

Малыгин Иван Вячеславович

**МЕТОДЫ МАШИННОГО ОБУЧЕНИЯ
В ГЕОФИЗИЧЕСКИХ ЗАДАЧАХ С ДЕФИЦИТОМ ДАННЫХ**

Специальность 25.00.10 Геофизика,
геофизические методы поисков полезных ископаемых

АВТОРЕФЕРАТ

диссертации на соискание ученой степени
кандидата технических наук

Работа выполнена в лаборатории геоинформатики Федерального государственного бюджетного учреждения науки Института физики Земли им. О.Ю. Шмидта Российской академии наук.

Научный руководитель:	Алешин Игорь Михайлович кандидат физико-математических наук, Федеральное государственное бюджетное учреждение науки Институт физики Земли им. О.Ю. Шмидта Российской академии наук, заведующий лабораторией геоинформатики
Официальные оппоненты:	
Ведущая организация:	

Защита диссертации состоится дд.мм.2022 г. в хх:хх часов на заседании диссертационного совета Д.002.001.01, созданном на базе Федерального государственного бюджетного учреждения науки Института физики Земли им. О.Ю. Шмидта Российской академии наук, по адресу 123242, г. Москва, ул. Большая Грузинская, д. 10, стр. 1, конференц-зал.

С диссертацией можно ознакомиться в библиотеке ИФЗ РАН и на сайте института www.ifz.ru. Автореферат размещен на официальном сайте Высшей аттестационной комиссии при министерстве образования и науки Российской Федерации www.vak.minobrnauki.gov.ru и на сайте ИФЗ РАН.

Отзывы на автореферат, заверенные печатью, в двух экземплярах, просьба направлять по адресу: 123242, Москва, Большая Грузинская ул., д. 10, стр.1, ИФЗРАН, ученому секретарю диссертационного совета Владимиру Анатольевичу Камзолкину.

Автореферат разослан « ____ » _____ 2022 г.

Ученый секретарь диссертационного совета,
кандидат геолого-минералогических наук

В.А. Камзолкин

ОБЩАЯ ХАРАКТЕРИСТИКА РАБОТЫ

Актуальность темы

Диссертация посвящена исследованию нескольких классических задач с недостатком данных: пространственная интерполяция (двумерная и трехмерная) и классификация на основе временных рядов. Задачи с пропуском данных являются традиционными для геофизических исследований. Это связано с недостаточным количеством измерений, непродолжительным промежутком времени наблюдений, большой протяженностью объектов. Подобная ситуация дефицита данных затрудняет обработку и интерпретацию результатов измерений геофизических полей. Несмотря на то, что в последнее время большое внимание уделяется задачам с большими данными (Big Data), большинство геофизических данных по-прежнему не являются таковыми.

Применение классических методов статистического анализа позволяет получать достоверные результаты при наличии большого объема наблюдений. Недостаток данных в геофизических исследованиях требует применения специальных методов анализа. С точки зрения математической постановки такие проблемы являются классическим случаем задач с пропуском данных. Одним из подходов при решении подобных задач, является группа методов, разработанных в рамках теории машинного обучения и распознавания образов. В практических исследованиях выделяются две группы методов для решения задач с пропусками данных: базовые и специализированные.

Базовые методы основаны на универсальных алгоритмах машинного обучения. Они могут применяться для анализа многомерной и разнородной информации, включая геофизические данные. Такие алгоритмы для обучения используют простые характеристики исходных данных, например, вычисляют расстояния между точками, строят линейные приближения, производят операции с множествами. Эти алгоритмы начали разрабатываться с 1950-х годов XX в., имеют широкое теоретическое и эмпирическое обоснование. В качестве примера задачи, для решения которой эффективно использовать базовые методы теории машинного обучения, можно привести задачу интерполяции. Она возникает при построении двумерных и трехмерных моделей физических характеристик горных пород и геофизических полей в ограниченной области по измеренным значениям. В таких исследованиях, как правило, измерения проводятся на нерегулярной сетке в небольшом количестве точек. Это обусловлено значительным расстоянием между точками наблюдений и особенностями методик измерений. Для построения распределений в таких случаях широко используются интерполяционные процедуры. Наибольшее распространение получил кригинг – метод разработанный Д. Криге. Применение кригинга имеет ряд недостатков. Этот

метод часто приводит к излишнему сглаживанию пространственного распределения исследуемых величин. Поэтому одним из актуальных направлений является разработка методов, которые позволяют улучшить контрастность получаемых образов. Эту задачу в постановке машинного обучения возможно рассматривать как задачу пространственной классификации: необходимо отнести значения интерполянта в промежуточных точках к одному из заданных классов, что позволяет построить выраженные границы в пространстве.

Специализированные методы разрабатываются для решения конкретной прикладной задачи. В их основе лежат общие алгоритмы машинного обучения, существенно адаптированные под специфику решаемой задачи. Центральной частью применения специально адаптированных методов машинного обучения является более сложная процедура обучения. Машинное обучение – систематическое обучение алгоритмов, в результате которого их знания и качество работы возрастают по мере накопления опыта. Под обучением понимается автоматизированная настройка параметров алгоритма на данных конкретной прикладной задачи. На практике это чаще всего означает численное решение некоторой оптимизационной задачи: конечный прикладной результат исследования представляют в виде формализованного функционала качества, который оптимизируется в процессе обучения на объектах исходных данных. Процедура обучения активно используется, например, в одном из методов прогнозирования мест сильных землетрясений ЕРА (Earthquake-Prone Areas recognition). Здесь обучение используется для поиска критериев, обеспечивающих наилучшее согласование с известными результатами наблюдений. Использование дополнительной информации позволяет получать карты распределения вероятности возникновения землетрясения в рассматриваемой сейсмоактивной территории (Карта Ожидаемых Землетрясений), улучшить прогнозирование мест возникновения сильных землетрясений (алгоритм Барьер). Другим примером специализированных методов является классификация на основе временных рядов. Задачи с временными рядами особо актуальны при мониторинге катастрофических процессов (магнитные бури, гидрометеопроцессы). Особенностью этой задачи в ситуации недостатка данных являются короткие временные ряды однородных наблюдений на исследуемой территории, возможно, содержащие пропуски в данных измеряемых значений. Одним из возможных подходов к решению подобной задачи классификации является использование в процессе обучения интегральных характеристик, построенных по этим рядам с предварительной интерполяцией пропусков.

Цель исследования

Цель исследования – разработка компьютерных систем и методов обработки данных в условиях ограниченного количества данных, не достаточных для проведения классического статистического анализа, на основе методов машинного обучения, включая интерпретацию результатов измерений геофизических полей и анализа многопараметрической информации, применение таких систем для построения геолого-геофизических моделей и решения задач охраны окружающей среды.

Основные задачи исследования:

1. Построение моделей двумерного и трехмерного распределения геофизических величин по данным полевых измерений на нерегулярной сетке с помощью базовых методов теории машинного обучения.
2. Построение прогнозной интеллектуальной системы, основанной на рядах коротких временных данных на основе специализированных методов теории машинного обучения и распознавания образов.

На защиту выносятся следующие положения:

1. Метод пространственной интерполяции двумерных и трехмерных нерегулярно распределённых геофизических измерений с пропусками данных, основанный на применении теории машинного обучения. Метод применим для данных с экстремальной анизотропией пространственного распределения.
2. Интеллектуальная система для краткосрочного прогноза образования ледовых заторов в весенний период на северных реках. Система позволяет прогнозировать возникновение ледовых заторов по ограниченному набору наблюдений на гидропостах и метеостанциях. Для участков рек Сев. Двина и Лена точность прогнозирования системы составляет 85% и 76% соответственно.

Методика исследований

Основные результаты исследования получены с применением методов машинного обучения и методов теории распознавания образов. Реализация алгоритмов обработки данных для пространственной интерполяции выполнена на языке Python 3 (Глава 2). Реализация логического алгоритма прогнозной системы выполнена на языке программирования C/C++ в среде Microsoft Visual Studio (Глава 3). Составление карт (Глава 2) проводилось в среде GoldenSoftware Surfer 15, визуализация данных прогнозной системы (Приложение 3) проводилась в среде Esri ArcGIS 10. Источниками информации являются

научные публикации, справочные издания, тематические электронные ресурсы, экспертные знания.

Научная новизна

Предложено решение научно-технической задачи, имеющей влияние на развитие геофизических и геоинформационных технологий, методов прогнозирования сложных, трудно-формализуемых процессов, методов анализа и обработки временной и пространственно-распределенной геофизической информации.

Показано, что применение базового метода машинного обучения (метод ближайших соседей) в задачах по построению пространственных распределений двумерных геофизических величин с ограниченным объемом исходных данных позволяет достичь лучших результатов в части уточнения границ объектов по сравнению с использовавшимся ранее методом кригинга на основе линейной регрессии. Метод ближайших соседей позволил лучше определять нелинейные зависимости в пространственном распределении геофизических величин, лучше выделять области пространственной неоднородности.

Показано, что в задаче построения трехмерной модели среды при проведении межскважинных исследований метод ближайших соседей позволяет оконтурить малые объекты даже при использовании синхронной схемы измерений.

Влияние пространственной анизотропии распределения данных можно исключить, если модифицировать пространственную метрику, определяющую расстояние между данными. Это достигается введением коэффициента скейлинга, который изменяет масштаб в горизонтальном направлении. Использованный подход позволяет получить контрастное изображение неоднородных областей, чьи геометрические размеры меньше расстояния между скважинами.

По данным инверсии приемных функций с помощью интерполяции методами машинного обучения построена новая трехмерная модель южной Фенноскандии с возможностью построения профилей вдоль произвольного направления.

Разработанная технология является универсальной: процесс построения трехмерной модели не зависит от физической модели, использованной для интерпретации измерений.

Для задачи анализа временных рядов ограниченной длины предложен оригинальный подход на основе специализированных методов теории машинного обучения и распознавания образов. Разработанный метод создания прогнозной системы сочетает используемые на практике принципы классификации явлений с экспертными и математическими методами прогнозирования. Разработанная на основе методов машинного обучения прогнозная система является универсальной в части требований к исходным

данным, алгоритмического обеспечения задачи прогноза и анализа результата его достоверности.

Выполнен прогноз ледового заторообразования для участка р. Северная Двина на несколько сезонов. Проведена валидация прогнозов системы. Оцененная достоверность составила 85%.

Выполнен прогноз ледового заторообразования по данным нового региона – бассейна р. Лена. Проведена валидация прогнозов системы. Оцененная достоверность составила 76%.

Предложен способ проведения факторного анализа в условиях дефицита данных на основе коротких временных рядов наблюдений характеристик процесса заторообразования. Результаты анализа подтвердили выдвинутые ранее теоретические гипотезы о важности признаков.

Практическая значимость результата работы

Представленные в работе результаты анализа геофизических данных могут применяться на практике для решения ряда геофизических задач.

Построена уточненная карта границы Мохоровичича для региона Фенноскандия. Толщина коры, определяемая как расстояние от поверхности до этой границы, является основной характеристикой при анализе строения региона, а также при изучении структуры европейской литосферы. Построенная карта может применяться для дальнейших исследований строения литосферы северной части Балтийского щита, изучения строения мантии северной и южной Финляндии, построения трехмерных сейсмических моделей южной Финляндии.

Построенная карта слоя с низкими скоростями поперечных сейсмических волн может применяться для продолжения исследований природы его возникновения. Приведенный анализ сейсмических данных показал эффективность методов машинного обучения для их анализа и обобщения. Достоинства такого подхода связаны с универсальностью применяемых методов. Особенно ярко преимущества алгоритмов теории машинного обучения проявляются в условиях недостатка данных, типичных для многих геофизических исследований.

Построена трехмерная модель проводимости среды при проведении межскважинных исследований. Использованный метод машинного обучения (метод ближайших соседей) позволяет построить трехмерную модель проводимости среды между скважин даже при использовании синхронной схемы измерений.

Построена трехмерная региональная модель коры для региона южной Фенноскандии. Разработанный метод является универсальным, и не зависит от способа получения

одномерных моделей. Необходимые вычисления выполняются непосредственно для нужных сечений, при этом форма сечений может быть произвольной.

Разработана интеллектуальная система для осуществления краткосрочного прогнозирования мощности процесса заторообразования для участка реки Северная Двина от г. Котлас до г. Великий Устюг, что является важной частью прогноза наводнений для данной территории. Достигнута точность прогнозирования на уровне 85%, результаты подтверждены проведенной валидацией прогнозов.

Разработанная интеллектуальная система применена на данных бассейна реки Лена, что подтверждает независимость разработанной технологии от региона применения. Достигнута точность прогнозирования 76%.

Реализована функциональность, которая позволяет применять интеллектуальную систему в качестве инструмента анализа данных: проверять гипотезы относительно влияния признаков на исследуемый процесс, количественно оценивать величину вклада конкретного признака в итоговый результат явления.

Соответствие паспорту специальности

Работа содержит решение задач, имеющих научно-практическую значимость в части совершенствования способов обработки и интерпретации данных измерений геофизических полей, интегрированного анализа многомерной, многопараметрической и разнородной информации, включающей геофизические данные, а также применение геофизических методов в решении задач охраны окружающей среды и соответствует пунктам №№ 14, 18, 25 Паспорта специальности ВАК 25.00.10 «Геофизика, геофизические методы поисков полезных ископаемых» (технические науки).

Апробация работы

Работа и отдельные результаты обсуждались на научных семинарах ИФЗ РАН, МГУ им. М.В. Ломоносова, а также на следующих конференциях: Научная конференция молодых ученых и аспирантов ИФЗ РАН (2021, 2020, 2019, 2018, 2017); Information Technologies in Earth Sciences and Applications for Geology, Mining and Economy (ITES&MP-2019); Всероссийская конференция с международным участием II Юдахинские чтения «Проблемы обеспечения экологической безопасности и устойчивое развитие арктических территорий»; VI Международная научно-практическая конференция «Индикация состояния окружающей среды: теория, практика, образование»; IV Школа-семинар «Гординские чтения»; Международный молодежный научный форум «ЛОМОНОСОВ» (2015, 2014); IV

Международная научно-практическая конференция «Научные перспективы XXI века. Достижения и перспективы нового столетия».

Получены 3 свидетельства о государственной регистрации программ для ЭВМ.

Публикации

По материалам диссертации опубликовано 13 работ, в том числе 9 статей в ведущих рецензируемых изданиях, рекомендованных ВАК РФ.

Структура работы

Диссертация состоит из введения, трех глав, заключения, списка литературы из 168 наименований, четырех приложений. Текст диссертации изложен на 156 страницах машинописного текста, содержит 16 таблиц и 39 рисунков.

Автор выражает благодарность научному руководителю к.ф.-м.н. Игорю Михайловичу Алешину (ИФЗ РАН) за поддержку на всех этапах проведения работы, а также коллективу лаборатории геоинформатики ИФЗ РАН.

СОДЕРЖАНИЕ РАБОТЫ

Во **введении** обоснована актуальность работы, сформулированы цель и задачи исследования, показаны научная новизна, практическая значимость работы, представлены основные защищаемые положения.

В соответствии с целью исследования в **первой главе** рассмотрены примеры геофизических задач в ситуации недостатка данных, в которых применение современной теории машинного обучения и методов распознавания образов привело к новым результатам. Введены понятия и определения теории машинного обучения, рассмотрена общая постановка задачи обучения с учителем, приведены основные функционалы качества. В задачах построения 2D-модели региона и построения 3D-модели среды рассмотрена ситуация недостатка исходных пространственных данных измерений, предложен способ решения на основе метода ближайших соседей. В задаче прогнозирования опасных геофизических явлений с ограниченным объемом временных данных предложен способ решения на основе комбинаторно-логического подхода теории распознавания образов. Сделан обзор основных методов решения задач с пропуском данных.

Выводы по главе 1

1. Приведены примеры использования методов машинного обучения в различных геофизических задачах: поиск полезных ископаемых, анализ месторождений, прогнозирование опасных явлений.
2. Проведен анализ методов машинного обучения. Предложены методы решения задач интерполяции на основе алгоритма ближайших соседей и анализа временных данных для создания прогнозной системы.
3. Приведена общая схема моделирования с использованием методов машинного обучения.

Во **второй главе** представлены результаты, полученные с помощью методов машинного обучения в применении к задаче пространственной интерполяции геофизических полей. Рассмотрено четыре примера геофизических приложений.

В задаче построения трехмерной модели среды при проведении межскважинных исследований предложена новая интерпретация данных радиоволнового просвечивания, позволяющая более точно выделить границы слоев по сравнению с методами, используемыми ранее (кригинг). В настоящий момент значительно возросла глубина работ по разведке кимберлитовых тел и рудных месторождений. Традиционные геологические методы поиска оказались неэффективными. На практике единственным прямым методом

поиска является бурение системы скважин до глубин, которые обеспечивают доступ к вмещающим породам. Из-за высокой стоимости бурения возросла роль межскважинных методов. Они позволяют увеличить среднее расстояние между скважинами без существенного снижения вероятности пропуска кимберлитового или рудного тела. Среднее расстояние между ближайшими скважинами составляет 200 м. Метод радиоволнового просвечивания особенно эффективен при поиске объектов, отличающихся высокой контрастностью электропроводящих свойств (Рис. 1).

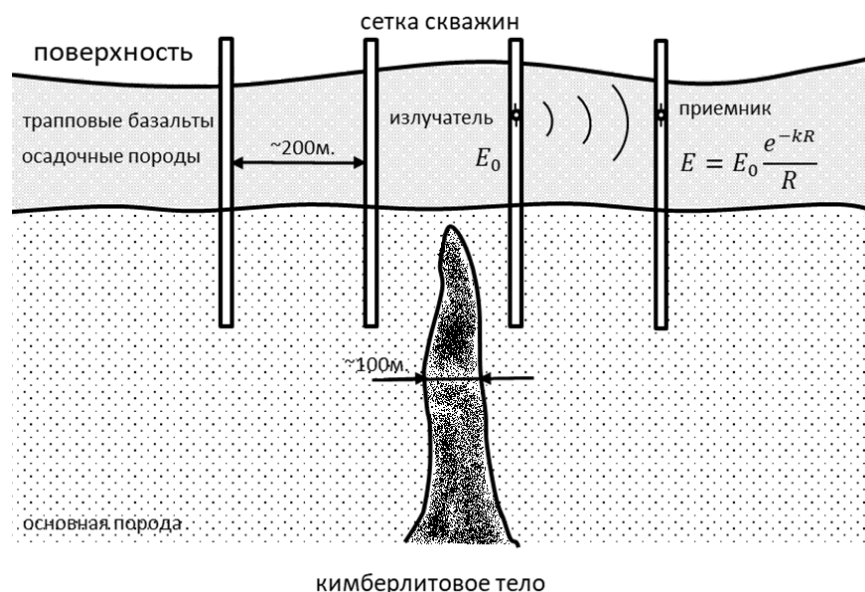


Рис. 1. Схема процесса радиоволнового просвечивания

Физическую основу метода составляет зависимость распространения электромагнитной волны от проводящих свойств среды распространения. Источником и приемником электромагнитного излучения является электрический диполь. При измерениях они размещаются в соседних скважинах, а расстояние между источником и приемником известно. Поэтому измерив величину уменьшения амплитуды электромагнитной волны при ее распространении между скважинами можно оценить коэффициент поглощения среды. Породе с низким электрическим сопротивлением соответствует высокое поглощение радиоволн. Поэтому данные межскважинных измерений позволяют оценить эффективное электрическое сопротивление породы. Обычно источник и приемник синхронно погружаются в соседние скважины. Измерение величины амплитуды электрического поля в приемнике позволяет оценить среднее значение коэффициента затухания на линии, соединяющей источник и приемник. Измерения проводятся во время остановок, приблизительно каждые 5 м. Расстояние между остановками значительно меньше расстояния между соседними скважинами. Это приводит к значительной пространственной анизотропии

в распределении данных. При проведении разведочного бурения скважины покрывают большую площадь. Задача состоит в построении трехмерной модели распределения электрических свойств межскважинного пространства на всем участке по результатам совокупности измерений. Существенная анизотропия пространственного распределения препятствует использованию стандартных методов геостатистики.

Основными направлениями, за счет которых стало возможным получить новые результаты, стали применение и адаптация метода ближайших соседей в ситуации дефицита исходных данных измерений, а также учет пространственной анизотропии геофизических процессов. Особенность пространственного распределения данных приводит к тому, что практически для всех точек пространства ближайшими оказываются данные, относящиеся к одной группе измерений. Масштабирование горизонтальных осей позволяет исправить эту ситуацию (Рис 2).

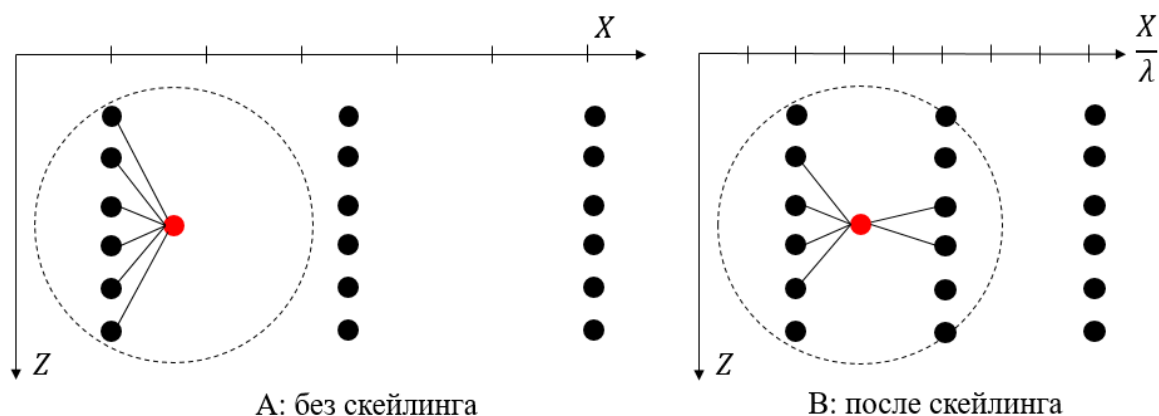


Рис. 2. Влияние масштабного множителя λ на распределение ближайших точек

Для определения гиперпараметров модели (масштабного множителя λ и числа ближайших соседей k) использован метод кросс-валидации. Исходные данные разбиваются на $M=5$ групп. Каждая из этих групп поочередно устраняется из процедуры обучения, и используется для проверки. Оценки качества решения производятся по функционалу качества R^2 (коэффициент детерминации) – доля дисперсии зависимой переменной, объясняемая моделью. Коэффициент детерминации был рассчитан на сетке $1 \leq k \leq 25$, $1 \leq \lambda \leq 25$ с единичным шагом по каждому из параметров. Результат расчетов приведен на Рис. 3. Полученное распределение имеет вид, типичный для задач многопараметрической оптимизации. Для выбора значений гиперпараметров использован уровень значений коэффициента $R^2 = 0.7$. Выбранные значения гиперпараметров $k = 11$ и $\lambda = 10$ соответствуют пересечению медианы треугольника, образованного осями координат и прямой, аппроксимирующей 70% уровень значений коэффициента детерминации.

После того, как значения гиперпараметров определены, задача построения образа состоит в определении интересующей величины – коэффициента затухания методом kNN с модифицированной метрикой в узлах трехмерной решетки. На Рис. 4 приведены результаты моделирования: набор вертикальных и горизонтальных сечений

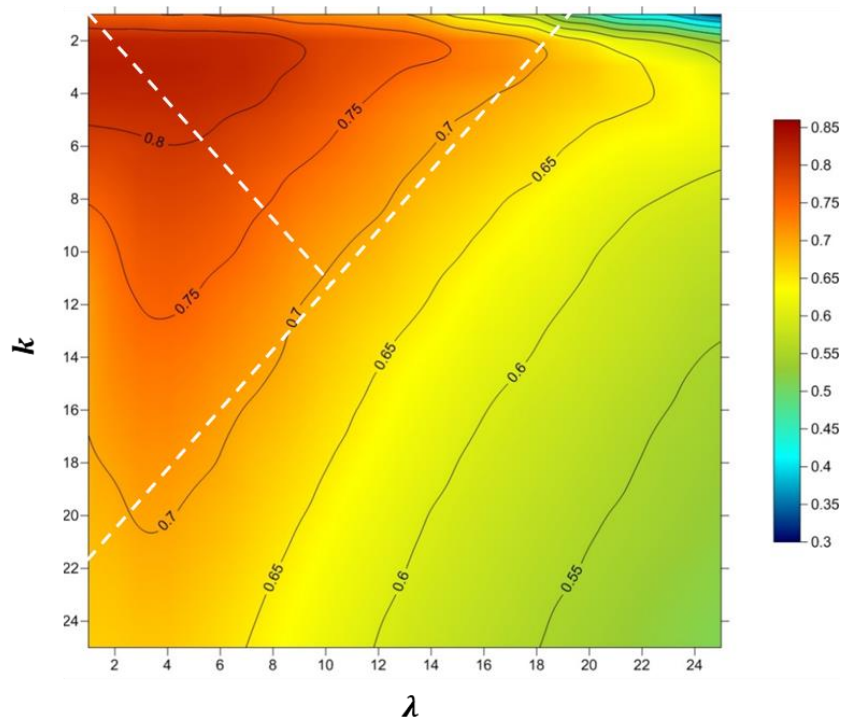


Рис. 3. Значения коэффициента детерминации $R^2(k, \lambda)$, рассчитанные на сетке параметров.

пространства. Глубина отложена от уровня моря, положение горизонтальных осей согласовано с геометрией участка. Расположение вертикального разреза соответствует линии $Y = 0$ на схеме расположения скважин. Горизонтальные сечения соответствуют глубинам $Z = -560$ и $Z = -250$ метров (черная и белая пунктирные линии на вертикальном разрезе). Из рисунка видно, что построенная модель позволяет локализовать объекты, чьи горизонтальные размеры существенно меньше расстояния между скважинами. В качестве примера можно привести области повышенных значений коэффициента затухания, расположенные на глубине -560 м., и горизонтальными координатами $X = 2950$, $X = 4750$ и $X = 5150$ метров. Для наглядности, на Рис. 4 приведены также короткие вертикальные сечения, соответствующие этим линиям, на которых соответствующие области диаметром менее 100 м. также отчетливо видны.

Использованный подход позволяет получить достаточно контрастное изображение неоднородных областей, что позволяет выделить объекты, чьи геометрические размеры меньше расстояния между скважинами. Следует заметить также, что процесс построения модели не зависит от физической модели, использованной для интерпретации измерений. Уточнение физической модели процесса распространения радиоволн между скважинами позволит улучшить качество построения образа. Кроме того, модель может быть улучшена, если привлечь дополнительные данные (геологические, сейсмические, магнитные) для их совместной интерпретации.

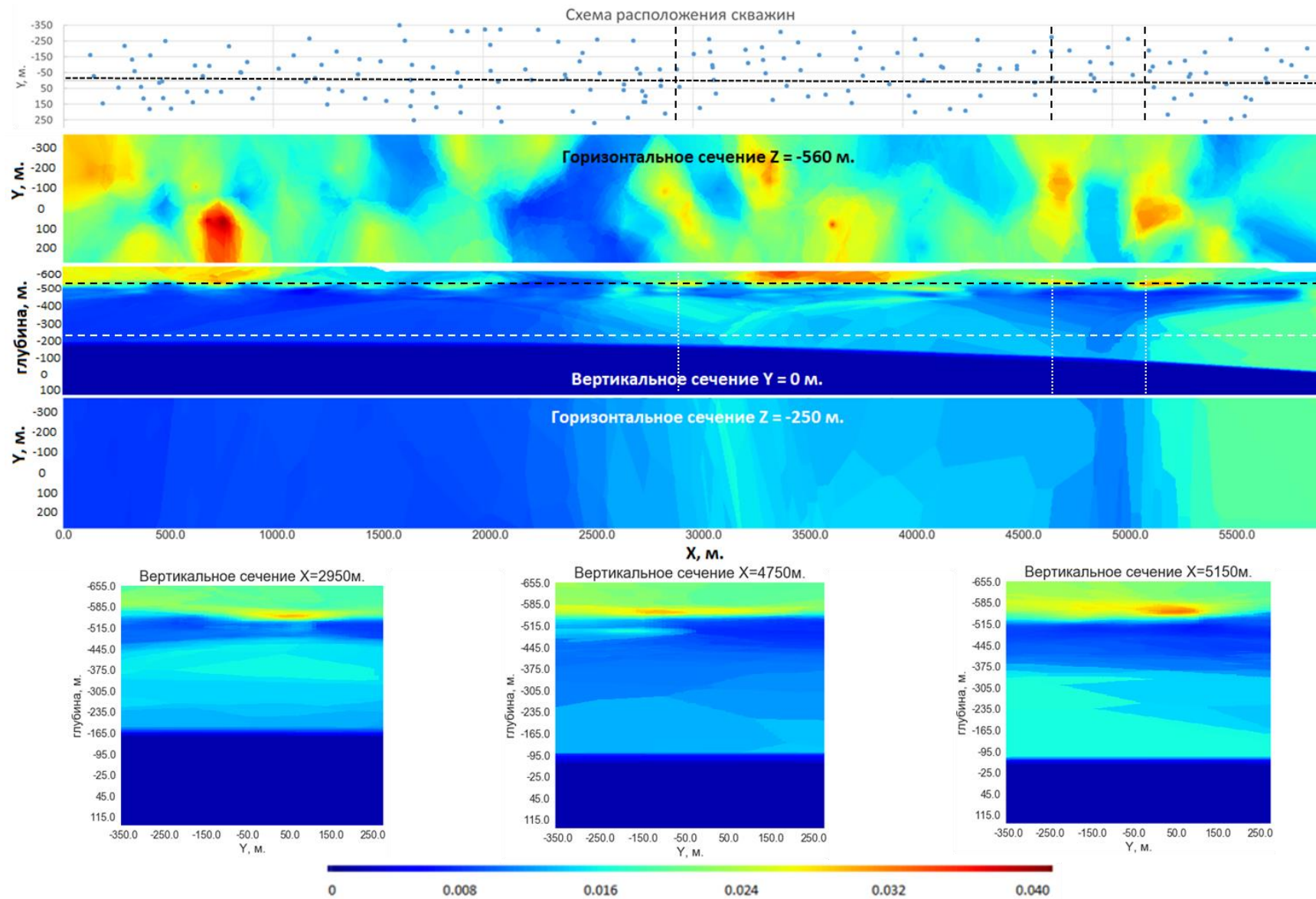


Рис. 4. Результаты 3D-моделирования.

В следующем геофизическом приложении – задаче анализа строения коры северной части Балтийского щита построена уточненная карта поверхности Мохоровичича (Рис. 5).

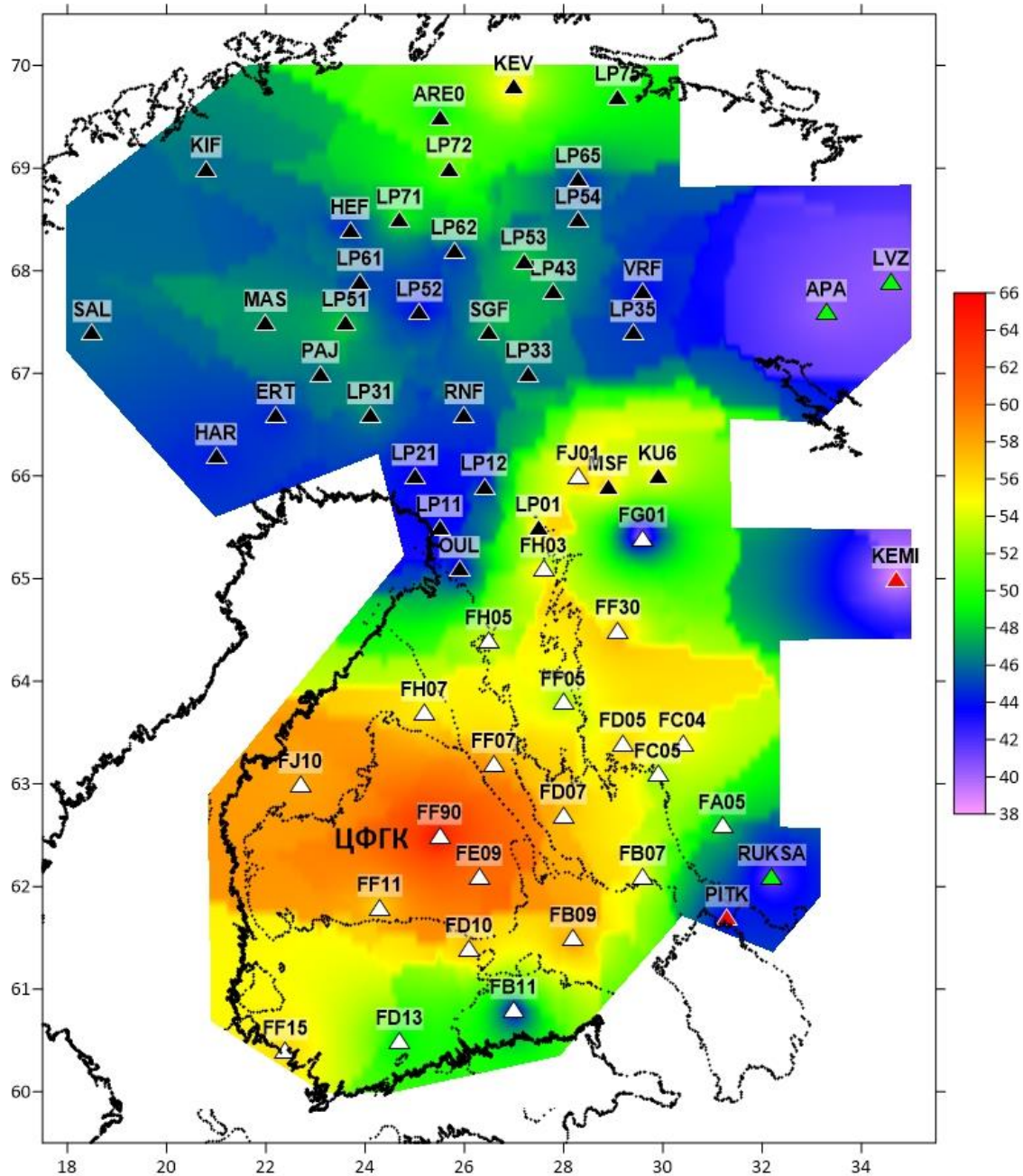


Рис. 5. Карта глубины границы Мохоровичича.

Основу исследования составляют данные, полученные методом приемных функций. Были использованы сведения, полученные в предыдущих исследованиях этого региона, дополненные новыми расчетами и измерениями. Исходные данные представляют собой набор зависимостей сейсмической скорости от глубины, рассчитанных для 61 постоянных и временно действующих геофизических станций. С точки зрения машинного обучения, данная задача является задачей регрессии. Для восстановления регрессионной зависимости глубины Мохоровичича от двумерных координат был использован метод k ближайших

соседей с необходимой адаптацией в части выбора сферической метрики. На исходных данных проведен подбор гиперпараметров и обучение метода ближайших соседей. Оптимальное значение числа соседей равно четырем ($k=4$). Средняя ошибка интерполяции при этом составляет 3.7 км. Треугольниками на карте отмечены точки, в которых положение границы было измерено. Буквенные коды означают названия сейсмических станций. Цвет треугольников определяется использованным литературным источником: черный – *Silvennoinen et al.*, 2014 (эксперимент POLENET/LAPNET), белый – *Kozlovskaya et al.*, 2008 (эксперимент SVEKALAPKO), зеленый – *Dricker et al.*, 1996 и *Aleshin et al.*, 2006, красный – данные получены в рамках настоящего исследования. Аббревиатура ЦФГК означает Центральный финский гранитоидный комплекс.

Проведено графическое сравнение ранее полученных результатов с результатами настоящей работы по четырем сейсмическим профилям. Построена глубина поверхности Мохоровичича вдоль соответствующих профилей. Взято сечение поверхности, построенной методом kNN , и построена зависимость глубины Мохоровичича от расстояния вдоль этой плоскости. Результаты графически наложены на иллюстрации из публикаций красной сплошной линией, красными пунктирными линиями показана оцененная ошибка интерполяции (Рис. 6-9).

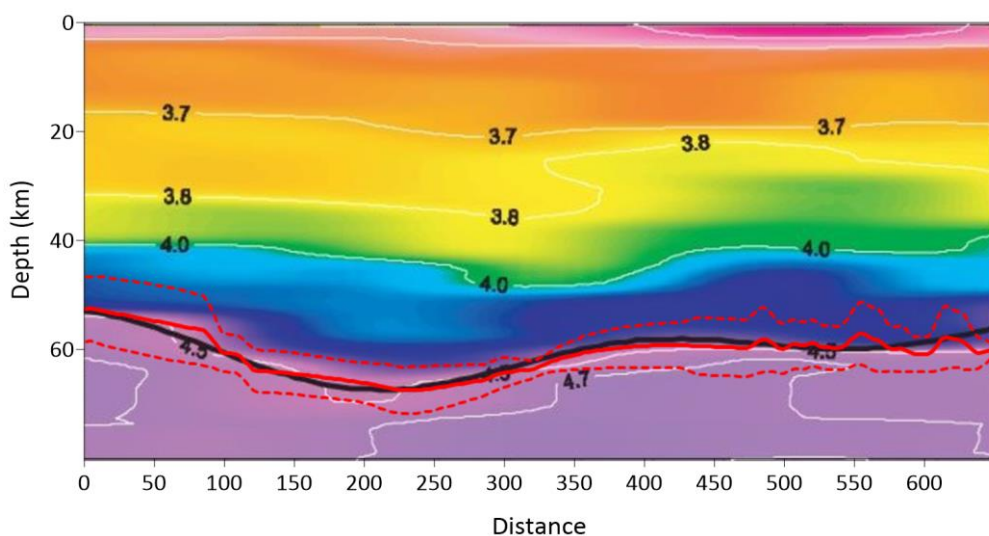


Рис. 6. Профиль FD13-FJ01. На профиль границы Мохо из работы *Kozlovskaya et al.*, 2008 (черная линия) нанесена граница Мохо, построенная методом kNN в рамках настоящего исследования (красная линия).

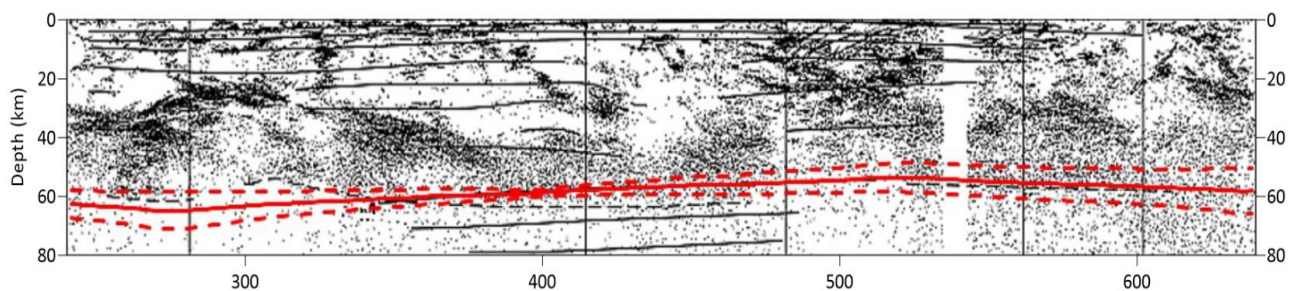


Рис. 7. Профиль FF30-A (аррх. FIRE 1). На профиль границы Мохо из работы *Janik et al., 2007*, нанесена граница Мохо, построенная методом *kNN* в рамках настоящего исследования – красная линия. Черная пунктирная линия – метод ГСЗ, черная сплошная линия – метод ОГТ

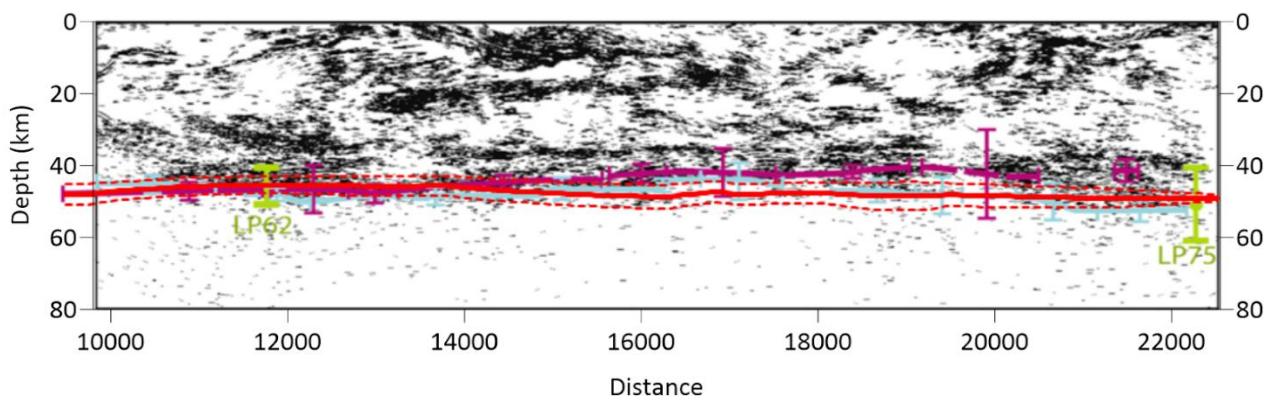


Рис. 8. Глубина Мохо вдоль профиля POLAR (LP51–LP75). На рис. 6 из статьи *Silvennoinen et. al., 2014* красной линией наложена глубина Мохо, полученная в рамках настоящего исследования методом *kNN*.

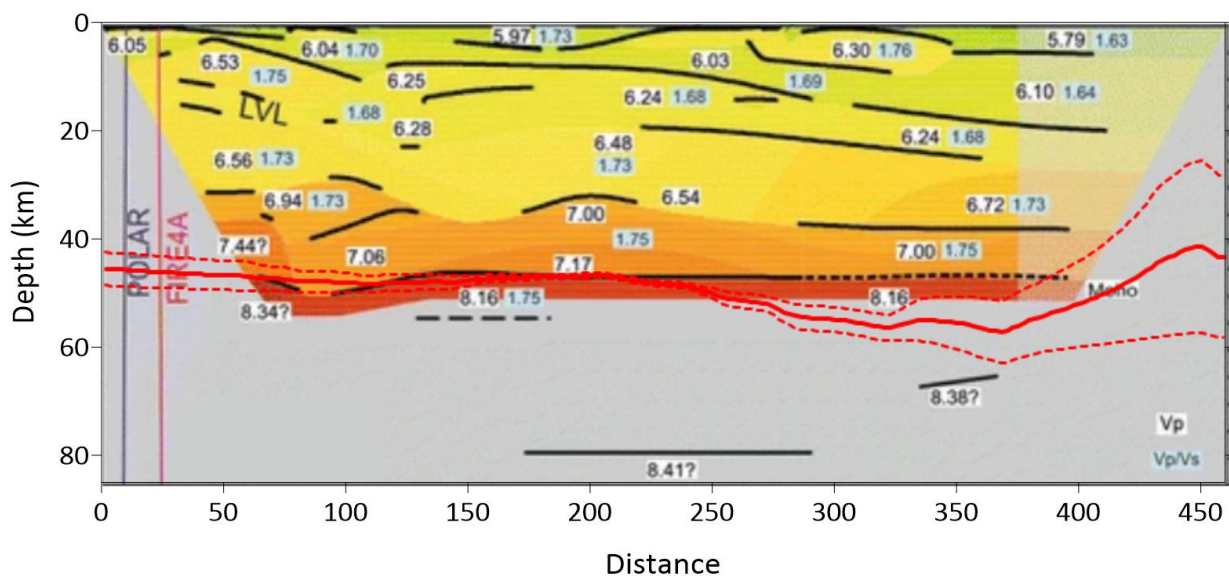


Рис. 9. Глубина Мохо вдоль профиля HUKKA2007 (аррх. LP62–FG01). На рис. 11 из статьи *Tiira et. al., 2014* красной линией наложена глубина Мохо, полученная в рамках настоящего исследования методом *kNN*.

Построенные профили хорошо согласуются с полученными ранее результатами. По всей длине профилей результаты интерпретации методом kNN попадают в границы доверительных интервалов предыдущих построений.

Еще одним результатом стало построение карты слоя с низкими значениями скорости поперечных сейсмических волн V_S (Рис. 10).

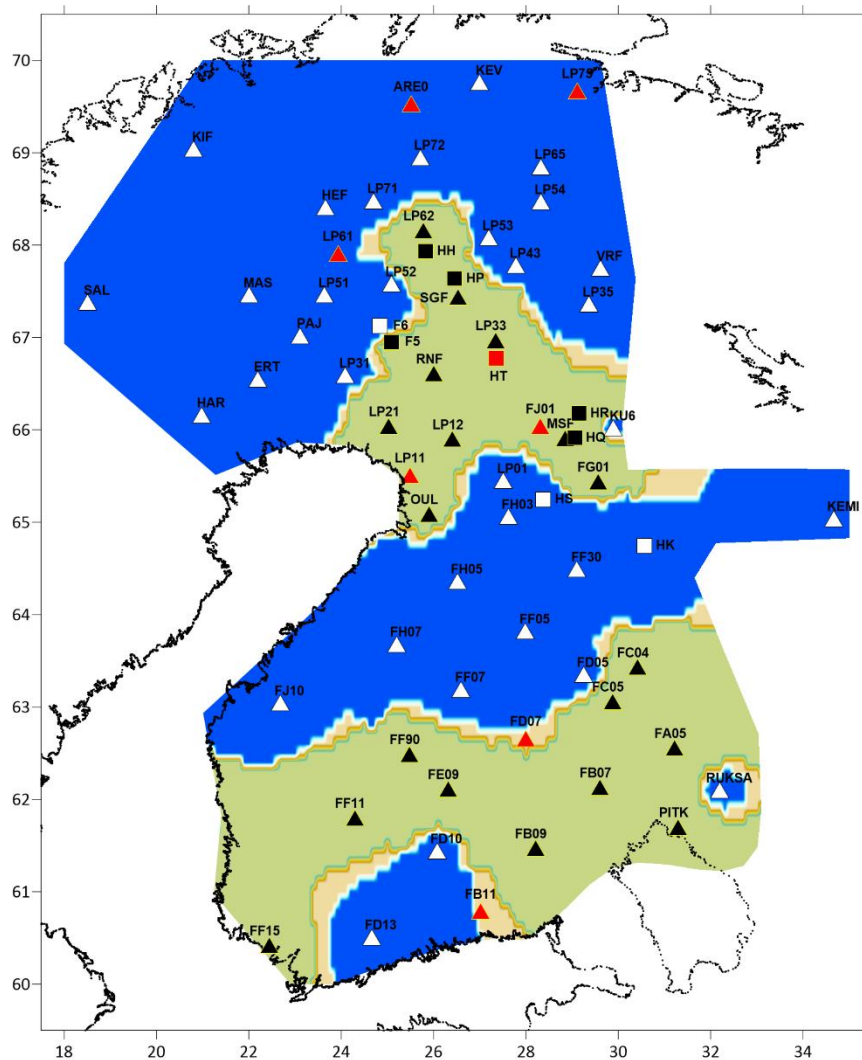


Рис. 10. Карта поверхностного слоя низкой скорости поперечных сейсмических волн

В исследуемом регионе практически отсутствует осадочный слой. Несмотря на это, имеются области, в которых присутствует слой с низкими значениями скорости V_S . Относительно низкие значения V_S обычно объясняют наличием в слое большого количества водонасыщенных трещин. Присутствие такого слоя не зависит от возраста пород. Эта задача относится к оценке принадлежности в рамках задачи бинарной классификации. Для 60 сейсмических станций известно наличие или отсутствие слоя низких скоростей. С помощью метода k ближайших соседей в каждой точке исследуемого региона оценена вероятность

наличия слоя низких скоростей. На исходных данных проведен подбор гиперпараметров и обучение метода ближайших соседей. Оптимальное значение числа соседей равно, как и в предыдущей задаче, четырем ($k=4$). Синим цветом на карте отображены области, в которых вероятность наличия слоя превышает 0.55, оливковым цветом окрашены области, в которых слой отсутствует (вероятность меньше 0.45). Промежуточным значениям вероятности соответствуют области цвета охра.

Построенная карта включает в себя классификацию по принципу наличия или отсутствия слоя низких скоростей, а также буферную область, в которой на основании имеющихся данных нельзя сделать однозначный вывод. Показано, что слой низких сейсмических скоростей на поверхности присутствует на значительной части региона, включая области с протерозойскими породами. В южной части Финляндии положение низкоскоростной области коррелирует с относительно низким значением толщины коры.

На основании приведенных результатов получен метод построения цифровой модели региона по совокупности одномерных зависимостей упругих параметров от глубины. Разработанный метод применен к данным инверсии волновых форм приемных функций, и построена цифровая модель южной части Фенноскандии. Это позволило детальнее проанализировать структуру верхней части коры под низкоскоростным слоем.

Идея метода перенесена из решения задачи анализа данных межскважинного электромагнитного просвечивания и состоит в масштабировании исходных данных по одному из измерений. При этом, масштабный коэффициент и другие параметры интерполяции определяются одним из методов машинного обучения.

В качестве входных данных использованы одномерные модели: зависимость скорости поперечных волн от глубины для 23 широкополосных станций пассивного сейсмического эксперимента SVEKALAPKO. Расположение станций приведено на Рис. 11. Градациями цвета показаны значения скоростей поперечных волн V_s непосредственно на поверхности. Знаками «+» замощена область, содержащая поверхностный слой низких значений V_s (аналогично Рис. 10). Отрезок, соединяющий точки А и В – проекция вертикального разреза полученной цифровой модели (см. далее Рис. 13). Одномерные модели были получены ранее совместным обращением волновых форм приемной функции, дисперсионной кривой рэлеевских волн и времени пробега обменных волн от верхней границы переходной зоны мантии. Для параметризации использовались однородные упругие слои переменной толщины.

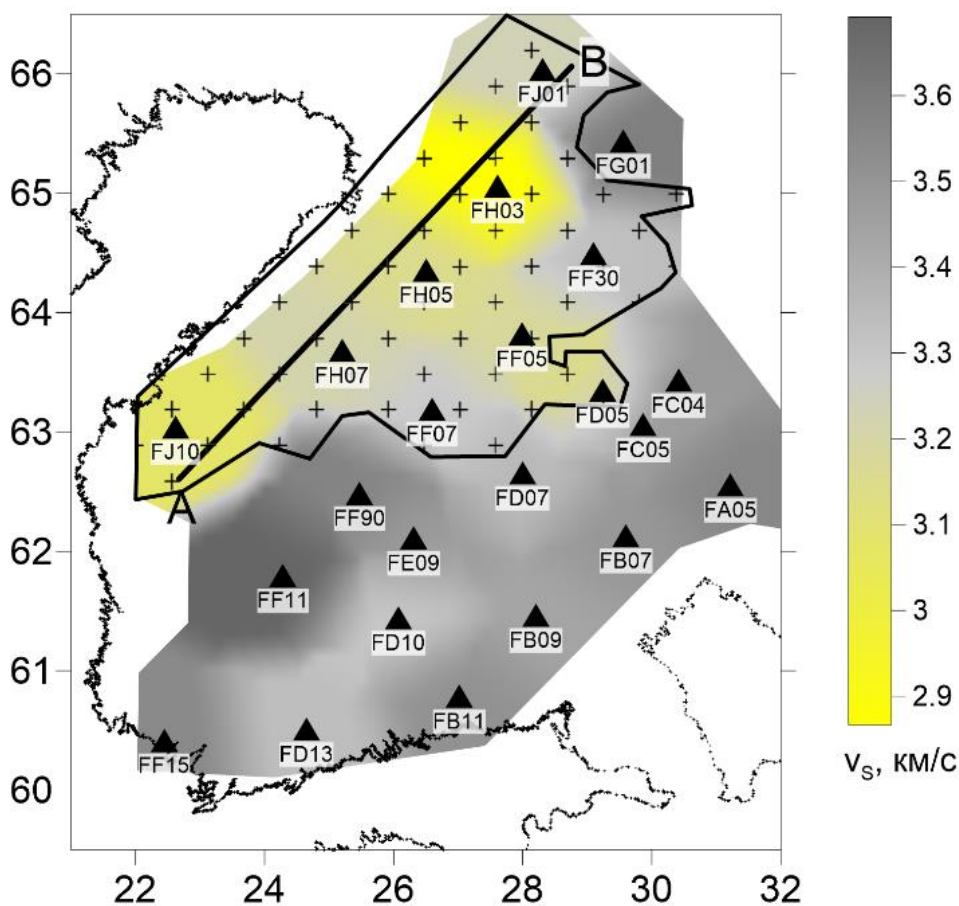


Рис. 11. Карта изучаемого региона.

В ситуации, когда механические свойства среды под станциями слабо меняются по латерали, для построения трехмерной цифровой модели возможно применить метод, изложенный ранее в решении задачи электромагнитного просвечивания. Его идея состоит в масштабном преобразовании горизонтальных координат, которое нивелирует различие между вертикальным и горизонтальным распределением данных. Если перейти к новой метрике, то при удачном выборе масштабного множителя пространственное распределение данных становится практически изотропным. Как и ранее, в качестве метода интерполяции выбран метод k ближайших соседей. Число ближайших соседей k вместе с масштабным множителем λ являются внешними параметрами (гиперпараметрами), требующими определения, методами, используемыми в машинном обучении. Для этого использован стандартный метод кросс-валидации с разбиением на 5 групп, каждая из этих групп поочередно устраняется из процедуры обучения и используется для проверки. Качество обучения определяется выбранным функционалом качества. В рассматриваемой задаче использована средняя квадратичная ошибка $RMSE$ интерполяции.

На Рис. 12 приведены значения функционала качества, рассчитанные для разных комбинаций значений числа соседей и масштабного множителя. Видно, что полученное распределение имеет вид, характерный для задач конкурентной оптимизации. В методе приемных функций типичная средняя ошибка определения скорости поперечных волн составляет 0.1 км/с. В качестве минимального уровня ошибки интерполяции выбрано вдвое меньшее значение 0.05 км/с. Дальнейшее уменьшение этого параметра приводит к эффекту переобучения – неоправданной детальности получаемого образа.

Очевидно, что все пары значений параметров k и λ , определяемых точками линии уровня 0.05 эквивалентны относительно выбранной оценки качества. Для дальнейших вычислений возможно выбрать любую из них, например, точку B , ближайшую к началу координат, как это принято в задачах многоцелевой оптимизации. Это соответствует равным значениям гиперпараметров $k = \lambda = 42$. Непосредственные вычисления показали, что параметры, определяемые точками A , B и C на Рис. 12, приводят к визуально неотличимым результатам.

Поскольку среда под станцией описывается набором упругих слоев разной ширины, то для проведения расчетов необходимо выполнить дискретизацию моделей. В результате для каждой станции получен набор значений упругих параметров, заданных для разных глубин. Строгий критерий выбора шага дискретизации Δh отсутствует, т.е. он является еще одним гиперпараметром задачи.

Максимальное значение шага дискретизации определяется толщиной самого тонкого слоя, которая равна 1.4 км. Также полученный набор дискретных значений должен обеспечивать необходимую точность интерполяции.

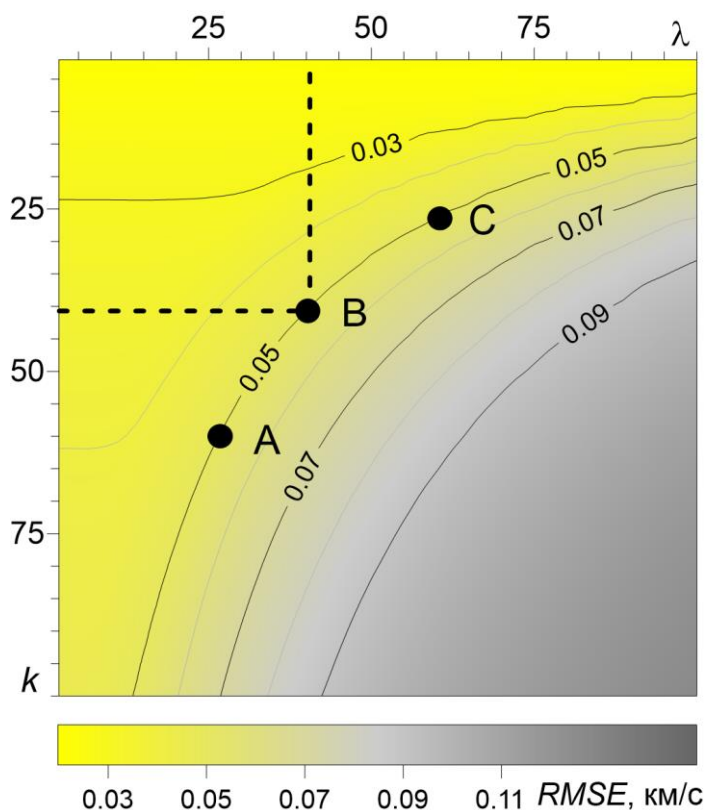


Рис. 12. Зависимость функционала качества $RMSE$ от значений гиперпараметров.

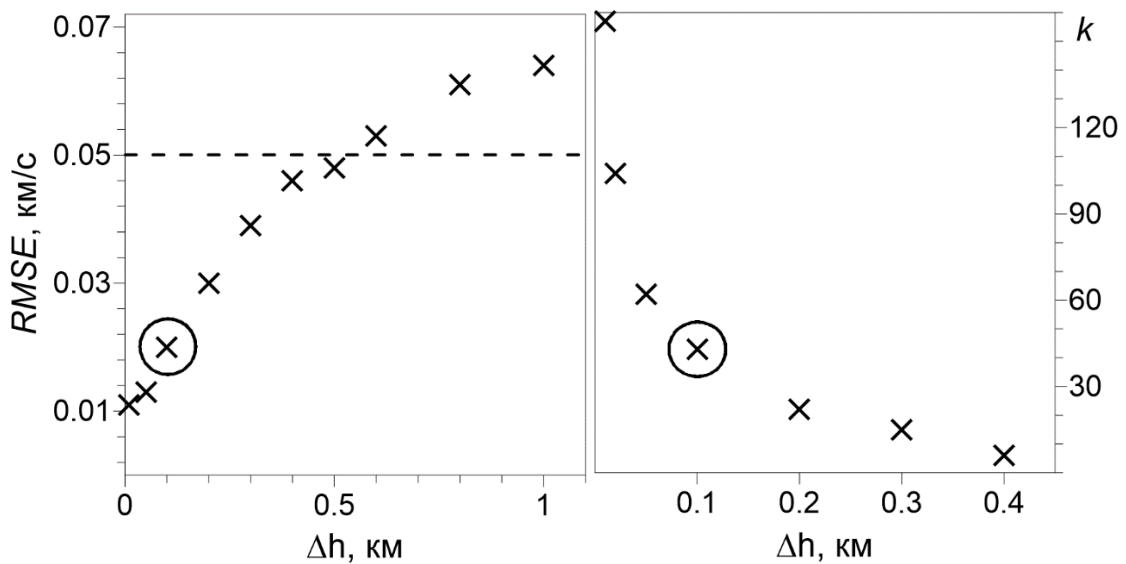


Рис. 13. Выбор шага дискретизации.

На Рис. 13 слева приведены минимальные значения функционала качества в зависимости от размера шага дискретизации. Вычисления показали, что минимальное значение функционала монотонно уменьшается с уменьшением Δh вплоть до величин порядка 0.01 км. Однако с увеличением качества интерполяции растет объем входных данных. На Рис. 13 справа для иллюстрации приведена оптимальное число ближайших соседей для разных значений Δh . Вместе с тем видно, что качество интерполяции удовлетворяет установленному критерию $RMSE = 0.05$ км/с. уже начиная с $\Delta h \approx 0.05$ км. Поэтому в дальнейших вычислениях использованы данные, полученные дискретизацией исходных значений с шагом $\Delta h = 0.1$ км.

Определение всех гиперпараметров решает задачу построения цифровой модели региона. При заданных k и λ модельное значение V_s может быть автоматически определено методом k ближайших соседей.

В качестве примера полученного результата рассмотрено строение верхней части коры в зоне перехода между архейскими и протерозойскими породами. На Рис. 11 приведено распределение сейсмических скоростей поперечных волн на поверхности. В северо-западной части региона отчетливо выделяется область низких значений V_s толщиной порядка одного километра.

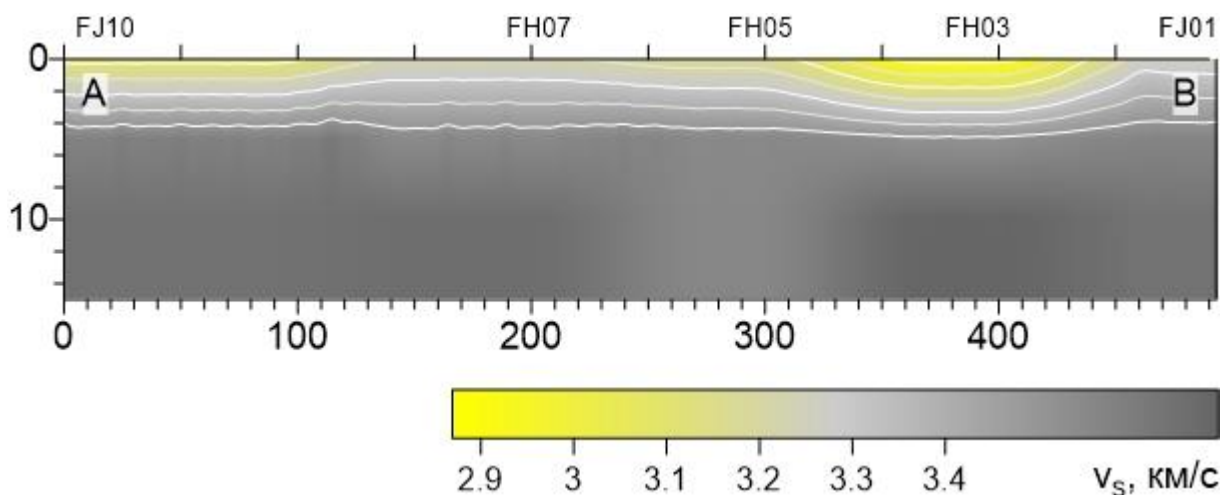


Рис. 14. Вертикальное сечение цифровой модели, проходящее через отрезок [A, B] на Рис. 11.

На Рис. 14 показано соответствующее вертикальное сечение верхних десяти километров. Понижение скорости S -волн в верхней части коры связано с наличием водонасыщенных трещин, образованных постледниковой релаксацией. Форма области содержащей такой слой была определена на основе анализа формы приемных функций на первых секундах после вступления основной фазы. Признаком слоя низких скоростей под станцией служило наличие обменной фазы, образованной контрастом упругих свойств в слое и в подстилающей породе. В отличие от качественного анализа, построенная цифровая модель позволяет определить не только область, содержащую низкоскоростной слой, но также его строение и пространственную структуру. На Рис. 11 и Рис. 14 видно, что форма поверхностного слоя низких скоростей V_s хорошо согласуется с результатом качественного анализа (Рис. 10). В то же время, толщина слоя и контрастный переход скорости в нем в архейских породах значительно выше, чем в протерозойских. Этот вывод согласуется с результатами, полученными ранее, в том, что параметр трещиноватости в архейской части региона почти в два раза выше, чем в протерозойской. Минимальные понижения скорости и толщины слоя относятся к переходной зоне архей-протерозой.

Описанный метод является универсальным. Он не зависит от способа получения одномерных моделей, поэтому в общем случае нет необходимости проводить расчеты для всех точек трехмерной регулярной решетки: удобнее выполнить вычисления непосредственно для нужных сечений. При этом форма сечений не сводится к плоскости, ориентированной горизонтально или вертикально, и может быть произвольной.

Выводы по главе 2

1. Построена трехмерная модель проводимости среды при проведении межскважинных исследований. Использованный метод машинного обучения (метод ближайших соседей) позволяет построить трехмерную модель проводимости среды между скважин даже при использовании синхронной схемы измерений. Влияние анизотропии распределения данных можно исключить, если модифицировать пространственную метрику, определяющую расстояние между данными, введением коэффициента скейлинга. Получено контрастное изображение неоднородных областей, позволяющее выделить неоднородности, геометрические размеры которых меньше расстояния между скважинами.
2. Построена уточненная карта границы Мохоровичича для региона Фенноскандия с использованием метода k ближайших соседей. Вертикальные сечения построенной модели хорошо согласуются с полученными ранее профилями.
3. Построена карта слоя с низкими значениями скорости поперечных сейсмических волн V_s для региона Фенноскандия с использованием метода k ближайших соседей. Показано, что слой низких сейсмических скоростей на поверхности присутствует на значительной части региона, включая области с протерозойскими породами. В южной части Финляндии положение низкоскоростной области коррелирует с относительно низким значением толщины коры.
4. На основании приведенных результатов получен метод построения цифровой модели региона по совокупности одномерных зависимостей упругих параметров от глубины. Разработанный метод применен к данным инверсии волновых форм приемных функций, и построена цифровая модель южной части Фенноскандии. Это позволило детальнее проанализировать структуру верхней части коры под низкоскоростным слоем. Приведенный анализ сейсмических данных показал эффективность методов машинного обучения для их анализа и обобщения. Достоинства такого подхода связаны с универсальностью применяемых методов. Особенно ярко преимущества алгоритмов теории машинного обучения проявляются в условиях недостатка данных, типичных для многих геофизических исследований.

В третьей главе рассматривается задача построения интеллектуальной системы на основе коротких рядов наблюдений для прогноза ледовой обстановки района с малой территорией и с коротким периодом наблюдения. Система разрабатывалась на данных р. Северная Двина, в дальнейшем была применена на данных другого региона – бассейне р. Лена.

Одним из подходов к решению задачи прогнозирования заторообразования является ее исследование в рамках построения прогнозных систем, основанных на алгоритмах теории машинного обучения и распознавания образов. Исходными данными являются наблюдения гидрологических постов и метеостанций различной временной глубины. Имеется статистика по явлению, т.е. фактический результат проявления (было ли явление, какой мощности). Разработанная система представляет прогноз по исследуемому явлению в будущий момент времени. Для исследования ледовой обстановки экспертным образом в качестве признаков выбран ряд гидрологических и метеорологических показателей. Общий список этих признаков представлен в Табл. 1.

Табл. 1. Признаковое пространство для региона бассейна р. Сев. Двины

№	Название признака	Характеристика признака	Единицы измерения
1	Предлежавший уровень воды	Гидрологический признак	сантиметры
2	Продолжительность осеннего ледохода	Гидрологический признак	сутки
3	Наличие зажоров	Гидрологический признак	есть (1) – нет (0)
4	Особенности температурного режима в период замерзания	Метеорологический признак Дата перехода температуры воздуха через ноль	количество суток с 1 сентября
5	Сумма отрицательных температур воздуха за холодный период	Метеорологический признак	градусы Цельсия
6	Сумма положительных температур воздуха за холодный период	Метеорологический признак	градусы Цельсия
7	Количество дней с положительными температурами воздуха за холодный период	Метеорологический признак	сутки
8	Сумма твердых осадков	Метеорологический признак	миллиметры
9	Особенности температурного режима в период вскрытия	Метеорологический признак Дата перехода температуры воздуха через 0.	количество суток с 1 марта
10	Толщина льда перед вскрытием	Гидрологический признак	сантиметры
11	Интенсивность роста уровней и расходов воды в период	Гидрологический признак	сантиметры в сутки

	подвижек	
--	----------	--

На Рис. 15 представлена схема района наблюдения с расположением речных постов: Каликино, Великий Устюг, Медведки, Котлас, Абрамково, Подосиновец.

Под классификацией понимается два возможных сценария ледохода:

- 1) наличие заторов с различными мощностью и продолжительностью на участке г. Великий Устюг – г. Котлас;
- 2) отсутствие заторов, либо их незначительное проявление на участке г. Великий Устюг – г. Котлас (в этот класс попадают и ситуации, когда затор произошел выше или ниже по течению, чем исследуемый участок).

Указанные сценарии ледохода определяют классы K_1 , K_2 периода наблюдения (Табл. 2.). Ситуация недостатка данных ограничивает возможность разделения на большее число классов, соответствующих более подробной классификации исследуемого явления.

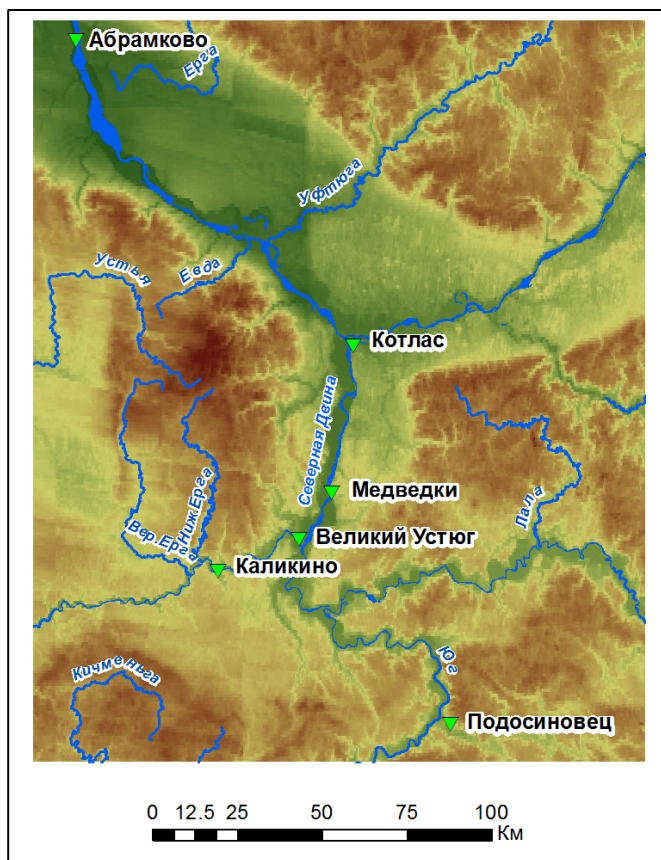


Рис. 15. Схема расположения речных постов на р. Северная Двина и ее притоках

Табл. 2. Классификация периода наблюдения

Сезон	91	92	93	94	95	96	97	98	99	00	01	02	03	04	05	06	07	08	09	10
№ класса	1	1	2	2	2	1	2	1	2	2	1	1	2	1	1	2	2	1	1	1

Процедура построения решения задачи состоит из двух этапов. На первом этапе производится обучение системы – определение наборов пороговых значений признаков, обеспечивающих наилучшую классификацию. Для этого используется обучающая выборка – набор объектов, для которых результирующее состояние известно. Затем составляется решающее правило, на основании которого каждый новый, поступающий на вход системы, набор может быть отнесен к одному из классов состояний. Решающее правило содержит зависимость от набора свободных параметров, и процесс обучения представляет процедуру определения таких значений этих параметров, которые обеспечивают наилучшую классификацию наборов состояний. На втором этапе решающее правило с подобранными на основе обучения параметрами используется для классификации наборов, не входящих в

обучающую выборку, т.е. для прогнозирования. Оценка качества осуществляется с помощью Leave-one-Out кросс-валидации.

Первоначально, описанная система была апробирована на периоде наблюдений 1991-2010, оценка качества прогнозирования составила 85% (Табл. 3.).

Табл. 3. Оценка качества прогнозирования за период 1991-2010

Сезон	1991	1992	1993	1994	1995	1996	1997	1998	1999	2000	2001	2002	2003	2004	2005	2006	2007	2008	2009	2010
№ класса	1	1	2	2	2	1	2	1	2	2	1	1	2	1	1	2	2	1	1	1
Результат прогнозирования	1	1	1	2	2	1	2	1	2	2	1	1	2	1	1	1	2	2	1	1

При получении новых данных за 2011-2016 гг. были проделаны два эксперимента имитирующие применение системы в реальных условиях. В первом случае для составления прогноза на каждый из шести добавленных сезонов использован набор параметров, полученных ранее (при обучении на данных 1991-2010 гг.). Во втором случае для составления прогноза на добавленный сезон производилось полное переобучение системы: заново строились наборы параметров, обеспечивающих необходимое качества прогноза. Затем процесс повторялся для следующего периода. Оба подхода дали одинаковый результат, состоящий в успешном прогнозировании образования заторов для всех шести новых сезонов. Результат представлен в Табл 4.

Построенная система позволяет не только прогнозировать мощность опасного явления, но и является также инструментом для исследований в части определения важности признаков, влияющих на итоговый результат. Прямое сравнение признаков в сложных задачах с небольшим объемом доступных наблюдений не отражает реальной картины зависимости. Для решения подобных проблем используется естественная мера значимости признака по вкладу в итоговый результат классификации – информационный вес признака. Упорядочивание информационных весов позволяет разбить признаки на группы: ведущие, значимые, незначимые.

Табл. 4. Оценка качества прогнозирования за период 2011-2016

Сезон	2011	2012	2013	2014	2015	2016
№ класса	1	2	1	1	1	1
Результат прогнозирования	1	2	1	2	2	1

Проведено исследование признакового пространства задачи прогнозирования ледового заторообразования на реке Северная Двина. По результатам обучения системы произведен расчет информационных весов признаков. Итоговый результат нормирован и приведен к 100%. Ранжирование признаков по убыванию вклада в итоговый результат классификации приведено на Рис. 16. Аналогичным образом исследованы зависимости признаков между собой. В результате видно, что, хотя и есть некоторые корреляции отдельных признаков, признаковое пространство подобрано хорошо. Все признаки влияют

на итоговый результат, нет сильно зависимых признаков. Прделана качественная экспертная работа, результаты согласуются с ранее выдвинутыми гипотезами. Признаки 11, 10, 9, 4, 5, 6 вносят наибольший вклад в итоговый результат, что, в целом, соотносится с теоретическими исследованиями.



Рис. 16. Вклад признаков в итоговый результат классификации.

Далее рассмотрено применение разработанной прогнозной интеллектуальной системы на данных нового региона – бассейна р. Лена. Прогнозная система решает задачу бинарной классификации методами машинного обучения в условиях неполноты исходных наблюдений, включая наличие пропусков данных. Применение разработанной технологии к данным другого региона обусловлено схожей постановкой задачи прогнозирования. Имеется возможность сформулировать гипотезы о зависимости ряда интегральных показателей (признаков) на итоговый результат явления (целевую переменную) по данным измерений гидрологических и метеорологических показателей на гидропостах и метеостанциях в бассейне р. Лена.

В качестве исходных данных для бассейна Лены использованы два новых набора гидрологических *Гидрология по реке Лена и притокам, 2021* и метеорологических *Метеорология района реки Лена и притоков, 2021* данных в регионе за период 1985-2019 гг. Указанные наборы содержат данные по 27 гидропостам и 38 метеостанциям в регионе. В

качестве основных факторов возникновения заторообразования выбрано 7 интегральных признаков, построенных на имеющихся исходных данных: среднесуточный уровень воды, среднесуточная температура воды, толщина льда перед вскрытием, среднесуточный расход воды, сумма твердых осадков, запас воды в снеге, среднесуточная температура воздуха. На основании анализа статистических данных доступна итоговая интегральная классификация сезонов по критерию наличия (1 класс) или отсутствия (2 класс) мощных и продолжительных заторов в регионе в указанный сезон.

Особенностью данной задачи является необходимость адаптации метео данных под координаты гидропостов. Поскольку пространственная привязка измерений на метеостанциях и гидропостах различная, возникает задача по переносу измерений с метеостанций на гидропосты, координатами которых индексированы входные данные для прогнозирования. Задача была решена с помощью интерполяции методами машинного обучения. Использован алгоритм случайного леса.

Для обучения системы было проведено 64 тыс. итераций Монте-Карло, оценена точность с использованием кросс-валидации. Лучший результат дал оценку точности классификации порядка 76% по функционалу качества accuracy, т.е. 26 из 34 сезонов корректно классифицированы системой к своим классам.

Была проведена оценка важности факторов. Основное влияние оказывает группа гидрологических факторов, что соответствует ранее полученным результатам для Северной Двины. Среди гидрологических факторов наименьшую оценку важности получил среднесуточный расход воды. Вероятно, это связано с существенной нестационарностью этой характеристики по времени, выявленной ранее при анализе многолетних рядов гидрологических наблюдений в регионе. Группа метеорологических факторов показала достаточно слабый отклик на мощность заторного явления.

Выводы по главе 3

1. Разработана интеллектуальная система, предназначенная для краткосрочного прогноза образования заторов в весенний период на северных реках на основе анализа коротких временных данных с применением специализированных методов теории машинного обучения и распознавания образов.
2. Система позволяет проверять гипотезы о влиянии признаков на итоговую мощность явления в условиях ограниченного набора исходных наблюдений на гидропостах и метеостанциях. Построена иерархия используемых признаков по

степени влияния результат, выделен группы ведущих и значимых признаков. Полученные результаты согласуются с ранее проведенными исследованиями.

3. Проведена оценка достоверности прогнозов на периоде разработки (20 сезонов), и валидация на расширенном периоде (26 сезонов). Оцененная достоверность прогнозирования для этих периодов согласуется и для региона бассейна р. Северная Двина составляет 85%.
4. Разработанная интеллектуальная система применена к другому региону – бассейну р. Лена. Оцененная достоверность прогнозирования для нового региона составляет 76%. Основные качественные выводы соответствуют полученным ранее результатам.

ЗАКЛЮЧЕНИЕ

В результате проведенных исследований и практических разработок **была достигнута цель диссертационного исследования** – осуществлена разработка компьютерных систем и методов обработки данных в условиях ограниченного количества данных, не достаточных для проведения классического статистического анализа, на основе методов машинного обучения, включая интерпретацию результатов измерений геофизических полей и анализа многопараметрической информации, применение таких систем для построения геолого-геофизических моделей и решения задач охраны окружающей среды.

Проведен анализ методов машинного обучения в аспекте применения к геофизическим задачам: использование базовых методов, разработка более глубоких подходов. Приведены примеры применения методов машинного обучения в геофизических задачах с недостатком данных. Введены базовые понятия и алгоритмы машинного обучения. Проведена их необходимая адаптация для возможности применения к исследуемым геофизическим примерам. Показаны основные методы решения задачи о восстановлении пропусков данных.

В задаче построения трехмерной модели среды при проведении межскважинных исследований предложена новая интерпретация данных радиоволнового просвечивания, позволяющая более точно выделить границы слоев по сравнению с методами, используемыми ранее (кригинг). Использованный метод k ближайших соседей позволяет построить трехмерную модель проводимости среды между скважин даже при использовании синхронной схемы измерений. Влияние анизотропии распределения данных можно исключить, если модифицировать пространственную метрику, определяющую расстояние

между данными. Это достигается введением коэффициента скейлинга, который изменяет масштаб в горизонтальном направлении. Используемый подход позволяет также получить достаточно контрастное изображение неоднородных областей, что позволяет выделить неоднородности, чьи геометрические размеры меньше расстояния между скважинами. Процесс построения трехмерной модели фактически не зависит от физической модели, использованной для интерпретации измерений. Уточнение физической модели процесса распространения радиоволн между скважинами позволит улучшить качество построения образа. Модель может быть улучшена, если привлечь дополнительные данные (геологические, сейсмические, магнитные) для их совместной интерпретации.

Разработан метод анализа пространственных данных для построения двумерных и трехмерных (на основе набора 1D-профилей) изображений на основе алгоритма k ближайших соседей. Разработанный метод применен в задаче построения двумерных моделей строения коры северной части Балтийского щита. Построена уточненная карта поверхности Мохоровичича. Основу исследования составляют данные, полученные методом приемных функций. Были использованы сведения, полученные в предыдущих исследованиях этого региона, дополненные новыми расчетами и измерениями. Исходные данные представляют собой набор зависимостей сейсмической скорости от глубины, рассчитанных для более чем 60 постоянных и временно действующих геофизических станций. С точки зрения машинного обучения, данная задача является задачей регрессии. Для восстановления регрессионной зависимости глубины Мохо от двумерных координат был использован метод k ближайших соседей с необходимой адаптацией в части выбора метрики.

Еще одним результатом в данной задаче стало построение карты слоя с низкими значениями скорости поперечных сейсмических волн V_s . В исследуемом регионе практически отсутствует осадочный слой. Несмотря на это, имеются области, в которых присутствует слой с низкими значениями скорости V_s . Относительно низкие значения V_s обычно объясняют наличием в слое большого количества водонасыщенных трещин. Присутствие такого слоя не зависит от возраста пород. Эта задача относится к оценке принадлежности в рамках задачи бинарной классификации. Для небольшого количества сейсмических станций (порядка 20) известно наличие или отсутствие слоя низких скоростей. С помощью метода k ближайших соседей в каждой точке исследуемого региона оценена вероятность наличия слоя низких скоростей. Построенная карта включает в себя классификацию по принципу наличия или отсутствия слоя низких скоростей, а также буферную область, в которой на основании имеющихся данных нельзя сделать однозначный

вывод. Показано, что слой низких сейсмических скоростей на поверхности присутствует на значительной части региона, включая области с протерозойскими породами. В южной части Финляндии положение низкоскоростной области коррелирует с относительно низким значением толщины коры.

В результате сочетания новых разработок для трехмерного локального и двумерного регионального моделирования геофизических полей был разработан метод пространственной интерполяции нерегулярно распределённых геофизических измерений с пропусками данных, основанный на применении теории машинного обучения. Метод применим для данных с экстремальной анизотропией пространственного распределения. Разработанный метод является универсальным, и не зависит от способа получения одномерных моделей. Необходимые вычисления выполняются непосредственно для нужных сечений, при этом форма сечений может быть произвольной. В качестве его реализации была построена трехмерная региональная модель коры для южной Фенноскандии.

Разработан метод для анализа многомерных временных рядов ограниченной длины. Техническая реализация метода выполнена на основе прогнозной интеллектуальной системы для процесса ледяного заторообразования на северных реках. Создание прогнозных систем, включающих в себя методы теории искусственного интеллекта, является актуальным развитием геоинформационных систем, а задача прогнозирования опасных природных явлений является востребованной в любой отрасли хозяйственной деятельности человека. Прогнозная система, основанная на разработанной методике, качественно проявила себя при решении трудно-формализуемой задачи прогнозирования опасного природного явления – образование заторов льда на участках р. Северная Двина и р. Лена. Система также допускает применение в качестве инструмента проверки гипотез относительно влияния тех или иных факторов на итоговые состояния опасных процессов в условиях дефицита данных, когда результатам традиционных методов анализа данных не хватает статистической значимости. Проведена валидация системы на новых, недоступных в момент первоначальной разработки, исторических данных. После этого, система применена на новом, отличном от первоначального использованного для разработки системы, регионе – бассейне р. Лена. Итоговая оцененная достоверность прогнозирования составила от 76% до 85%.

ОСНОВНЫЕ ПУБЛИКАЦИИ ПО ТЕМЕ ДИССЕРТАЦИИ

Публикации в журналах, включенных в перечень российских рецензируемых научных журналов и изданий для опубликования основных научных результатов диссертации

1. Алешин И.М., Козловская Е.Г., **Малыгин И.В.** Применение методов теории машинного обучения в томографии приемных функций // Геофизические исследования. — 2022. — Т.23, №1. — С. 49–61.
2. Алешин И.М., Ваганова Н.В., Косарев Г.Л., **Малыгин И.В.** Свойства коры Фенноскандии по результатам kNN-анализа инверсии приемных функций // Геофизические исследования. — 2019. — Т.20, №4. — С. 25–39.
3. Алешин И.М., **Малыгин И.В.** Интерпретация результатов радиоволнового просвечивания методами машинного обучения // Компьютерные исследования и моделирование. — 2019. — Т. 11, № 4. — С. 675–684.
4. Алешин И.М., **Малыгин И.В.** Верификация экспертной системы прогноза заторообразования на Северной Двине // Геофизические процессы и биосфера. — 2018. — Т. 17, № 2. — С. 48–60.
5. **Малыгин И.В.** Логический подход к созданию экспертных систем прогнозирования опасных природных явлений // Естественные и технические науки. — 2015. — № 2. — С. 102–112.
6. **Малыгин И.В.** Методика прогноза образования ледовых заторов на реках на основе теории распознавания образов // Вестник Московского университета. Серия 5: География. — 2014. — № 3. — С. 43–47.
7. **Малыгин И.В.** О задаче прогнозирования ледовых заторов // Интеллектуальные системы. Теория и приложения. — 2014. — Т. 18, № 3. — С. 73–80.
8. Aleshin I.M., **Malygin I.V.** Machine learning approach to inter-well radio wave survey data imaging // Russian Journal of Earth Sciences. — 2019. — V. 19, no. ES3003. — P. 1–6.
9. Aleshin I.M., **Malygin I.V.** Verification of an expert system for forecasting ice-block-formation: The case of the Northern Dvina river // Izvestiya - Atmospheric and Oceanic Physics. — 2018. — V. 54, №8. — P. 898–905.

Авторские свидетельства

10. **Малыгин И.В.** Свидетельство о государственной регистрации программы для ЭВМ №2014614960 Экспертная система прогнозирования ледового заторообразования. Дата гос. регистрации в Реестре программ для ЭВМ 14.05.2014.
11. **Малыгин И.В.,** Алешин И.М. Свидетельство о государственной регистрации программы для ЭВМ № 2020617961 Программа расчёта и построения региональных карт геофизических свойств методом k-ближайших соседей. Дата гос. регистрации в Реестре программ для ЭВМ 15.07.2020.

12. **Малыгин И.В.**, Алешин И.М. Свидетельство о государственной регистрации программы для ЭВМ № 2020617962 Программа расчёта и построения трехмерной модели проводимости среды по данным межскважинных измерений методом к-ближайших соседей. Дата гос. регистрации в Реестре программ для ЭВМ 15.07.2020.

Прочие публикации

13. **Малыгин И.В.** Формирование параметров обучения в прогнозных экспертных системах // Наука и мир. — 2013. — № 3. — С. 34–35.