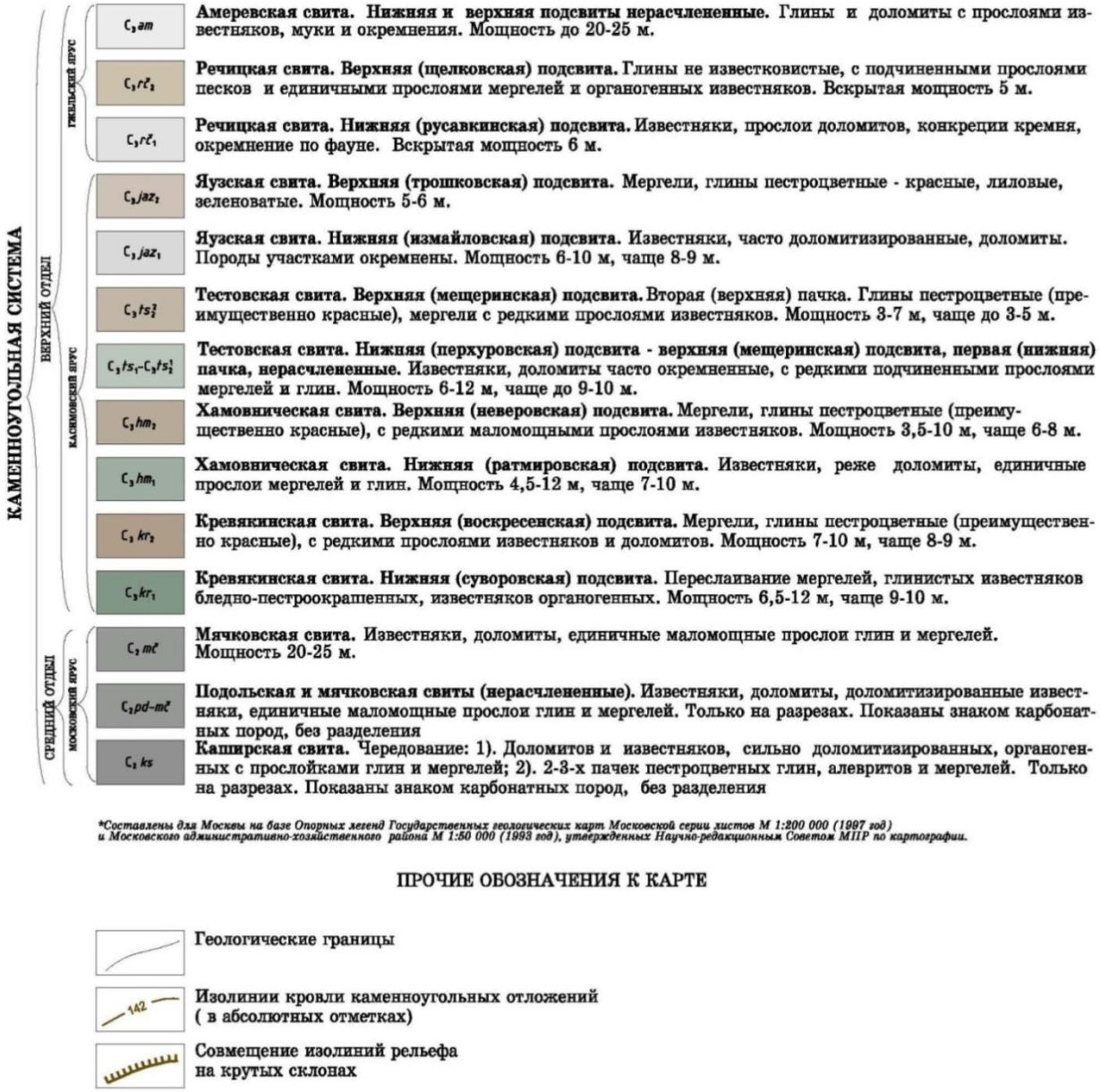
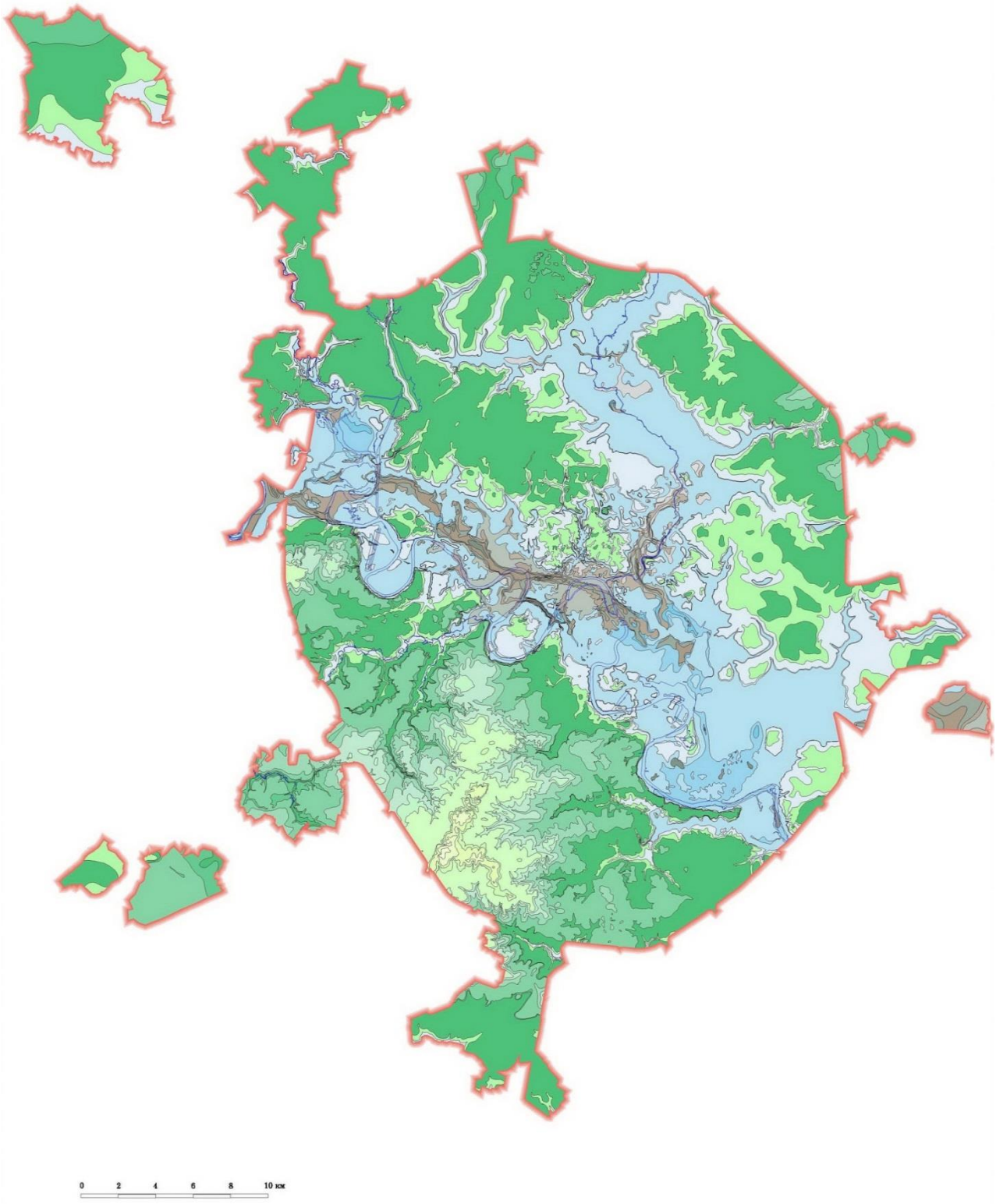


Рисунок 4.3. Карта каменооугольных отложений г. Москвы [Геологический Атлас Москвы].


0 2 4 6 8 10 км

Тенътиковская свита. Трепелы, опоки, пески палево- и зеленовато-серые. Мощность до 3 м.

Дмитровская свита. Пески зеленовато-серые, песчаники разнозернистые, кремнистые. Мощность до 3 м.

Загорская свита. Пески желтые мелко- и среднезернистые с линзами опоковидных глин, слюдистые. Мощность до 10 м, чаще 6-8 м.

Яхромская свита. Пески зеленовато- и желтовато-серые мелкозернистые. Мощность 5-8 м.

Парамоновская свита. Глины черные алевритистые, вверху - алевриты глинистые, внизу - пески глинистые с прослоями глин. Мощность до 30 м, чаще 23-25 м.

Гаврилковская свита. Пески серо-зеленые разнозернистые, местами глинистые, иногда с песчаными фосфоритами. Мощность 3-7 м.

Волгушинская свита. Вверху - пески зеленовато-серые, глинистые мелкозернистые с прослойками глин, внизу - глины алевритовые темно-серые с тонкими прослойками песков мелкозернистых. Мощность более 6-9 м.

Ворохобинская свита. Вверху - пески буровато- или зеленовато-серые мелкозернистые слюдистые, с прослойками глин,внизу - глинистые алевриты или глины алевритовые темно-серые со слабым зеленоватым оттенком с тонкими прослойками песков тонкозернистых. Мощность 4,5-9 м, чаще 6-8 м.

Бутовская толща баррема и икшинская свита апта, объединенные. Пески белые и желтые кварцевые мелкозернистые, слюдистые с линзами (стяжениями) слабых железистых песчаников, с тонкими прослойками каолиновых глин жирных и алевритовых преимущественно белого и сиреневого цветов. Мощность 9-18 м, чаще 10-16 м.

Котельниковская свита. Глины алевритовые и алевриты сильно глинистые черные, с темно-фиолетовым оттенком, плотные с тонкими линзами и гнездами ("рябец") белых песков тонкозернистых и алевритов. Мощность 0-6 м, чаще 2-3 м.

Кунцевская и люберецкая толщи берриаса, ростовская свита нижнего готерива, дьяковская толща, савельевская и гремячевская свиты верхнего готерива объединенные. Вверху - пески и песчаники желтовато-серые разнозернистые, ожелезненные ($K_1 gr$), ниже серые пески и алевриты слабоглинистые, сильно слюдистые ($K_1 sv$), пески серые тонко- и мелкозернистые ($K_1 rs$), пески ярко зеленые сильно глинистые тонко-среднезернистые железистые, глауконитовые (K_1 dk), пески серые тонко-мелкозернистые (K_1kn), пески светло-серые до белых (K_1lbr). Мощность 16-39 м, чаще 18-24 м.

Лопатинская свита. Пески глинистые зелено-черные глауконитовые, тонко-мелкозернистые, прослои глинистых алевритов; внизу часто фосфоритовая галька, вверху - стяжения песчаных фосфоритов. Мощность до 7,5 м.

Филевская свита. Глинистые алевриты (до глин алевритовых), вверху тонкопесчаные супеси. Породы темно-серые и черные с зеленоватым оттенком, сильно слюдистые. Мощность 9,5-13 м, чаще 4-9 м.

Егорьевская свита. Пески глинистые глауконит-кварцевые, зелено-черные, мелкозернистые, часто насыщенные конкрециями и галькой фосфоритов. Мощность 2-5 м, чаще 1-1,5 м.

Великодворская - ермолинская свиты нерасчлененные. Вверху - глины темно-серые алевритистые, редко сланцеватые, преимущественно тугопластичные, известковистые, слюдистые с обильным раковинным детритом, внизу (2-4 м) - глины серые тонко алевритистые до жирных, массивные, весьма плотные. Мощность 10-32 м, чаще 8-17 м.

Криушская свита. Песчаные глины и глинистые пески, разнозернистые буро-серые, известковые с раковинным детритом и обильными железистыми оолитами, с прослоями песчаного мергеля. Мощность 2,5-9,5 м, в долинах до 17 м.

Люблинская толща. Алевриты песчано-глинистые и глины тонкопесчаные буровато-серые, тонкослоистые с прослоями тонкозернистых песков, с углистыми включениями и отпечатками растений. Мощность до 9 м.

Москворецкая толща (верхняя пачка). Глины болотные черные, жирные, алевритовые, с подчиненными прослойками песков, с углистыми включениями. Мощность до 6,5 м.

Москворецкая толща (нижняя пачка). Пески, от тонко- до среднезернистых, внизу редко - гравий карбонатных пород и кремня. Мощность до 10 м.

Кудиновская толща. Глины зелено-серые жирные, весьма плотные с единичными, тонкими линзами песка и углистыми отпечатками растений, внизу часто со щебнем карбонатных пород и кремня. Мощность до 8,5 м.

Примечание: на крутых склонах поверхности дочетвертичных пород некоторые подразделения мезозоя объединены

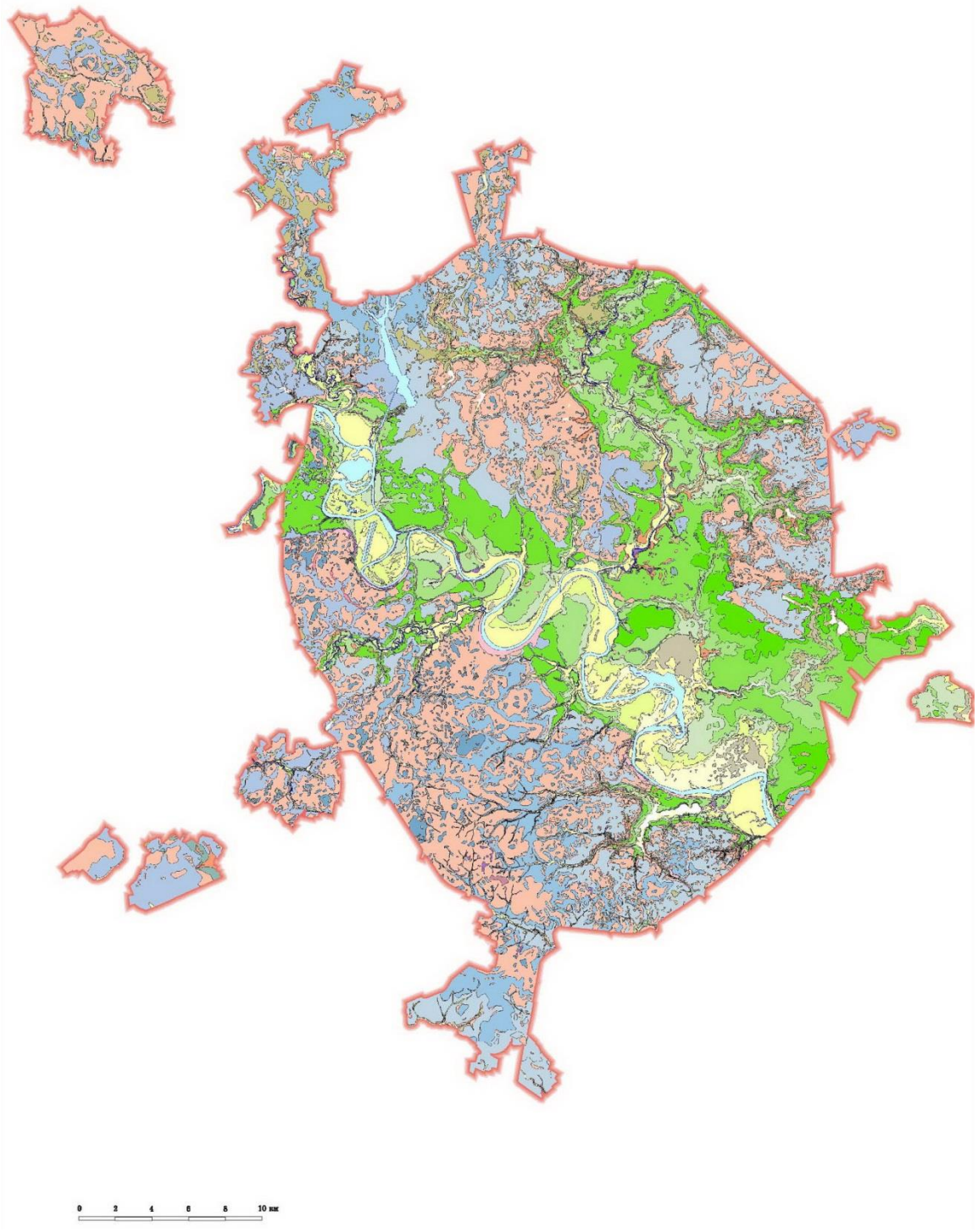
ПРОЧИЕ ОБОЗНАЧЕНИЯ К КАРТЕ

Геологические границы

Изолинии кровли дочетвертичных отложений (в абсолютных отметках)

Совмещение изолиний рельефа на крутых склонах

Рисунок 4.4. Карта дочетвертичных отложений г. Москвы [Геологический Атлас Москвы].


0 2 4 6 8 10 км

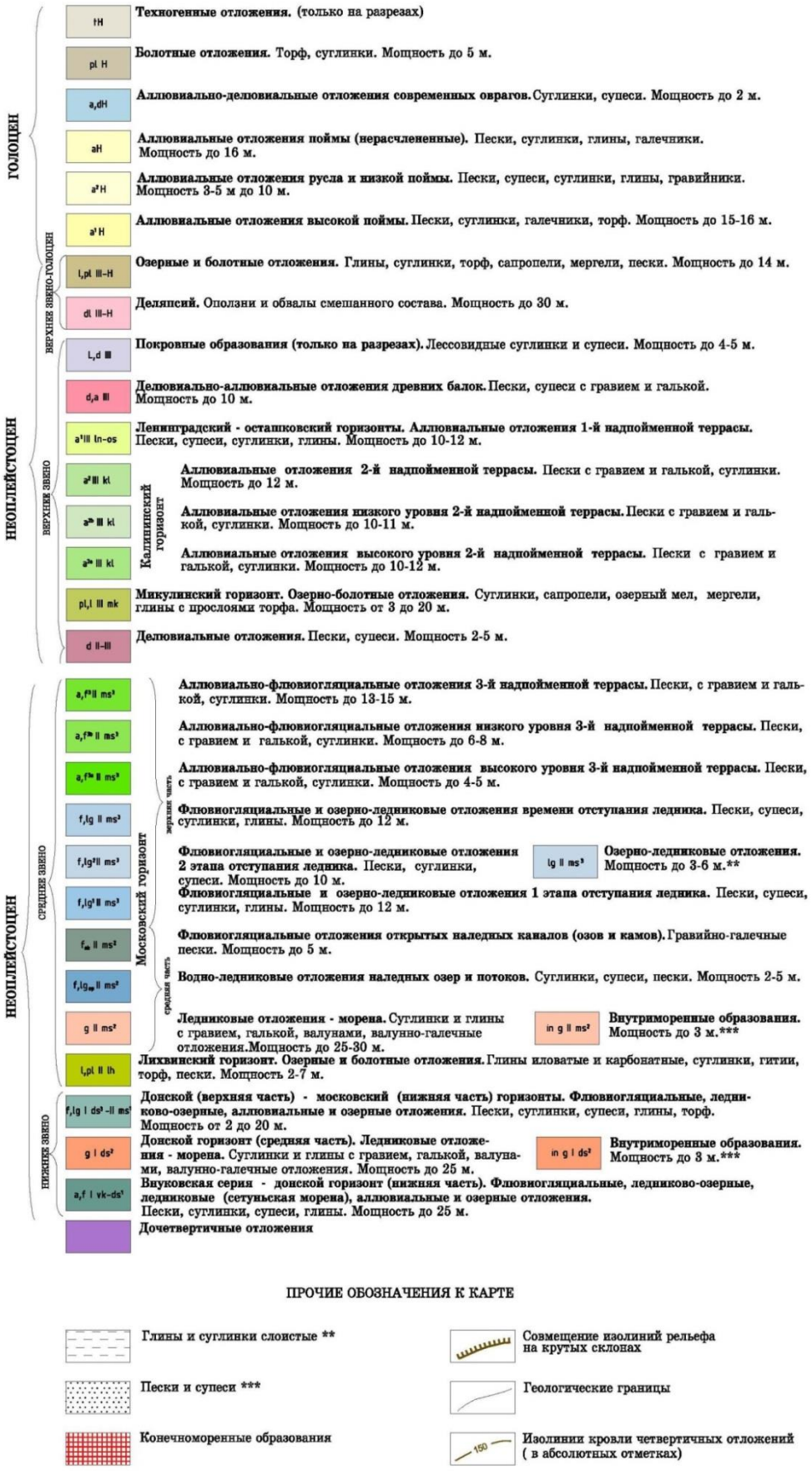


Рисунок 4.5. Карта четвертичных отложений г. Москвы [Геологический Атлас Москвы].

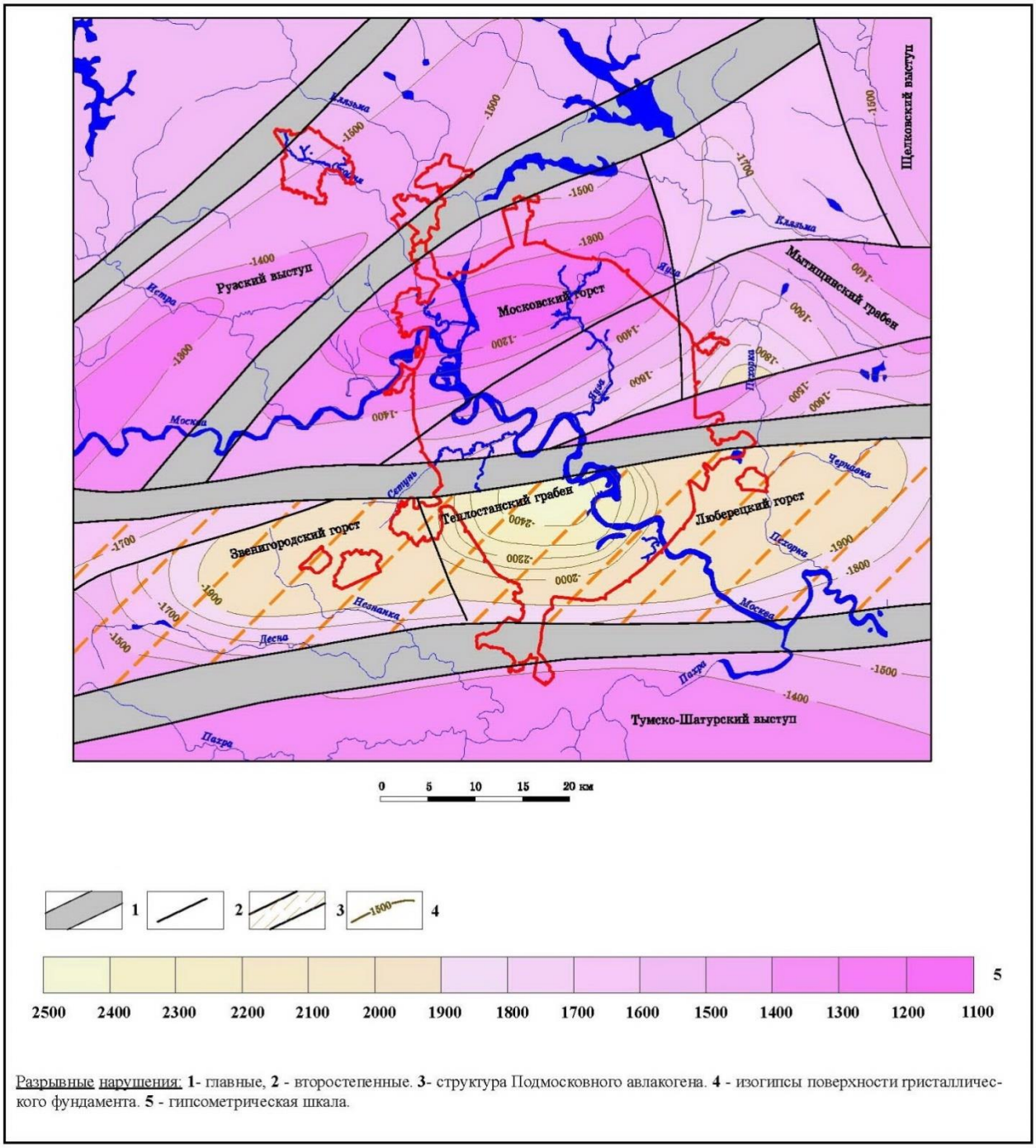
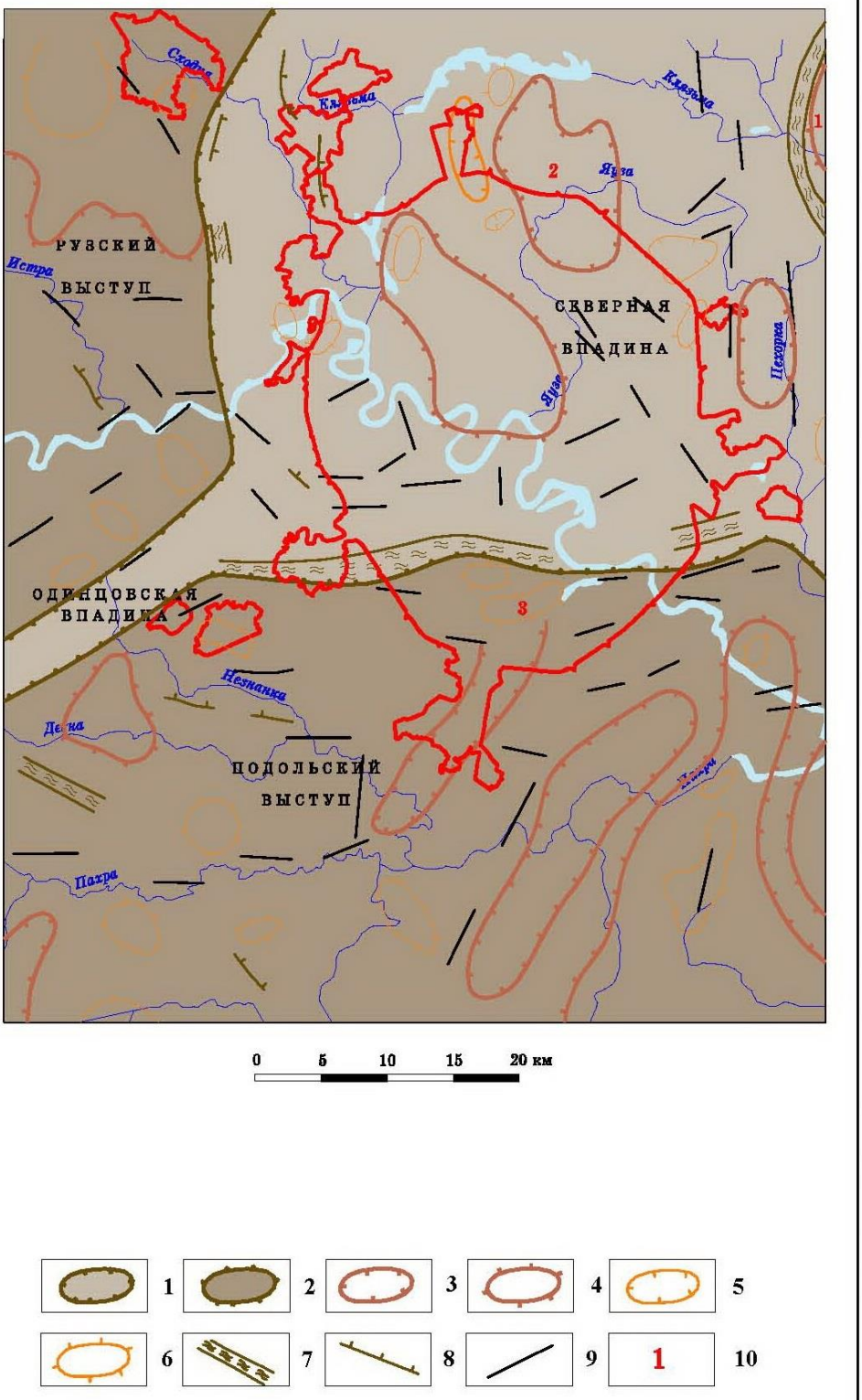


Рисунок 4.6. Тектоническая схема кристаллического фундамента [Геологический Атлас Москвы].



Структуры II порядка :**1** - отрицательные. **2** - положительные. Структуры III порядка: **3** - отрицательные. **4** - положительные. Структуры IV порядка: **5** - отрицательные. **6** - положительные. **7** - линейные зоны флексур. **8** - структрурные уступы. **9** - линиаменты. **10** - номера структур, упомянутые в тексте.

Рисунок 4.7. Тектоническая схема девонско-каменноугольных и юрских комплексов [Геологический Атлас Москвы].

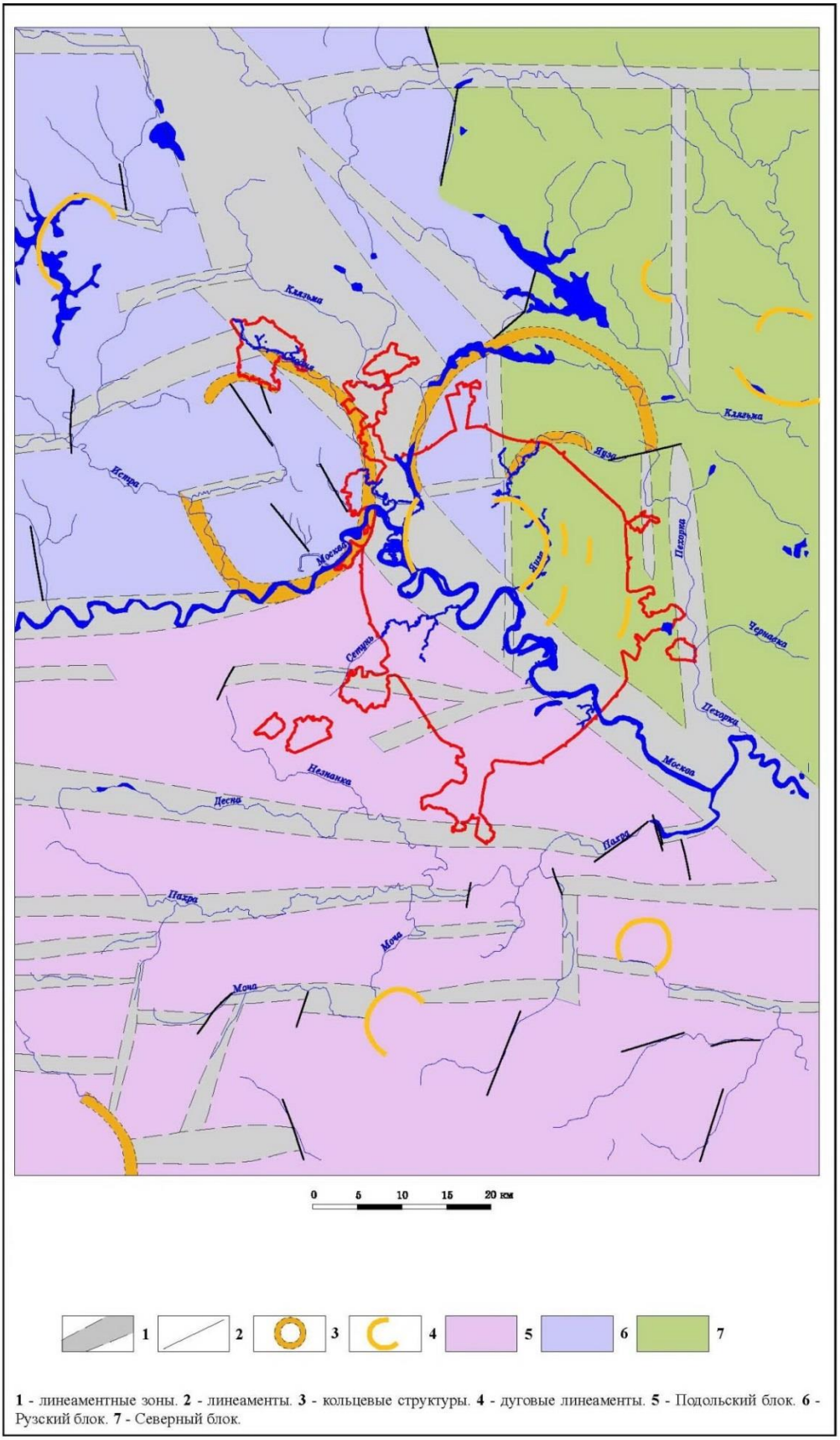
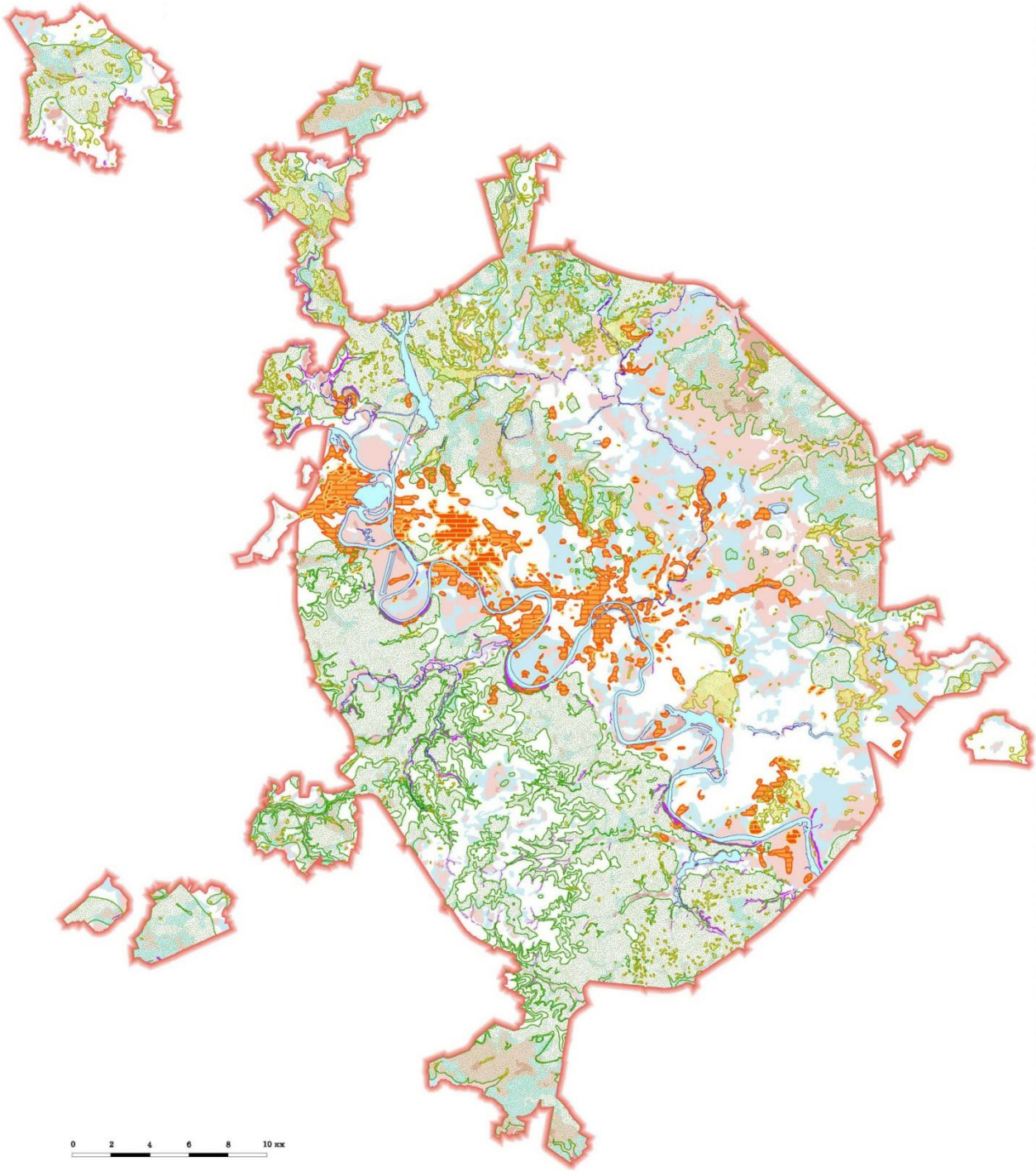


Рисунок 4.8. Схема линеаментных зон, линеаментов и кольцевых структур [Геологический Атлас Москвы].


0 2 4 6 8 10 км

4.9. Карта распространения опасных геологических процессов в Москве [Геологический Атлас Москвы].

Представленное общее геологическое и тектоническое описание помогает в общих чертах охарактеризовать геолого-геофизические условия на изучаемых территориях, что дает возможность более детально определять достоверность построенной модели. Также расположение разломов, границ тектонических платформ и геологическое строение, в случае удачного моделирования, должно частично коррелировать с границами выделенных зон устойчивости геологической среды, так как эти факторы в своей основе определяют рельеф и экзогенные процессы.

5. Методика оценки устойчивости геологической среды на основе машинного обучения

При анализе устойчивости геологической среды была рассмотрена территория города Москвы. Этот объект исследования является сложной природно-техногенной системой, состоящей из множества уровней, сложно связанных между собой. В качестве показателя устойчивости геологической среды был использован геологический риск. Он оценивается вероятными социальными и экономическими потерями в городе в случае наступления неблагоприятных геологических процессов и выражается в снижении устойчивости городской инфраструктуры и безопасности проживания людей. Под устойчивостью при этом понималась стабильность функционирования и долговечность жилой и промышленной застройки, транспортных магистралей, инженерных сетей, зеленых насаждений и других объектов городской среды.

В данной работе для этого анализа использовалось так называемое «обучение с учителем». Оно заключается в использовании результатов проведенных специалистами исследований (в данном случае группой во главе с Осиповым В.И. [Осипов В.И. 1997]) так, чтобы на их основе «научить» машину, копировать логику принятия решений при определении уровня устойчивости геологической среды г. Москвы. После чего, когда возникнет задача сделать подобный анализ на новой неизвестной, но со сходными инженерно-геологическими условиями территории, машина, отталкиваясь от полученных знаний, могла бы провести подобный анализ уже самостоятельно без участия специалистов.

Первоначально, определение устойчивости геологической среды производилось специалистами при многолетнем анализе территории г. Москвы. Результаты этого анализа были взяты за основу для обработки при помощи машинного обучения. В итоге у компьютера была задача выявить

критерии оценки территории устойчивости геологической среды на основе данных, используемых экспертами, и получить модель, описывающие реальную ситуацию на изучаемой площади.

Критериями оценки территории при экспертном анализе являлись данные о развитии трех важнейших опасных геологических процессов, развивающихся на территории города (карстово-суффозионном, подтоплении, образовании оползней), интенсивности и месте их проявления.

Задача исследования состояла в том, чтобы показать, что при помощи машинного обучения возможно повторить подобное моделирование, а следовательно оптимизировать и унифицировать его для новых изысканий.

Конечно, полностью заменить работу специалиста невозможно. Это объясняется тем, что даже в случае совершенного алгоритма обработки конкретного вида информации во время компьютерного моделирования постоянно появляется новая информация, вносящая корректировки в исходные условия. Эта информация нуждается в новом понимании. После экспертного анализа новой информации, интеллектуальная модель улучшается и испытывается ещё раз. После улучшения вводятся новые данные для экспертного анализа, которые он мог бы использовать для дальнейшего улучшения системы и так далее. Метод оптимизации позволяет выполнять его работу быстрее и проще. В процессе оптимизации всё многообразие вопроса исследования, описанное выше, должно быть уложено в логику компьютера и выражено при помощи небольшого числа параметров, являющихся ключевыми при оценке уровня устойчивости геологической среды. Отсюда можно сделать вывод, что одной из основных проблем данного исследования является определение правильных параметров оценки геологической среды для интеллектуального анализа данных и поиск лучшего алгоритма их обработки. Логика выбора критериев описана далее. Выбор алгоритма осуществляется экспериментально.

Как было показано ранее, в предыдущих исследованиях не было использовано большое количество алгоритмов для решения экологических

проблем при помощи машинного обучения. Больше всего были использованы алгоритмы дерева J48, опорных векторов, логистической регрессии и многослойного персептрона. В некоторых случаях использовались в большей степени деревья решений и статистические методы. В данной работе проводилось сравнение большего числа алгоритмов, что позволило подобрать однозначно лучший вариант для решения сложной проблемы типизации. При использовании большего числа алгоритмов вероятность увеличения точности прогноза возрастает.

На рисунке 5.1 ниже показана схема подготовки, обработки данных и всего процесса моделирования.

Рисунок 5.1. Схема машинного обучения при определении устойчивости геологической среды. Белыми стрелками указан процесс обучения, зелеными стрелками указан процесс тестирования.

Как показано на рисунке 5.1. эту общую методологию можно разделить на две части: обучение и тестирование. Обучение необходимо для создания базовой модели на обучающей выборке. Тестирование производится для проверки корректности модели, её корректировки и оценки возможности решения реальных задач. После опробования модели, она проверялась решением реальной задачи типизации территории. Таким образом, после прохождения всех уровней была получена работающая система оптимизации принятия решения для определения устойчивости геологической среды ниже, указанная методология описана более подробно на каждом этапе.

5.1. Методика создания модели устойчивости геологической среды

Машинное обучение в своей основе представляет собой анализ влияния группы факторов на некий признак. В нашем случае это анализ влияния взаимосвязанных между собой природных и техногенных факторов на устойчивость геологической среды. По сути это смесь многофакторного анализа, метода аналогии и статистики. В инженерной геологии многофакторный анализ был описан Бондариком Г.К. в его теории геологического поля, Пендиным В.В. в комплексном количественном анализе информации в инженерной геологии и многими другими исследователями.

Рассмотрим, какие, в таком случае, факторы влияют на оценку устойчивости геологической среды?

Так как в качестве примера применения машинного обучения для определения уровня устойчивости геологической среды был выбран город Москва, исходные данные были взяты именно по этому городу. Для этого были использованы результаты исследований группы ученых, опубликованных под редакцией Осипова В.И., Медведевой О.П. [Осипов В.И. 1997]. Как было указано выше, эти авторы в качестве факторов, влияющих на устойчивость геологической среды, выделяли три важнейших опасных геологических процесса, развивающихся на территории города: карстово-суффозионный, подтопление, образование оползней. Однако, для получения информации по этим процессам исследователями предварительно были обработаны данные по геологическому строению, гидрогеологии, тектонике, гидрологии, ландшафтно-геоморфологическим условиям и др. Т.е. следует учитывать, что данные по опасным геологическим процессам сами по себе уже являлись результатом обработки огромного числа исходных параметров, описывающих инженерно-геологические условия, т.е. были результатом промежуточного моделирования. Такие данные не могут быть

получены на первоначальном этапе. А, как указывалось выше, с точки зрения полезности оценки устойчивости геологической среды, намного лучше было бы ее получить в самом начале исследований.

Поэтому для оценки устойчивости геологической среды при помощи машинного обучения в качестве критериев оценки использовались следующие:

1. Рельеф;
2. Разломы;
3. Линиаментные зоны;
4. Уровень грунтовых вод;
5. Гидрологическая сеть;
6. Мощность техногенных образований.

Компьютеру была поставлена задача определения устойчивости геолгической среды на базе вышеперечисленных критериев, как наиболее универсальных, достаточно отражающих инженерно-геологические условия и доступных на самых ранних стадиях работ.

В качестве обучающей выборки использовались данные рельефа поверхности земли (абсолютная высотная отметка местности, м), уровня грунтовых вод (абсолютная отметка, м), мощность техногенных образований (м), наличие главных и второстепенных разломов фундамента, наличие линиаментов и гидрологическая сеть на территории г. Москвы.

Данные по этим критериям, представленные, главным образом, в виде карт по г. Москва оцифровывались и представлялись в виде таблиц.

В таблице, описывающей каждый критерий, содержалась информация о координатах каждой условной точки на оцифрованной карте и значении параметра в этой точке. В случае рельефа, например, в таблице приводились данные по абсолютной высотной отметке в каждой точке, а в случае разломов указывалось просто его наличие или отсутствие в указанной точке.

Затем было необходимо эти данные обработать. При обработке указанные таблички сливались в одну таким образом, чтобы в координаты

1ой точки первой таблички совпадали с координатами 1ой точки второй таблички и т.д. При этом получалась единая матрица значений, содержащая сведения сразу обо всех значимых факторах в каждой условной точке. Далее эта матрица была проверена на наличие шумов, отсутствующих значений и других дефектов данных для более быстрой и качественной обработки.

После этого идентичным образом оцифровывалась карта устойчивости геологической среды г. Москвы, являющаяся результатом многолетней работы группы специалистов.

В итоге получилась такая таблица, содержащая координаты и значения категории устойчивости геологической среды в каждой условной точке, совпадающей с точками уже описанной матрицы. После чего эти данные так же добавлялись в матрицу, завершая ее. Матрица описывала 40 000 точек.

Благодаря этим действиям была получена так называемая «обучающая выборка», т.е. выборка, содержащая в себе и значения факторов, характеризующих устойчивость геологической среды, и сами значения устойчивости в каждой условной точке исследуемой территории. В случае если бы выборка содержала только значения критериев, а устойчивость геологической среды была бы нам неизвестна, то такая выборка была бы тестовой. Стоит отметить, что для обучающей выборки были использованы данные по территориям, находящимся исключительно внутри МКАД.

Данные имели табличный вид, пример которого представлен в таблице 5.1.1.

Таблица 5.1.1. Пример таблицы обучающей выборки.

№ п/п	X	У	Высотная отметка (м)	УГВ (м)	Мощность техногенных образований (м)	Разломы	Линиаменты	Поверхностные воды	Уровень геологического риска
1	0	0	191,50	170	0,14	0	1	0	1
…	…	…	…		…			.	…
40000	3300	3300	155,04	152, 89	1	0	0	0	3

Все данные были разделены на 5 категорий: чрезвычайно опасная (1), весьма опасная (2), опасная (3), малоопасная (4) и неопасная (5).

До компьютерного моделирования типизация проводилась на основании анализа карт опасных геологических процессов. На картах были представлены карстово-суффозионные процессы, процессы подтопления и образования оползней. Данные карты накладывались друг на друга, и эксперты определяли, к какому классу в итоге относится тот или иной участок города.

Далее результаты этого моделирования прикладывались к выбранным критериям.

Обработка данных для использования машинным обучением

После решения проблемы поиска критериев, данные были готовы к обработке машинным обучением. Это означает, что все данные критериев и значений классов были объединены в один файл. Затем этот файл был изменен для применения в программном обеспечении и машинного обучения. В данной работе главным образом использовался программный продукт WEKA. Это означает, что файл должен быть преобразован в .arff формат.

В итоге, все данные представляются в формате таблицы, включающей описание всех 40000 значений. Вид таблицы представлен в таблице 5.1.2.

Таблица 5.1.2. Вид данных, для обработки машинным обучением

Высотная отметка (м)	УГВ (м)	Мощность Техногенных образований	Разломы	Линиаменты	Поверхностные воды	Уровень геологического риска
191,50	170	0,14	0	1	0	1
…		…		..	…	…
155,04	152,89	1	0	0	0	3

В итоговой таблице отсутствуют координаты точек и номер п/п, т.к. это не нужные данные для определения класса. Впоследствии, координаты подставлялись в анализируемые данные для визуализации результата.

Далее перед началом поиска алгоритма типизации и упрощения дальнейшей обработки данные были предварительно почищены. Для этого в пакете WEKA имеется набор фильтров данных, позволяющих минимизировать значения шумов, отсутствующих значений и других дефектов выборки. После компьютерной обработки выборка была ещё раз проверена вручную на предмет наличия ошибок в значениях.

После обработанной таким образом обучающей выборки, были получены подготовленные данные для анализа информации при помощи машинного обучения.

Поиск алгоритма машинного обучения

На этом этапе необходимо было найти лучший алгоритм обработки данных и его параметры. На данный момент существуют десятки алгоритмов. Из всего предоставленного множества, в данной работе использовались только некоторые алгоритмы, которые демонстрируют себя как наиболее универсальные и точные. Среди них использовались разнообразные деревья решений, метод опорных векторов, логистическая

регрессия, алгоритм Байеса, Bagging и др., описанные в главе 3.4. Поиск алгоритма осуществляется экспериментально при помощи программы WEKA. Суть этого этапа заключается в том, чтобы найти те алгоритмы, которые лучше всего справляются с проблемой определения устойчивости геологической среды и способны повторить логику эксперта, анализирующего подобные данные. Однако из-за опасности переобучения нельзя считать результат, полученный таким образом абсолютно точным.

Точность алгоритма определялась при помощи 10-кратной перекрестной проверки. Разбиение на 10 частей является оптимальным для обучения, поскольку позволяет сформировать сильно отличающие друг от друга выборки, каждая из которых при этом имела бы достаточный для анализа набор значений. При этом, так как проверка выполняется 10 раз, такой анализ не сильно перегружает компьютер, поскольку, например, при 100 кратной проверке обучение будет длиться слишком долго без увеличения его результативности.

После нахождения оптимального алгоритма или группы алгоритмов, показавших себя наилучшим образом в решении указанной проблемы, был сделан вывод, что основа модели для определения устойчивости геологической среды построена. Подготовка модели включала в себя поиск способа формирования данных для обработки, т.е. выбор и обоснование критериев, и определение алгоритма, позволяющего определять класс по этим критериям.

5.2 Методика тестирования модели устойчивости геологической среды

Тестирование полученной при обучении основы модели проводилось для того, чтобы уточнить её параметры и окончательно завершить её разработку.

Создание тестовой выборки

В качестве тестовых данных снова использовались данные указанных критериев оценки устойчивости геологической среды. Однако в данном случае выборки были меньшего размера. Были использованы данные, взятые по районам Куркино и Можаниновскому, в то время как обучающая выборка включала в себя данные, описывающую остальную территорию города Москва. Такая тренировочная выборка позволила в лучшей степени оценить универсальность модели. Данные имели тот же исходный формат, что и обучающая выборка, но не содержали информации о геологическом риске на данной территории.

Адаптация данных в соответствии с разработанным списком критериев оценки

Тестовая выборка формировалась в соответствии со списком критериев оценки, определенных на этапе обучения. При обработке машинным обучением данные были идентичны обучающим данным по структуре. Так же, соответственно, были произведены идентичные преобразования для использования в дальнейшей апробации модели.

Обработка данных для использования в машинном обучении

Данные тестовой выборки с критериями собирались в единый файл в том же порядке и в том же формате, что и при создании обучающей выборки. Файл был обработан в WEKA, после чего он был переведен в формат .arff. Данные та же были очищены от шумов и отсутствующих значений, чтобы их обработка не была более сложной, чем при обучении модели.

Данные не содержали информации об устойчивости геологической среды, однако компьютер определял его самостоятельно. Таблица имела тот же формат, что и в обучающей выборке. Вследствие этого, вместо класса ставился знак вопроса «?». Данные имели вид, представленный в таблице 5.2.1.

Таблица 5.2.1. Вид данных, для обработки машинным обучением

Высотная отметка (м)	УГВ (м)	Мощность техногенных образований (м)	Разломы	Линиаменты	Поверхностные воды	Уровень геологического риска
181	168	0,14	0	1	0	?
…		…		…	…	…
105	104	1	0	0	0	?

Проверка найденного алгоритма при помощи обработки подготовленной тестовой выборки

После формирования тестовых данных проверялся алгоритм или алгоритмы, которые были определены при обучении. При проверке обучающая выборка оставалась неизменной, но вместо перекрестной проверки использовалась независимая база данных идентичного вида. Обученная модель определяла устойчивость геологической среды на территории, где подобная оценка ещё не проводилась.

Типизация

Далее тренировочная выборка была разбита на классы, вследствие чего можно было увидеть результаты модели. Каждой точке территории был присвоен один из пяти классов, основываясь на логике определения классов в обучающей выборке. При адекватной типизации на тестовой выборке можно было говорить об актуальности модели и о том, что при помощи машинного обучения удалось повторить логику анализа устойчивости геологической среды на выбранной территории. Для оценки адекватности подобной

типизации необходимо было сравнить результаты моделирования с выбранным критерием и оценить сходство.

Визуализация результата

Для оценки качества модели удобнее всего сравнивать визуализированные результаты моделирования с визуализированными данными критерия оценки. То есть необходимо было создать карту, которая характеризует исследуемую территорию. С помощью визуализации можно было легко увидеть категорию исследуемой территории и сравнить ее с реальным геологическим риском. Для построения указанной карты в данной работе использовалось такое программное обеспечение ГИС, как Surfer11. С помощью данного программного продукта можно было преобразовывать числовые данные в картографические изображения.

6. Моделирование устойчивости геологической среды при помощи машинного обучения на примере реальной инженерно-геологической информации

Пространственно-временное развитие геологической среды происходит вследствие влияния множества процессов. Очевидно, что любой физический объект может быть описан целым набором физических характеристик.

Геологическая среда может быть охарактеризована цифровыми многоканальными космическими снимками, картами: гравитационного, магнитного, электрического полей, поверхности рельефа, изогипс литофациальных границ и пр. Каждая из этих характеристик описывает её на своём уровне. Именно анализ тех характеристик, которые максимально точно соотносятся с её устойчивостью, позволяет делать компьютерное моделирование. Это позволяет значительно приблизиться к расшифровке реальной картины её интегрального состояния и процессов её пространственно-временной эволюции.

Комплекс характеристик, которые описывают устойчивость геологической среды, и созданные на их основе модели позволяют:

- изучать влияние изменения состояния геологической среды, на зарождение и перемещение основных опасных процессов;
- исследовать влияние системных изменений на геологической среды другие компоненты окружающей среды,
- объединять все многообразие геологических и природных факторов в системы, моделировать процессы изменчивости геологической среды;
- моделировать процессы природных системных изменений;
- выделить структурные и функциональные неоднородности среды;

- получить картину динамики геологических процессов.

Масштаб исходной картографической информации определяет точность оценочных и прогнозных моделей устойчивости геологической среды. При этом комбинированная обработка и корреляция разномасштабных данных позволяет значительно повышать их точность.

Однако в представленной работе эта обработка заменена компьютерным анализом исходных данных. Это позволяет ускорить обработку информации с сохранением логики принятия решений.

Инженерно-геологическое районирование территории позволяет оценить опасные генетические изменения геологической среды. С помощью него также можно определить проблемные ситуации территорий и экономических объектов, которые оказывают большое влияние на строительство и здоровье населения.

Результаты такого рода инженерно-геологической типизации, экспертных оценок и систем поддержки принятия решений, связанных с уменьшением промышленного воздействия на окружающую среду, являются комплексом мер, которые необходимы для перспективного планирования городских и областных структур. Эти меры используются при распределении и размещении промышленных, транспортных и социальных объектов на данной территории.

В данной работе разработана аналитическая модель состояния геологической среды на базе машинного обучения. Созданная модель имеет возможность типизировать территорию, оценивать уровень опасных изменений геологической среды и распознавать проблемные ситуации и области их локализации. Модель показывает неоднородность условий формирования территории и оценивает устойчивость геологической среды.

6.1. Создание модели геологической среды на основе машинного обучения

В данной работе была создана модель определения устойчивости геологической среды. Детальное описание процесса создания модели представлено далее.

При обработке машинным обучением показано только одно из возможных применений этой модели. Алгоритмы оценивались по способности нахождения взаимосвязи между устойчивостью геологической среды и базовыми инженерно-геологическими критериями ее оценки. До сих пор подобный анализ проводился исключительно экспертами, а использование инструментов машинного обучения является лучшим способом для автоматизации этого процесса.

Построение модели производится в соответствии с методикой, изложенной в предыдущей главе.

Как было указано выше, вначале необходимо выбрать исходные данные и обработать их с учетом необходимых критериев.

В качестве обучающей выборки для создания модели были использованы данные по г. Москве, за исключением двух районов: Коркино Можаниновский. Для создания обучающей выборки были собраны и типизированы данные по каждому описанному выше критерию, которые содержали 40000 значений по каждому из них. Первая типизация была сделана экспертом. Эксперты разделили территорию на различные зоны по уровню устойчивости геологической среды. Разделение на зоны основывалось на типизации всех процессов по степени опасности и их распространенности. По степени опасности процессы изначально были разделены на три группы. К наиболее опасным относили процессы, приводящие к карстово-суффозионным провалам, глубоким оползням и мульдам оседания. К опасным относили процессы мелких оползневых подвижек и подтопления. К потенциально-опасным – ещё не проявившие

себя процессы, но которые могут активизироваться при определенных условиях. На основе анализа распространенности этих групп процессов на территории г. Москва и было произведено изначальное моделирование.

Так как карта устойчивости геологической среды являлась продуктом экспертного анализа, необходимо было совместить её данные с данными по анализируемым критериям. На основе этого совмещения и была получена выборка, показанная в таблице 5.1.2.

Каждый процесс обработки машинным обучением подразумевает предварительную подготовку данных. Это означает, что необходимо выбрать правильные критерии для модели. Критерий является частью каждого вектора значений, который использовался для определения класса. Определение наиболее правильных критериев оценки является одной из наиболее сложных задач при решении прикладных вопросов типизации при помощи машинного обучения [The Machine Learning Dictionary].

В качестве критериев были выбраны наиболее точно описывающие состояние геологической среды и позволяющие производить автоматизацию моделирования более универсально. Смысловая нагрузка каждого из критериев описана в предыдущей главе.

Ниже приведен список этих критериев оценки устойчивости геологической среды:

1. Высотная отметка (м)
2. Уровень грунтовых вод (м)
3. Мощность техногенных образований (м)
4. Наличие главных и второстепенных разломов фундамента
5. Наличие линиаментов
6. Наличие поверхностных вод
7. Уровень геологического риска

При типизации были использованы 5 категорий, определяющих устойчивость геологической среды территории: чрезвычайно опасный, весьма опасный, опасный, малоопасный, неопасный. Таким образом, при

формировании значений критериев была сформирована обучающая выборка, показанная в таблице 5.1.2, где были изъяты значения координат.

После того, как критерии были выбраны, была начата предварительная обработка данных. Первоначально, инженерно-геологические данные содержали много шумов и проблем, которые должны были быть решены до проведения анализа машинным обучением. Изначально данные состояли только из данных выбранных критериев, которые были типизированы и использованы для построения карты устойчивости геологической среды. Т.е. каждый критерий был взят с отдельной карты, которые нужно было привести к единой системе координат. Все изначальные параметры были обработаны таким образом, чтобы значение каждого из них определялось в одних и тех же точках. Кроме того, некоторые данные имели пропущенные значения. Поэтому в тех местах, где данных одного из параметров недоставало, данные других параметров были исключены. В случае незначительного объема пропущенных значений допускалась интерполяция. Также данные содержали незначительное количество шумов, которые были устранены встроенными системными фильтрами данных. В итоге это позволило на выходе получить готовую для обработки обучающую выборку.

Однако значение классов в этой выборке было представлено неравномерно, вследствие неравномерного их распределения в реальной ситуации. В таком случае обучить компьютер с одинаковой точностью распознавать все классы было сложнее, так как для каждого из них у машины было разное количество примеров. Принцип же машинного обучения с учителем заключается в обучении ее именно на примерах. И чем больше таких примеров будет в процессе обучения, тем точнее потом окажутся результаты моделирования.

Ниже на рисунке 6.1.1. приведено распределение значений в обучающей выборке по классам.

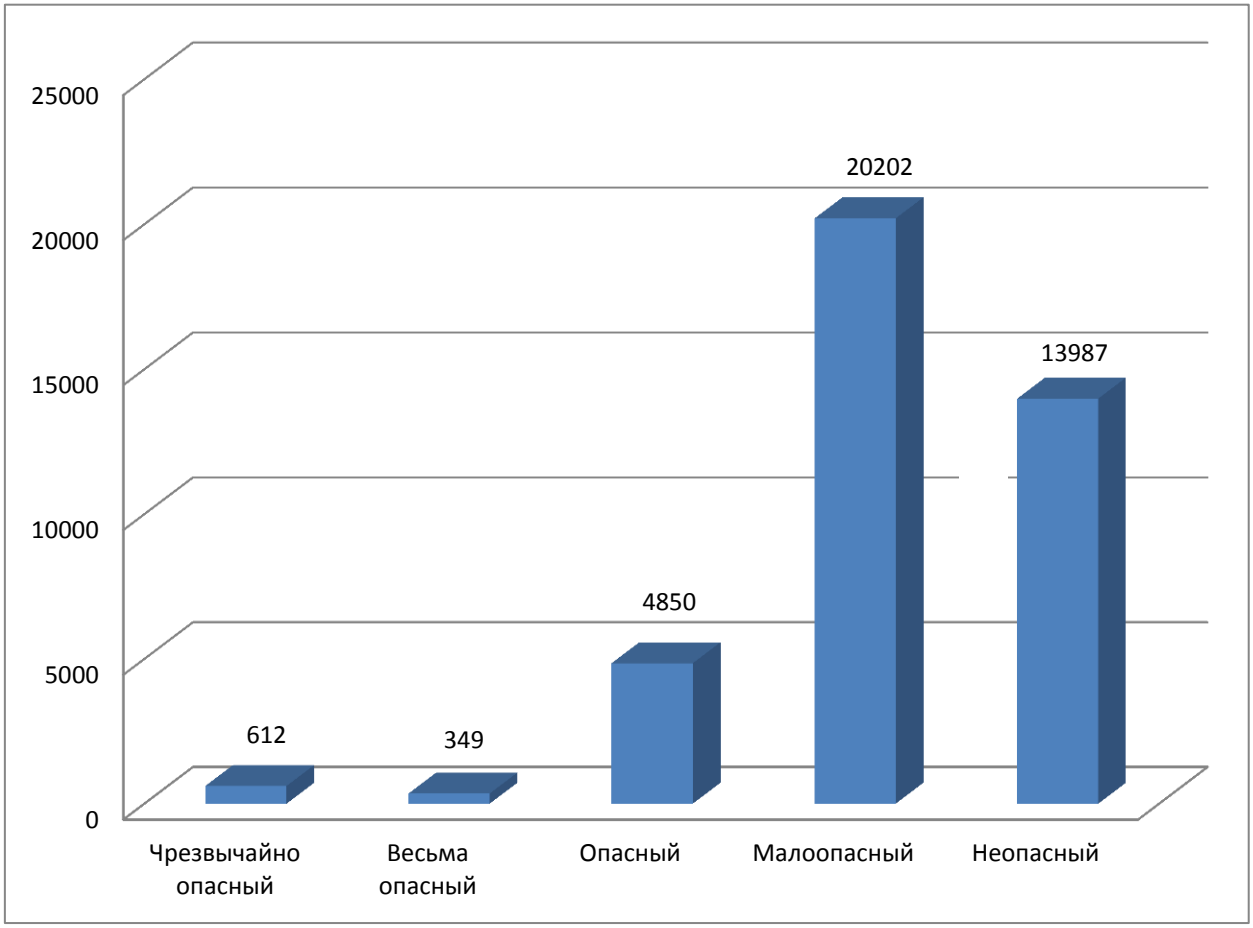


Рисунок 6.1.1. Распределение значений классов в данных с разбиением по 5 классам

При использовании 5 классов, минимальное количество точек для одного класса было 349, что является достаточным для обучения, но характеристика этого класса компьютеров будет хуже, чем других. На этой стадии предварительная обработка данных была закончена. Следующим шагом было создание модели. Для решения этой задачи необходимо было найти наиболее подходящий алгоритм.

Экспериментальные поиски алгоритмов для моделирования устойчивости геологической среды проводились при использовании программного обеспечения WEKA и See5.Для поиска алгоритма была использована перекрестная проверка с разбивкой на 10 частей.

В первую очередь необходимо определить, какая группа алгоритмов является наиболее подходящей для решения проблемы моделирования. Эта проблема была решена путем оценки точности алгоритма при обработке указанных данных. Точность алгоритма определялась исходя из рассчитанной ошибки типизации. Точность алгоритмов оценивается автоматически при помощи указанного программного обеспечения. Примеры результатов расчета точности типизации указаны на рисунке 6.1.2.

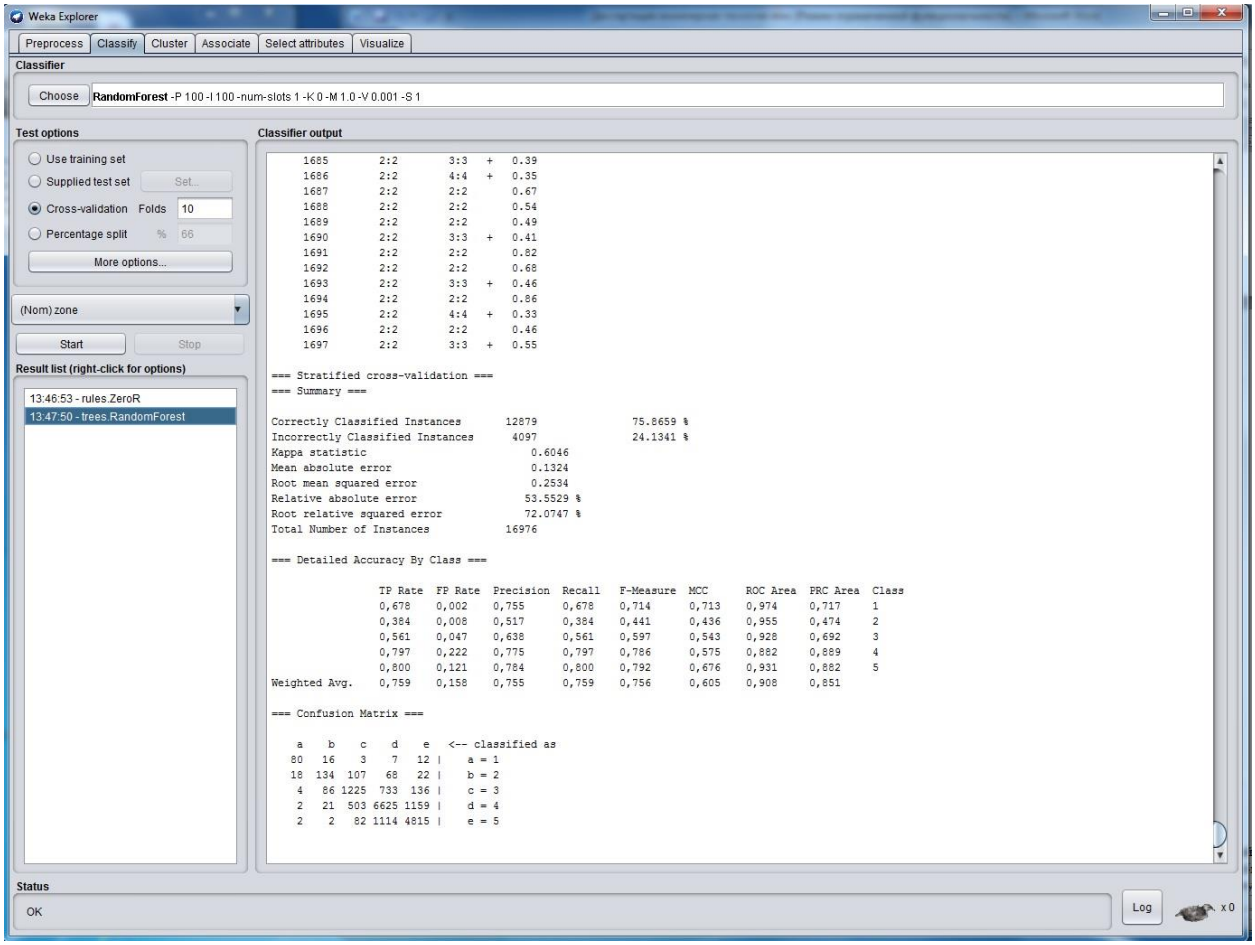


Рисунок 6.1.2. Вид результата расчета WEKA алгоритмом Random Forest.

Из расчета видно количество правильно и неправильно определенных значений и их процентное соотношение. Также программа показывает среднюю абсолютную ошибку вычислений, среднеквадратичную ошибку,

относительную ошибку и среднеквадратичную относительную ошибку. Далее следует общее описание базы данных с указанием количества значений, класса, а также распределение правильно определенных значений по классам. Например, из всех значений, которые должны были быть определены, как опасные области, 1225 были определены правильно, 733 значение было определено как малоопасные, 136 значений - как неопасные, 86 – как весьма опасные и 4 – как чрезвычайно опасные..

Алгоритмы тестировались с базовыми настройками, поскольку смыслом первой проверки было просто найти группу наиболее подходящих алгоритмов, с которыми уже можно работать более детально.

Результат первой проверки пригодности алгоритмов машинного обучения для решения проблемы определения устойчивости геологической среды указан в таблице 6.1.1.

Таблица 6.1.1. Точность алгоритмов при обработке обучающей выборки

Алгоритмы		**Точность (%)**
Trees	BF Tree	74,46
	Des Stump	49,99
	FT	74,3
	J48	74,76
	LAD Tree	68,99
	LMT	73,92
	NBTree	74,54
	Random Forest	75,87
	RandomTree	74,92
	REP Tree	74,35
	Simple Cart	74,63
	See5	75,74
Bayes	Bayes Net	67,62
	Naive Bayes	59,38
	Naive Bayes Simple	59,43
Functions	Logistic	67,88
	Multlayer Perceptron	70,6
	RBF Network	62,24
	Simple Logistic	67,86
	SMO	69,38
Lazy	IBk	74,16
Meta	Ada Boos tM1	49,99
	Attribute Selected Classifier	71,2
	Bagging	75,16
	Classification Via Clustering	18,08
	Classification Via	74,87

Алгоритмы		**Точность (%)**
	Regression	
	CV Parameter Selection	20,9
	Dagging	68,53
	Decorate	75,11
	END	75,19
	Grading	20,9
	Loget Boost	71,64
	Multi Boost AB	49,99
	Multi Class Classififier	60,03
	Multi Schrme	20,9
	Class Balanced ND	74,76
	Data Near Balanced ND	74,77
	ND	74,77
	Original Class Classassifier	74,45
	Raced Incremental Logit Boost	71,11
	Randome Committee	74,99
	Random Subspace	75,19
	Rotation Forest	74,75
	Stacking	20,9
	Stacking C	20,9
	Vote	20,9
Miscellaneous	HiperPipes	43,35
	VFI	40,78
Rules	ConjunctiveRule	49,85
	DecisionTable	71,85
	DTNB	71,98
	JRip	74,34

Алгоритмы		**Точность (%)**
	OneR	63,23
	ZeroR	20,9

Как видно из таблицы 6.1.1, машинное обучение, в общем, хорошо типизирует имеющуюся обучающуюся выборку. Многие алгоритмы давали точность более 75%, что означает, что эти алгоритмы могут создать модель, которая с достаточной достоверностью скопирует логику эксперта, а значит, сможет определить уровень устойчивости геологической среды.

Для дальнейшей работы были выбраны 15 алгоритмов обработки данных, которые показывали себя наилучшим образом при предыдущей проверке их качества. Это следующие алгоритмы:

1) J48
2) Random Forest
3) Random Tree
4) Simple Cart
5) See5
6) IBk
7) Bagging
8) Decorate
9) END
10) Classification Via Regression
11) Data Near Balanced ND
12) ND
13) Class Balanced ND
14) Random Subspace

15) Rotation Forest

Для каждого из этих алгоритмов были выбраны наилучшие параметры, подходящие конкретно для имеющейся обучающей выборки. Параметры подбирались также экспериментально и для каждого алгоритма отдельно до получения лучшего результата типизации.

В приведенной ниже таблице указаны результаты поиска лучших параметров выбранных алгоритмов.

Таблица 6.1.2. Лучшие алгоритмы и их параметры.

Алгоритмы	**Результат(%)**	**Лучшие параметры**
J48	74,74	confidenceFactor 0,01; minNumObj 2; numFolds2; seed 1
Random Forest	75,87	maxDepth 0; numFeatures 0; seed 1
Random Tree	74,9	KValue 4; maxDepth 0; minNum 1; num Folds 0; seed 5
Simple Cart	74,62	minNumObj 2; numFoldsPruning 5; seed 3; sizePer 0,9
See5	75,72	Amount of boosts 10
IBk	74,52	KNN 5; nearestNeighbourSearchAlgorithm BallTree; windowSize 0
Bagging	75,19	bagSizePercent 100; classifier Random Forest; numIterations 10; seed 1
Decorate	75,11	artificialSize 1,0; classifier J48; desiredSize 20; numIterations 10; seed 1
END	75,18	classifier J48; numIterations 10; seed 1
Classification Via Regression	74,87	Classifier Random Tree

Data Near Balanced ND	74,78	classifier Random Fores; seed 2
ND	74,78	classifier J48; seed 1
Class Balanced ND	74,76	classifier Random Fores; seed 5
Random Subspace	75,18	classifier J48; numIterations 10; seed 1; subSpaceSize 0,5.
Rotation Forest	74,82	classifier J48; maxGroup 3; minGroup 3; numIterations 10; projectionFilter PrincipalComponents; removedPercentage 50; seed 1

Параметры записаны в том виде, в котором они определяются в программном продукте WEKA для удобства настройки модели при её использовании.

Таким образом, необходимо было выбрать наилучшую модель определения устойчивости геологической среды с наиболее подходящим для этого алгоритмом.

Как видно из приведенной выше таблицы 6.1.2, лучший результат был показан алгоритмом Random Forest (точность более чем 75,87%). Это означает, что компьютерная модель может хорошо скопировать логику человека-эксперта при проведении такой типизации. В общем, деревья решений показали себя с лучшей стороны, однако проверять только деревья решений было бы не верно. Поэтому при тестировании модели оценивались 2 алгоритма: Random Forest и Bagging. Bagging был выбран вследствие того, что он показал наилучший результат из алгоритмов, не относящихся к группе деревьев решений. При описании экспериментально определенных параметров выбранных алгоритмов следует отметить, что лучше всего алгоритм Random Forest показал себя при использовании 12 деревьев. Т.е. для построения модели алгоритм создавал 12 случайных деревьев решений,

на основе которых и проводилась типизация. Увеличение или уменьшение этого числа давало худшие результаты. Алгоритм Bagging генерирует новые тренировочные выборки и анализирует их при помощи других типизаторов, как бы изменяя структуру исходных данных. В качестве лучших его параметров были определены следующие:

- Тренировочная выборка не разбивалась на части. Т.е. алгоритм использовал одну тренировочную выборку, сгенерированную на основе исходной.

- В качестве использованного при анализе данных типизатора был использован Random Forest.

Поэтому в дальнейшем речь идет о сравнении способа применения алгоритма Random Forest для наиболее точного определения устойчивости геологической среды.

Одна из главных проблем обработки данных с помощью средств машинного обучения состоит в том, что даже если компьютер проанализировал одну ситуацию, и перекрестная проверка показала хорошие результаты, в других обстоятельствах результат может быть хуже. Это происходит потому, что каждая природная ситуация уникальна, и очень трудно создать абсолютно адекватную и абсолютно универсальную модель оценки состояния природной среды. Настолько большая точность модели при обработке обучающей выборки показывает, что компьютер скопировал логику эксперта при типизации этой конкретной территории с этими конкретными условиями. Реальная значимость модели была оценена на этапе тестирования модели.

На данный момент можно было сказать, что алгоритмы деревьев решений являются лучшими для решения проблемы определения устойчивости геологической среды.

Однако, качество оценки моделей можно улучшить при помощи анализа критериев оценки устойчивости геологической среды. Для этого использовались различные методы их анализа. Механизмы анализа

критериев также включены в программное обеспечение WEKA. Результаты указаны в таблице 6.1.3. По результатам проверки можно выкинуть наименее значимые критерии, как лишние и затрудняющие проведение типизации при помощи машинного обучения.

Таблица 6.1.3. Значимость критериев оценки устойчивости геологической среды

Метод анализа	**Анализатор атрибутов**	**Порядок значимости атрибутов**
Ranker	Info Gain Attribute Eval	Уровень грунтовых вод
		Рельеф
		Мощность техногенных образований
		Наличие поверхностных вод
		Наличие линиаментов
		Наличие разломов
	One R Attribute Eval	Мощность техногенных образований
		Наличие линиаментов
		Наличие разломов
		Наличие поверхностных вод
		Уровень грунтовых вод
		Рельеф
	Symmetrical Uncert Attribute Eval	Уровень грунтовых вод
		Рельеф
		Мощность техногенных образований
		Наличие поверхностных вод
		Наличие линиаментов
		Наличие разломов

Для лучшего понимания механизма определения значимости критериев ниже приведено описание методов их классификации. Был использован метод поиска Ranker, поскольку он может легко сортировать

критерии по их значимости. Ranker сортирует критерии по их индивидуальной оценке значимости для принятия решения. В таблице 6.1.3 критерии указаны в порядке из значимости для системы с точки зрения компьютера сверху вниз в порядке от наиболее значимого к наименее значимому. В качестве алгоритмов определения значимости изначальных критериев были использованы GainAttributeEvaluator, OneRAttributeEvaluator и SymmetricalUncertAttributeEvaluator.

InfoGainAttributeEvaluator оценивает значимость критериев путем измерения прироста информации по отношению к классу.

$$InfoGain(Class, Attribute) = H(Class) - H(Class \mid Attribute)$$

OneRAttributeEvaluator оценивает значимость критериев с помощью алгоритма одного логического правила.

SymmetricalUncertAttributeEvaluator оценивает значимость критериев путем измерения симметричных неопределенностей по отношению классу.

$$SymmU(Class, Attribute) = \frac{2(H(Class) - H(Class \mid Attribure))}{H(Class)} + H(Attribute)$$

Все классификаторы, кроме OneRAttributeEvaluator, показывают одну и ту же картину. УГВ является наиболее значимым критериям; Рельеф является вторым, а затем идут мощность техногенных образований и так далее. В конце идут разломы и линиаменты, однако один их классифиакторов поставил их в начало списка по значимости, поэтому выкинуть их из модели нельзя, без ухудшения её качества.

Таким образом, подготовка обучающей выборки и обучение модели было завершено. По его итогам были точно выбраны оптимальные критерии, определены два алгоритма и их параметры, которые показали наивысшую точность моделирования. Это алгоритмы Bagging и Random Forest. Именно

эти алгоритмы далее тестировались на независимой тестовой выборке для определения достоверности модели и окончательного выбора лучшей модели определения устойчивости геологической среды. С учетом того, что в основе алгоритма Bagging также лежит алгоритм Random Forest, поскольку именно с этим алгоритмом Bagging показывал наилучшие результаты, сравнение двух алгоритмов сводится к определению лучшего способа применения алгоритма Random Forest: в «сыром» виде или в сочетании с алгоритмом Bagging.

6.2. Оценка результатов моделирования устойчивости геологической среды по выбранным критериям

В данной главе пойдет речь о проверке полученной модели на других наборах данных и окончательном определении лучшего алгоритма типизации. Для этого была создана тренировочная выборка. Были взяты данные по двум районам г. Москва: Куркино и Можаниновскому. Данные по ним были интерпретированы экспертами, однако компьютер не получил эти результаты для обработки и должен был сам определить степень устойчивости геологической среды. Этот способ позволяет проверить модель в реальной работе. Все оценки тестовых данных должны производиться только с помощью компьютера, а затем результаты могут быть проанализированы отдельно. Дело в том, что метод оценки геологической среды должен быть относительно универсальным и одинаково хорошо работать в разных областях. Для оценки точности прогнозирования были использованы результаты многолетних исследований. В результате сравнивалась предсказанная устойчивость геологической среды с реальной ситуацией с точки зрения эксперта, на основе чего и оценивалась точность модели.

Для проведения моделирования устойчивости геологической среды г. Москвы были собраны и подготовлены соответствующие данные. Исходные данные тестовой выборки, так же как и в случае с обучающей выборкой, содержали информацию по выбранным критериям оценки этой территории [Геологический Атлас Москвы], но без проведенной типизации. Данные были взяты при помощи оцифровки существующих карт, создания карт в векторном формате и последующего перевода их в матричный вид, что обеспечило требуемую точность пространственного расположения картографических показателей. Все карты были привязаны к одной координатной сетке, для правильного их наложения. Оцифровка проводилась в единой системе координат.

В итоге, после обработки карт и подготовки матриц, данные содержали 540 значений.

После подготовки данных были рассчитаны описанные ранее критерии, после чего было произведено моделирование. Обучающая выборка по данным г. Москвы была использована как основа модели, а данные по двум её северо-западным районам являлись тестовой выборкой. Т.е. так как для проведения типизации в любом случае необходима обучающая выборка была использована ранее анализируемая, но на этот раз она не подвергалась перекрестной проверке, а использовалась только для того, чтобы алгоритм «знал» на чем основывать логику типизации при проведении анализа тестовой выборки. При проведении анализа машинным обучением, обучающая выборка как бы ставит задачу, которую предстоит решать, а алгоритм является просто механизмом моделирования, который по-разному себя проявляет и по-разному строит модель при наличии различных обучающих выборок. Таким образом, совокупность обучающей выборки с правильно выбранными критериями и наиболее подходящим для данной конкретной проблемы алгоритмом с оптимальными параметрами и создают модель, которая определяет уровень устойчивости геологической среды на территории, при анализе данных с этой территории.

Для моделирования были использованы 2 алгоритма: Random forest, так как он показал наивысшую точность моделирования и Bagging, так как он показал наилучший результат из алгоритмов, не относящихся к группе деревьев решений.

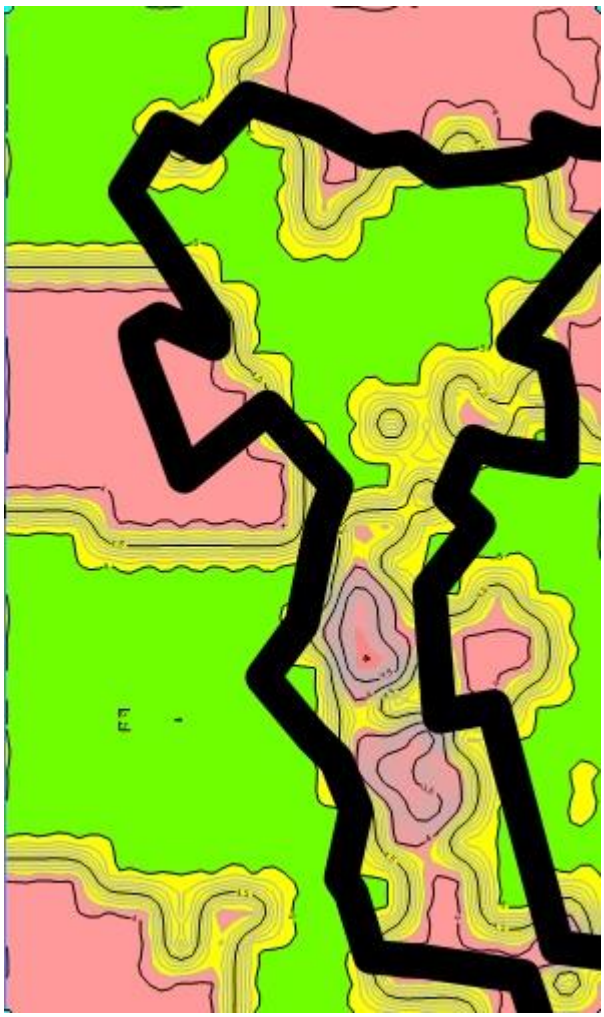
По результатам обработки тестовой выборки при помощи машинного обучения были построены таблицы на подобии таблицы 6.2.1 для каждого изучаемого алгоритма.

Таблица 6.2.1. Уровень устойчивости геологической среды г. Москвы.

№ п/п	X	Y	Уровень устойчивости геологической среды
1	811.56	2403.52	4
…	…	…	…
540	1093.47	2884.42	3

В таблице по итогам вычислений содержались условные координаты точек и значение предполагаемой устойчивости геологической среды в этих точках.

На основе этих таблиц в Surfer11 были построены карты изолиний, описывающие устойчивость геологической среды. Результаты показаны на рисунках 6.2.1 и 6.2.2.



Категории устойчивости геологической среды:

Весьма опасная

Опасная

Малоопасная

Неопасная

Рисунок 6.2.1. Результаты анализа модели устойчивости геологической среды по тестовой выборке при использовании алгоритма Random Forest.

Категории устойчивости геологической среды:

Опасная

Малоопасная

Неопасная

Рисунок 6.2.2. Результаты анализа модели устойчивости геологической среды по тестовой выборке при использовании алгоритма Bagging.

Как видно из сравнения рисунков 6.2.1 и 6.2.2, ни один из алгоритмов не обнаружил чрезвычайно опасных территорий. Алгоритм Bagging также не смог обнаружить весьма опасную территорию, что говорит о том, что в обучающей выборке было не достаточно значений этого типа для нормальной работы указанного алгоритма. Результаты данных расчетов были

сравнены с реальной ситуацией, оцененной группой экспертов (Рисунок 6.2.3).

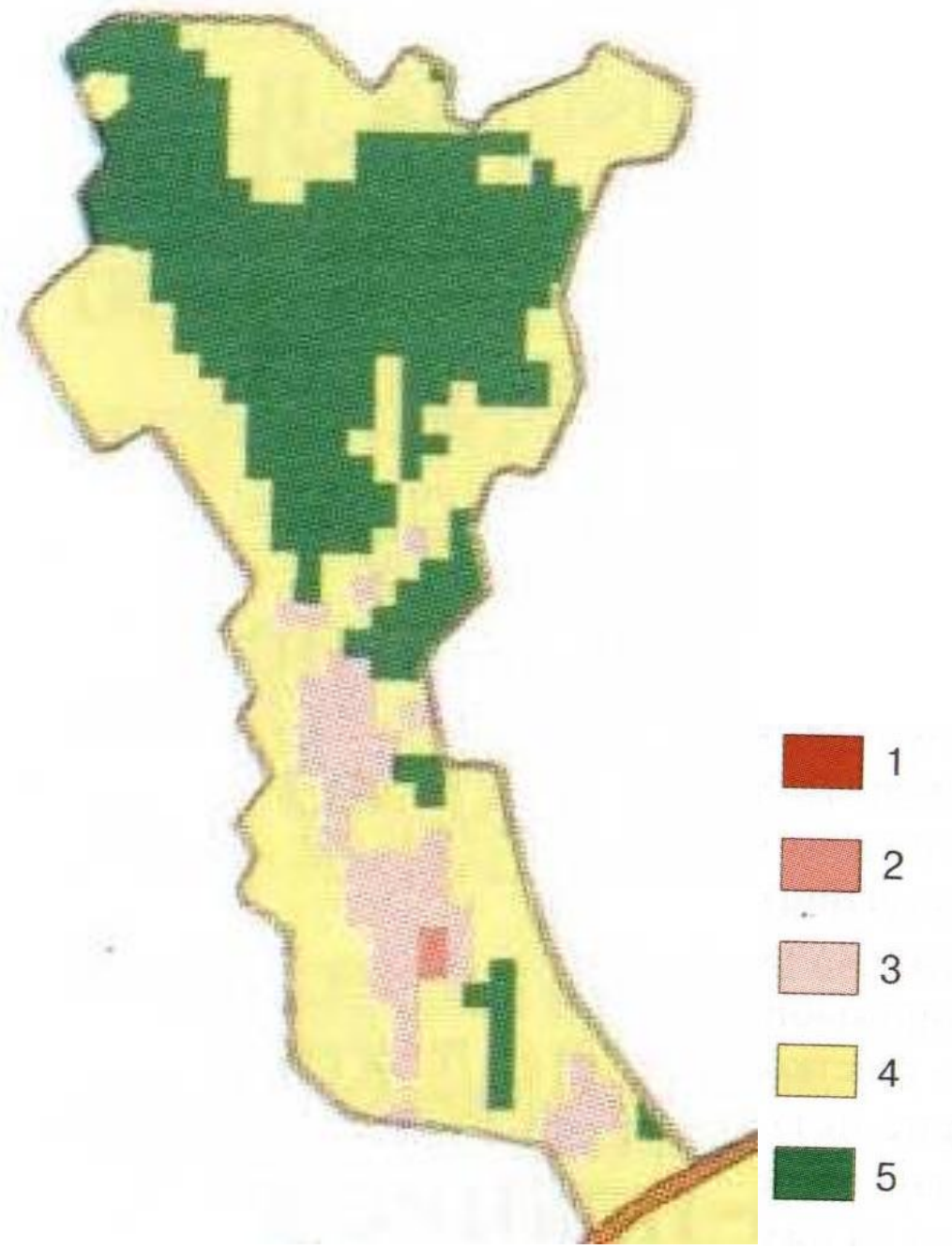


Рисунок 6.2.3. Фрагмент карты устойчивости геологической среды г. Москвы. Категории устойчивости геологической среды: 1 – чрезвычайно опасная, 2 – весьма опасная, 3 – опасная, 4 – малоопасная, 5 – неопасная [Осипов В.И. 1997].

В действительности на данной территории не было обнаружено чрезвычайно опасных территорий с точки зрения экспертов, что и показали результаты компьютерного моделирования. Однако участок с весьма опасной категорией всё-таки присутствовал. Определить его наличие смог только Random Forest, но ошибившись в точном его местоположении. Также алгоритм Bagging неправильно определил местоположение исследуемых территорий. Опасные территории распространены довольно сильно на изучаемом пространстве, но найти их в нужных местах смог только алгоритм Random Forest. Всё это говорит о том, что алгоритм Random Forest справился с реальной задачей лучше, чем Bagging и, следовательно, в наибольшей степени подходит для определения устойчивости геологической среды.

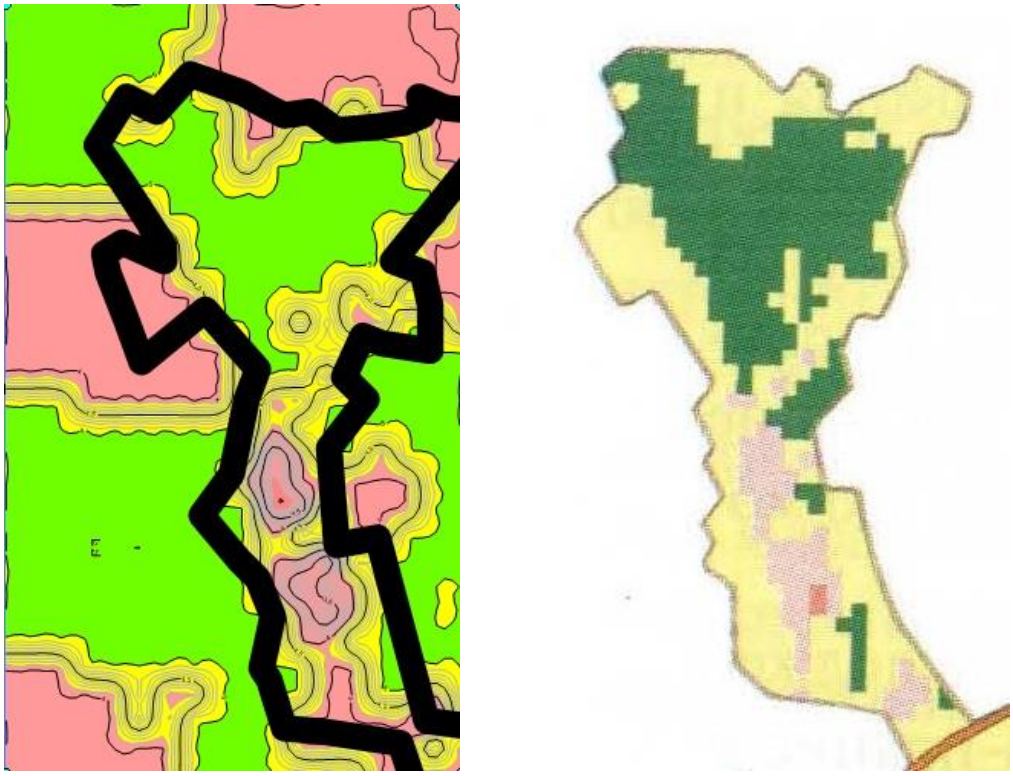


Рисунок 6.2.4. Сравнение результатов компьютерного моделирования при помощи алгоритма Random Forest и экспертного моделирования [Осипов В.И. 1997].

Как видно из рисунка 6.2.4. результаты проверки модели на тестовой выборке приблизительно соответствуют результатам перекрестной проверки, относительно алгоритма Random Forest, т.е. точность моделирования составляет приблизительно 75%. Рассчитать точную корреляцию между изображениями с неодинаковыми контурами не представляется возможным, поскольку достоверные оценки корреляции изображений доступны только при разнице контрастов. При смещении контуров известные методы расчета выдают необоснованно большие ошибки. Однако для количественного определения сходства моделей был проведен анализ процентного соотношения территорий каждой категории в первой и второй модели. В результате получилось следующее соотношение категорий устойчивости в модели, построенной компьютером:

Весьма опасная – 0,04%

Опасная – 17,57%

Малоопасная – 48,82%

Неопасная – 33,57%

В то время как в модели, построенной специалистами, соотношение было такое:

Весьма опасная – 0,28%

Опасная – 8,94%

Малоопасная – 52,97%

Неопасная – 37,81%

Суммарная ошибка компьютерной модели составляет при этом 17,26%, что даже меньше, чем при обучении. Совпадение границ изолиний при этом можно увидеть визуально.

Если учитывать результаты перекрестной проверки, анализа точности обучающей выборки, информацию, полученную из сопоставления данных результатов моделирования устойчивости геологической среды и выбранных критериев оценки в Москве, то с уверенностью можно сказать, что модель дает достоверные количественные результаты и может выполнять свою

задачу. Также из анализа следует, что алгоритм Random Forest показал себя с лучшей стороны и модель, построенная на его основе, является более точной.

Согласно проведенному тестированию модели можно полагать, что компьютерная модель хорошо себя показывает при выполнении прогнозов. Несмотря на трудную задачу, тестирование было успешным.

По итогам моделирования лучшим себя показал алгоритм Random Forest. Это не удивительно, поскольку его вложенные особенности полностью соответствуют решаемой задаче. Эти особенности следующие:

1.Он имеет хорошее качество модели и работает лучше, чем нейронные сети.

2. Он работает эффективно для больших баз данных

3. Random forest может эффективно обрабатывать данные с большим количеством критериев оценки.

4. Он содержит методы для балансировки ошибки в значениях класса и несбалансированных наборах данных

5. Он работает с той же точностью, несмотря на масштабные изменения данных.

6. Он имеет высокую параллелизуемость и масштабируемость.

7. Random Forest создает модели с большим объемом данных [Caruana. 2006].

Все характеристики соответствуют поставленным задачам и их эффективному решению. Были проанализированы большие несбалансированные наборы данных на основе критериев, позволяющими ее масштабировать и применять на ранних стадиях проектирования. Для рассматриваемого типа анализа хорошая обработка большого количества критериев оценки и классов также является необходимой особенностью. Размер модели является относительно крупным, и время обработки модели также не мало, но все преимущества алгоритма компенсировали эти недостатки. Алгоритм Random Forest был скорректирован именно для

данного вида моделирования, что было детально показано в описанных выше экспериментах.

В итоге была получена модель, позволяющая оценить устойчивость геологической среды без непосредственного анализа опасных процессов. Параметры процессов были определены компьютером за счет имеющейся информации, описывающей рельеф, гидрогеологические условия и тектонику изучаемого района. Благодаря этому проведение оценки устойчивости геологической среды и геологического риска становится намного быстрее, что позволяет специалисту в большей степени сконцентрироваться на анализе всех возможных опасностей для строительства и лишь скорректировать построенную компьютером модель.

В работе было показано, что такое моделирование может быть осуществлено компьютером. Соответственно при увеличении обучающей выборки, введении корректур в структуру модели и проведении работ по созданию сбалансированной полноценной базы данных исходных параметров моделирования, данные для которой будут взяты из зон с различными инженерно-геологическими условиями, возможно построение универсальной модели, способной выдавать прогнозы о наличии геологического риска в любых масштабах и на любых территориях с достаточно большой точностью.

6.3. Применение методики и развитие исследования

Оценка устойчивости геологической среды на основе анализа исходной инженерно-геологической информации была успешно проведена при помощи машинного обучения. Машинное обучение может во многом помочь человеку-эксперту в решении этой конкретной проблемы. Построенная модель позволила провести типизацию территории и смоделировать геологическую среду тестовых территорий. Эта модель в целом совпала с проведенными многолетними исследованиями и трудно отрицать корреляцию между этими двумя параметрами.

Полученная модель может быть использована для определения качественных рисков для проживания населения и строительства жилых кварталов. Риски, предполагаемые неустойчивым состоянием геологической среды вызваны условиями, носящими объективный характер, т.е. не зависят непосредственно от характеристик предполагаемого проекта и от другого антропогенного влияния. Модель позволяет оперативно сделать качественный анализ рисков и определить территории с их повышенным значением для организации жилых районов и других видов деятельности. Опасность проявления негативных процессов изменяется в зависимости от времени их активизации. Активизация же происходит в соответствии с природными циклами различного порядка, а также в связи с антропогенным воздействием. В свою очередь локализация этих процессов совпадает с зонами высокого геологического риска, поэтому определение устойчивости геологической среды во многом снижает риски негативных последствий для населения при активизации опасных процессов.

Вследствие того, что устойчивости геологической среды влияет и на общее проявление природных процессов, построенная модель очень полезна при выборе мест расположения крупных промышленных объектов, таких как атомные электростанции (АЭС). Неправильное базирование АЭС и неучтение опасных факторов при выборе её местоположения может иметь

очень серьёзные последствия. Определение устойчивости геологической среды поможет быстрее провести анализ территории для выбора пункта расположения АЭС при сохранении качества этого анализа. Таким образом, выбор пункта происходит быстрее и, при определенном направлении проведения изысканий, зоны, требующие более детального изучения при выборе пункта расположения АЭС, определяются ещё на начальном этапе. Далее, при выборе площадки в пределах пункта, определение устойчивости геологической среды не менее полезно для уточнения зон, требующих более внимательного изучения и уточнения данных изысканий.

Применение машинного обучения при анализе устойчивости геологической среды позволяет выполнять этот анализ быстро и с использованием меньшего количества человеческих ресурсов, что делает его прекрасным оперативным инструментом при проведении рекогносцировки территории.

В будущем это исследование может быть продолжено в разных направлениях. Дело в том, что уровень устойчивости геологической среды имеет связи со многими природными параметрами, которые могут быть рассчитаны и спрогнозированы.

Новые точки зрения на комплексную оценку инженерно-геологических факторов позволяют понять и описать закономерности в структуре и состоянии компонентов геологических систем. Использование координатной сетки в области анализа естественных и техногенных факторов позволяет создавать многофакторные инженерно-геологические модели территории. Также модель помогает оценить риски хозяйственного использования территории, качество территории для жизни. Машинное обучение является хорошей частью интеллектуального анализа модели и системы поддержки принятия решений.

Модель может специализироваться на природных процессах. Это означает, что могут быть построены сложные модели геологической системы. Эти модели основаны на геофизических, геохимические,

геологических, гидрогеологических данных и данных о поверхности земли. Они позволяют:

1.получить и представить реальную информацию о напряженном состояние земной поверхности,

2. типизировать районы с опасными экзогенными процессами (оползни, эрозия и др.) и прогнозировать динамику их развития.

3. показать природные зоны с высоким уровнем риска промышленной аварии.

Все описанные способы исследования развития являются очень перспективным и полезным. Это становится особенно актуальным, если учесть, что все модели в потенциале универсальны. Это означает, что с одной обучающей выборкой можно анализировать любые территории в любом масштабе. С одним хорошо подготовленным набором данных и одной хорошо рассчитанной моделью можно решить множество разных задач без внесения каких-либо изменений.

Заключение

Методика определения устойчивости геологической среды путем анализа исходных инженерно-геологических данных была оптимизирована при помощи средств машинного обучения.

Для этого использовались различные методики их предварительной обработки, выбор необходимых критериев оценки и анализ группы алгоритмов машинного обучения.

Проведенное исследование дает возможность построить пространственную модель устойчивости геологической среды на территории любого масштаба.

В наши дни существует множество различных методов прогноза. Многие из них хорошо работают в конкретной ситуации. Однако, когда речь идет о большом объеме данных и сложной типизации, лучшим решением для

оптимизации моделирования является использование машинного обучения. Ранее применялось много различных методов машинного обучения в инженерно-геологическом моделировании, но определение устойчивости геологической среды такими методами не осуществлялось.

Была разработана методика применения машинного обучения для решения проблемы определения устойчивости геологической среды.

Целью работы была разработка методики оптимизации решения задачи определения устойчивости геологической среды при помощи машинного обучения.

Для ее осуществления было выполнено:

- Обоснование применения машинного обучения для оптимизации анализа устойчивости геологической среды

- Разработка методики определения устойчивости геологической среды при помощи машинного обучения;

- Выбор базовых параметров, характеризующих устойчивость геологической среды;

- Создание представительной базы данных, характеризующей устойчивость геологической среды посредствам выбранных базовых параметров.

- Адаптация исходных данных, для дальнейшей их обработки при помощи машинного обучения;

- Определение устойчивости геологической среды реальной территории при помощи машинного обучения и сравнение результатов с данными экспертного анализа;

- Интерпретация результатов и обоснование применимости машинного обучения для определения устойчивости геологической среды на предпроектной стадии строительства.

На основе проделанных исследований сделаны выводы и выработаны рекомендации по практической реализации полученных результатов:

1. Разработанная методика определения устойчивости геологической среды на основе методов машинного обучения действительно позволяет осуществлять данный анализ на предпроектной стадии с достаточной для оптимизации инженерных изысканий точностью.

2. Метод машинного обучения наилучшим способом подходит для оптимизации анализа устойчивости геологической среды. Были описаны различные методики моделирования с указанием положительных и отрицательных сторон каждой, и на основе их анализа было установлено, что метод машинного обучения является наилучшим для решения указанной задачи.

3. Машинное обучение можно рассматривать, как один из основных методов математического моделирования и прогнозирования устойчивости геологической среды, который позволяет обрабатывать большие объемы разнородной и сложной по составу инженерно-геологической информации в режиме реального времени с выдачей оперативных рекомендаций по принятию проектно-технических решений на ранних стадиях функционального освоения территории и по выбору альтернативного местоположения объекта строительства.

4. В работе проведен анализ предыдущих исследований, который выявил, что определение устойчивости геологической среды ранее не было подходящим образом оптимизировано, и при этом подобный анализ является очень полезным и перспективным направлением.

5. Была построена модель, позволяющая определять устойчивости геологической среды на различных территориях, на основе одной обучающей выборки.

Метод объединяет способы предварительной обработки данных для решения этой проблемы и использования конкретных алгоритмов для лучшего моделирования процесса. Однако точность модели при итоговом тестировании отличалась от 100%. Это произошло, вероятно, из-за несовершенства обучающей выборки и малого ее размера. Полученная

точность моделирования, составляющая примерно 75% при тестировании показывает, что эта модель работает, и все критерии оценки и алгоритм были выбраны правильно. Но анализ точности модели при оценке отдельных категорий показал, что проблема была в несбалансированной обучающей выборке. Кроме того, модель основана на данных из одной локальной области. Когда речь идет об инженерно-геологическом моделировании, лучше собирать информацию из различных стандартных объектов со всего мира, что не представлялось возможным в рамках данного исследования.

При условии создания большой сбалансированной выборки, которая содержит информацию о многих различных областях, точность модели может приблизиться к 99%.

Также была проведена и продемонстрирована непосредственно оптимизация. При использовании машинного обучения, оценка производится гораздо быстрее и проще. Оптимизация решения указанной проблемы сохранила результативность выходной информации, сделала её решение более быстрым и позволила проводить оценку устойчивости геологической среды более унифицировано.

Выбранные критерии оценки и способы их получения являются основополагающими для дублирования логики эксперта при помощи машинного обучения при решении задачи определения устойчивости геологической среды;

4. Была разработана система адаптации инженерно-геологических данных для определения устойчивости геологической среды методами машинного обучения. При помощи обработки данных был получен необходимый набор параметров для построения адекватной модели, определения устойчивости геологической среды.

5. Найден наилучший алгоритм обработки данных и наилучшие его параметры для компьютерного определения устойчивости геологической среды, что позволяет оптимизировать сложный процесс экспертной типизации территории. Этим алгоритмом является Random Forest.

6. Был создан механизм, частично дублирующий логику эксперта при решении проблемы типизации территории по устойчивости геологической среды, что позволяет решать подобные задачи быстрее и применять решения подобных проблем более широко.

7. Построенная система, позволяет определять устойчивость геологической среды на различных территориях, сходных по инженерно-геологическим условиям с г. Москва. Однако, при расширении обучающей выборки, география потенциального применения модели может быть также расширена.

8. Направление дальнейших исследований связано с расширением базы данных по другим территориям с различными базовыми условиями и устойчивостью геологической среды для улучшения точности дальнейшего моделирования. Этот вид моделирования очень полезен не только для инженерно-геологических исследований. Он может успешно применяться: для планирования новых городов и для выбора места расположения жилых зданий и промышленных объектов; для оценки качества территории; для принятия рациональных решений об экономическом и промышленном использовании территории.

Литература

1) Азроянц Э.А.; Харитонов А.С.; Яницкий И.Н. Способ обнаружения возможности наступления катастрофических явлений. Патент Российской Федерации № 2030769. 1995 г.

2) Балдын Бат. Оценка и прогнозирование устойчивости геологической среды к природным и техногенным воздействиям территории города Улан-Батор. Диссертации о Земле. Иркутск, 2004. URL: http://earthpapers.net/otsenka-i-prognozirovanie-ustoychivosti-geologicheskoy-sredy-k-prirodnym-i-tehnogennym-vozdeystviyam-territorii-goroda-ul#ixzz4uWL64Ge3 (дата обращения15.05.2017)

3) Безноздрева Е.А., Воробьев Д.С., Емельянова Л.Г., Землянова О.И., Каргашина М.А., Касимов Д.В., Кривонос Е.В., Михеева А.И., Пинаев В.Е., Протопопов Н.Ф., Резников И.С., Рукавицын В.В., Рыбкина Г.И., Скворцова Е.А., Сосновских С.Л., Сысоева И.А., Хуршудов А.Г., Шахин Д.А., Якунин С.А. Сборник инновационных решений по сохранению биоразнообразия для нефтедобывающего сектора. РА ИЛЬФ. Москва, 2015. 274 с.

4) Болтунов В.А., Дудлер И.В., Болтунов В.В.. Геологические и инженерно-геологические аспекты геопатогенных зон. Мелиорация и водное хозяйство, № 2, 1999. С.24-27.

5) Бахирева Л.В., Кофф Г.Л., Мамонтова С.А. Оценка геологического и геохомического риска в схемах охраны геологической среды культурно-исторических зон (на примере Москвского региона). Инж.геол.№6. 1989. С.36-47.

6) Белозеров Б. В., А.С. Бочков, О.С. Ушмаев , О.М. Фукс. Использование метода ближайших соседей при восстановлении обстановки осадконакопления. Машинное обучение и анализ данных Том 1, номер 9. ISSN 2223-3792. 2014. С 1319-1329

7) Бондаренко Я.И. Составление космофотокарт и многофакторное прогнозирование риска детской заболеваемости в г. Днепропетровске. Отчет в двух книгах (договор 58-96 от 8.07.1996г.), г. Днепропетровск. 2000.

8) Бондаренко Я.Я, Зайонц И.О. Картография и мультифакторный прогноз риска детской заболеваемости в г. Днепропетровске. Днепропетровск. 2000.

9) Бондаренко Я.И., Зайонц И.О. Создание мультифакторных моделей экономического использования территории и риска проживания населения на примере г. Киева. Сборник конференции "проблемы экологического менеджмента, устойчивого развития и технологической безопасности. Днепропетровск. 2001.

10) Бондарик Г.К. Методика инженерно-геологических исследований. Недра. 1986. 333 с.

11) Бондарик Г.К. Теория геологического поля (филосовские и методологические основы геологии). КДУ. 2009. 128 с.

12) Бондарик Г.К. Системный подход при инженерно-геологических прогнозах. Доклад 27-го Международного геол. конгресса. Инж.-геол. секц. С-17, т.17. Москва. Наука. 1984.

13) Вагнер Б.Б., Б.О. Манучарянц. Геология, рельеф и полезные ископаемые московского региона. учебное пособие по курсу «География и экология Московского региона». Москва. 2003. 81 с.

14) Геологический Атлас Москвы. URL:http://www.georesurs.su/Georesurs/AtlasMos (дата обращения 04.03.2017)

15) Герасимова А.С., Королев В.А. Проблемы устойчивости геологической среды к техногенным воздействиям. М., АО «Геоинформмарк». 1994. 47 с.

16) Гридин В.И., Гак Е.З. Физико-геологическое моделирование природных явлений. – М.: Наука, том 1. 1994. 204с.

17) Демьянов Г.В., Назарова Н.Г., Никольский Н.Г., Никольский Ю.И., Таранова В.А. Карта аномалий гравитационного поля России и прибрежных территорий. Масштаб 1:10 000 000. Рудгеофизика. 1995.

18) Дзекцер Е.С. Гашвина З.П., Казенков С.М. Вопросы методики масштабного районирования застраиваемых территорий по степени потенциальной подтопляемости грунтовыми водами. Инж.геол №5. 1984. С.96-104.

19) Зайонц И.О. Бондаренко Я.И. Виділення територій з різним рівнем природно-техногенного ризику господарчої діяльності та проживання населення та експертна оцінка сучасного стану об'єктів навколишнього природного середовища м. Києва. Звіт, (договор № 51/02, от30.09.2002 р.), г. Киев. 2003.

20) Зайонц И.О., БондаренкоЯ.И, СлипсенкоБ.В., ЛысиченкаГ.В. New approaches to the problem of geoecological risk for urbanized territories. ECO-INFORMA 2001. Chicago, USA. 2001.

21) Иванов И.П., Тржцинский Ю.Б. Инженерная геодинамика. СПб.: Наука. 2001. 416 с.

22) Иголкина Н.С., Кириков В.П., Кривская Т.Д.// Основные этапы формирования осадочного покрова Русской платформы // Советская геология. 1970. № 11. С. 15–35

23) Инамов А.Н. Закономерности изменений инженерно-геологических условий орошаемых территорий (на примере Каршинской степи). Проблемы инженерной геолоии, гидрогеологии и геокриологии. Тез. докл. 1-ой Всесоюзн. Съезда инж.-геол, гидрогеологов, геокриологов. Ч.3. Киев, 1989. С.98-100.

24) Казначеев В.П. Очерки теории и практики экологии человека. Москва. Наука. 1981. 260 с.

25) Казначеев П.Ф., Самойлова Р.Г., Курчиски Н.В. Применение методов искусственного интеллекта для повышения эффективности в нефтегазовой и других сырьевых отраслях // Экономическая политика. 2016.

№5. URL: https://cyberleninka.ru/article/n/primenenie-metodov-iskusstvennogo-intellekta-dlya-povysheniya-effektivnosti-v-neftegazovoy-i-drugih-syrievyh-otraslyah (дата обращения: 21.12.2017)

26) Карты Москвы. URL:http://www.etomesto.ru (дата обращения 09.04.2017)

27) Киссель М. А. Философская эволюция Ж.-П.Сартра, Лениздат. 1976. 238 с.

28) Кобринский Б.А. К вопросу учета образного мышления и интуиции в экспертных медицинских системах. V Нац. конф. с межд. уч. "Искусственный интеллект-96": Сб. науч.тр. - Казань. - Т.2. 1996. С. 207-210.

29) Котлов В.Ф. Геопатогенные зоны и их роль в формировании чрезвычайных ситуаций. В кн.: Лобацкая Р.М., Кофф Г.Л. Разломы литосферы и чрезвычайные ситуации. – М., РЭФИЛ, гл.4, 1997. С.84-104.

30) Котлов В.Ф., Кофф Г.Л. Методические аспекты оценки состояния геологической среды. Инж.геол №1. 1987. С.60-68.

31) Кофф Г. Л., Медведев О. П., Зиангиров Р. С. и др. Опыт оценки уязвимости и риска ущерба для зданий и сооружений жилищно-гражданского назначения в связи с воздействием опасных инженерно-геологических процессов (на примере Москвы) // Прикладная геоэкология, чрезвычайные ситуации, земельный кадастр и мониторинг: Сб. тр. (выпуск 3). М., 1999. С. 64–81.

32) Критерии оценки экологической обстановки территорий для выявления зон чрезвычайной экологической ситуации и зон экологического бедствия. Утверждено Министерством природных ресурсов Российской Федерации 30 ноября 1992 года. Москва. 1992. URL: http://docs.cntd.ru/document/901797511 (дата обращения 15.12.2017)

33) Лагутина Н.П., Остовская О.Е. Карта рельефа России. Масштаб 1:10 000 000. «Картографическая фабрика Омска». 2005.

34) Лисенков А.Б., Попов Е.В. Методика решения эколого-гидрогеологических задач при помощи аппарата теории информации. Учебное пособие. МГГРУ. 2003. 23 с.

35) Лисенков А.Б., Фисун Н.В., Малков А.В., Королев И.Б., Иванов А.А. под ред. Гавич И.К. Техногенные процессы в подземных водах. Научный мир. 2003. 216-244 с.

36) Литвинова Т.П., Шмиярова Н.П. Карта аномалий магнитного поля России и прибрежных территорий. Масштаб 1:10 000 000. ВСЕГЕИ. 1995.

37) Ленихен Дж., Флетчер У. (Ред.). Здоровье и окружающая среда. Москва. Мир. 1979. 231 с.

38) Лобацкая Р.М., Кофф Г.Л. Разломы литосферы и чрезвычайные ситуации. Москва. РЭФИА. 1997. 196 с.

39) Мавлянов Э. В., Инамов А. Н. Крупномасштабная карта изменения геологической среды под воздействием мелиоративного строительства. Изменение гидрогеологических и инженерно-геологических условий под влиянием водохозяйственной деятельности. Ташкент: Фан, 1986. С. 5–10.

40) Мешалкин В.П., Бутусов О.Б., Гнаук А., Панарин В.М. Основы информатизации и математического моделирования экологических систем. Учебное пособие. Интернет-институт ТулГУ. 2009. 358 с.

41) Мельников Е.К., Рудник В.А. Геоактивные зоны и их влияние на здоровье человека. Разведка и охрана недр, № 7-8. 1998. С. 54 - 57.;

42) Мельников Е.К., Рудник В.А., Мусийчук Ю.И., Рымарёв В.И. Патогенное воздействие зон активных разломов земной коры Санкт - Петербургского региона. Геоэкология, № 4. 1994. С. 50-69.

43) Мельников Е.К., Скакун А.П. К вопросу о методике выделения геопатогенных зон. Проблемы геодинамической безопасности (доп. сб). СПБ. МНЦ ВИМИ. 1997. с.282-289.

44) Минаев Ю.Н., Тонкаль С.С., Филимонова О.Ю., Куклин В.Ю., Мруе М.Л. Экологический мониторинг на предприятиях нефтеперерабатывающей промышленности на основе методов и моделей искусственного интеллекта. Программные продукты и системы №1. 1996. URL: http://www.swsys.ru/index.php?page=article&id=1064 (дата обращения 16.12.2017)

45) Михаленко Ю.П. Античные учения об индукции и их современные интерпретации // Зарубежное философское антиковедение. Критический анализ. М., 1990. С.58-75

46) Назарова О.В. Методика оценки устойчивости геологической среды урбанизированных территорий в сейсмоактивных областях. Диссертации о Земле. Москва, 2001 URL: http://earthpapers.net/metodika-otsenki-ustoychivosti-geologicheskoy-sredy-urbanizirovannyh-territoriy-v-seysmoaktivnyh-oblastyah#ixzz59o2O9jWs (дата обращения12.04.2017)

47) Осипов В.И. Зоны геологического риска на территории Москвы. Вестник РАН. , т.64, № 1. 1994. С.33-45

48) Осипов В.И., Медведева О.П. ред. Москва. Геология и город. Москва. Московские учебники и Картолитография. 1997. 398 с.

49) Паклин Н. Б., Орешков В. И. Бизнес-аналитика: от данных к знаниям (+ CD). — СПб.: Изд. Питер. 2009. 624 с.

50) Пендин В.В. Комплексный количественный анализ информации в инженерной геологии. КДУ. 2009. 350 с.

51) Петренко А.С. Анализ и оценка риска ущерба от последствий опасных геологических процессов на территории крупного города (на примере г.Москвы). Диссертация на соискание ученой степени кандидата геолого-минералогических наук. Российская академия наук институт литосферы окраинных и внутренних морей. Москва 2001. 171 с.

52) Потылицына Е.Н., Липинский Л.В., Сугак Е.В. Использование искусственных нейронных сетей для решения прикладных экологических задач // Современные проблемы науки и образования. – 2013. – № 4. URL: http://www.science-education.ru/ru/article/view?id=9779 (дата обращения: 15.03.2017).

53) Рукавицын В.В. Анализ состояния экосистем крупных городов методами машинного обучения. Геоэкология, инженерная геология, гидрогеология, геокриология №1. «Наука». Москва. 2018. С 79-88.

54) Рукавицын В.В. Машинное обучение в экологическом прогнозировании. Номер: XI год: наука о земле на современном этапе издательство: ООО "Издательство "Спутник+" .Москва. 2014. С 62-64.

55) Рукавицын В.В.; Применение искусственного интеллекта для оценки степени риска проживания населения. Сборник трудов Международной научно-практической конференции «Развитие экологической экономики и образования в Украине и в мире». 25 ноября 2011 г. г. Киев, Украина. С 54-56

56) Рукавицын В.В., Экзарьян В.Н. Оценка устойчивости экосистем россии основная экологическая характеристика здоровья населения. Здоровье - основа человеческого потенциала: проблемы и пути их решения. 2014. Т. 9. № 2. Федеральное государственное автономное образовательное учреждение высшего образования "Санкт-Петербургский политехнический университет Петра Великого" .Санкт-Петербург. С. 827-829.

57) Серкова В.И. Использования НС для решения задач классификации и прогноза геологической интерпретации геофизических данных. Сборник научных трудов [Текст]: материалы научно-технической конференции (14-17 апреля 2009 г.): в 2 ч.; ч. II / под ред. Н. Д. Цхадая. – Ухта: УГТУ – 2009. 421 с.

58) СНиП 2.02.01-83. Основания зданий и сооружений. Москва. ФГУП ЦПП. 2006. 48 с.

59) СНиП 2.06.15-85. Инженерная защита территории от затопления и подтопления. Москва. ГУП ЦПП, 2001

60) СНиП 2.07.01-89 Градостроительство. Планировка и застройка городских и сельских поселений. Минрегион России. Москва. ОАО "ЦПП". 2011

61) СП 116.13330.2012 Инженерная защита территорий, зданий и сооружений от опасных геологических процессов. Основные положения. Актуализированная редакция СНиП 22-02-2003. Москва. Минрегион России. 2012

62) СП 14.13330.2014. Строительство в сейсмических районах. Москва. Минстрой России. 2016.

63) Степановских А.С. Экология. ЮНИТИ-ДАНА. 2001. 703 с.

64) Строкова Л.А. Использование алгоритмов Data Mining для решения прогнозных задач при строительстве метрополитена // Известия ТПУ. 2009. №1. URL: https://cyberleninka.ru/article/n/ispolzovanie-algoritmov-data-mining-dlya-resheniya-prognoznyh-zadach-pri-stroitelstve-metropolitena (дата обращения: 19.03.2017).

65) Строкова Л.А. Прогноз оседания земной поверхности при строительстве городских туннелей // Вестник МГСУ. 2009. №4. URL: https://cyberleninka.ru/article/n/prognoz-osedaniya-zemnoy-poverhnosti-pri-stroitelstve-gorodskih-tunneley-1 (дата обращения: 21.12.2017)

66) Татаринов В.Н. Геологическая среда как активная иерархически устроенная термодинамическая система с внутренними источниками энергии. Москва, 2000. URL: http://www.wdcb.ru/mining/geolog.html#chaper_2 (дата обращения 02.12.2017)

67) Трофимов В. Т.. Инженерно-геологическое прогнозирование. 2015. URL: https://bigenc.ru/geology/text/2011706 (дата обращения 04.07.2017)

68) Федеральный закон "Об охране окружающей среды" от 10.01.2002 N 7-ФЗ. СПС КонсультантПлюс

69) Фомин, С. В., Беркинблит, М.Б. Математические проблемы в биологии. Москва. Наука, Гл. ред. физ-мат литературы. 1973. 200 с.

70) Фролова В. И. под ред. Методы обоснования программ устойчивого развития сельских территорий: моногр. /; СПб. гос. архит.-строит. ун-т. СПб. 2011. 464 с.

71) Хай Г.А. Как избежать потерь смысла на пути от содержательного знания к автоматизированной экспертной системе. Вопросы применения экспертных систем. Минск. НПО "Центрсистем". 1988. С.53-61

72) Цуриков К.О. Некоторые принципы оценки величины техногенного воздействия на геологическую среду Крымской области. Инж.геол. №3. 1991. С.52-58.

73) Чульский Л.А. Устойчивость в природе. Москва. Знание, 1965. 94 с.

74) Шейдеггер А.Е. Физические аспекты природных катастроф. Москва. Недра. 1981. 232 с.

75) Шестопалов В.М., Бондаренко Я.И., Зайонц И.О., Руденко Ю.Ф., Бохуславский А.С. Complexation of Structural-Geodynamical and Hydrogeological Methods of Studying Areas to Reveal Geological Structural Perspectives for Deep Isolation of Radioactive Wastes. Field Testing and Associated Modeling of Potential High-Level Nuclear Waste Geologic Disposal Sites, Berkeley, USA. 1997.

76) Шеко А.И., Круподерова В.С. Методы долговременных региональных прогнозов экзогенных геологических процессов. ВСЕГИНГЕО. Изд. Недра. 1984. 167 с.

77) Экзарьян В.Н. Геоэкология и охрана окружающей среды. Москва. «Щит-М», 2009. 208 с.

78) Экзарьян В.Н., Рукавицын В.В. Методика оценки уровня устойчивости экосистем при помощи машинного обучения. Известия ВУЗов «Геология и разведка», №1, М., 2015, с.38-42

79) Экзарьян В.Н., Рукавицын В.В. Разработка критериев оценки сохранности окружающей среды при нефтедобыче. Известия высших учебных заведений. Геология и разведка, №4, Москва. 2016. С.70-74.

80) Экзарьян В.Н., Рукавицын В.В., Зюляева М.В. Автоматизация комплексного экологического мониторинга территории аэропорта с использованием метода машинного обучения. Известия высших учебных заведений. Геология и разведка. № 4. Москва. 2017. С. 72-78.

81) Эшби У.Р. Введение в кибернетику. Москва. Издательство иностранной литературы, под редакцией В. А. Успенского. 1959. URL: http://www.booksshare.net/index.php?id1=4&category=cybern&author=ashbiwr&book=1959 (дата обращения 01.12.2017)

82) Aha D., Kibler D. Instance-based learning algorithms. Machine Learning.6. 1991. p 37-66.

83) Bayes, Thomas, and Price, Richard An Essay towards solving a Problem in the Doctrine of Chance. By the late Rev. Mr. Bayes, communicated by Mr. Price, in a letter to John Canton, M. A. and F. R. S. Philosophical Transactions of the Royal Society of London 53: 1963 . 370—418 pp.

84) Bondarenko Y.I. The Multifactor PredictiveSeis (GIS) Model of Ecological, Genetical and Population Health Risk in Connection with Dangerous Bio-Geodynamical Process in Geopatogenic Hazard Zones. ECO-INFORMA 2001, Chicago, USA. 2001.

85) Bramer M. Principles of Data Mining. Springer, 2007. 344 p. DOI:10.1007/978-1-84628-766-4.

86) Breiman Leo. Bagging predictors. Machine Learning. 24. 1996. 123-140 pp.

87) Breiman Leo. Random Forests. Machine Learning.45 (1). 2001. 5-32 pp.

88) Buettner Mason. SGS Geostat wins Integra Gold Corp’s Gold Rush Challenge. 2015. URL:

http://www.miningandexploration.ca/exploration/article/sgs_geostat_wins_integra_gold_corps_gold_rush_challenge/ (дата обращения13.08.2017)

89) Caruana, Rich, Niculescu, Alex, Crew, Geoff, and Ksikes, Alex. Ensemble Selection from Libraries of Models. The International Conference on Machine Learning (ICML'04). 2004. p. 18

90) De'ath G., Fabricius K. E. Classification and Regression Trees: A Powerful Yet Simple Technique for Ecological Data Analysis. Ecology. 2000. p 3178-3192.

91) Delphi method. Principia cybernetica web URL= http://cleamc11.vub.ac.be/ASC/DELPHI_METHO.html (дата обращения 15.04.2017)

92) Dymitr Ruta, Bogdan Gabrys. A Framework for Machine Learning based on Dynamic Physical Fields. Natural Computing: an international journal. Volume 8 Issue 2. 2009. 219 p.

93) Dzeroski Sas. Aclications of symbolic machine learning to ecological modeling. Ecological Modelling, Volume 146, Issues 1-3. 2001. p. 263-273.

94) Forecasting methods. URL: http://www2.uiah.fi/projekti/metodi/190.htm#analenn (дата обращения 13.04.2016)

95) Hsieh W. William. Machine Learning Methods in the Environmental Sciences. Nueral Networks and Kernels. Cambridge University Press. New Deihi. 2009. p. 314-317.

96) Horváth, Tamás; Yamamoto, Akihiro, eds. Inductive Logic Programming. Lecture Notes in Computer Science. 2835. DOI:10.1007/b13700.

97) Iba Wayne, Langley Pat (1992). Induction of One-Level Decision Trees. ML92. San Francisco, CA: Morgan Kaufmann. 2003. p. 233–240

98) Is See5/C5.0 Better Than C4.5? Rulequest research 2017. URL: http://www.rulequest.com/see5-comparison.html (дата обращения 13.12.2017)

99) Jeraj Marjeta, Dzeroski Saso, Todorovski Ljupco and Debeljak Marko. Aclication of machine learning methods to palaeoecological data. Ecological Modelling, Volume 191, Issue 1. 2005. p. 159-169.

100) Kanevski M., Parkin R., Pozdnukhov A., Timonin V., Maignan M., Demyanov V.and Canu S. Environmental data mining and modeling based on machine learning algorithms and geostatistics.Environmental Modelling & Software, Volume 19, Issue 9. 2004. p. 845-855.

101) Krasnopolsky Vladimir M. and Fox-Rabinovitz Michael S. A new synergetic paradigm in environmental numerical modeling: Hybrid models combining deterministic and machine learning components.Ecological Modelling, Volume 191. 2006. p. 5-18.

102) Matthew J.Cracknell, Anya M.Reading. Geological mapping using remote sensing data: A comparison of five machine learning algorithms, their response to variations in the spatial distribution of training data and the use of explicit spatial information Computers & Geosciences Volume 63, February. 2014. p. 22-33

103) Mitchell T. Machine Learning. McGraw Hill. 1997. p. 2

104) Petkos Georgios. Aclying machine learning techniques to ecological data. School of Informatics University of Edinburgh. 2003. 81 p.

105) Rodriguez J., L. Kuncheva and C. Alonso, "Rotation Forest: A New Classifier Ensemble Method," in IEEE Transactions on Pattern Analysis & Machine Intelligence, vol. 28, no. 10, pp. 1619-1630, 2009. DOI:10.1109/TPAMI.2006.211

106) Rukavitsyn V., Zizka J.; Modern Data-Mining Methods Applied to Real-World Agricultural/Economic Data Sets. The abstract in the conference proceedings MendelNET PEF 2009. November 27, 2009. Brno. P. 59. ISBN 978-80-7375-351-1.

107) Rukavitsyn V., Zizka J.; Opinion Classification in Text Entries Using Machine-Learning Approach. Сборник статей международной конференции «Интеллектуальные системы принятия решений и проблемы

вычислительного интеллекта» (ISDMCI 2010). 17-21 Мая, 2010. Евпатория, Украина. С 283-288

108) Rukavitsyn V.; Real Text Data Analysis Using Machine-Learning Approach. The article in the conference proceedings MEKON 2011. February 2-3, 2011. Ostrava. CD-ROM. ISBN 978-80-248-2372-0.

109) Rukavitsyn V., Zizka J.; Automatic Categorization of Reviews and Opinions of Internet E-Shoping Customers. International Journal of Online Marketing. April-June 2011, Vol.1, No.2. p 70-79. ISSN 2156-1753.

110) Rukavitsyn V., Zizka J.; Inductive Formation of Ecosystem Condition Models. 9th Summer School of Applied Informatics, Bedřichov, Czech Republic. 7-9 September 2012. p. 143-151. ISBN 978-80-210-6058-6

111) The Machine Learning Dictionary URL: http://www.cse.unsw.edu.au/~billw/mldict.html (дата обращения 09.01.2017)

112) Weka 3: Data Mining Software in Java URL: http://cs.waikato.ac.nz/ (дата обращения 21.05.2016)

113) Witten H. Ian,Eibe Frank . Data Mining: Practical Machine Learning Tools and Techniques with Java Implementations. Morgan Kaufmann Publishers. 1999. 629 p.

114) Zizka J., Rukavitsyn V. Automatic Categorization of Reviews and Opinions of Internet E-Shopping Customers. In: Transdisciplinary Marketing Concepts and Emergent Methods forVirtual Environments. Hershey, Pennsylvania (USA): IGI Global, 2012. pp.154-163, ISBN 978-1-4666-1861-9.