

Федеральное государственное автономное образовательное учреждение
высшего образования «Уральский федеральный университет имени первого
Президента России Б.Н. Ельцина»

Уральский энергетический институт

Кафедра автоматизированных электрических систем

На правах рукописи



Снегирев Денис Алексеевич

**ПОВЫШЕНИЕ ЭФФЕКТИВНОСТИ
КРАТКОСРОЧНОГО ПРОГНОЗИРОВАНИЯ ГЕНЕРАЦИИ
ВЕТРОЭЛЕКТРИЧЕСКИХ СТАНЦИЙ**

2.4.3. Электроэнергетика

Диссертация на соискание ученой степени

кандидата технических наук

Научный руководитель
доктор технических наук,
профессор А.В. Паздерин

Екатеринбург, 2025

ОГЛАВЛЕНИЕ

ВВЕДЕНИЕ.....	5
ГЛАВА 1. АКТУАЛЬНЫЕ ВОПРОСЫ ПРОГНОЗИРОВАНИЯ	
 ГЕНЕРАЦИИ ВЕТРОЭЛЕКТРИЧЕСКИХ СТАНЦИЙ.....	14
1.1. Динамика развития ветроэлектростанций в России и в мире	14
1.2. Интеграция ветроэлектростанций в энергосистемы	18
1.3. Область применения прогнозов генерации ВЭС	24
1.4. Особенности краткосрочного прогнозирования выработки ВЭС	31
1.4.1. Постановка задачи и основные этапы.....	31
1.4.2. Критерии для оценки эффективности краткосрочного прогнозирования выработки ВЭС	36
1.4.3. Влияние различных факторов на мощность и выработку ветроэлектрической установки.....	39
1.4.4. Основные источники погрешности прогноза.....	44
1.5. Обзор основных подходов к повышению эффективности прогнозирования генерации ВЭС	47
1.6. Выводы по главе.....	59
ГЛАВА 2. ОПТИМИЗАЦИЯ ГИПЕРПАРАМЕТРОВ МОДЕЛЕЙ	
 МАШИННОГО ОБУЧЕНИЯ ДЛЯ КРАТКОСРОЧНОГО	
 ПРОГНОЗИРОВАНИЯ ГЕНЕРАЦИИ ВЕТРОЭЛЕКТРОСТАНЦИЙ	62
2.1. Постановка задачи оптимизации гиперпараметров	62
2.2. Модели машинного обучения и их применение для краткосрочного прогнозирования генерации ВЭС.....	64
2.2.1. Регрессия на основе гауссовских процессов.....	64
2.2.2. Нейронная сеть прямого распространения.....	73
2.2.3. Метод опорных векторов	78
2.2.4. Регрессионное бинарное дерево решений	83

2.2.5. Ансамблевые методы: случайный лес и градиентный бустинг .	87
2.3.Подходы к оптимизации гиперпараметров и их применение для краткосрочного прогнозирования генерации ВЭС.....	94
2.4.Вычислительные эксперименты	106
2.5.Выводы по главе.....	114
ГЛАВА 3. ОПРЕДЕЛЕНИЕ СПОСОБА АГРЕГАЦИИ МОЩНОСТИ ОТДЕЛЬНЫХ ВЕТРОУСТАНОВОК ПРИ КРАТКОСРОЧНОМ ПРОГНОЗИРОВАНИИ ГЕНЕРАЦИИ ВЕТРОЭЛЕКТРОСТАНЦИЙ	116
3.1.Общие подходы к агрегации мощности отдельных ветроустановок ..	116
3.2.Методики объединения ветроустановок в группы	120
3.3.Вычислительные эксперименты	131
3.4.Выводы по главе.....	138
ГЛАВА 4. ПРЕДВАРИТЕЛЬНАЯ ОБРАБОТКА ДАННЫХ ДЛЯ КРАТКОСРОЧНОГО ПРОГНОЗИРОВАНИЯ ГЕНЕРАЦИИ ВЕТРОЭЛЕКТРОСТАНЦИЙ	141
4.1.Идентификация выбросов на характеристиках мощности ветроустановок и ветроэлектростанций	141
4.2.Методики идентификации выбросов по характеристике мощности...	149
4.3.Методики учета состава работающих ветроустановок.....	165
4.4.Вычислительные эксперименты	168
4.5.Итоговые результаты исследований	175
4.6.Выводы по главе.....	181
ЗАКЛЮЧЕНИЕ	183
ПЕРЕЧЕНЬ СОКРАЩЕНИЙ И УСЛОВНЫХ ОБОЗНАЧЕНИЙ	188
СПИСОК ЛИТЕРАТУРЫ	195
ПРИЛОЖЕНИЕ А ОПТИМИЗАЦИЯ ГИПЕРПАРАМЕТРОВ	228
ПРИЛОЖЕНИЕ Б АКТ О ВНЕДРЕНИИ РЕЗУЛЬТАТОВ ДИССЕРТАЦИОННОЙ РАБОТЫ	234

ПРИЛОЖЕНИЕ В СВИДЕТЕЛЬСТВА О ГОСУДАРСТВЕННОЙ РЕГИСТРАЦИИ ПРОГРАММ ДЛЯ ЭВМ.....	235
---	------------

ВВЕДЕНИЕ

Актуальность темы исследования.

Развитие возобновляемых источников энергии (ВИЭ) во всем мире обосновывается соображениями экологии, энергетической безопасности, энергетической эффективности, всеобщей доступности, а также интенсификацией функционирования розничных и оптовых рынков электроэнергии с вовлечением большого количества субъектов. Одним из наиболее быстро растущих секторов возобновляемой энергетики является использование ветроэнергетических ресурсов. Всего в мире на конец 2024 года установленная мощность ветроэлектрических станций (ВЭС) составила 1 136 ГВт. За 2024 год прирост составил рекордные 114,9 ГВт. Практически половина установленной мощности ВЭС приходится на Китай, около 521 ГВт.

Установленная мощность ВЭС в России достигла 2 556 МВт на конец 2024 года, чуть более 1% от установленной мощности всех электростанций Единой энергетической системы (ЕЭС) России. За последние 5 лет было введено более 2 ГВт установленной мощности ВЭС в рамках программы поддержки развития возобновляемой энергетики с помощью договоров о предоставлении мощности (ДПМ) ВИЭ. По разным оценкам, до 2035 года может быть введено до 5 ГВт ВЭС для достижения целевого показателя объема производства электрической энергии с использованием возобновляемых источников энергии (кроме гидроэлектростанций установленной мощностью более 25 МВт) в 4,5 %, закрепленного в распоряжении Правительства Российской Федерации от 08.01.2009 №1-р.

В объединенной энергосистеме Юга доля установленной мощности ВЭС уже достигла практически 8 %, а в отдельных регионах, например Республике Калмыкия, около 45 %. При такой значительной доле стохастической генерации традиционный подход к планированию балансов мощности, в том числе при краткосрочном планировании режимов работы электроэнергетических энергосистем (ЭЭС), должен дополняться прогнозированием выработки ВЭС. Чем

больше доля ВЭС, тем большее влияние будет оказывать точность прогноза на надежную и экономически эффективную работу ЭЭС.

Под краткосрочным прогнозированием генерации ВЭС понимается определение средней выдаваемой мощности на каждый час следующих суток, что эквивалентно выработке электрической энергии за соответствующий час суток. Основной областью применения таких прогнозов являются расчеты планов выработки для участия ВЭС в оптовом рынке электроэнергии «на сутки вперед» (РСВ) в соответствии с регламентами оптового рынка. На сегодняшний день основную часть дохода собственники ВЭС получают на рынке мощности, при этом требуется обеспечивать определенный коэффициент использования установленной мощности за отчетный период. При работе на РСВ ВЭС подают ценопринимающие заявки, поэтому при выполнении оптимизационного расчета для формирования планового диспетчерского графика данный тип генерации оказывается более приоритетным, чем генерация традиционного типа, которая подает ступенчатые ценовые заявки. Отклонение фактической выработки электроэнергии (ЭЭ) от плановой приводит к возникновению убытков, в том числе упущенной выгоды, при работе ВЭС на оптовом рынке. Отклонения от планового почасового производства могут приводить к убыткам и для покупателей ЭЭ из-за возникновения незапланированного стоимостного небаланса между суммарными требованиями и обязательствами поставщиков и потребителей. Он распределяется между всеми участниками рынка по достаточно сложным правилам.

В случае, если фактическая выработка ВЭС отклоняется выше заявленного на РСВ объема, то за отклонения ВЭС получает оплату по индикаторам балансирующего рынка (БР) вместо цены РСВ. Причем за значительные отклонения более 10% от установленной мощности оплата может быть снижена до 1 руб/МВт·ч, кроме того увеличивается доля стоимостного небаланса, который будет распределен на ВЭС.

Если же фактическая выработка ВЭС отклоняется ниже заявленного на РСВ объема, то для компенсации отклонений ВЭС вынуждена покупать электроэнергию

по индикаторам БР. Значительные отклонения также увеличивают долю стоимостного небаланса.

Таким образом, разработка и внедрение методик повышения точности краткосрочного прогнозирования генерации ВЭС, обеспечивает для:

- собственников ВЭС – снижение убытков и упущенной выгоды, связанных с отклонением фактической выработки от планового почасового производства;
- участников оптового рынка, в том числе потребителей – снижение затрат на компенсацию стоимостного небаланса;
- системного оператора – повышение надежности функционирования энергосистем с ВЭС.

Степень научной разработанности темы исследования.

Анализ мирового опыта в области краткосрочного прогнозирования выработки ВЭС показывает, что единого универсального способа для решения данной задачи на текущий момент нет. Используемые методики прогнозирования во многом носят индивидуальный характер, который определяется составом и качеством доступных исходных данных, полученных от метеопровайдеров и собранных с помощью SCADA на самой ВЭС. Значительное внимание в научных публикациях по данной теме уделяется выбору моделей и алгоритмов машинного обучения. В меньшей степени уделяется внимание трансформации и декомпозиции исходных данных, отбору информативных признаков, идентификации выбросов. Ещё меньшее количество исследований посвящено изучению стратегий агрегации мощности отдельных ветроустановок (ВЭУ), а также оптимизации гиперпараметров моделей машинного обучения применительно к задаче краткосрочного прогнозирования выработки ВЭС. Практически отсутствуют исследования, связанные с учетом состава работающих ветроустановок.

Объект исследования представляет собой энергосистемы, к которым присоединены ВЭС, а также временные ряды, используемые для обучения и тестирования моделей прогнозирования: измерений генерации отдельных ВЭУ, групп ВЭУ и всей ВЭС, прогнозных и фактических метеоданных.

Предметом исследования является повышение точности краткосрочного прогнозирования выработки ветроэлектрических станций с помощью оптимизации гиперпараметров моделей машинного обучения, объединения турбин в группы, исключения недостоверных наблюдений и учета состава работающих ВЭУ.

Целью работы является совершенствование методик краткосрочного прогнозирования выработки ветроэлектрических станций для снижения погрешности прогноза.

Для достижения данной цели в работе решаются следующие **задачи**:

1. Анализ существующих подходов к оценке и повышению эффективности краткосрочного прогнозирования генерации ВЭС.
2. Сравнение существующих и разработка новых методик оптимизации гиперпараметров моделей машинного обучения применительно к задаче краткосрочного прогнозирования выработки ВЭС.
3. Сравнение стратегий агрегации мощности отдельных ветроустановок и разработка нового способа объединения ветроустановок в группы.
4. Повышение эффективности методов предварительной обработки данных (исключение недостоверных наблюдений).
5. Повышение эффективности учета состава работающих ВЭУ.

Научная новизна работы состоит в следующем:

1. Предложено два новых критерия для оценки эффективности краткосрочного прогнозирования генерации ВЭС: нормализованная суммарная абсолютная ошибка, показывающая суммарную ошибку прогноза в выработке электрической энергии, и вероятность охвата интервала ошибки в заданном диапазоне, которая может использоваться для оценки надежности получаемых прогнозов.
2. Предложены новые методики подбора гиперпараметров моделей машинного обучения, основанные на использовании байесовской оптимизации с функциями выбора «вероятность улучшения» и «нижний доверительный интервал», обеспечивающие снижение погрешности краткосрочного прогнозирования генерации ВЭС.

3. Разработан альтернативный способ объединения турбин в группы с помощью метода иерархической агломеративной кластеризации по расстоянию, определенному с помощью алгоритма динамической трансформации временной шкалы. Предложенный способ позволил снизить погрешность прогнозирования выработки ВЭС на уровне групп турбин.

4. Предложена новая более эффективная с точки зрения снижения погрешности прогноза генерации ВЭС методика идентификации выбросов по характеристикам мощности с использованием алгоритма глобально-локальной оценки выбросов.

5. Разработана новая методика учета состава работающих ВЭУ, основанная на трансформации отклика регрессионных моделей, обеспечивающая снижение погрешности прогнозирования генерации на уровне ВЭС и групп турбин.

Теоретическая значимость работы состоит в развитии методик повышения эффективности краткосрочного прогнозирования выработки ВЭС в области оптимизации гиперпараметров моделей машинного обучения, определения способа агрегации мощности отдельных ВЭУ, идентификации выбросов по характеристике мощности и учета состава работающих ВЭУ.

Практическая значимость работы подтверждается положительными результатами оценки эффективности предложенных методик в задаче краткосрочного прогнозирования генерации ВЭС при использовании реальных данных, полученных с находящейся в промышленной эксплуатации ВЭС. Исследования по данной теме выполнялась в рамках совместной научно-исследовательской работы УрФУ и компании «Прософт-Системы» (договор № 21-53810 и договор 22-59892).

Практическая ценность исследования также подтверждается актом внедрения. Результаты работы были использованы в ООО "Прософт-Системы" при разработке модуля краткосрочного прогнозирования выработки ветроэлектрических станций для программного комплекса «Энергосфера».

Положения, выносимые на защиту.

1. Методики подбора гиперпараметров моделей машинного обучения для краткосрочного прогнозирования выработки ВЭС, основанные на использовании

байесовской оптимизации с функциями выбора «вероятность улучшения» и «нижний доверительный интервал».

2. Стратегия прогнозирования суммарной мощности ВЭС по группам ВЭУ, при этом объединение в группы выполняется методом иерархической агломеративной кластеризации по расстоянию, определенному с помощью алгоритма динамической трансформации временной шкалы.

3. Методика идентификации выбросов по характеристикам мощности ВЭС, групп турбин и отдельных турбин с использованием алгоритма глобально-локальной оценки выбросов по иерархической основанной на плотности пространственной кластеризации для приложений с шумами.

4. Методика учета состава работающих ветроустановок с помощью трансформации отклика регрессионных моделей.

Методология и методы исследования.

При выполнении настоящей работы были применены методы математического моделирования, математического анализа, математической статистики и машинного обучения. Для обработки данных, проведения вычислительных экспериментов (ВЭ) и визуализации результатов используется программное обеспечение, разработанное автором на языках *MATLAB* с использованием пакетов *MATLAB*, *Statistics and Machine Learning Toolbox*, *Signal Processing Toolbox*, и *Python* с использованием библиотек *matplotlib*, *kneed*, *hdbscan*.

Соответствие диссертации паспорту научной специальности.

Диссертационная работа соответствует специальности

2.4.3. Электроэнергетика, так как область исследования соответствует пункту:

«20. Разработка методов использования информационных и телекоммуникационных технологий и систем, искусственного интеллекта в электроэнергетике, включая проблемы разработки и применения информационно-измерительных, геоинформационных и управляющих систем для оперативного и ретроспективного мониторинга, анализа, прогнозирования и управления электропотреблением, режимами, надежностью, уровнем потерь энергии и качеством электроэнергии», паспорта специальности.

Обоснованность и достоверность полученных в работе результатов обеспечена корректной реализацией и применением методов машинного обучения, а также методологией проведения вычислительных экспериментов.

Апробация диссертационной работы.

Основные положения диссертационной работы докладывались и обсуждались на конференциях:

1. II Международная научно-техническая конференция «*Smart Energy Systems 2021*» (*SES-2021*), Казань – 2021;
2. Международная научно-практическая конференция студентов, аспирантов и молодых ученых – Даниловские чтения, Екатеринбург – 2022, 2023;
3. Международная научно-техническая конференция *Belarusian-Ural-Siberian Smart Energy Conference*, Екатеринбург – 2023;
4. VI Международная научно-техническая конференция «Релейная защита и автоматика», Москва – 2023;
5. *22nd International Conference on Renewable Energy and Power Quality (ICREPQ'24)*, Бильбао, Испания – 2024.

Публикации по теме исследования.

Основное содержание работы отражено в 8 научных работах, из них 5 – в рецензируемых научных журналах, определенных ВАК РФ и Аттестационным советом УрФУ, в том числе 2 в изданиях, входящих в международную базу цитирования *Scopus*; 2 свидетельства о регистрации программы для ЭВМ.

Личный вклад автора заключается в разработке новых методик повышения эффективности краткосрочного прогнозирования выработки ветроэлектрических станций, а также разработке программного обеспечения, с помощью которого были проведены вычислительные эксперименты и подтверждена эффективность предложенных методик.

Объем и структура диссертации

Диссертация состоит из введения, четырёх глав, заключения, списка литературы и приложений. Объём работы включает в себя 227 страниц печатного текста, 21 таблицу, 49 рисунков и 4 приложения.

В первой главе приведены данные о динамике развития ВЭС в России и в мире. Показано, что прогнозирование выработки ВЭС является одной из наименее затратных мер, позволяющих повысить эффективность планирования режимов в энергосистемах со значительной долей установленной мощности ВЭС. Рассмотрены основные особенности краткосрочного прогнозирования выработки ВЭС, в том числе постановка задачи, этапы и основные источники погрешности прогнозов. Проанализированы критерии для оценки эффективности прогнозов. Выполнен обзор основных подходов к повышению эффективности прогнозирования генерации ВЭС.

Во второй главе приведена постановка задачи оптимизации гиперпараметров. Кратко описаны рассматриваемые модели машинного обучения, а также существующие подходы к оптимизации гиперпараметров. Более подробно рассмотрен способ определения гиперпараметров с использованием байесовской оптимизации. Предложены две новые методики подбора гиперпараметров, основанные на использовании байесовской оптимизации с функциями выбора «вероятность улучшения» и «нижний доверительный интервал» применительно к задаче краткосрочного прогнозирования выработки ВЭС. Проведены вычислительные эксперименты, по результатам которых определено, что модель градиентного бустинга над деревьями решений обеспечивает наименьшую погрешность. Показано, что для ряда моделей машинного обучения предложенные методики оптимизации гиперпараметров позволяют снизить время оптимизации и значение целевой функции, а также получить набор гиперпараметров, обеспечивающий снижение ошибки прогноза по сравнению с другими рассмотренными подходами.

В третьей главе приведены основные причины различия в характере выработки мощности ветроустановок внутри одной ВЭС. Рассмотрены существующие стратегии агрегации мощности отдельных ВЭУ при получении суммарного прогноза по ВЭС: прогноз на уровне ВЭС, на уровне групп турбин и на уровне отдельных турбин. Показано, что в существующих исследованиях нет единого мнения о том, какой из подходов обеспечивает наибольшее снижение

погрешности прогноза. Описана существующая методика объединения турбин в группы с использованием кластеризации k -средних по усредненным данным. Предложены три новых методики разделения турбин на группы только по временным рядам мощности ветроустановок. Проведены вычислительные эксперименты, показавшие, что наибольшей точности прогноза позволяет добиться стратегия прогнозирования на уровне групп турбин. Наибольшего снижения погрешности удалось добиться с использованием предложенной методики объединения турбин в группы методом иерархической агломеративной кластеризации по расстоянию, определенному с помощью алгоритма динамической трансформации временной шкалы

В четвертой главе приведена существующая классификация типов выбросов на характеристике мощности. Описана существующая методика идентификации выбросов по суммарной характеристике мощности с использованием метода межквартильного размаха. Предложена новая методика, основанная на использовании алгоритма глобально-локальной оценки выбросов по иерархической основанной на плотности пространственной кластеризации для приложений с шумами. Также предложена новая методика для учета состава работающих ВЭУ с помощью трансформации откликов. Проведены вычислительные эксперименты, результаты которых позволяют судить об эффективности предложенных методик. Приведены итоговые результаты исследования, позволяющие оценить частный и суммарный эффект от предложенных в данном исследовании методик повышения точности прогнозирования генерации ВЭС.

ГЛАВА 1. АКТУАЛЬНЫЕ ВОПРОСЫ ПРОГНОЗИРОВАНИЯ ГЕНЕРАЦИИ ВЕТРОЭЛЕКТРИЧЕСКИХ СТАНЦИЙ

1.1. Динамика развития ветроэлектростанций в России и в мире

За последние 10 лет рост доли выработки электроэнергии электростанций, использующих ВИЭ, таких как ветроэлектрические, фотоэлектрические, гидроэлектрические, приливные станции и др. составил около 10%. На конец 2024 года выработка достигла 9 848 ТВт·ч, около 30% от всей выработки электроэнергии в мире [1, 2]. Традиционно среди основных причин бурного развития ВИЭ выделяют: экологичность, энергетическую безопасность, энергетическую эффективность, всеобщую доступность первичных энергоресурсов, а также интенсификацию функционирования розничных и оптовых рынков электроэнергии с вовлечением большего количества субъектов. Соображения экологичности производства электрической энергии обусловлены необходимостью снижения выбросов в атмосферу для противодействия эффекту глобального потепления. Целью мирового сообщества в этой области является ограничение уровня повышения средней температуры климатической системы Земли на величину не более 1,5 °C по отношению к доиндустриальным значениям. Данная цель была закреплена Парижскими соглашениями, определенными в рамках конвенции ООН по изменению климата. В течение 2023 г. и 2024 г. впервые в истории два года подряд наблюдалось превышение порога 1,5°C в течение каждого месяца года [3].

Одним из наиболее быстро растущих секторов возобновляемой энергетики является использование ветроэнергетических ресурсов. Считается, что первая установка, которая позволила преобразовывать энергию ветра в электрическую энергию, была разработана шотландским инженером Джеймсом Блитом в 1887 году. К сети переменного тока ВЭУ впервые была подключена в 1919 году компанией «Agrisco» в Дании. Установка имела мощность 40 кВт и представляла собой одну из модификаций турбин, разработанных датским инженером Полом ла Куром. Первая группа из восьми турбин, представляющая собой

ветроэлектрическую станцию, была построена в середине 1950-х годов в Германии компанией «Allgaier» [4].

Как правило, период с 1940-х по 1970-е годы рассматривается как период снижения темпов развития ветроэнергетики, что обусловлено, в первую очередь, широким распространением и повышением экономической эффективности централизованного электроснабжения, обеспечивающего потребителей электроэнергией вне зависимости от погодных условий по умеренной стоимости. При этом продолжалось развитие конструкций ветроэнергетических установок [5].

Возникший в 1973 году энергетический кризис привел к возвращению интереса к ветроэнергетике, показав зависимость от импорта энергетических ресурсов во многих странах, и, как следствие, уязвимость в энергетической безопасности. В странах Европы и США возникли программы стимулирования развития ветроэнергетики и ветроэнергетических технологий [6].

В начале 21 века проблемы энергетической безопасности, экономической эффективности и экологичности производства электрической энергии становятся все более актуальными. На рисунке 1.1 представлены ретроспективные данные по установленной мощности ветроэлектрических станций (ВЭС) в мире за 2001-2024 гг. [7–9], а также планируемые к вводу мощности до 2030 года [1, 2, 10]. Кроме того, на графике представлена фактическая и прогнозная доля выработки ВЭС в суммарной выработке электроэнергии в мире [1].

На рисунке 1.1 показано, что в 2001 году установленная мощность ВЭС в мире составляла 23,8 ГВт. Но на конец 2024 года установленная мощность ВЭС в мире уже превысила один тераватт и составила 1 136 ГВт. За последние 10 лет установленная мощность ВЭС увеличилась на 766 МВт. За 2024 год прирост составил рекордные 114,9 ГВт [8]. Наибольшая доля, практически половина, установленной мощности ВЭС на 2024 год приходится на Китай, около 521 ГВт [11]. В структуре выработки электроэнергии в мире доля ВЭС с 2014 года выросла почти в 3 раза и в 2024 году составила 8,2% [1].

По прогнозам Международного энергетического агентства (МЭА, *IEA*) до 2030 года установленная мощность ВЭС должна вырасти в 2,5 раза и достичь

2 051 ГВт [2] с целью достижения сценария, при котором количество производимых парниковых газов или углерода и количество удаляемых из атмосферы выбросов оказываются равными (*Net Zero scenario*). Доля выработки ВЭС при этом может достигнуть 15% от суммарной выработки электроэнергии в мире.

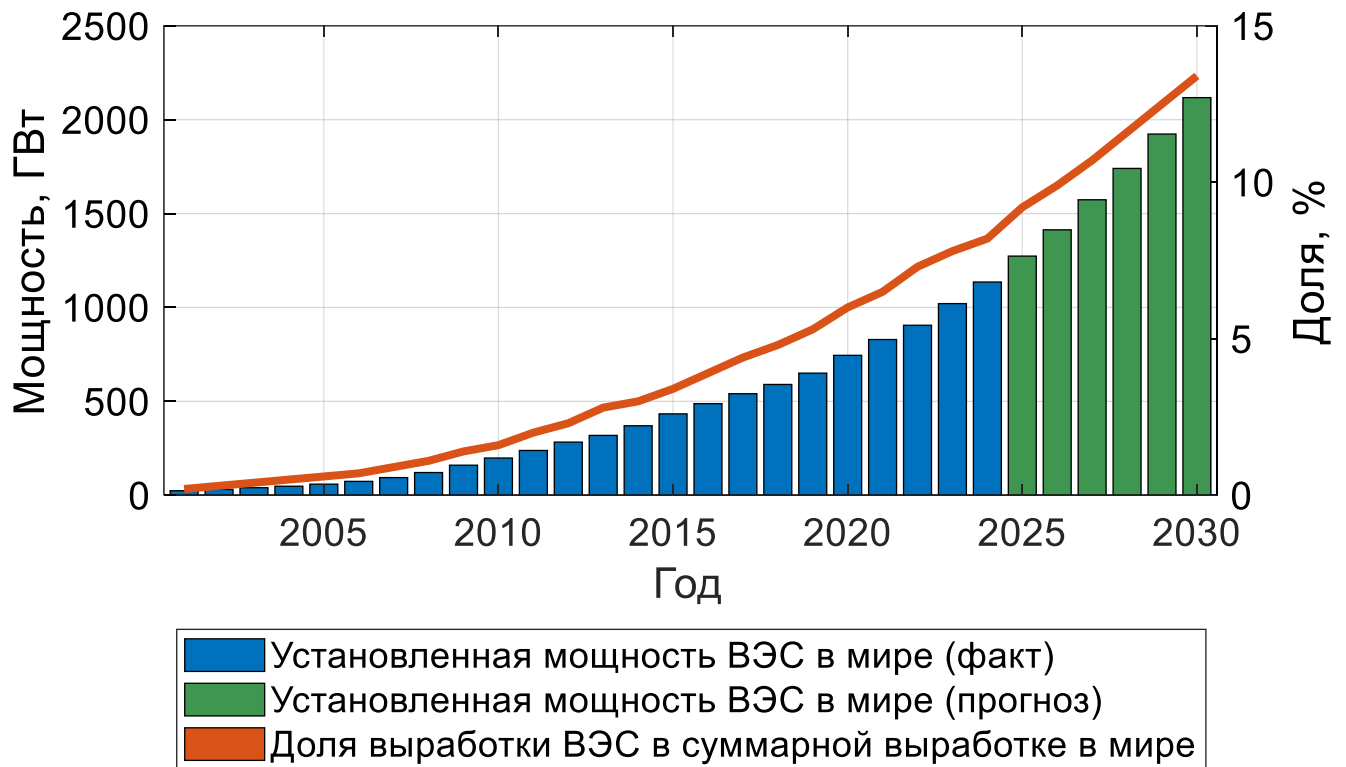


Рисунок 1.1 – Установленная мощность ВЭС в мире

В России первый ветрогенератор был разработан и сконструирован в 1924 г. сотрудниками Центрального аэрогидродинамического института (ЦАГИ) под руководством начальника отдела ветряных двигателей Николая Васильевича Красовского. Мощность установки составляла 36,8 кВт и использовалась для питания насосов на нефтяных промыслах. В 1931 г. впервые в СССР под г. Севастополем к сети переменного тока была подключена ветроустановка – Балаклавская ВЭС мощностью 100 кВт. До 1940 г. данная ветроустановка обладала наибольшей мощностью в мире. Технические решения, использованные при создании этой ветроустановки, во многом схожи с конструкцией современных ВЭУ [12].

В связи с массовой электрификацией и обеспечением централизованного электроснабжения потребителей в СССР развитие ветроэнергетики было приостановлено. Первая крупная ВЭС на постсоветском пространстве была построена в Крыму 1993 г. – Донузлавская ВЭС мощностью 6,78 МВт. Первая сетевая ВЭС в современной России – Куликовская ВЭС с установленной мощностью 5 МВт, была построена в Калининградской области [13].

Современный этап развития ветроэнергетики в России во многом связан с распоряжением Правительства РФ от 08.01.2009 N 1-р «Об основных направлениях государственной политики в сфере повышения энергетической эффективности электроэнергетики на основе использования возобновляемых источников энергии на период до 2035 года», которое устанавливает целевые показатели объема производства электрической энергии с использованием ВИЭ (кроме гидроэлектростанций установленной мощностью более 25 МВт): 4,5% в 2024 г. и не менее 6% в 2035 г [14]. Для достижения данного целевого показателя были разработаны меры поддержки развития ВИЭ в России как на оптовом рынке электроэнергии и мощности (ОРЭМ), так и розничном рынке электроэнергии.

Механизм стимулирования инвестиций в ВИЭ на ОРЭМ был определен в Постановлении Правительства РФ от 28.05.2013 N 449 «О механизме стимулирования использования возобновляемых источников энергии на оптовом рынке электрической энергии и мощности» [15], в соответствии с которым после отбора инвестиционных проектов с собственником квалифицированного генерирующего объекта заключается договор о предоставлении мощности, гарантирующий возврат инвестиционного капитала и получение доходности на ОРЭМ. О результатах выполнения программы ДПМ ВИЭ можно судить уже сейчас. За период 2024 г. удалось достигнуть значения показателя доли производства электроэнергии ВИЭ, равного 1,25% [16, 17]. Всего с 2014 по 2024 гг. установленная мощность ВИЭ выросла на 4 488 МВт, в том числе 2 465 МВт ВЭС [16]. В рамках продолжения программы ДПМ ВИЭ ожидается ввод еще минимум 6 ГВт мощности ВИЭ [18].

На рисунке 1.2 представлены ретроспективные данные по установленной мощности ВЭС в России за 2014-2024 гг. [19, 20], а также планируемые к вводу мощности до 2035 года [21, 22]. Установленная мощность ВЭС в России достигла 2 556 МВт на конец 2024 года, чуть более 1% от установленной мощности всех электростанций Единой энергетической системы (ЕЭС) России [23]. По результатам конкурсных отборов инвестиционных проектов ВИЭ к 2031 г. ожидается ввод в эксплуатацию ветроэлектрических станций с суммарной установленной мощностью 3,7 ГВт [21]. По прогнозам Ассоциации развития возобновляемой энергетики (АРВЭ) России к 2035 гг. установленная мощность ВЭС в России вырастет на 4,9 ГВт и составит 7,5 ГВт [22].

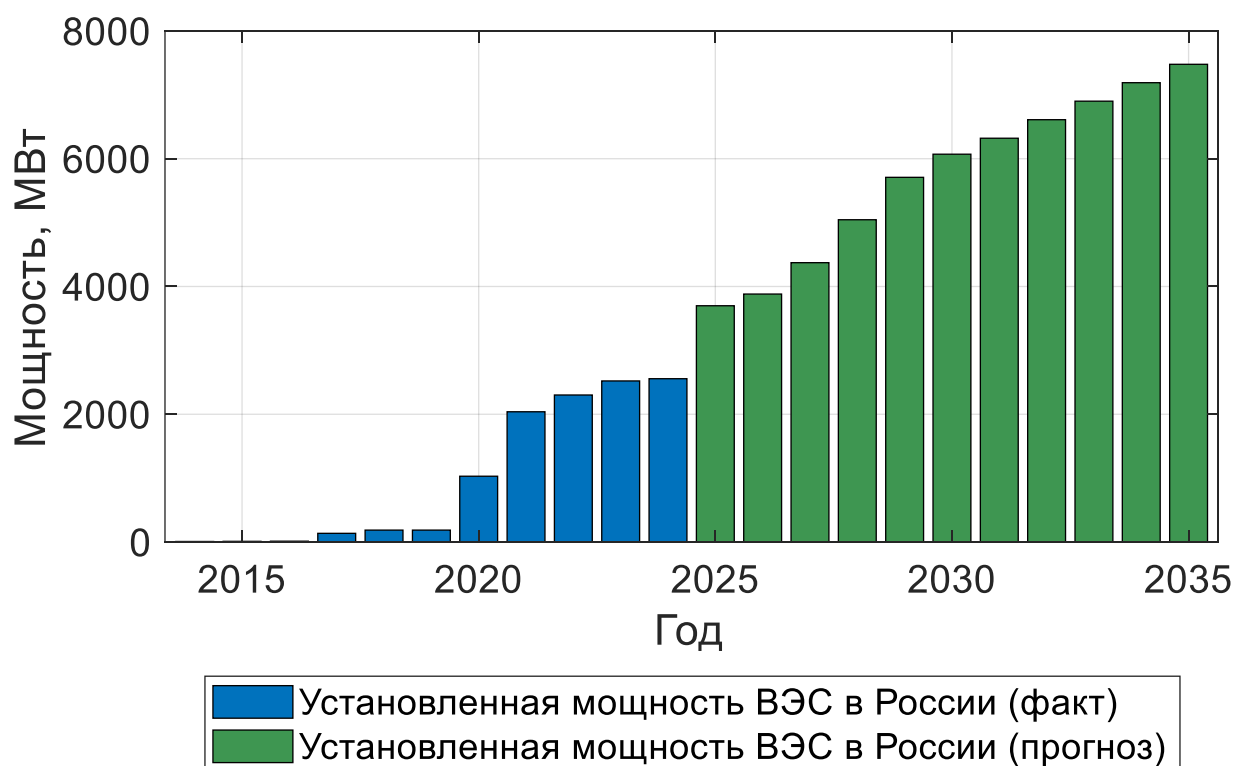


Рисунок 1.2 – Установленная мощность ВЭС в России

1.2. Интеграция ветроэлектростанций в энергосистемы

Внедрение ВИЭ с переменным характером выработки, в том числе ВЭС, в энергосистемы требует решения целого комплекса различных задач, как технического, так и организационного характера. Среди наиболее важных вопросов функционирования ВИЭ в составе энергосистем выделяют [24–29]:

1. Актуализация существующей и разработка новой нормативно-технической документации (НТД), регламентирующей требования к оборудованию ВИЭ, технологическому присоединению, функционированию и управлению режимами энергосистем, работе ВИЭ на рынке электроэнергии [30–33].

2. Обеспечение качества электроэнергии в прилегающей сети (коэффициентов гармонических составляющих и искажения синусоидальности напряжения, дозы фликера, установившегося отклонения напряжения) при подключении ВИЭ через устройства силовой электроники [34–37].

3. Поддержание напряжения на требуемом уровне при нормативных возмущениях для обеспечения устойчивой работы ВЭУ с помощью корректной настройки функции *LVRT* (*Low Voltage Ride Through*) инвертора [38–41].

4. Обеспечение динамической устойчивости энергосистемы при высокой доле генерации ВИЭ, подключенных к энергосистеме с помощью устройств силовой электроники и приводящих к снижению постоянной времени энергосистемы [42–46].

5. Адаптация релейной защиты в прилегающей сети для обеспечения селективности отключения ЛЭП, через которые осуществляется выдача мощности электростанций ВИЭ [47–51].

С увеличением доли установленной мощности стохастической генерации ВИЭ особое внимание уделяется вопросам обеспечения надежного функционирования энергосистем: балансу потребления и генерации, отсутствию нарушения режимных ограничений, в том числе перетоков мощности в контролируемых сечениях, наличию объемов доступных резервов активной мощности. Во многом успех решения данной задачи определяется точным планированием электроэнергетических режимов [50]. Экспертным сообществом выделяется пять фаз интеграции ВИЭ с переменным характером выработки в энергосистемы [24, 31, 52, 53]. В таблице 1.1 представлен обзор приблизительных воздействий и ответных мер при растущих долях ВИЭ с переменным характером выработки [54].

Первая фаза характерна для доли генерации ВИЭ от 0 до 3% (Индонезия, Россия, Египет и др.) [55], при этом, как правило, отсутствуют выраженные эффекты, связанные с воздействием данного типа генерации на работу энергосистемы в целом.

Вторая фаза наступает при доле ВИЭ от 3 до 13% (ЮАР, Мексика, Индия и др.) [55]. На данном этапе увеличение неопределенности выработки электростанций становится заметным, основными задачами является актуализация НТД в области управления и планирования режимов, разработка систем прогнозирования генерации ВИЭ.

На третьем этапе интеграции, когда доля достигает от 13 до 25 % (Италия, Австралия, Чили и др.) [55], влияние генерации с неопределенным характером выработки по отношению к другим типам электростанций и энергосистеме в целом становится значительным. Важной задачей становится повышение гибкости энергосистемы за счет ввода новых маневренных электростанций, усиления межсистемных связей и повышения эффективности систем управления режимами.

Четвертая фаза характеризуется долей порядка 25–50% (Великобритания, Германия, Испания и др.) [55]. В некоторых режимах генерация ВИЭ с переменным характером может покрывать все текущее потребление. На данном этапе наиболее приоритетной задачей оказывается обеспечение динамической устойчивости энергосистем в условиях снижения суммарной инерции.

Пятая фаза соответствует доли генерации с переменным характером выработки более 50% (Дания) [55] и возможности возникновения избытка генерации ВИЭ в часы минимума потребления, который необходимо искусственно ограничивать. Для решения данной проблемы может потребоваться разработка новых способов хранения электроэнергии.

Таблица 1.1 – Обзор приблизительных воздействий и ответных мер при растущих долях ВИЭ с переменным характером выработки

Доля генерации ВИЭ		0 % – 3 %	3 % – 13 %	13 % – 25 %	25 % – 50 %	более 50 %
Воздействия		Нет ощутимого воздействия	Небольшое увеличение неопределенности выработки заметно на уровне работы энергосистемы	Растущая неопределенность выработки оказывает значительное воздействие на работу энергосистемы	Повышенная неопределенность выработки имеет серьезное влияние на работу энергосистемы	Структурный избыток генерации ВИЭ с переменным характером выработки и сезонные энергетические дисбалансы
Необходимые меры		Нет дополнительных мер	Некоторые корректировки в работе энергосистемы и сетевой инфраструктуры	Значительные изменения в работе системы, необходимо усиление сети, повышение гибкости спроса и предложения	Сильные изменения в работе системы, существенное усиление сети	Дополнительные шаги для управления балансом спроса и предложения
Меры	Прогноз выработки					
	Усиление ЭС					
	Аккумуляирование					
	Управление спросом					
Примеры технологических и операционных мер		Разработка стандартов для ВИЭ с переменным характером выработки	Создание системы прогнозирования генерации ВИЭ, совершенствование систем планирования режимов	Улучшение методов прогнозирования, усиление межсистемных связей, уменьшение интервала планирования режимов	Значительное повышение масштаба управления спросом и хранения электроэнергии	Разработка новых технологий хранения электроэнергии

Несмотря на то, что в целом ЕЭС России находиться лишь на первом этапе внедрения генерации ВИЭ, в ряде регионов уже сейчас наблюдается довольно значительная установленная мощность генерации с переменным характером выработки, в частности, ветроэлектростанций. В таблицах 1.2 и 1.3 приведены данные об установленной мощности ВЭС в МВт и в процентах относительно установленной мощности энергосистем и регионов. Информация об установленной мощности ВЭС приведена по данным Ассоциации развития возобновляемой энергетики (АРВЭ) [19] и Реестра квалифицированных генерирующих объектов, функционирующих на основе использования возобновляемых источников энергии, формируемого «Центром энергосертификации» [20]. При расчете доли мощности ВЭС в суммарной мощности энергосистем использовались данные об установленной мощности объединенных энергосистем (ОЭС) и ЕЭС России из Отчетов о функционировании ЕЭС АО «СО ЕЭС» [23], а также информация об установленной электрической мощности электростанций по регионам России из информационно-статистической системы Росстата [56].

[illegible]

Таблица 1.3 – Установленная мощность ВЭС в региональных, объединенных и Единой энергосистемах России 2020-2024 гг.

Энергосистема	2020		2021		2022		2023		2024	
	МВт	%	МВт	%	МВт	%	МВт	%	МВт	%
ЕЭС России	1030,1	0,4	2039,0	0,8	2301,4	0,9	2521,4	1,0	2556,4	1,0
ОЭС Средней Волги	85,4	0,3	85,4	0,3	85,4	0,3	85,4	0,3	85,4	0,3
Ульяновская область	85,4	7,9	85,4	7,8	85,4	7,9	85,4	7,9	85,4	7,8
ОЭС Урала	5,3	0,0	5,3	0,0	5,3	0,0	5,3	0,0	5,3	0,0
Республика Башкортостан	1,7	0,0	1,7	0,0	1,7	0,0	1,7	0,0	1,7	0,0
Оренбургская область	3,6	0,1	3,6	0,1	3,6	0,1	3,6	0,1	3,6	0,1
ОЭС Северо-Запада	5,1	0,0	5,1	0,0	207,5	0,8	207,5	0,8	207,5	0,8
Калининградская область	5,1	0,3	5,1	0,3	5,1	0,3	5,1	0,3	5,1	0,3
Мурманская область	0,0	0,0	0,0	0,0	202,4	5,3	202,4	5,1	202,4	5,1
ОЭС Юга	934,4	3,6	1943,3	7,2	2003,3	7,3	2223,3	8,0	2258,3	7,9
Астраханская область	0,0	0,0	340,2	28,7	340,2	29,7	340,2	22,9	340,2	22,9
Республика Адыгея	150,0	70,3	150,0	67,1	150,0	67,1	150,0	67,6	150,0	69,2
Волгоградская область	0,0	0,0	88,2	2,0	88,2	2,0	88,2	2,0	88,2	1,9
Республика Калмыкия	219,0	58,0	219,0	48,7	219,0	45,4	219,0	46,7	219,0	46,6
Республика Крым	88,6	4,9	88,6	4,9	88,6	5,2	88,6	5,0	88,6	5,0
Ростовская область	346,8	4,5	607,3	7,7	607,3	7,7	607,3	7,6	607,3	7,6
Ставропольский край	130,0	2,7	450,0	9,1	510,0	10,1	730,0	15,8	765,0	16,5

В таблицах 1.2 и 1.3 проиллюстрировано, что уже в 2015 г. в Республике Калмыкия доля ВЭС оказалась более 15 %. В 2020 г. доля генерации ВЭС впервые превысила величину 3% в ОЭС Юга, а в 2024 г. достигла практически 8 %. Также в 2024 г. доля установленной мощности ВЭС превысила 1 % по всей ЕЭС России. На конец 2024 г. доля ВЭС находится в пределах:

- 3–13 % – в пяти регионах;
- 13–25 % – в двух регионах;
- 25–50 % – в одном регионе;
- Более 50 – % в одном регионе.

ВЭС с наибольшей установленной мощностью, которая составляет 210 МВт, на конец 2024 г. является Кочубеевская ВЭС в Ставропольском крае. Чуть меньшая установленная мощность у Кольской ВЭС, расположенной за Полярным кругом в Мурманской области – 202 МВт.

Исходя из анализа проблемных вопросов интеграции ВЭС в энергосистемы, можно сказать, что ключевым фактором, сдерживающим развитие данного типа генерации с переменным характером выработки, является ограниченные возможности регулирования активной мощности. Максимальная величина активной мощности, которую может выдать ветроэлектростанция в конкретный момент времени, определяется текущими погодными условиями, в наибольшей степени скоростью ветра, которые могут иметь переменный, резко изменяющийся, а иногда даже непредсказуемый характер [31, 50, 51].

Как показано в таблице 1.1, среди мер, которые могут быть приняты для интеграции ВЭС в энергосистемы, обычно выделяют управление спросом на электропотребление, создание дополнительных связей между энергосистемами (расширение зон свободных перетоков), внедрение накопителей энергии и снижение погрешности прогнозирования выработки ВЭС [52]. Среди этих мер особо выделяют задачу прогнозирования генерации ВЭС как одну из наиболее эффективных и наименее затратных мер. Разработка системы прогнозирования генерации ВЭС рекомендуется на самых первых этапах внедрения ВЭС в энергосистемы. С ростом доли установленной мощности ВЭС обычно используют все четыре способа.

1.3. Область применения прогнозов генерации ВЭС

Основной областью применения прогнозирования генерации ВИЭ с переменным характером выработки, в том числе ВЭС, являются процессы планирования режимов электроэнергетических систем и тесно связанные с ними процессы функционирования рынков в электроэнергетике. Процессы планирования электроэнергетических режимов в ЕЭС России различаются по горизонту планирования и интервалу планирования. Горизонт планирования, также называемый интервалом упреждения, представляет собой период времени, на который осуществляет планирование (на час, сутки или год вперед, и т.д.). Интервал планирования, или дискретность планирования – это фиксированный

период времени, определяющий границы применимости запланированного режима (30 минут, 1 час, 1 месяц и т.д.).

Существующая система планирования электроэнергетических режимов включает в себя долгосрочное, среднесрочное, краткосрочное и оперативное планирование. На каждом из этих этапов требуется обеспечение сбалансированности объемов потребления и производства электрической энергии и мощности, а также нахождение параметров электроэнергетического режима в допустимых пределах с учетом технологических ограничений оборудования [57].

Понятия долгосрочного и среднесрочного планирования, а также соответствующие им горизонт и дискретность планирования, определены Правилами разработки и утверждения документов перспективного развития электроэнергетики [58].

Долгосрочное планирование электроэнергетического режима осуществляется с горизонтом планирования от 1 до 12 лет, а также на 18 лет вперед с интервалом планирования 1 год. Основной задачей на данном этапе планирования режимов является разработка генеральной схемы размещения объектов электроэнергетики. В рамках этой работы выполняется размещение генерирующих мощностей и объектов электросетевого хозяйства, относимых к межсистемным связям, по критерию минимизации совокупных дисконтированных затрат на производство, передачу и распределение электроэнергии в долгосрочном периоде. При этом необходимо обеспечить нормативный уровень балансовой надежности и предотвратить прогнозируемые дефициты электроэнергии и мощности [58, 59].

Среднесрочное планирование электроэнергетического режима осуществляется с интервалом упреждения от 1 до 6 лет с дискретностью планирования, определяемой характерными режимно-балансовыми условиями (максимум/минимум потребления, температура наружного воздуха, паводки и т.д.). Данный этап планирования связан с разработкой схем и программ развития (СиПР) электроэнергетических систем России, в рамках которых формируется состав объектов по производству электроэнергии и мощности для удовлетворения прогнозного потребления, а также определяются решения по размещению линий

электропередач и подстанций 110 кВ и выше, с учетом необходимости обеспечения параметров электроэнергетических режимов в допустимых пределах и предотвращения дефицитов электроэнергии и мощности [58, 59].

Планирование электроэнергетического режима на интервалы упреждения менее 1 года регламентируются Правилами технологического функционирования электроэнергетических систем [57], на интервалы менее одной недели также Правилами оптового рынка электрической энергии и мощности [61] и Регламентами оптового рынка [62–70]. Считается, что горизонт планирования от 1 суток до 1 недели соответствует краткосрочному планированию, а на периоды внутри суток – оперативному.

Среднесрочное планирование электроэнергетического режима на календарный год также осуществляется для интервалов планирования, определяемых характерными режимно-балансовыми условиями. Результаты планирования с данным интервалом упреждения используются для планирования ремонтов владельцами объектов электроэнергетики, определения потребности в объемах топлива для электростанций, расчетам необходимых объемов управляющих воздействий противоаварийной автоматики, в том числе автоматической частотной разгрузки, а также графиков аварийного ограничения потребления, определения объемов услуг по обеспечению системной надежности [57, 60].

Среднесрочное планирование электроэнергетического режима на предстоящий месяц осуществляется для режимно-балансовых условий, характерных для данного месяца. На данном этапе планирования решается задача составления сводных месячных графиков ремонтов и технического обслуживания объектов диспетчеризации [57].

Краткосрочное планирование электроэнергетического режима на горизонт планирования от 2 до 4 суток вперед с дискретностью планирования 1 час выполняется для выбора состава включенного генерирующего оборудования (ВСВГО). Решаемая при этом оптимизационная задача позволяет при соблюдении всех технических ограничений на состояние оборудования и значения параметров

электроэнергетического режима обеспечить минимизацию суммарной стоимости выработки электрической энергии с учетом стоимости резервов активной мощности на загрузку, определяемой на основании уведомлений участников оптового рынка о составе и параметрах генерирующего оборудования, и поданных участниками оптового рынка ценовых заявок с учетом заявленных величин расходов на осуществление пусков генерирующего оборудования (стоимости включения единицы генерирующего оборудования) [61, 64, 71–73]. Также на данном этапе планирования осуществляется подготовка предварительного электроэнергетического режима (ПЭР) на 2 суток вперед.

При краткосрочном планировании электроэнергетического режима на сутки вперед дискретность планирования составляет 1 час. На данном этапе планирования формируется прогнозный диспетчерский график (ПДГ) и актуализированная расчетная модель по обновленным данным о топологии электрической сети, состоянии и параметрах генерирующего оборудования, прогнозе потребления и других системных условиях. С использованием этой модели, а также переданных участниками оптового рынка ценовых заявок, проводится конкурентный отбор заявок на рынке электроэнергии «на сутки вперед». По сформированному в результате отбора торговому графику определяется предварительный план балансирующего рынка (ППБР), который доводится до электростанций как плановый диспетчерский график [61, 63, 66, 67, 68, 74, 75].

При оперативном планировании электроэнергетического режима на 1 час – 23 часа вперед с интервалом планирования 1 час осуществляется конкурентный отбор ценовых заявок на балансирование системы, при этом используются актуализированные данные о топологии электрической сети, состоянии и параметрах генерирующего оборудования, прогнозе потребления и системных условиях. В результате формируются план балансирующего рынка (ПБР), который доводится до электростанций как новый плановый диспетчерский график. Расчет ПБР выполняется 12 раз в сутки [61, 69, 76]. Оперативное планирование электроэнергетического режима осуществляется также и на горизонт

планирования менее 1 часа для задач управления электроэнергетическим режимом в реальном времени.

Основные этапы планирования режимов в ЕЭС России сведены в таблицу 1.4.

Таблица 1.4 – Этапы планирования режимов в ЕЭС России

Тип	Горизонт	Дискретность	Решаемые задачи
Долгосрочное	От 1 до 12 лет вперед, 18 лет вперед	1 год	Генеральная схема размещения объектов электроэнергетики
Среднесрочное	От 1 года до 6 лет вперед	Периоды характерных режимно-балансовых условий (до 9 типов условий)	СиПР
Среднесрочное	От 1 месяца до 12 месяцев вперед (календарный год)		Сводный годовой график ремонтов, обеспеченность топливом
Среднесрочное	На 1 месяц вперед (календарный месяц)	Соответствующие месяцу характерные режимно-балансовые условия	Сводный месячный график ремонтов
Краткосрочное	От 2 до 4 суток вперед	1 час	ВСВГО
Краткосрочное	На 1 сутки вперед	1 час	ППБР, РСВ, ПДГ
Оперативное	От 1 до 23 часов вперед	1 час	ПБР
Оперативное	Менее 1 часа	30 минут и менее	Диспетчерское управление

Традиционно при определении прогнозного баланса мощности и электроэнергии для каждого этапа планирования режимов выполняется прогноз потребления, по которому определяется генерация, необходимая чтобы его покрыть. С увеличением доли генерации ВИЭ в энергосистемах для более точного планирования электроэнергетических режимов также должен выполняться прогноз генерации ВИЭ с переменным характером выработки, в том числе ВЭС, для каждого этапа планирования режимов.

В международной практике, применительно к задаче прогнозирования генерации ВЭС, также выделяют четыре типа горизонтов прогнозирования [77]:

- Долгосрочный (более 10 дней вперед);
- Среднесрочный (от 2 до 10 дней вперед)
- Краткосрочный (от 6 до 48 часов вперед)
- Сверхкраткосрочный (менее 6 часов вперед).

При этом диапазоны горизонтов и дискретность, соответствующие различным типам прогнозирования, различаются в разных источниках [78-81]. Во многом это связано с различиями в функционировании рынков электроэнергии и правилами управления режимами в энергосистемах различных стран.

В данной работе решается задача повышения эффективности краткосрочного прогнозирования генерации ВЭС на сутки вперед, под которым понимается определение средней выдаваемой мощности на каждый час следующих суток. Основной областью применения таких прогнозов являются расчеты планов выработки для участия ВЭС в оптовом рынке электроэнергии на сутки вперед в соответствии с регламентами ОРЭМ.

Определение нагрузки генерирующего оборудования ВЭС при формировании ПДГ и ППБР выполняется в процессе оптимизационного расчета ПК Барс и ПК Барс-МДП, исходя из заявленных собственником максимальных и минимальных диапазонов мощности и поданных собственником ВЭС через ПАК *MODES-Terminal* ценовых заявок [63]. При работе на оптовом рынке электроэнергии «на сутки вперед» собственники ВЭС могут подавать ценовые заявки на планирование объемов производства, состоящие только из ценопринимающих одноступенчатых подзаявок на каждый час следующих суток [65]. Поэтому при выполнении оптимизационного расчета данный тип генерации окажется более приоритетным [66], чем тепловые электростанции (за исключением генерации, необходимой для обеспечения надежности, производства тепла и т.д.), которые, как правило, подают ступенчатые ценовые заявки для покрытия расходов на выработку электроэнергии и получения прибыли. Иллюстрация данного процесса представлена на рисунке 1.3.

Изменение графика работы, заявленного собственником объекта ВЭС, может производиться только при наличии технологических ограничений по выдачи мощности в сеть. В зависимости от схемно-режимной ситуации и складывающегося баланса мощности генерирующее оборудование ВИЭ может быть разгружено до нуля [68].

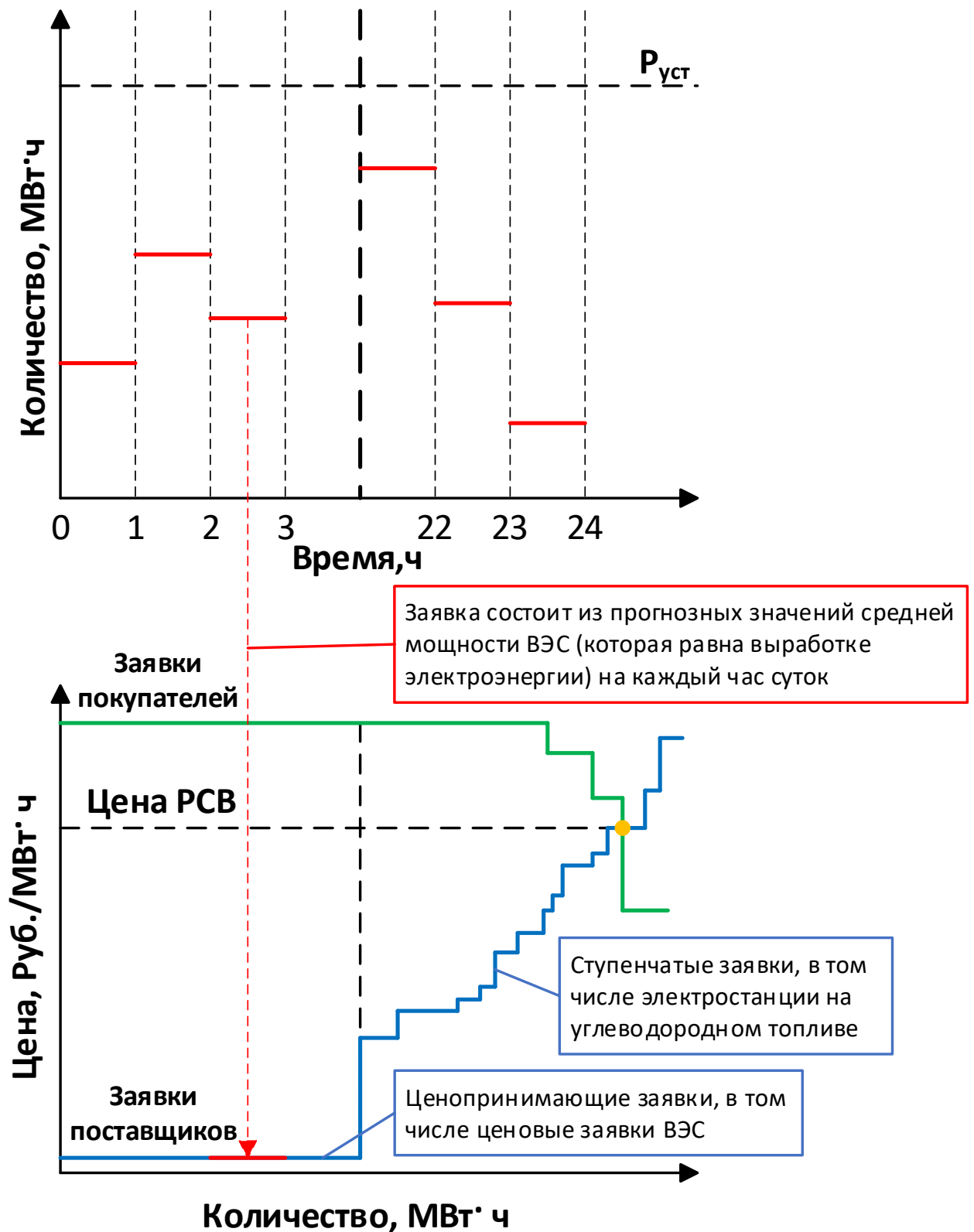


Рисунок 1.3 – ВЭС на ОРЭМ

Отклонение фактической выработки электроэнергии от плановой приводит к возникновению убытков, в том числе упущенной выгоды, при работе ВЭС на оптовом рынке. Отклонения от планового почасового производства могут

приводить к убыткам и для покупателей ЭЭ из-за возникновения незапланированного стоимостного небаланса между суммарными требованиями и обязательствами поставщиков и потребителей. Он распределяется между всеми участниками рынка пропорционально отклонениям по собственной инициативе и величине исполнения внешних инициатив.

В случае, если фактическая выработка ВЭС отклоняется выше заявленного на РСВ объема, то за отклонения ВЭС (участник оптового рынка) получает оплату по индикаторам балансирующего рынка (БР) вместо цены РСВ. Причем за значительные отклонения более 10% от установленной мощности оплата может быть снижена до 1 руб/МВт·ч, кроме того увеличивается доля стоимостного небаланса, который будет распределен на ВЭС.

Если же фактическая выработка ВЭС отклоняется ниже заявленного на РСВ объема, то для компенсации отклонений ВЭС (участник оптового рынка) вынуждена покупать электроэнергию по индикаторам БР. Значительные отклонения также увеличивают долю стоимостного небаланса.

Таким образом, разработка и внедрение методик повышения точности краткосрочного прогнозирования генерации ВЭС, обеспечивает для:

- собственников ВЭС – снижение убытков, в том числе упущенной выгоды, связанных с отклонением фактической выработки от планового почасового производства;
- участников оптового рынка, в том числе потребителей – снижение затрат на компенсацию стоимостного небаланса;
- системного оператора – обеспечение надежного функционирования энергосистем с ВЭС.

1.4. Особенности краткосрочного прогнозирования выработки ВЭС

1.4.1. Постановка задачи и основные этапы

Ввиду того, что регламенты ОРЭМ предусматривают работу с детерминированными данными, в настоящей работе прогноз выработки ВЭС осуществляется в детерминистической постановке.

Задачу определения детерминистического, также называемого точечным, прогноза генерации ВЭС можно обозначить следующим образом [82, 83]:

$$\hat{P}(t + k|t) = f(x(t + k|t), \beta), \quad (1.1)$$

где $\hat{P}(t + k|t)$ – прогнозная мощность на момент времени $t + k$, рассчитанная в момент времени t ; $x(t + k|t)$ – вектор прогнозных значений признаков (объясняющих переменных) на момент времени $t + k$, определенных в момент времени t ; f – модель прогнозирования; β – параметры модели; k – горизонт прогнозирования (интервал упреждения), в данной работе рассматриваются краткосрочные прогнозы на сутки вперед.

Основные этапы процесса прогнозирования представлены на рисунке 1.4. Эти этапы включают в себя создание и применение модели.

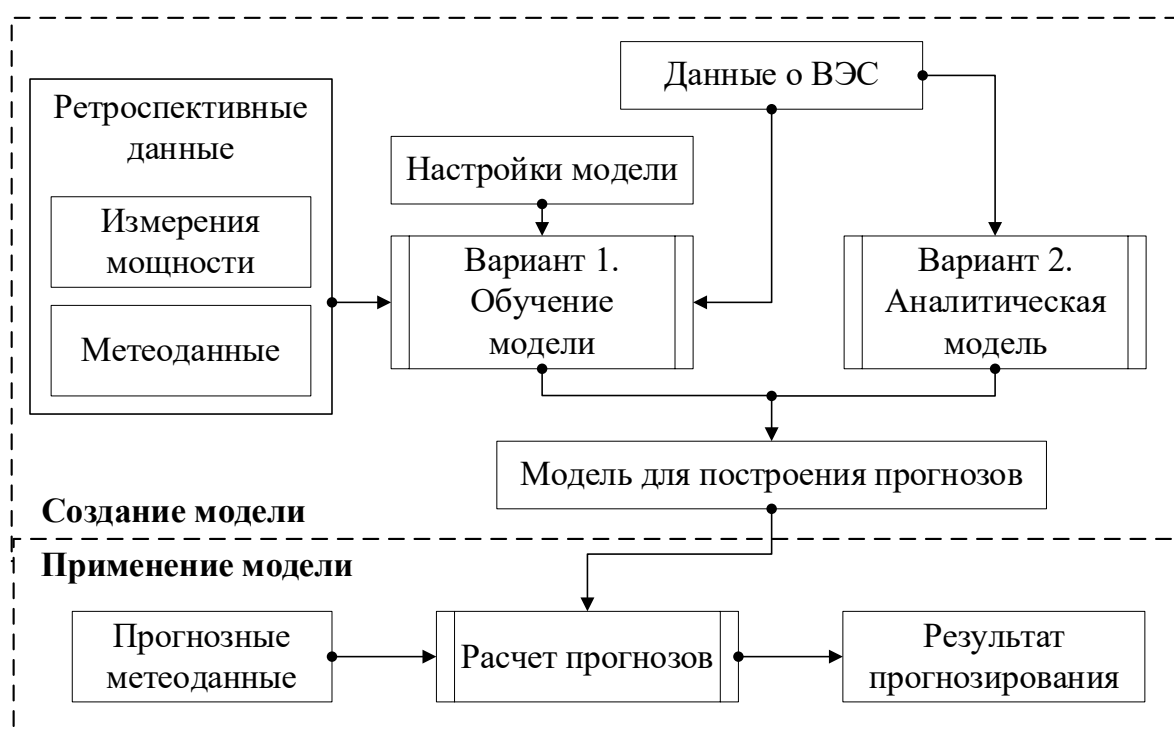


Рисунок 1.4 – Основные этапы процесса прогнозирования генерации ВЭС

На этапе создания модели главной задачей является определить модель прогнозирования f . Существуют две основных категории: модели, требующие ретроспективной информации для обучения (статистические модели, модели машинного обучения, гибридные модели), и аналитические модели, для которых требуются лишь данные о ВЭС: модели оборудования, расположение, высота

турбин, топографические данные и т.д. В ходе обучения модели по ретроспективным данным выполняется определение параметров модели β , позволяющих наилучшим образом с точки зрения функции потерь L описать связь между вектором зависимых переменных P и матрицей независимых переменных X :

$$\beta^* = \underset{\beta}{\operatorname{argmin}} L(P, f_{\beta}(X)), \quad (1.2)$$

где β^* – искомые значения параметров модели f ; L – функция потерь, обычно квадратичная; P – ретроспективные наблюдения мощности, представляющие собой вектор откликов; X – ретроспективные наблюдения метеоданных, полученные, как правило, с помощью численного прогноза погоды (ЧПП), представляющие собой матрицу признаков.

На этапе применения модели определяется значение $\hat{P}(t + k|t)$ с помощью прогнозных метеоданных $x(t + k|t)$, также полученных с помощью ЧПП.

Методы численного прогноза погоды позволяют определять будущие значения метеорологических параметров с помощью гидродинамической модели, состоящей из дифференциальных уравнений, описывающих изменение состояния атмосферы: уравнения движения, притока тепла, неразрывности, переноса влаги и атмосферных примесей, выражающие законы сохранения импульса, энергии и массы, а также уравнения состояния [84, 85].

Основным способом получения ЧПП является численное решение данной системы уравнений, при этом выполняется численное интегрирование переменных состояния атмосферы в конечном числе точек пространства на заданный горизонт прогнозирования, при этом в качестве начальных условий используются текущие метеорологические измерения, данные спутников, метеозондов и т.д. [85, 86].

При определении расположения и расстояния между такими точками используется метод сеток [85, 87]. Масштаб сетки определяет, какие метеорологические феномены позволяет описать модель ЧПП. Горизонт прогнозирования при этом определяет скорость протекания процессов. Обычно выделяют три вида процессов: крупномасштабные, среднемасштабные и

мелкомасштабные. Характеристики данных процессов приведены в таблице 1.5 [82, 88].

Таблица 1.5 – Виды атмосферных процессов

Виды	Масштаб сетки	Горизонт прогнозирования	Процесс
Крупномасштабные	2000 км и более	От 1 дня до 1 месяца и более	Перемещение областей высокого и низкого давления, осадков, атмосферных фронтов
Среднемасштабные (Альфа)	200 км – 2000 км	От 1 дня до 1 месяца	Перемещение линий шквала, тропических циклонов
Среднемасштабные (Бета)	20 км – 200 км	От 1 часа до 24 часов	Морской бриз
Среднемасштабные (Гамма)	2 км – 20 км	От 1 минуты до 60 минут	Грозы, перемещение потоков воздуха вдоль поверхности
Мелкомасштабные	Менее 2 км	Менее 1 минуты	Перемещение небольших облаков, турбулентность

Подход к получению ЧПП, использующий численное интегрирование системы уравнений гидродинамической модели из начальных условий, определяющих текущее состояние атмосферы, считается детерминированным. Тем не менее, получаемые прогнозы погодных факторов имеют погрешность, поскольку [89]:

- поведение атмосферы содержит хаотичную составляющую, и используемая гидродинамическая модель не позволяет описать все процессы;
- масштаб описываемых процессов ограничен масштабом сетки моделирования;
- начальные условия могут иметь пропуски в связи с отсутствием данных о метеорологических параметрах на определенной высоте или в определенной географической точке;
- начальные условия также имеют погрешность определения, которая связана в том числе с погрешностью измерительных устройств.

Для снижения влияния данных факторов применяется ансамблевый подход к прогнозированию погоды. Задачей ансамблевого прогноза является учет данных факторов неопределенности и получение вероятностного прогноза, который при

необходимости может быть преобразован в детерминированный. Существует два основных типа ансамблей [89]:

1. Мультимодельный ансамбль (учет неопределенности системы уравнений и параметров гидродинамической модели атмосферы);
2. Ансамбль, получаемый для множества разных начальных условий (учет неопределенности текущего состояния атмосферы).

Большинство современных систем прогнозирования, используемых в Гидрометеорологическом центре России, Европейском центре среднесрочных прогнозов погоды, Национальном метеорологическом центре США и т.д., основаны на ансамблевом подходе к получению ЧПП [90]. Кроме того, мультимодельный ансамблевый подход к получению ЧПП применяется также различными погодными сервисами и метеопровайдерами, использующими методы машинного обучения для агрегирования прогнозов различных метеорологических служб.

В рамках настоящей работы прогнозные метеоданные используются в качестве исходных данных для обучения регрессионных моделей, используемых для прогнозирования генерации ВЭС. При обучении моделей общепризнанным подходом, обеспечивающим повышение эффективности прогнозирования генерации ВЭС, является использование в качестве признаков метеоданных нескольких независимых метеопровайдеров [77]. С точки зрения повышения эффективности прогнозирования генерации ВЭС за счет увеличения количества признаков, используемых для обучения моделей, также применяется пространственный подход, подразумевающий использование метеоданных, полученных сразу для нескольких точек в районе расположения ВЭС [77], а также использование лаговых признаков, смещенных по оси времени (особенно при оперативном прогнозировании) [77].

1.4.2. Критерии для оценки эффективности краткосрочного прогнозирования выработки ВЭС

При описании критериев для оценки точности и надежности прогноза используются следующие условные обозначения:

$P_{уст}$ – установленная мощность ВЭС, кВт;

$k = 1, 2 \dots k_{max}$ – горизонт прогнозирования (интервал упреждения);

k_{max} – максимальный горизонт прогнозирования;

t – момент расчета прогноза;

N – количество наблюдений, использованных для определения значений метрики;

$P(t + k)$ – фактическая измеренная мощность в момент времени $t + k$, кВт;

$\hat{P}(t + k|t)$ – прогнозная мощность на момент времени $t + k$, рассчитанная в момент времени t , кВт;

$e(t + k|t)$ – ошибка, соответствующая прогнозу мощности на момент времени $t + k$, рассчитанному в момент времени t , кВт.

В качестве основных критериев для оценки погрешности краткосрочных прогнозов генерации ВЭС мировым экспертным сообществом используются следующие общепринятые показатели [82, 91, 92, 93]: нормализованная средняя абсолютная ошибка (*Normalized Mean Absolute Error, NMAE*), нормализованный квадратный корень из средней квадратичной ошибки (*Normalized Root Mean Squared Error, NRMSE*), нормализованное смещение (*Normalized Bias, NBias*).

При прогнозировании временных рядов ошибка прогноза обычно определяется как разница между фактическим и прогнозным значением параметра. Поэтому при прогнозировании мощности ветроэлектрических станций для горизонта прогнозирования k ошибка прогноза определяется в соответствии с выражением:

$$e(t + k|t) = P(t + k) - \hat{P}(t + k|t). \quad (1.3)$$

$NMAE$ используется для оценки общей погрешности прогноза, является линейной мерой. Все рассматриваемые ошибки вносят пропорциональный своей

величине вклад в значение показателя. Нормализация относительно установленной мощности позволяет сравнивать данный показатель, рассчитанный для различных ВЭС. Чем меньше значение показателя, тем точнее прогноз. Для определения данного показателя используется выражение [91, 92, 94, 95]:

$$NMAE = \frac{1}{N} \sum_{i=1}^N \frac{|e_i|}{P_{уст}} \cdot 100\%. \quad (1.4)$$

NRMSE позволяет оценить наличие грубых ошибок прогноза, является квадратичной мерой. Чем больше значение ошибки, тем больше окажется вклад квадрата этой ошибки в величину показателя. Использование квадратного корня позволяет получить значение в тех же единицах измерения, что и *NMAE*. В случае, если величина всех ошибок одинакова, значения *NMAE* и *NRMSE* будут совпадать. Чем больше разница между *NMAE* и *NRMSE*, тем большая доля ошибок прогноза в рассматриваемой выборке отличается от средней величины ошибки. Как правило, величина *NRMSE* больше *NMAE*, и количество малых ошибок прогноза не позволяет компенсировать вклад грубых ошибок в величину показателя. В редких случаях возможна обратная ситуация, и тогда *NRMSE* меньше *NMAE*. Аналогично, для универсализации метрики используется нормализация относительно установленной мощности. Чем меньше значение показателя, тем точнее прогноз. Для определения данного критерия используется выражение [91, 92, 96, 97]:

$$NRMSE = \sqrt{\frac{1}{N} \sum_{i=1}^N \left(\frac{e_i}{P_{уст}} \right)^2} \cdot 100\%. \quad (1.5)$$

NBias позволяет выявить систематическую ошибку, когда используемый алгоритм в среднем по рассматриваемой выборке завышает или занижает результаты прогноза. В рамках данной работы используется нормализованная версия показателя, аналогично предыдущим критериям. Близкое к нулю значение метрики показывает, что положительные ошибки компенсируются отрицательными. Если *NBias* больше нуля, то прогнозные значения в среднем меньше, чем фактические. Если же *NBias* меньше нуля, то, соответственно,

прогнозные значения в среднем больше, чем фактические. Нормализация показателя также выполняется по установленной мощности ВЭС. Для определения данного критерия используется выражение [91, 93, 98, 99]:

$$NBias = \frac{1}{N} \sum_{i=1}^N \frac{e_i}{P_{уст}} \cdot 100\%. \quad (1.6)$$

В рамках данной работы также предложены два новых критерия: нормализованная сумма абсолютных ошибок (*Normalized Sum of Absolute Errors, NSAE*) и вероятность охвата интервала ошибки в диапазоне от -20% до +20% от установленной мощности (*Error Interval Coverage Probability, EICP_20*).

NSAE показывает суммарную ошибку прогноза в выработке электрической энергии за рассматриваемый период времени, при этом в качестве фактора нормализации для данного параметра используется сумма фактических значений. Чем меньше значение показателя, тем точнее прогноз. Для определения данного критерия используется выражение:

$$NSAE = \frac{\sum_{i=1}^N |e_i|}{\sum_{i=1}^N P_i} \cdot 100\%. \quad (1.7)$$

EICP_20 используется в качестве метрики, позволяющей оценить надежность получаемых прогнозов. Эта величина показывает долю ошибок, находящихся в интервале $\pm 20\%$ от установленной мощности ВЭС. Данная метрика близка по своей сути к показателю вероятности охвата доверительного интервала прогноза, но не требует расчета доверительных интервалов и более проста для восприятия. Чем больше значение показателя, тем надежнее прогноз. Для определения данного критерия используется выражение:

$$EICP_{20} = \frac{1}{N} \sum_{i=1}^N c_i \cdot 100\%, \quad (1.8)$$

где значение c_i определяется в соответствии с правилом

$$c_i = \begin{cases} 1, & \text{если } |e_i| < 0,2 \cdot P_{уст} \\ 0, & \text{в остальных случаях} \end{cases}. \quad (1.9)$$

В качестве основной метрики, на основании которой выполняется сравнение эффективности методик прогнозирования мощности ВЭС, в данной работе используется $NRMSE$. При этом выполняется сравнение рассматриваемой методики с референсной с помощью фактора улучшения SS , определяемого по формуле [91, 92, 100, 101]:

$$SS = \left(1 - \frac{NRMSE_{eval}}{NRMSE_{ref}} \right) \cdot 100\%, \quad (1.10)$$

где $NRMSE_{eval}$ – показатель $NRMSE$ для рассматриваемой методики, %; $NRMSE_{ref}$ – показатель $NRMSE$ для референсной методики, %.

1.4.3. Влияние различных факторов на мощность и выработку ветроэлектрической установки

Ветроэлектрическая установка (ветроустановка, ветрогенератор, турбина, ВЭУ) представляет собой устройство, преобразующее кинетическую энергию потока воздуха (ветра) в механическую энергию вращения ротора, а затем в электрическую энергию [102–105]. Наиболее широкое распространение получили горизонтально-осевые ветроустановки. Большая часть ВЭУ, входящих в состав ВЭС и подключаемых к ЕЭС России, имеют такую конструкцию, поэтому именно они рассматриваются в данной работе.

Горизонтально-осевая ветроустановка состоит из следующих основных узлов: ветроколесо, представляющее собой ротор, на котором расположены лопасти турбины; гондола (кабина), в которой расположены трансмиссия (редуктор мультипликатор), генератор, а также может находиться другое механическое и электрическое оборудование (элементы системы управления, преобразовательное оборудование, повышающий трансформатор и т.д.); башня (мачта) на фундаменте, на вершине которой расположена гондола [102, 103]. Пример [106] и схема устройства горизонтально-осевой ветроустановки представлены на рисунке 1.5.

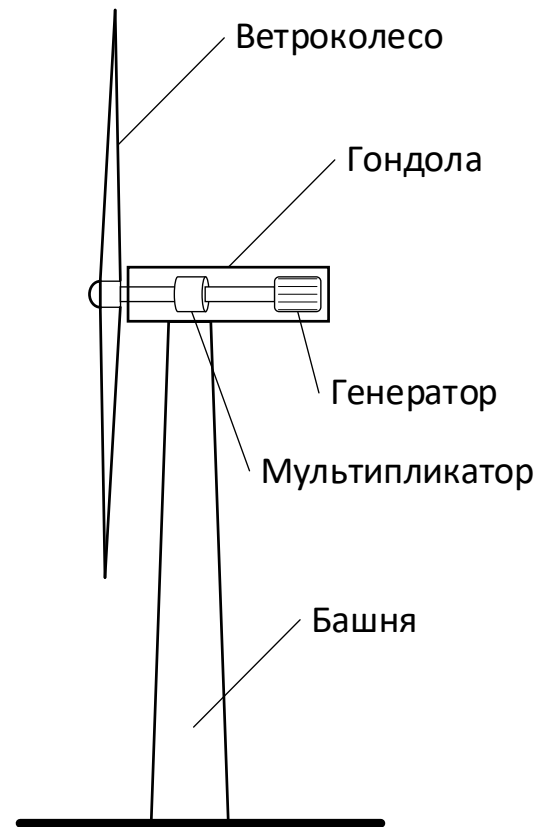


Рисунок 1.5 –Горизонтально-осевая ветроэлектрическая установка
(пример ВЭУ – слева, устройство ВЭУ – справа)

Кинетическая энергия движения воздушных масс m со скоростью V может быть определена по известной формуле [107, 108]:

$$E = \frac{1}{2} \cdot m \cdot V^2, \quad (1.11)$$

где E – кинетическая энергия ветра; V – скорость ветра; m – масса воздуха в объеме цилиндра с основанием, равным ометаемой площади ветроколеса A и длиной V .

Кинетическая энергия ветрового потока со скоростью V и плотностью ρ , проходящая через поперечное сечение площадью A в единицу времени, представляет собой мощность ветрового потока [107, 109]:

$$P_{\text{в}} = \frac{dE}{dt} = \frac{1}{2} \cdot (A \cdot \rho \cdot V) \cdot V^2 = \frac{1}{2} \cdot A \cdot \rho \cdot V^3, \quad (1.12)$$

где $P_{\text{в}}$ – мощность ветрового потока; V – скорость ветра; ρ – плотность воздуха; A – ометаемая площадь.

Ветроустановка может преобразовать лишь часть кинетической энергии ветра в механическую энергию вращения ротора. Эта доля является

коэффициентом эффективности использования мощности ветрового потока или коэффициентом мощности, обозначается C_p . Максимальное теоретическое значение коэффициента мощности называется критерием Бетца и равняется $16/27 \approx 0,593$. Максимальные значения коэффициентов мощности промышленных ВЭУ находятся в диапазоне 0,40 – 0,45 [102, 103].

Итоговое выражение для определения электрической мощности ВЭУ можно представить в виде [103, 110]:

$$P = \frac{1}{2} \cdot C_p \cdot \eta \cdot A \cdot \rho \cdot V^3, \quad (1.13)$$

где P_B – электрическая мощность, выдаваемая ВЭУ; C_p – коэффициент мощности; η – коэффициент эффективности электромеханического преобразования энергии; V – скорость ветра; ρ – плотность воздуха; A – ометаемая площадь.

Зная зависимость C_p от скорости ветра, можно построить характеристику мощности ветроустановки. Данная величина может быть определена с помощью испытаний, методика которых описана в ГОСТ Р 54418.12.2-2012 «Возобновляемая энергетика. Ветроэнергетика. Установки ветроэнергетические». Суть испытаний сводится к получению усредненных на 10-минутных интервалах времени значений электрической мощности, скорости ветра и плотности воздуха, усреднению полученных данных с помощью метода бинов на интервалах 0,5 м/с, и выполнению кусочно-линейной интерполяции значений в бинах для получения характеристики мощности [111–114]. Пример характеристики мощности ВЭУ представлен на рисунке 1.6 [82, 102, 103]. На рисунке обозначены три характерных значения скорости ветра: минимальная, номинальная и максимальная.

Минимальная скорость ветра $V_{\text{мин}}$ (*cut-in wind speed*) – значение скорости ветра, при котором кинетическая энергия потока воздуха недостаточна для того, чтобы заставить ветроколесо вращаться, для большинства турбин находится в диапазоне 3-5 м/с. При скоростях ветра большей, чем минимальная, но меньше номинальной, турбина ВЭУ начинает вырабатывать электроэнергию, при этом выдаваемая мощность ниже номинальной. Лопasti ветроколеса поворачиваются

таким образом, чтобы создать максимальное аэродинамическое сопротивление [102, 103].

Номинальная скорость ветра $V_{\text{ном}}$ (*rated wind speed*) – значение скорости ветра, при котором выдаваемая мощность ветроустановки достигает номинальной. Под номинальной мощностью понимается максимальное значение активной мощности, которую может выдавать ВЭУ в длительном режиме. Диапазон значений скорости ветра для данной характерной точки 10-13 м/с. При дальнейшем увеличении скорости ветра лопасти ветроколеса поворачиваются для снижения аэродинамического сопротивления, при этом уменьшается величина коэффициента эффективности использования мощности ветрового потока C_p [102, 103].

Максимальная скорость ветра $V_{\text{макс}}$ (*cut-out wind speed*) – значение скорости ветра, при котором невозможно обеспечить длительную работу ВЭУ, для большинства турбин находится в диапазоне 20-25 м/с. При преодолении данного значения лопасти ветроколеса разворачиваются параллельно ветровому потоку для минимизации аэродинамического сопротивления, срабатывает система торможения ветроколеса, и ВЭУ отключается от сети [102, 103].

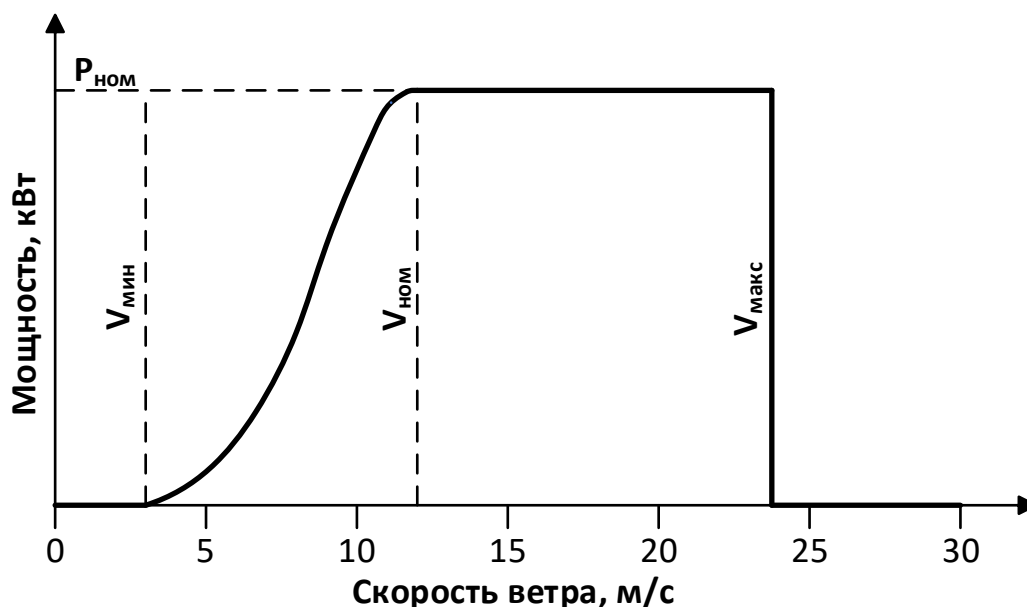


Рисунок 1.6 –Пример характеристики мощности ВЭУ

На рисунке не обозначена буревая скорость, при которой должна быть обеспечена устойчивость конструкции ВЭУ. Данная величина находится в

диапазоне 60-80 м/с в зависимости от ожидаемых максимальных скоростей ветра в конкретном регионе [102, 103].

Выражение (1.13) показывает основные факторы, влияющие на мощность, выдаваемую ВЭУ, они приведены далее [107, 115–118].

1. Скорость ветра V является определяющим фактором. Она позволяет ветроустановке достигать номинальной мощности, при этом при меньших скоростях ветра в диапазоне от $V_{\min}$ до $V_{\max}$ выдаваемая мощность имеет кубическую зависимость от скорости ветра. Сама же скорость ветрового потока зависит от высоты башни ВЭУ, окружающего рельефа, на котором расположена ветроустановка, препятствий на пути воздушного потока, и других атмосферных явлений.

2. Направление ветра WD , хотя и не присутствует в (1.13) в явном виде, тем не менее может оказывать влияние, поскольку выражение (1.13) справедливо в том случае, когда направление ветра перпендикулярно ометаемой площади ветроколеса. Как правило, горизонтально-осевые ВЭУ оснащены системой ориентации, позволяющей максимально эффективно использовать энергию воздушного потока. Тем не менее, данные системы могут работать с некоторой погрешностью, которая возрастает при резких изменениях направления ветра.

3. Плотность воздуха ρ , которая в свою очередь зависит от температуры, атмосферного давления и относительной влажности, и может быть определена по упрощенной зависимости [119, 120]:

$$\rho = \frac{\frac{P_B}{R_B} + \frac{P_{\Pi}}{R_{\Pi}}}{T}, \quad (1.14)$$

$$P_{\Pi} = \varphi \cdot 6,1078 \cdot 10^{\frac{7,5 \cdot T - 2048,625}{T - 35,85} + 3} \quad (1.15)$$

$$P_B = B - P_{\Pi} \quad (1.16)$$

где T – температура; B – атмосферное давление; φ – влажность воздуха; $R_B = 287,058 \text{ Дж} \cdot \text{кг}^{-1} \cdot \text{К}^{-1}$ – газовая постоянная для сухого воздуха; $R_{\Pi} = 461,495 \text{ Дж} \cdot \text{кг}^{-1} \cdot \text{К}^{-1}$ – газовая постоянная для водяного пара; P_B – парциальное давление сухого воздуха; P_{Π} – давление водяного пара.

4. Коэффициент эффективности использования мощности ветрового потока C_p , определяемый быстроходностью колеса (*tip-speed ratio*), то есть отношением угловой скорости вращения конца лопасти к скорости ветра, и угла поворота лопастей (*pitch angle*). Таким образом, величина C_p также является функцией скорости ветра.

5. Коэффициент эффективности электромеханического преобразования энергии η , определяемый коэффициентами полезного действия мультипликатора, генератора, системы подключения ВЭУ к сети через инвертор, и других электрических и механических узлов ветроустановки. Данный параметр может считаться условно постоянным, тем не менее он может снижаться в процессе эксплуатации в результате старения оборудования. Так же он может быть снижен искусственно при введении ограничений на выходную мощность ВЭУ с помощью системы управления инвертором.

6. Ометаемая площадь A , зависящая от диаметра ветроколеса, данный фактор может считаться постоянным.

1.4.4. Основные источники погрешности прогноза

На первый взгляд может показаться, что задачу прогнозирования генерации ВЭС можно свести только к задаче прогнозирования скорости ветра и пересчета прогнозного значения в мощность по уравнению (1.13). Несмотря на то, что такой подход иногда используется на практике, особенно при отсутствии ретроспективных данных, например, сразу после ввода ВЭС в эксплуатацию, фактически выдаваемая мощность может значительно отличаться от теоретических значений [82]. На рисунке 1.7 показано сравнение аналитической модели (1.13) с реальными измерениями мощности ВЭУ и скорости ветра на высоте турбины, усредненных на часовых интервалах. Лишь малое количество точек оказывается вблизи аналитической модели.



Рисунок 1.7 – Сравнение аналитической модели и реальных наблюдений

Основные факторы, затрудняющие процесс краткосрочного прогнозирования генерации ВЭС, и оказывающих наибольшее влияние на погрешность прогнозов, приведены далее [121–126].

1. Погрешность самого ЧПП, возрастающая при увеличении горизонта прогнозирования, обусловленная использованием упрощенной гидродинамической модели атмосферы, ограниченностью возможностей моделирования атмосферных процессов масштабом сетки, частичным отсутствием данных о текущем состоянии атмосферы в каждой точке сетки для определения начальных условий для решения дифференциальных уравнений, погрешностью измерений при определении текущего состояния атмосферы [121].

2. Характерные для конкретного расположения ВЭС факторы, резко изменяющие скорость или направление воздушного потока, попадающего на лопасти турбин [124, 125]:

- устойчивость атмосферы, связанная с явлением конвекции и вертикальным перемещением воздушных масс,

- турбулентность, представляющая собой хаотичное вихревое движение воздуха, вызванное препятствиями на пути воздушного потока,
- сдвиг ветра по вертикальной оси, при котором скорость ветра у верхней границы ветроколеса значительно выше, чем у нижней.

Часто данные факторы не могут быть учтены при ЧПП на интервалы упреждения 24 часа и более, но при этом могут оказывать существенное влияние на выработку ВЭС. Среди основных причин возникновения резких изменений скорости и направления ветра можно выделить: взаимное влияния ВЭУ друг на друга, также называемое «затенением» турбин, наличием препятствий на пути воздушных масс, наклоном поверхности, на которой расположены ВЭУ, и другие особенности ландшафта.

3. Обледенение лопастей ветроустановки и другие экстремальные погодные условия [121, 122].

4. Погрешности в работе систем управления ВЭУ: системы ориентации ВЭУ перпендикулярно ветровому потоку, системы отслеживания точки максимальной мощности и т.д. [124, 126].

5. Низкая информационная обеспеченность ретроспективными или прогнозными данными при обучении моделей и определении прогнозов [121]:

- недостаточное количество наблюдений в обучающей выборке;
- наличие метеорологической информации: только по одному параметру – скорости ветра, от одного метеопровайдера, только по одной точке на ВЭС;
- наличие пропущенных наблюдений мощности и/или метеопараметров в обучающей выборке;
- отсутствие ретроспективной информации о состоянии ВЭУ, а также искусственных ограничениях мощности отдельных ВЭУ и всей ВЭС, а также прогнозной информации при незапланированных отключениях ВЭУ и вводе ограничений.

1.5.Обзор основных подходов к повышению эффективности прогнозирования генерации ВЭС

Задаче прогнозирования генерации ВЭС посвящено огромное количество исследований в зарубежных научных изданиях. Первые значительные работы по данной тематике были опубликованы в середине 80-х в США и Великобритании [127–129]. Рост количества исследований по теме прогнозирования генерации ВЭС пропорционален росту установленной мощности и выработки электроэнергии ВЭС в мире [130], поэтому можно говорить о том что за последние 10 лет количество исследований в этой области возросло по меньшей мере в 2 раза.

В исследованиях, посвященных прогнозированию генерации ВЭС, как правило, решается одна из следующих задач (или сразу обе): снижение погрешности прогнозов и снижение затрат вычислительных ресурсов на получение прогнозов. При этом предлагается определенная методика повышения эффективности прогнозирования генерации ВЭС, которая по выбранным критериям оценки эффективности прогнозов сравнивается с референсной. Классификация основных подходов к повышению эффективности прогнозирования генерации ВЭС представлена на рисунке 1.8 [131–133].

Все существующие методики повышения эффективности прогнозирования генерации ВЭС можно разделить на две большие группы. Первая группа методик связана с обработкой исходных данных, используемых для выполнения прогнозов. Вторая группа связана непосредственно с построением моделей для прогнозирования.

Модели прогнозирования генерации ВЭС, как было отмечено в п.1.4.1, можно разделить на две основные категории. Первая категория представляет собой аналитические модели (также часто называемые в литературе физическими). Данные модели не требуют ретроспективных данных о мощности и метеопараметрах, и используют методы ЧПП и сложные математические модели изменения воздушного потока для максимально точного представления различных атмосферных процессов, турбулентности, рельефа, «затенения» турбин и т.д. Для

определения генерации ВЭС при этом обычно используется характеристика мощности, предоставляемая производителем ВЭУ. Данная категория моделей в упрощенном виде может применяться при отсутствии достаточного количества наблюдений для обучения моделей, например, сразу после запуска ВЭС в эксплуатацию [134].

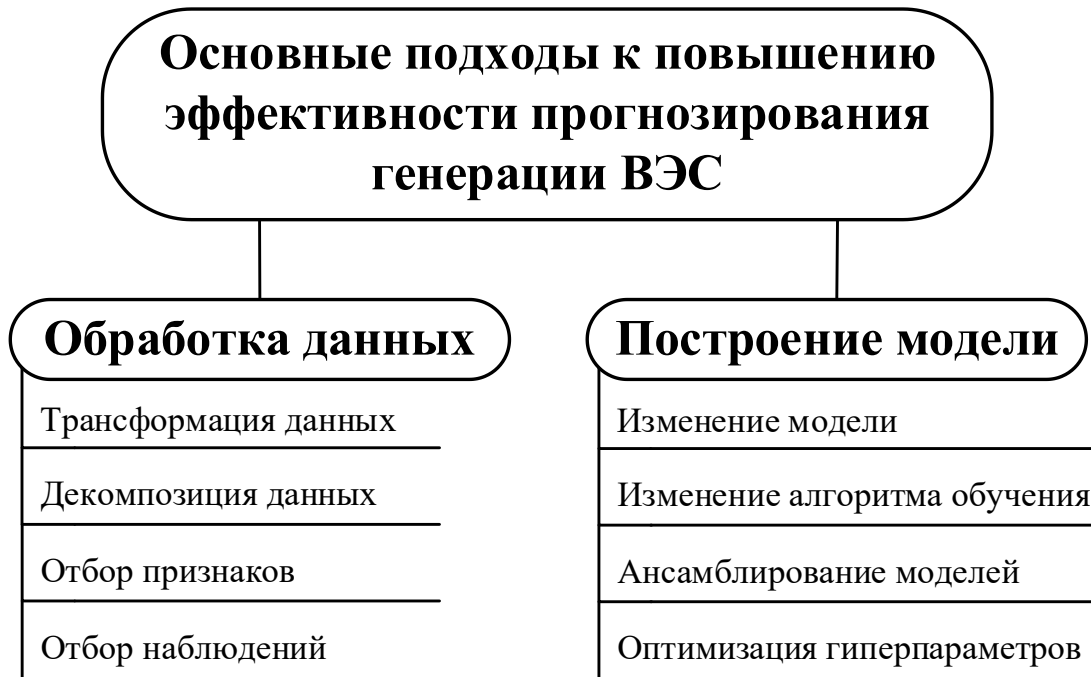


Рисунок 1.8 – Основные подходы к повышению эффективности прогнозирования генерации ВЭС

Вторая категория моделей, требующая ретроспективных данных, традиционно разделяется на три группы: статистические модели, модели машинного обучения и гибридные модели [79, 135].

Под статистическими моделями обычно подразумеваются методы анализа временных рядов, позволяющие определить прогнозные значения мощности по предыдущим значениям: модель авторегрессии (AR), авторегрессии скользящего среднего ($ARMA$), авторегрессии интегрированного скользящего среднего ($ARIMA$) и т.д. Также распространены авторегрессионные модели с внешним входом (ARX), которые могут использовать данные о метеопараметрах в модели. К группе статистических моделей также обычно относят вероятностные модели, построенные с помощью квантильной регрессий, ядерной оценки плотности и др.

Область применения статистических моделей обычно ограничена оперативным прогнозированием с горизонтом менее 6 часов [136, 137].

Широчайшее распространение в последнее десятилетие, особенно применительно к задаче краткосрочного прогнозирования на интервалы упреждения 24 часа и более, получили различные модели машинного обучения. Основным преимуществом данной группы методов является возможность восстановления сложных нелинейных регрессионных зависимостей, характерных для задачи прогнозирования генерации ВЭС, что позволяет значительно снизить погрешность прогнозов. Среди основных недостатков большинства методов машинного обучения выделяют высокую вычислительную сложность обучения моделей и чувствительность к выбору задаваемых перед началом обучения параметров [131, 136].

Отдельно выделяют гибридные методы [131, 136], комбинирующие аналитические модели, статистические модели или модели машинного обучения. Также под гибридными моделями часто подразумевают использование нескольких моделей машинного обучения для получения прогноза, представляющее собой ансамблирование моделей, но в данной работе такой подход рассматривается как отдельная методика повышения эффективности прогнозирования генерации ВЭС.

На рисунке 1.9 представлены соотношения используемых в научных работах по прогнозированию выработки ВЭС моделей и методов машинного обучения.

В таблице 1.6 представлен обзор ряда исследований, посвященных эффективности методик повышения точности прогнозирования генерации ВЭС, связанных с использованием различных моделей и алгоритмов обучения этих моделей. Описание и обзор эффективности методов машинного обучения, использованных в данной работе для прогнозирования генерации ВЭС, более подробно представлены далее в п.2.2.



Рисунок 1.9 – Соотношения используемых в научных работах моделей и методов машинного обучения

Таблица 1.6 – Обзор методик повышения эффективности прогнозирования генерации ВЭС, связанных с использованием различных моделей и алгоритмов обучения этих моделей

Авторы	Методика	Иссл. объект	Гор. k	Результаты
<i>Artipoli</i> и др. [138]	Аналитическая модель	4 ВЭС с разным количеством турбин	От 1 ч до 72 ч	$NRMSE$ 14-16%, улучшение в 2 раза по сравнению с «наивной» моделью
<i>Bhaskar</i> и др. [139]	Нейронная сеть прямого распространения ($FFNN$)	1 ВЭС, мощность не указана	От 1 ч до 30 ч	$NRMSE$ 9%, улучшение в 3 раза по сравнению с «наивной» моделью
<i>Mandzhieva</i> и др. [140]	Сверточная нейронная сеть с долгой краткосрочной памятью ($CNN-LSTM$)	1 ВЭС, мощность 8,2 МВт	12 ч и 24 ч	$RMSE$ 0,168 МВт, улучшение на 7,7% по сравнению с рекуррентной нейронной сетью с памятью ($RNN-LSTM$)
<i>Kisvari</i> и др. [141]	Рекуррентная нейронная сеть с управляемым рекуррентным блоком ($RNN-GRU$)	1 ВЭУ, мощность 7 МВт	Не указан	MSE 0,01 МВт ² , улучшение в 7 раз по сравнению с рекуррентной нейронной сетью с памятью ($RNN-LSTM$)

Продолжение таблицы 1.6

Авторы	Методика	Иссл. объект	Гор. k	Результаты
Huang и др. [142]	Рекуррентная нейронная сеть с долгой краткосрочной памятью (<i>RNN-LSTM</i>)	1 ВЭС, мощность 113 МВт	1 ч, 3 ч и 6 ч	<i>RMSE</i> 24,77 МВт, улучшение на 8% по сравнению с методом опорных векторов (<i>SVM</i>)
Kassa и др. [143]	Адаптивная сеть на основе системы нечеткого вывода (<i>ANFIS</i>)	1 ВЭУ, мощность 2,5 МВт	24 ч	<i>MAE</i> 60,84 кВт, Улучшение в 3,2 раза по сравнению с «наивной» моделью
Amroune [144]	Метод опорных векторов (<i>SVM</i>)	1 ВЭС, мощность 17,56 МВт	48 ч	<i>RMSE</i> 1,26 МВт, улучшение на 38% по сравнению с регрессионным деревом решений (<i>RDT</i>)
Нao и др. [145]	Алгоритм случайного леса (<i>RF</i>)	1 ВЭУ, мощность 2,5 МВт	Не указан	<i>MSE</i> 0,096 МВт ² , улучшение в 2,3 раз по сравнению с градиентным бустингом над деревьями решений (<i>GBDT</i>)
Tian и др. [146]	Авторегрессия интегрированного скользящего среднего с условной гетероскедастичностью (<i>ARIMA-GARCH</i>)	1 ВЭС, мощность около 100 МВт	От 1 ч до 12 ч	<i>MAE</i> 0,013 МВт, улучшение в 10 раз по сравнению с авторегрессией скользящего среднего (<i>ARMA</i>)
Pessanha и др. [147]	Квантильная регрессия (<i>QR</i>)	1 ВЭС, мощность 17,56 МВт	24 ч	<i>RMSE</i> 2,6 МВт

С точки зрения построения модели часто используемым способом повышения точности прогнозирования генерации ВЭС является ансамблирование, то есть обучение нескольких базовых моделей и агрегирование полученных результатов прогнозов по выбранному правилу для получения итогового прогноза. Подходы к ансамблированию моделей отличаются однородностью данных, используемых для обучения отдельных моделей, и однородностью самих моделей. Существует ряд классических подходов к построению ансамблей, такие как бэггинг, бустинг и стэкинг. Бэггинг представляет собой обучение каждой отдельной модели независимо с помощью случайно отобранного подмножества из тренировочного множества наблюдений. Бустинг позволяет обучать модели последовательно, каждая новая модель уменьшает среднеквадратичную ошибку предыдущих моделей. Бэггинг и бустинг обычно используют с однородными

моделями, например, деревьями решений. Стекинг обеспечивает возможность применения разнородных моделей, при этом итоговый прогноз получают с помощью отдельной модели, обучаемой по результатам, получаемых из базовых моделей. Также можно выделить несколько подходов к ансамблированию, специфичных для задачи прогнозирования генерации ВЭС: использование данных от разных метеопровайдеров для обучения отдельных моделей; использование данных по отдельным турбинам или группам турбин для обучения отдельных моделей и последующее агрегирование результатов для получения суммарного прогноза.

В таблице 1.7 представлен обзор ряда исследований, посвященных ансамблированию моделей, применяемому для прогнозирования генерации ВЭС. Обзор методов ансамблирования, связанных с различными способами агрегации мощности отдельных турбин, более подробно представлен далее в п.3.1.

Таблица 1.7 – Обзор методик повышения эффективности прогнозирования генерации ВЭС, связанных с использованием ансамблирования моделей

Авторы	Методика	Иssl. объект	Гор. <i>k</i>	Результаты
<i>Alessandrini</i> и др. [148]	Ансамбль из трех нейронных сетей прямого распространения (<i>FFNN</i>), обученных по разным источникам метеоданных	1 ВЭС, мощность 68,85 МВт	72 ч	<i>NRMSE</i> 15,6%, улучшение на 0,6% по сравнению с использованием одного источника метеоданных
<i>Wang</i> и др. [149]	Ансамбль нейронной сетью прямого распространения (<i>FFNN</i>), нейронной сети радиально базисных функций (<i>RBFNN</i>) и метода опорных векторов (<i>SVM</i>)	1 ВЭЕ, мощность 1,5 МВт	Не указан	<i>RMSE</i> 27,89 кВт, улучшение на 51% по сравнению с методом опорных векторов (<i>SVM</i>)
<i>He</i> и др. [150]	Ансамбль сверточной нейронной сети (<i>CNN</i>) и рекуррентной нейронной сети с долгой краткосрочной памятью (<i>RNN-LSTM</i>)	1 ВЭС, мощность 249,9 МВт	24 ч	<i>NRMSE</i> 12,15%, улучшение на 16% по сравнению с нейронной сетью прямого распространения (<i>FFNN</i>)

Продолжение таблицы 1.7

Авторы	Методика	Иssl. объект	Гор. <i>k</i>	Результаты
<i>Kumar</i> и др. [151]	Ансамбль сверточной нейронной сети (<i>CNN</i>), рекуррентной нейронной сети с долгой краткосрочной памятью (<i>RNN-LSTM</i>) и градиентного бустинга над деревьями решений (<i>GBDT</i>)	1 ВЭС, мощность 27 МВт	24 ч	<i>RMSE</i> 0,04 МВт, улучшение на 66% по сравнению с градиентным бустингом над деревьями решений (<i>GBDT</i>)
<i>Li</i> и др. [152]	Ансамбль нейронных сетей прямого распространения (<i>FFNN</i>), обученных по данным, полученным с помощью вейвлет-преобразований с разными параметрами	2 ВЭС, мощность 171,8 МВт и 1003,7 МВт	24 ч	<i>NRMSE</i> 6,29%, улучшение в 6 раз по сравнению с «наивной» моделью
<i>Xiaosheng</i> и др. [153]	Ансамбль нейронных сетей с архитектурой вложенных автокодировщиков (<i>SDAE</i>)	1 ВЭС, мощность не указана	12 ч	<i>NRMSE</i> 0,18 %, улучшение на 16% по сравнению с методом опорных векторов (<i>SVM</i>)
<i>Jalali</i> и др. [154]	Ансамбль сверточных нейронных сетей с долгой краткосрочной памятью (<i>CNN-LSTM</i>)	1 ВЭС, мощность 1756 МВт	Не указан	<i>RMSE</i> 0,07 МВт, улучшение на 13% по сравнению с одной сверточной нейронной сетью с долгой краткосрочной памятью (<i>CNN-LSTM</i>)
<i>Harrou</i> и др. [155]	Алгоритм ансамбля деревьев решений – случайный лес (<i>RF</i>)	1 ВЭУ, мощность 2,05 МВт	Не указан	<i>RMSE</i> 80 кВт, улучшение в 2 раза по сравнению с линейной регрессионной моделью
<i>Duan</i> и др. [156]	Ансамбль глубокой сети доверия (<i>DBN</i>) и рекуррентной нейронной сети с долгой краткосрочной памятью (<i>RNN-LSTM</i>)	1 ВЭУ, мощность 1,2 МВт	1 ч	<i>RMSE</i> 42,9 кВт, улучшение в 5 раз по сравнению с нейронной сетью прямого распространения (<i>FFNN</i>)
<i>Wang</i> и др. [157]	Ансамбль сверточных нейронных сетей (<i>CNN</i>)	1 ВЭС, мощность 21 МВт	3 ч	<i>RMSE</i> 0,08 МВт, улучшение в 11 раз по сравнению с рекуррентной нейронной сетью с долгой краткосрочной памятью (<i>RNN-LSTM</i>)

Наиболее редко среди способов повышения эффективности прогнозирования генерации ВЭС, связанных с построением модели, рассматривается оптимизация гиперпараметров моделей машинного обучения. Более подробно данная тема

рассмотрена в главе 2, обзор основных исследований по данной теме применительно к задаче прогнозирования генерации ВЭС приведен в разделе 2.3.

Для повышения эффективности прогнозирования генерации ВЭС также широко применяются различные способы предварительной обработки данных. Эти способы можно разделить на четыре основные группы: трансформация данных, декомпозиция данных, отбор признаков и отбор наблюдений.

Методики трансформации данных связаны с изменением исходного временного ряда признаков или откликов по некоторому правилу. Наиболее часто встречающимися способами трансформации можно считать нормализацию данных между 0 и 1 с помощью максимального и минимального значений временного ряда, а также стандартизацию данных с помощью Z-оценки. Эффективность данных двух методик оценивается достаточно редко, обычно они применяются по умолчанию. Более часто в исследованиях используется трансформация признаков, временные ряды откликов чаще остаются без изменений.

В таблице 1.8 представлен обзор ряда исследований, посвященных использованию трансформации данных при прогнозировании генерации ВЭС.

Таблица 1.8 – Обзор методик повышения эффективности прогнозирования генерации ВЭС, связанных с использованием трансформации данных

Авторы	Методика	Иссл. объект	Гор. k	Результаты
<i>Tasdemir</i> и др. [158]	Нормализация признаков с помощью сигмоидной функции	1 ВЭУ, синтетические данные	Не указан	R^2 0,95 кВт, улучшение на 3% по сравнению с отсутствием нормализации
<i>Gao</i> и др. [159]	Стандартизация признаков с помощью z-оценки	1 ВЭС, мощность 200 МВт	Не указан	$RMSE$ 13,0 МВт, улучшение на 16 % по сравнению с отсутствием стандартизации
<i>Makhloufi</i> и др. [160]	Сглаживание шумов во временных рядах признаков с помощью вейвлет-функции (WTD)	1 ВЭУ, мощность 15 МВт	Не указан	$RMSE$ 1,01 МВт, улучшение на 2 % по сравнению с отсутствием сглаживания
<i>Kahraman</i> и др. [161]	Трансформация откликов с помощью преобразования Фурье	1 ВЭС, мощность не указана	Не указан	$RMSE$ 0,96 МВт, улучшение на 2 % по сравнению с отсутствием преобразования
<i>Han</i> и др. [162]	Реконструкция фазового пространства (PSR) признаков	2 ВЭС, мощность не указана	От 1 ч до 48 ч	$RMSE$ 6 МВт, улучшение в 2.5 раза % по сравнению с отсутствием трансформации

Нередко как способ повышения эффективности прогнозов генерации ВЭС применяются методики декомпозиции временных рядов, то есть разложение исходных временных рядов на составляющие по заданным правилам. Декомпозиция может применяться как по отношению к временным рядам признаков, так и откликов. При декомпозиции временного ряда откликов каждая составляющая временного ряда используется для обучения отдельных моделей. При декомпозиции признаков каждая составляющая может использоваться в качестве отдельного признака. Наиболее часто встречающимися методиками декомпозиции временных рядов являются разложение на эмпирические моды (*Empirical Mode Decomposition, EMD*) и методы, основанные на вейвлет-преобразовании.

В таблице 1.9 представлен обзор ряда исследований, посвященных использованию декомпозиции данных при прогнозировании генерации ВЭС.

Таблица 1.9 – Обзор методик повышения эффективности прогнозирования генерации ВЭС, связанных с использованием декомпозиции данных

Авторы	Методика	Иссл. объект	Гор. k	Результаты
Leng и др. [163]	Декомпозиция временных рядов признаков с использованием риджлет-преобразования (<i>ridgelet-transform</i>)	2 ВЭС, мощность 171,8 МВт и 1003,7 МВт	24 ч	<i>NRMSE</i> 4,12-5,89%, улучшение на 6% по сравнению с отсутствием декомпозиции
Liu и др. [164]	Декомпозиция временного ряда мощности с помощью дискретного-вейвлет преобразования (<i>DWT</i>)	3 ВЭС, мощность не указана	1 ч	<i>RMSE</i> 35-64 МВт, улучшение на 22-24% по сравнению с отсутствием декомпозиции (<i>WPD</i>)
Wang и др. [165]	Декомпозиция временного ряда мощности на две составляющие с помощью метода анализа сингулярного спектра (<i>SSA</i>)	1 ВЭС, мощность не указана	12 ч	<i>NRMSE</i> 15%, улучшение на 6% по сравнению с отсутствием декомпозиции
Liu и др. [166]	Декомпозиция временного ряда скорости ветра при совместном использовании разложения на вариационные моды (<i>VMD</i>) и анализа сингулярного спектра (<i>SSA</i>)	1 ВЭС, мощность 200 МВт	5 ч	<i>RMSE</i> 0,96 м/с, улучшение в 2 раза по сравнению с вейвлет декомпозицией (<i>WPD</i>)

Продолжение таблицы 1.9

Авторы	Методика	Иssl. объект	Гор. <i>k</i>	Результаты
Wang и др. [167]	Декомпозиция временного ряда мощности с помощью метода ансамблевого разложения на эмпирические моды (<i>EEMD</i>)	1 ВЭС, мощность 49 МВт	Не указан	<i>RMSE</i> 2,16 МВт, улучшение в 4 раза по сравнению с отсутствием декомпозиции
Shi и др. [168]	Декомпозиция временного ряда мощности с помощью преобразования Гильберта-Хуанга (<i>HHT</i>)	1 ВЭС, мощность 353, 1 МВт	1 ч	<i>NRMSE</i> 2,2 %, улучшение в 2 раза по сравнению с отсутствием декомпозиции
Abedinia и др. [169]	Декомпозиция временного ряда мощности с помощью улучшенного вейвлет преобразования (<i>IWT</i>)	1 ВЭС, мощность 17,56 МВт	1 ч	<i>NRMSE</i> 2,42 %, улучшение в 2,5 раза по сравнению с отсутствием декомпозиции
Zheng и др. [170]	Декомпозиция временного ряда мощности с помощью метода разложения на эмпирические моды (<i>EMD</i>)	1 ВЭУ, мощность 3 МВт	Не указан	<i>MSE</i> 2,5 кВт ² , улучшение на 75% раза по сравнению с отсутствием декомпозиции
Bokde и др. [171]	Декомпозиция временного ряда мощности с помощью метода ансамблевого разложения на эмпирические моды (<i>EEMD</i>)	Несколько ВЭС, мощность около 2500 МВт	24 ч	<i>MSE</i> 184,8 МВт, улучшение на 30% раза по сравнению с отсутствием декомпозиции
Zhou и др. [172]	Декомпозиция временного ряда скорости ветра при совместном использовании разложения на вариационные моды (<i>VMD</i>) и анализа сингулярного спектра (<i>SSA</i>)	1 ВЭС, мощность 100 МВт	24 ч	<i>RMSE</i> 13,8 МВт, улучшение на 8,8% раза по сравнению с отсутствием декомпозиции

Для сокращения времени обучения моделей и борьбы с «переобучением» при прогнозировании генерации ВЭС также применяются различные методики отбора значимых признаков. Среди часто встречающихся методик можно выделить метод главных компонент (*Principal Component Analysis, PCA*), критерий минимальной избыточности – максимальной релевантности (*Minimum-redundancy-maximum-relevance, MRMR*), оптимизационные алгоритмы и т.д.

В таблице 1.10 представлен обзор ряда исследований, посвященных использованию методик отбора признаков при прогнозировании генерации ВЭС.

Таблица 1.10 – Обзор методик повышения эффективности прогнозирования генерации ВЭС, связанных с использованием отбора признаков

Авторы	Методика	Иссл. объект	Гор. k	Результаты
<i>Qin</i> и др. [173]	Методика <i>RRelief</i>	3 ВЭС, суммарная мощность 603 МВт	24 ч	<i>NRMSE</i> 8,5%, улучшение на 23% по сравнению с отсутствием отбора признаков
<i>Liu</i> и др. [174]	Метод главных компонент (<i>PCA</i>)	2 ВЭС, мощность 171,8 МВт и 1003,7 МВт	24 ч	<i>NRMSE</i> 5 %, улучшение на 32,5% по сравнению с методом оценки взаимной информации (<i>CMI</i>)
<i>Cali</i> и др. [175]	Метод поиска по сетке (перебор комбинаций признаков)	1 ВЭС, мощность 17,56 МВт	24 ч	<i>NRMSE</i> 10,8%, улучшение на 14% по сравнению с отсутствием отбора признаков
<i>Zhang</i> и др. [176]	Критерий максимальной значимости – минимальной избыточности (<i>MRMR</i>)	1 ВЭС, мощность 47 МВт	Не указан	<i>RMSE</i> 1,15 МВт, улучшение на 23% по сравнению с отсутствием отбора признаков
<i>Shi</i> и др. [177]	Использование индекса важности переменных (<i>VIM</i>)	1 ВЭС, мощность 49,5 МВт	24 ч	Снижение времени обучения в 4 раза без увеличения погрешности
<i>Xue</i> и др. [178]	Рекурсивное исключение признаков (<i>RFE</i>)	1 ВЭС, мощность 49,5 МВт	15 мин	<i>NRMSE</i> 3,95%, улучшение на 42% по сравнению с отсутствием отбора признаков
<i>Peng</i> и др. [179]	Методика последовательного прямого плавающего отбора (<i>SFFS</i>)	20 ВЭС, мощность от 49,5 до 297 МВт	От 1 до 72 ч	<i>NRMSE</i> 12 %, улучшение на 3% по сравнению с отсутствием отбора признаков
<i>Carrillo</i> и др. [180]	Отбор признаков с помощью алгоритма оптимизации подражанием муравьиной колонии (<i>ACO</i>)	2 ВЭС, мощность 36,7 МВт и 30 МВт	1 ч	R^2 0,6, улучшение в 2 раза по сравнению с отсутствием отбора признаков
<i>Wu</i> и др. [181]	Метод главных компонент (<i>PCA</i>)	1 ВЭС, мощность 49,5 МВт	Не указан	<i>RMSE</i> 4,1 МВт, улучшение на 7% по сравнению с отсутствием отбора признаков
<i>Shao</i> и др. [182]	Графовый метод отбора признаков (<i>Inf-FS</i>)	1 ВЭС, мощность не указана	1 ч	Снижение времени обучения на 40% без увеличения погрешности

Помимо отбора признаков могут применяться так же и различные методики отбора наблюдений. Под отбором наблюдений можно понимать процесс выделения из исходной обучающей выборки некоторой подвыборки по определенному правилу. Применительно к задаче прогнозирования генерации ВЭС наиболее часто применяется два подхода: идентификация и исключение выбросов,

а также разделение обучающей выборки на подвыборки по схожести наблюдений (например, поиск наблюдений с погодными условиями, похожими на прогнозные). Обзор методов идентификации выбросов с помощью характеристики мощности представлен далее в п.4.2.

В таблице 1.11 представлен обзор ряда исследований, посвященных использованию методик отбора наблюдений при прогнозировании генерации ВЭС.

Таблица 1.11 – Обзор методик повышения эффективности прогнозирования генерации ВЭС, связанных с использованием отбора наблюдений

Авторы	Методика	Иssl. объект	Гор. k	Результаты
Zheng и др. [183]	Отбор наблюдений с похожими погодными условиями при помощи разделения исходной выборки методом кластеризации k -средних	1 ВЭС, мощность не указана (не менее 180 МВт)	Не указан	$RMSE$ 20,8 МВт, улучшение на 15% по сравнению с отсутствием поиска похожих погодных условий
Peng и др. [184]	Отбор наблюдений с похожими погодными условиями при помощи евклидова расстояния с заданным пороговым значением	52 ВЭС, суммарная мощность 3694 МВт	12 ч	$NRMSE$ 6,4%, улучшение на 20% по сравнению с отсутствием поиска похожих погодных условий
Sun и др. [185]	Отбор наблюдений с похожими погодными условиями при помощи евклидова расстояния и коэффициента Отиаи	2 ВЭС, мощность 42 МВт и 90 МВт	24 ч	$RMSE$ 4,8-5,4 МВт, улучшение в 2 раза по сравнению с поиска похожих погодных условий с использованием только евклидова расстояния
Сао и др. [186]	Отбор наблюдений с похожими погодными условиями при помощи расстояния Хэмминга	1 ВЭС, мощность не указана (не менее 60 МВт)	6 ч	$RMSE$ 0,8 МВт, улучшение в 3 раза по сравнению с отсутствием поиска похожих погодных условий
Мiao и др. [187]	Отбор наблюдений с похожими погодными условиями при помощи разделения исходной выборки методом кластеризации k -средних и манхэттенского расстояния	2 ВЭС, мощность 806,7 МВт и 80 МВт	4 ч	$RMSE$ 0,12-0,2 МВт, улучшение на 2% по сравнению с отсутствием поиска похожих погодных условий

Анализ мирового опыта в области краткосрочного прогнозирования выработки ВЭС показывает, что единого универсального способа для решения данной задачи на текущий момент нет. Используемые методики прогнозирования во многом носят индивидуальный характер, который определяется составом и

качеством доступных исходных данных, полученных от метеопровайдеров и собранных с помощью SCADA на самой ВЭС.

Значительное внимание в научных публикациях по данной теме уделяется выбору моделей и алгоритмов машинного обучения. В меньшей степени уделяется внимание трансформации и декомпозиции исходных данных, отбору информативных признаков, идентификации выбросов. Ещё меньшее количество исследований посвящено изучению стратегий агрегации мощности отдельных ветроустановок, а также оптимизации гиперпараметров моделей машинного обучения применительно к задаче краткосрочного прогнозирования выработки ВЭС. Практически отсутствуют исследования, связанные с учетом состава работающих ветроустановок.

1.6. Выводы по главе

1. Развитие возобновляемых источников энергии во всем мире обосновывается соображениями экологичности, энергетической безопасности, энергетической эффективности и всеобщей доступности к первичным энергоносителям, а также интенсификацией функционирования розничных и оптовых рынков электроэнергии с вовлечением большего количества субъектов.

2. Установленная мощность ВЭС в России достигла 2 556 МВт на конец 2024 года, чуть более 1% от установленной мощности всех электростанций единой энергетической системы России. За последние 5 лет было введено более 2 ГВт установленной мощности ВЭС. По разным оценкам до 2035 года может быть введено еще 5 ГВт ВЭС.

3. Внедрение ВИЭ с переменным характером выработки, в том числе ВЭС, в энергосистемы требует решения целого комплекса различных задач, как технического, так и организационного характера. С увеличением доли установленной мощности стохастической генерации ВИЭ особое внимание уделяется вопросам обеспечения надежного функционирования энергосистем: балансу потребления и генерации, отсутствию нарушения режимных ограничений, в том числе перетоков мощности в контролируемых сечениях, наличию объемов

доступных резервов активной мощности. Во многом успех решения данной задачи определяется точным краткосрочным планированием электроэнергетических режимов, в рамках которого должно выполняться прогнозирование генерации ВИЭ с переменным характером выработки, в том числе ВЭС.

4. При работе на оптовом рынке электроэнергии на сутки вперед ВЭС подают ценопринимающие заявки, поэтому при выполнении оптимизационного расчета при формировании планового диспетчерского графика данный тип генерации окажется более приоритетным, чем генерация традиционного типа. При отклонениях фактической выработки от плановой для ВЭС (участника оптового рынка) могут возникать дополнительные убытки, связанные с: необходимостью покупки электроэнергии на балансирующем рынке при фактической выработке ниже плановой, упущенной выгоде от продажи электроэнергии при фактической выработке выше плановой, а также увеличением объемов собственных инициатив, пропорционально которым распределяется стоимостной небаланс.

5. Для работы ВЭС на рынке электроэнергии и для планирования режимов работы энергосистем требуется решение задачи определения детерминистического, также называемого точечным, прогноза генерации ВЭС. Основные этапы прогнозирования включают обучение прогнозной модели по ретроспективным данным и применение модели для получения прогнозов генерации ВЭС с помощью численного прогноза погоды.

6. В качестве основного критерия, на основании которого выполняется оценка эффективности методик прогнозирования генерации ВЭС, в данной работе используется нормализованный корень из средней квадратичной ошибки. При этом сравнение методик осуществляется с помощью фактора улучшения.

7. Среди основных факторов, затрудняющих процесс прогнозирования генерации ВЭС, можно выделить: погрешности в прогнозных метеоданных; случайные составляющие ветрового потока, резко изменяющие скорость или направление воздушного потока; взаимное затенение турбин; наличие препятствий на пути воздушных масс; наклон поверхности, на которой расположены ВЭУ и другие особенности ландшафта; обледенение лопастей ветроустановки и другие

экстремальные погодные условия; погрешности в работе систем управления ВЭУ; низкая информационная обеспеченность ретроспективными или прогнозными данными при обучении моделей и определении прогнозов.

8. Анализ мирового опыта в области краткосрочного прогнозирования генерации ВЭС показывает, что единого универсального способа для решения данной задачи на текущий момент нет. Используемые методики прогнозирования во многом носят индивидуальный характер, который определяется составом и качеством доступных исходных данных, полученных от метеопровайдеров и собранных с помощью *SCADA* на самой ВЭС. При прогнозировании на сутки вперед нормализованный корень из средней квадратичной ошибки прогноза может составлять 10 – 15%.

9. По результатам анализа литературы было выявлено, что значительное внимание в научных публикациях по данной теме уделяется выбору моделей и алгоритмов машинного обучения. В меньшей степени уделяется внимание трансформации и декомпозиции исходных данных, отбору информативных признаков, идентификация выбросов. Ещё меньшее количество исследований посвящено изучению стратегий агрегации мощности отдельных ветроустановок, а также оптимизации гиперпараметров моделей машинного обучения применительно к задаче краткосрочного прогнозирования выработки ВЭС. Практически отсутствуют исследования, связанные с учетом состава работающих ветроустановок.

ГЛАВА 2. ОПТИМИЗАЦИЯ ГИПЕРПАРАМЕТРОВ МОДЕЛЕЙ МАШИННОГО ОБУЧЕНИЯ ДЛЯ КРАТКОСРОЧНОГО ПРОГНОЗИРОВАНИЯ ГЕНЕРАЦИИ ВЕТРОЭЛЕКТРОСТАНЦИЙ

2.1. Постановка задачи оптимизации гиперпараметров

Для выполнения прогноза генерации ВЭС необходимо решить задачу восстановления регрессии. Исходными данными для решения задачи регрессии являются матрица признаков X , состоящая из N строк наблюдений x_i , а также вектор откликов y . Каждое наблюдение x_i представляет собой вектор размера M , где M – количество признаков. Каждому наблюдению x_i соответствует значение отклика y_i . Результатом решения задачи регрессии (1.2) является некоторая модель (регрессор) f с параметрами β , которая по вектору признаков x_* , соответствующих новому наблюдению, позволяет определять новые значения отклика y_* :

$$Y = f(X, \beta) + E, \quad (2.1)$$

$$y_* = f(x_*, \beta), \quad (2.2)$$

где $Y = (y_i)_{i=1}^N$ – вектор откликов (зависимая переменная) из обучающей выборки; $X = (x_i)_{i=1}^N$ – матрица признаков (независимая переменная) из обучающей выборки; f – модель регрессии (регрессор); β – параметры модели регрессии f ; E – вектор остатков регрессии, x_* – вектор признаков, соответствующий одному наблюдению из тестовой выборки; y_* – прогнозное значение отклика, полученное по наблюдению из тестовой выборки.

Часть параметров модели определяется непосредственно в процессе восстановления регрессии (значение весов в нейронной сети, значения коэффициентов в полиномиальной модели), а часть задаются заранее, до начала данного процесса (количество слоев в нейронной сети, порядок полиномиальной модели). Такого рода параметры называются гиперпараметрами.

Итак, имеется некоторый набор данных D , состоящий из признаков X и откликов Y . Набор данных D может быть разбит на обучающую D^{train} , валидационную D^{valid} и тестовую D^{test} выборки. Для определения модели f

используется алгоритм A , такой, что $A(D^{train}) = f$. Количество гиперпараметров алгоритма A и регрессионной модели f равно K . Область допустимых значений k -го гиперпараметра обозначается Λ_k , а все пространство гиперпараметров – $\Lambda = \Lambda_1 \times \Lambda_2 \times \Lambda_3 \times \dots \times \Lambda_K$. Модель и алгоритм с некоторым конкретным набором гиперпараметров λ обозначается как A_λ и f_λ , причем $A_\lambda(D^{train}) = f_\lambda$.

Задача определения набора гиперпараметров λ модели f_λ и алгоритма A_λ для набора данных D сводится к минимизации ожидаемой ошибки прогноза E с точки зрения меры оценки эффективности L [188]:

$$\lambda_* = \operatorname{argmin}_{\lambda \in \Lambda} \mathbb{E}_{(D_{train}, D_{valid}) \sim D} (\Psi(L, A_\lambda, D_{train}, D_{valid})) , \quad (2.3)$$

где λ_* – искомый набор гиперпараметров; Λ – пространство гиперпараметров; $\mathbb{E}$ – математическое ожидание (в данном случае определяется с использованием кроссвалидации); D_{train} – обучающая выборка; D_{valid} – валидационная выборка; D – полный набор данных; Ψ – функция оценки эффективности валидации (функция потерь валидации); L – функция оценки эффективности (функция потерь); A_λ – алгоритм A с некоторым набором гиперпараметров λ .

Функция оценки эффективности валидации может быть определена как:

$$\Psi(L, A_\lambda, D_{train}, D_{valid}) = \frac{1}{N_{valid}} \sum_{(X, Y) \in D_{valid}} L(Y, f_\lambda(X)), \quad (2.4)$$

где N_{valid} – количество наблюдений в валидационной выборке; f_λ – модель f с некоторым набором гиперпараметров λ ; X – матрица признаков; Y – вектор откликов.

Целью задачи оптимизации гиперпараметров является определение набора гиперпараметров λ_* , обладающей наибольшей эффективностью с точки зрения функции оценки эффективности L .

2.2. Модели машинного обучения и их применение для краткосрочного прогнозирования генерации ВЭС

В данном разделе описываются алгоритмы и модели машинного обучения, которые рассматривались в качестве возможной модели регрессии для прогнозирования выработки ветроэлектрических станций.

2.2.1. Регрессия на основе гауссовских процессов

Пусть есть некоторый набор обучающих данных, состоящий из матрицы признаков X и вектора откликов Y . При этом требуется получить модель, которая позволяла бы определять новое значение отклика y_* для нового вектора признаков x_* . Описать зависимость Y от X можно, например, с помощью линейной регрессионной модели следующего вида [189]:

$$Y = X^T \beta + E, \quad (2.5)$$

где E описывается нормальным распределением с нулевым матожиданием и дисперсией σ^2 , то есть $E \sim \mathcal{N}(0, \sigma^2)$.

Модель регрессии на основе гауссовских процессов позволяет описывать взаимосвязь между признаками и откликами с помощью скрытых переменных $f(x_i)$, $i = 1, 2, \dots, n$, представляющими собой значения некоторого скрытого гауссовского процесса в точках обучающей выборки x_i , а также заданных в явном виде базисных функций h . Ковариационная функция скрытых переменных позволяет описывать регулярность, то есть гладкость модели, а базисные функции проецируют матрицу обучающих признаков X в пространство большей размерности.

Гауссовский процесс представляет собой случайный процесс (набор некоторых индексированных случайных величин), такой, что любой конечный набор этих случайных величин имеет многомерное нормальное распределение [190]. Если $f(x)$, где $x \in R$ – это гауссовский процесс, то для заданного количества наблюдений n из обучающей выборки $x_1, x_2, \dots, x_n$ совместное распределение случайных величин $f(x_1), f(x_2), \dots, f(x_n)$ будет иметь многомерное нормальное распределение.

Гауссовский процесс $f(x)$ описывается функцией матожидания $m(x)$ и ковариационной функцией $k(x, x')$ (ядерной функцией) [191, 192]:

$$m(x) = \mathbb{E}(f(x)), \quad (2.6)$$

$$k(f(x), f(x')) = k(x, x') = \mathbb{E}[(f(x) - m(x))(f(x') - m(x')))]. \quad (2.7)$$

При заданных $m(x)$ и $k(x, x')$ гауссовский процесс можно определить как:

$$f(x) \sim GP(m(x), k(x, x')). \quad (2.8)$$

Тогда модель регрессии на основе гауссовских процессов можно представить в виде:

$$Y = h(x)^T \beta + f(x), \quad (2.9)$$

где $f(x)$ представляет собой гауссовский процесс $f(x) \sim GP(0, k(x, x'))$; $h(x)$ – набор базисных функций, трансформирующих исходный вектор признаков x в пространство большей размерности; β – вектор коэффициентов базисной функции.

Вид базисной функции определяет, каким образом будет получена преобразованная матрица признаков H из исходной матрицы признаков X [192]. В данной работе рассматривались следующие виды базисных функций.

1. Отсутствие базисной функции (*none*) означает, что H представляет собой пустую матрицу.
2. Константная базисная функция (*constant*) означает, что H представляет собой вектор размерностью N , состоящий только из единиц.
3. Линейная базисная функция (*linear*) определяет матрицу H как матрицу признаков X , которая расширена слева за счет вектора размерностью N , состоящего только из единиц:

$$H = [1, X]. \quad (2.10)$$

4. Квадратичная базисная функция (*pureQuadratic*) определяет матрицу H как матрицу признаков X , которая расширена слева за счет вектора размерностью N , состоящего только из единиц, а также справа с помощью матрицы X_2 , каждый элемент которой представляет собой квадрат соответствующего элемента из матрицы X :

$$H = [1, X, X_2]. \quad (2.11)$$

Ковариационная функция позволяет определить, каким образом отклик, соответствующий наблюдению x_i зависит от откликов, соответствующих другим наблюдениям x_j , $i \neq j, j = 1, 2, \dots, n$ [193]. Она обозначается $k(x_i, x_j | \theta)$, где θ – параметры функции, x_i – вектор признаков, соответствующий i -му наблюдению, x_j – вектор признаков, соответствующий j -му наблюдению, причем:

$$r = \sqrt{(x_i - x_j)^T (x_i - x_j)}. \quad (2.12)$$

В качестве параметров θ ковариационной функции в большинстве случаев используются характеристика масштаба длины процесса σ_l и среднеквадратичное отклонение колебаний шума σ_f :

$$\theta = [\theta_1 \ \theta_2] = [\log \sigma_l \ \log \sigma_f]. \quad (2.13)$$

В данной работе рассматриваются следующие ковариационные функции.

1. Квадратная экспоненциальная функция (*squaredexponential*):

$$k(x_i, x_j | \theta) = \sigma_f^2 \cdot \exp\left(-\frac{r^2}{2 \cdot \sigma_l^2}\right). \quad (2.14)$$

2. Экспоненциальная функция (*exponential*):

$$k(x_i, x_j | \theta) = \sigma_f^2 \cdot \exp\left(-\frac{r}{\sigma_l}\right). \quad (2.15)$$

3. Функция Матерна 3/2 (*matern32*):

$$k(x_i, x_j | \theta) = \sigma_f^2 \cdot \left(1 + \frac{\sqrt{3} \cdot r}{\sigma_l}\right) \cdot \exp\left(-\frac{\sqrt{3} \cdot r}{\sigma_l}\right). \quad (2.16)$$

4. Функция Матерна 5/2 (*matern52*):

$$k(x_i, x_j | \theta) = \sigma_f^2 \cdot \left(1 + \frac{\sqrt{5} \cdot r}{\sigma_l} + \frac{5 \cdot r^2}{3 \cdot \sigma_l^2}\right) \cdot \exp\left(-\frac{\sqrt{5} \cdot r}{\sigma_l}\right). \quad (2.17)$$

При отсутствии наблюдений, как показано на рисунке 2.1, модель гауссовской регрессии сводится к бесконечному количеству априорных (безусловных) ковариационных функций, функция матожидания равна 0. При появлении первого наблюдения (x_0, y_0) ковариационные функции становятся апостериорными (условными). Более вероятными в распределении функций

становятся те, значения которых в точке x_0 близко к y_0 , как показано на рисунке 2.2. При добавлении следующего наблюдения (x_1, y_1) более вероятными становятся функции, которые «проходят» вблизи обеих точек, как показано на рисунке 2.3. Визуализации регрессии на основе гауссовских процессов для рисунков получены с помощью инструмента, представленного в работе [194].

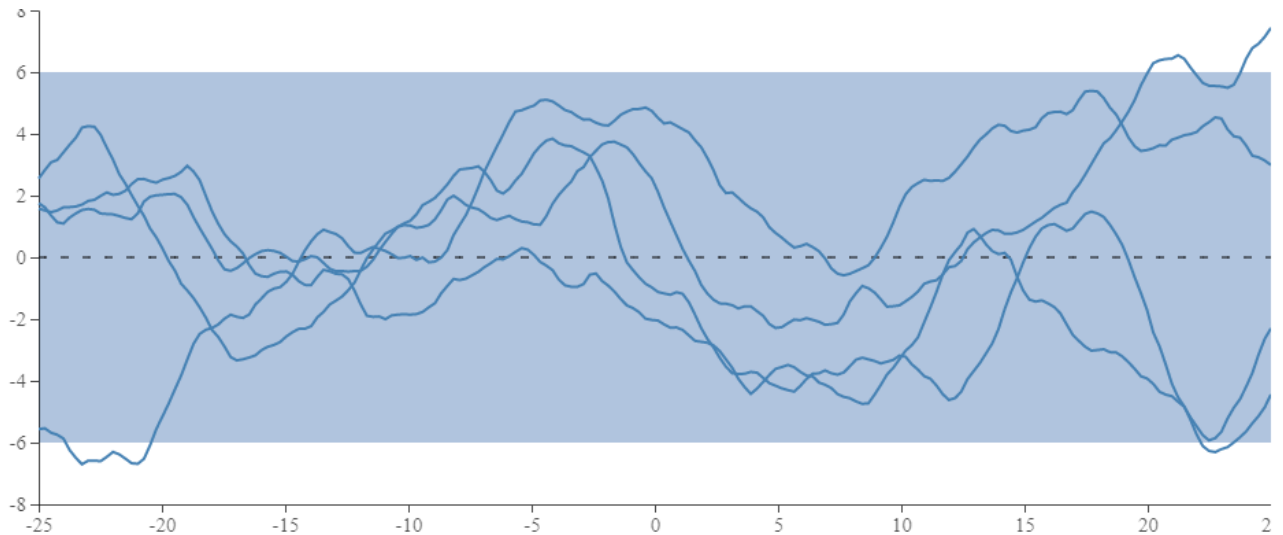


Рисунок 2.1 – Априорные ковариационные функции

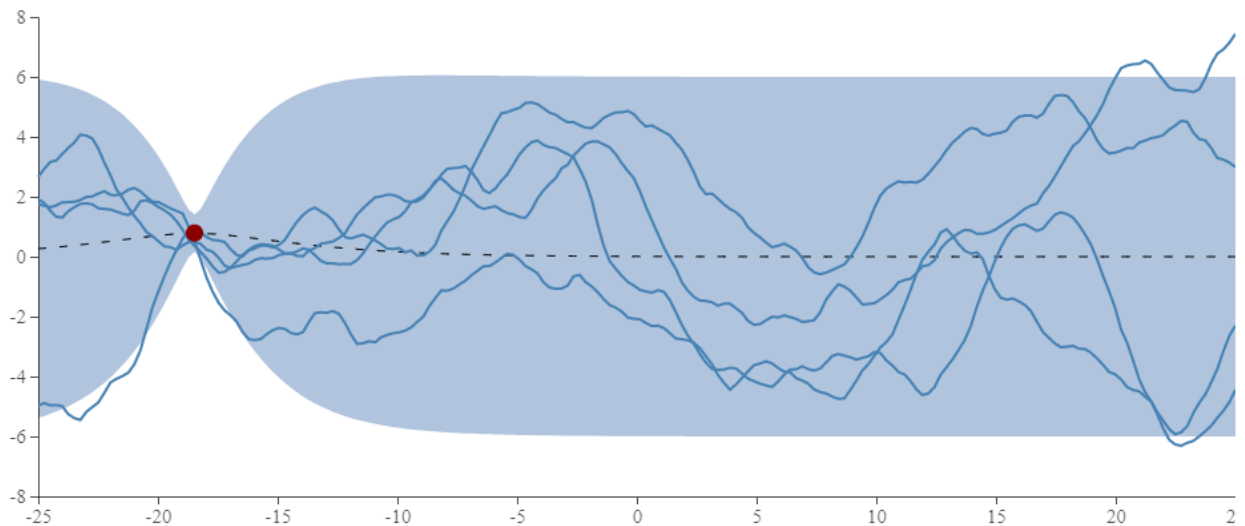


Рисунок 2.2 – Апостериорные ковариационные функции при наличии одного наблюдения

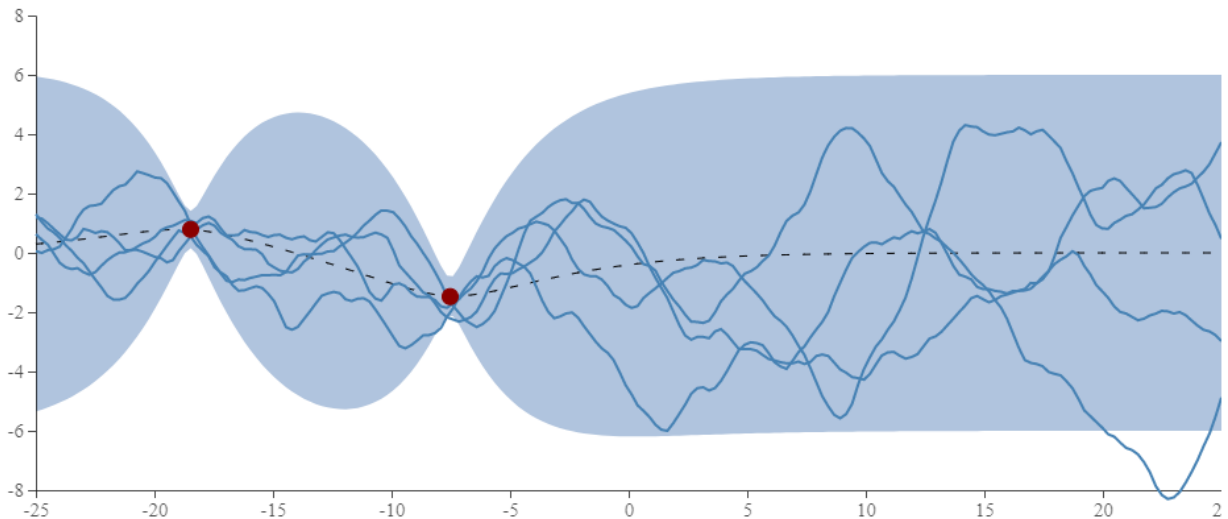


Рисунок 2.3 – Апостериорные ковариационные функции при наличии двух наблюдений

Модель регрессии на основе гауссовских процессов позволяет даже при значительном количестве признаков эффективно описывать зависимости при наличии противоречащих друг другу, «зашумленных» наблюдений, что продемонстрировано на рисунке 2.4.

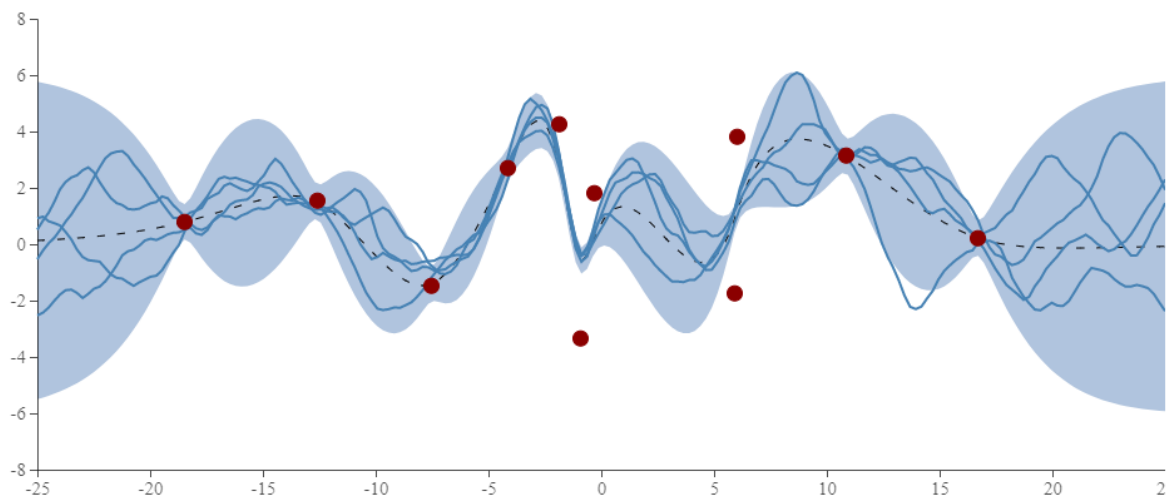


Рисунок 2.4 – Апостериорные ковариационные функции при наличии противоречащих наблюдений

В данной работе в рамках подбора гиперпараметров для модели регрессии на основе гауссовских процессов осуществляется выбор следующих значений [195]:

1. *Sigma* – начальное значение среднеквадратичного отклонения остатков регрессии σ_n , диапазон значений $[10^{-4}; \max(10^{-3}, 10 \cdot \sqrt{D[Y]})]$.

2. *BasisFunction* – вид базисной функции, через которую определяется функция матожидания, диапазон значений [*constant, none, linear, pureQuadratic*].

3. *KernelFunction* – вид ковариационной функции, диапазон значений [*squaredexponential, exponential, matern32, matern52, rationalquadratic*].

4. *KernelScale* – параметр масштабирования для ковариационной функции θ_1 , диапазон значений [$\max(\max(X) - \min(X)) \cdot 10^{-3}; \max(\max(X) - \min(X))$].

Далее представлено описание алгоритма регрессии на основе гауссовских процессов, основанное на алгоритме, разработанном авторами в [192], и соответствующее реализации данной модели машинного обучения в *Statistics and Machine Learning Toolbox* ПК *MATLAB* [195].

Алгоритм 1: Алгоритм обучения модели регрессии на основе гауссовских процессов (GPR) и получения прогнозов

Вход:

X – матрица признаков из обучающей выборки

Y – вектор откликов из обучающей выборки,

BasisFunction – вид базисной функции,

KernelFunction – вид ковариационной функции,

σ_n^0 (*Sigma*) – начальное приближение среднеквадратичного отклонения гауссовского белого шума,

θ_1^0 (*KernelScale*) – начальное приближение характеристики масштаба длины процесса,

x_* – вектор тестовых признаков (одно тестовое наблюдение).

1. Максимизация логарифма функции предельного правдоподобия для определения значений параметров ковариационной функции и среднеквадратичного отклонения гауссовского шума $\eta = [\theta_1, \theta_2, \sigma_n]$ (шаги 2-13):

$$\eta_* = \underset{\eta}{\operatorname{argmax}} (\log p(Y|X, \eta))$$

2. В качестве начального приближения искомых параметров используются θ_1^0 и σ_n^0 . θ_2^0 определяется в соответствии с выражением $\theta_2 = \frac{\sqrt{D[Y]}}{\sqrt{2}}$.

3. Функция логарифма предельного правдоподобия представляет собой:

$$\log p(Y|X, \theta) = -\frac{1}{2} \cdot L_{Yadj}^T \cdot L_{Yadj} - \sum_i \log(L_{ii}) - \frac{N}{2} \cdot \log(2\pi)$$

Значения элементов данной целевой функции для заданных η определяются следующим образом.

4. Расчет значений ковариационной функции для каждой пары наблюдений из обучающей выборки $k(x_i, x_j)$ и параметров функции θ в соответствии с заданным видом ковариационной функции *KernelFunction*. Формирование ковариационной матрицы K .
5. Добавление к ковариационной матрице K шума σ_n^2 , выполнение разложения Холецкого для полученной матрицы.

$$L = \text{cholesky}(K + \sigma_n^2 I), \text{ где } L - \text{нижняя треугольная матрица, такая, что } LL^T = A.$$
6. Формирование преобразованной матрицы признаков H по заданному виду базисной функции *BasisFunction*.
7. Если матрица H пустая
8.
$$L_Y = L^{-1}Y$$
9.
$$L_H = L^{-1}H$$
10.
$$\beta = L_H^{-1}L_Y$$
11.
$$L_{Yadj} = L_Y - L_H\beta$$
12. Иначе β – пустая матрица, $L_{Yadj} = L^{-1}Y$
13. Выполнение шагов 4-12 по полученным на шаге 1 значениям $\eta = [\theta_1, \theta_2, \sigma_n]$, соответствующего максимуму функции логарифма предельного правдоподобия.
14.
$$Y_{H\beta} = Y - H\beta$$
15.
$$L_{YH\beta} = L^{-1}Y_{H\beta}$$
16.
$$\alpha = (L^T)^{-1}L_{YH\beta}$$
17. По результатам обучения модели получены $\alpha, \beta, \theta_1, \theta_2, \sigma_n$. С использованием данных параметров определяются математическое ожидание прогноза $\bar{y}_*$ и дисперсия прогноза $D[y_*]$ для тестового наблюдения x_* (шаги 18-21):
18.
$$k_* = K(X, x_*)$$
19.
$$\bar{y}_* = k_*^T \alpha$$
20.
$$v = L^{-1}k_*$$
21.
$$D[y_*] = k(x_*, x_*) - v^T v$$

Выход:

$\bar{y}_*$ – математическое ожидание прогноза,

$D[y_*]$ – дисперсия прогноза.

Применительно к задаче прогнозирования генерации ВЭС использованию регрессии на основе гауссовских процессов посвящен ряд работ. В статьях [196, 197] было предложено использовать данную модель машинного обучения для краткосрочного прогноза на сутки вперед. Авторами было предложено две

методики. Первая методика заключалась в использовании регрессии на основе гауссовских процессов напрямую для получения по данным ЧПП величины прогнозной мощности ВЭС. Вторая методика предполагала использование двух моделей *GPR*: на первом этапе выполнялось построение модели для коррекции значений скорости ветра, полученных по данным ЧПП, на следующем этапе с помощью второй модели выполнялся прогноз генерации ВЭС по скорректированным значениям скорости ветра. Обе методики были протестированы на наборах данных с четырех реальных ВЭС установленной мощностью порядка 50-100 МВт. Предложенные авторами методики показали на 4-17% более высокую эффективность, чем референсная модель нейронной сети прямого распространения, для всех рассматриваемых наборов данных. Нормализованная среднеквадратичная ошибка прогноза *NRMSE* находилась в диапазоне 15-17% для рассмотренных наборов данных.

В работах [198, 199] предложено использование модификации регрессии на основе гауссовских процессов. Основным отличием данной модификации является добавление трансформирующей функции, изменяющей значения откликов, используемых для обучения. Также авторами была применена методика динамического моделирования процесса прогнозирования, когда наблюдения в наборе данных становятся доступны последовательно друг за другом. Разработанные методики были проверены на данных, полученных с ВЭС установленной мощностью около 300 МВт, и сопоставлены с моделью, основанной на квантильной регрессии. Рассмотренные модели показали примерно одинаковые результаты при краткосрочном прогнозировании на 12 часов вперед со средней абсолютной ошибкой прогноза *MAPE* порядка 20-21%.

Коллективом авторов в статьях [200, 201] была рассмотрена методика скользящего окна для выделения подвыборки из обучающего набора данных, по которой выполнялась обучение модели регрессии на основе гауссовских процессов. Для валидации предложенной методики использовалось два набора данных с ВЭС установленной мощностью 300 МВт и 60 МВт. При выполнении прогнозов на 12 часов вперед предложенная методика позволила снизить

среднеквадратичную ошибку $NRMSE$ на 10% по отношению к модели нейронной сети радиально-базисных функций.

Глубокий анализ эффективности применения регрессии на основе гауссовских процессов в задаче краткосрочного прогнозирования выработки ВЭС был проведен в работе [202]. Валидация разработанной методики прогнозирования, основанной на применении модели регрессии GPR , выполнялась на наборах данных с 22 ветроэлектрических станций установленной мощностью от 12 МВт до 322 МВт, расположенных как на суше, так и на море. В рамках работы рассматривались прогнозы на 3-72 часа вперед. Для прогнозов на 24 часа вперед показатель средней ошибки прогноза $NMAE$ составил 7-13%.

В статье [203] была проанализирована возможность использования модели регрессии на основе гауссовских процессов на наборах данных со значительным количеством отсутствующих наблюдений, потерянных в результате сбоев систем передачи данных и измерительных устройств. Разработанная методика позволила расширить обучающую выборку, и в условиях отсутствия 30% исходных наблюдений показала на 10% большую точность прогноза с точки зрения показателя $NRMSE$, чем метод опорных векторов.

Исследование [204] посвящено анализу возможностей модели регрессии на основе гауссовских процессов применительно к задаче прогнозирования на 24 часа вперед мощности не только отдельных ВЭС, но и суммарной мощности нескольких ВЭС, расположенных на одной территории. Предложено несколько стратегий обучения моделей GPR : прямое, не прямое, статическое и динамическое. Также авторами разработана методика автоматического определения наиболее подходящей ковариационной функции. В работе рассматриваются наборы данных с 6 ВЭС, расположенных на одной территории. Среднеквадратичная ошибка прогноза $NRMSE$ при использовании модели регрессии на основе гауссовских процессов оказалась на 20% ниже, чем референсных моделей: нейронной сети прямого распространения и случайного леса.

В работе [205] предложено использование ансамбля из нескольких моделей регрессии на основе гауссовских процессов применительно к задаче

прогнозирования выработки отдельно стоящей ВЭУ номинальной мощностью 16 МВт. В предложенной методике данные для обучения компонентов ансамбля получают из исходной выборки с помощью бутстрэппинга, а сам итоговый прогноз ансамбля определяется с помощью генетического алгоритма оптимизации. Также авторами сравнивается эффективность различных ковариационных функций применительно к разработанному алгоритму. Среднеквадратичная ошибка *NRMSE* для предложенной ансамблевой модели примерно на 5% ниже, чем для обычной модели регрессии *GPR*.

2.2.2. Нейронная сеть прямого распространения

Искусственная нейронная сеть представляет собой математическую модель, построенную по принципу организации сетей нервных клеток. Основными составляющими нейронной сети являются слои, входящие в эти слои узлы (нейроны) и связи между этими узлами (веса). Наиболее простым типом нейронных сетей являются нейронные сети прямого распространения, в структуре которых связи между узлами не образуют цикл [206]. В рамках данной работы рассматриваются нейронные сети прямого распространения, в том числе глубокие (до трех скрытых слоев). Структура нейронной сети прямого распространения представлена на рисунке 2.5.

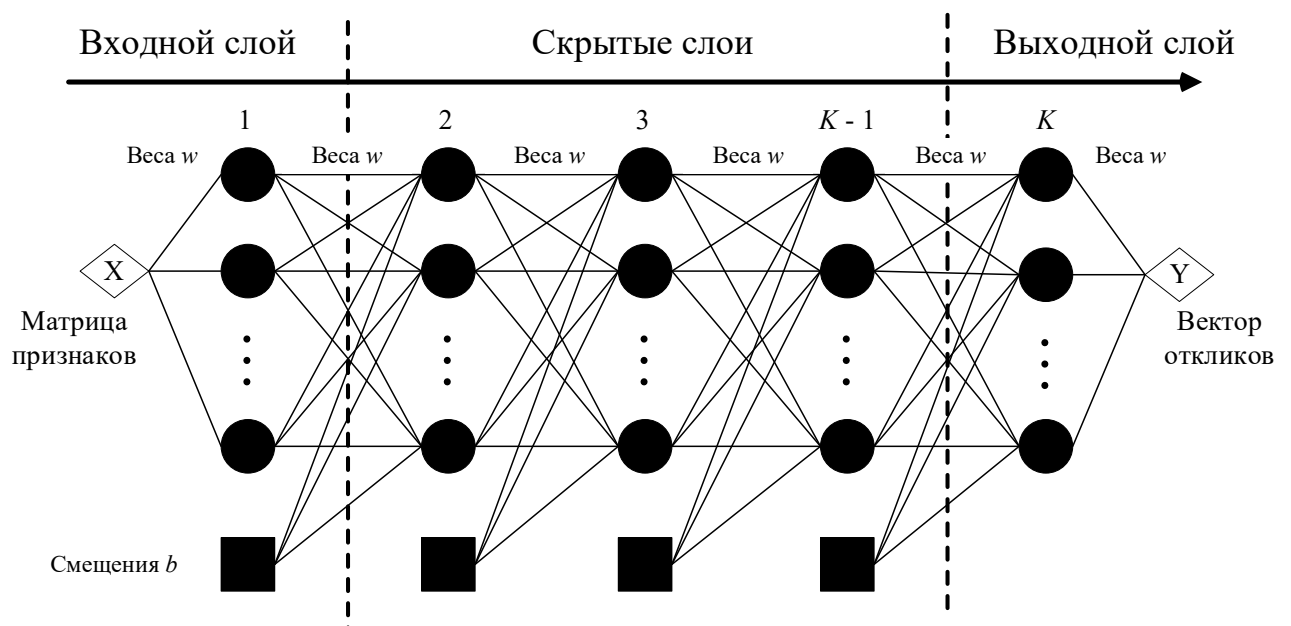


Рисунок 2.5 – Нейронная сеть прямого распространения

Связь между входным сигналом o_j^{k-1} и выходным сигналом o_i^k нейрона может быть представлена в виде выражения [207, 208]:

$$o_i^k = g \left(b^k + \sum_{j=1}^{r_{k-1}} w_{ji}^k o_j^{k-1} \right), \quad (2.18)$$

где o_i^k – выход i -го нейрона в k -м слое; g – функция активации; r_k – количество нейронов в k -м слое; w_{ij}^k – вес между j -м нейроном в k -м слое и i -м нейроном в слое $k - 1$; b^k – смещение для k -го слоя.

Функция активации $g(x)$ определяет способ изменения значения (сигнала), попадающего на вход нейрона. После этого измененный выходной сигнал используется в качестве входного в следующем слое нейронной сети. Тип функции активации нейронов определяет ее вычислительную эффективность и способность восстанавливать нелинейные зависимости. [209]. В данной работе рассматриваются следующие функции активации.

1. Линейный выпрямитель (*relu*):

$$g(x) = \begin{cases} x, & x \geq 0 \\ 0, & x < 0 \end{cases}. \quad (2.19)$$

2. Гиперболический тангенс (*tanh*):

$$g(x) = \frac{e^x - e^{-x}}{e^x + e^{-x}}. \quad (2.20)$$

3. Логистическая функция или сигмоида (*sigmoid*):

$$g(x) = \frac{1}{e^{-x} + 1}. \quad (2.21)$$

4. Тожественная функция (*none*):

$$g(x) = x. \quad (2.22)$$

В качестве начальных приближений весов w слоя k нейронной сети, как правило, используются случайные значения. В данной работе рассматриваются два способа определения исходных значений весов [195].

1. Способ *glorot* – использование равномерного распределения с нулевым математическим ожиданием и дисперсией, определяемой как

$2/(r_k + r_{k+1})$, где r_k – количество нейронов в данном слое, r_{k+1} – количество нейронов в следующем слое (для выходного слоя равно 1).

2. Способ *he* – использование нормального распределения с нулевым математическим ожиданием и дисперсией, определяемой как $2/r_k$, где r_k – количество нейронов в данном слое.

В качестве начальных приближений смещений b нейронной сети используются также два способа – смещения принимаются либо равными единице (*ones*), либо равными нулю (*zeros*) [195].

В данной работе в рамках подбора гиперпараметров для модели нейронной сети прямого распространения осуществляется выбор следующих значений [195]:

1. *Activations* – вид функции активации нейрона, диапазон значений [*relu, tanh, sigmoid, none*].

2. *Lambda* – коэффициент скорости обучения (мера регуляризации) λ , диапазон значений $[10^{-5}/N; 10^5/N]$.

3. *LayerBiasesInitializer* – способ инициализации смещений в слоях нейронной сети, диапазон значений [*zeros, ones*].

4. *LayerWeightsInitializer* – способ инициализации весов между нейронами нейронной сети, диапазон значений [*glorot, he*].

5. *NumLayers* – количество скрытых слоев, диапазон значений [1; 3].

6. *Layer_1_Size* – количество нейронов в первом скрытом слое, диапазон значений [0; 300].

7. *Layer_2_Size* – количество нейронов во втором скрытом слое, диапазон значений [0; 300].

8. *Layer_3_Size* – количество нейронов в третьем скрытом слое, диапазон значений [0; 300].

Далее представлено описание алгоритма обучения нейронной сети прямого распространения и получения прогнозов с помощью обученной модели [210], соответствующее реализации данной модели машинного обучения в *Statistics and Machine Learning Toolbox* ПК *MATLAB* [195].

Алгоритм 2: Алгоритм обучения модели нейронной сети прямого распространения (FFNN) и получения прогнозов

Вход:

X – матрица признаков из обучающей выборки,

Y – вектор откликов из обучающей выборки,

$Activations$ – вид функции активации нейрона,

λ ($Lambda$) – коэффициент скорости обучения,

$LayerBiasesInitializer$ – способ инициализации смещений,

$LayerWeightsInitializer$ – способ инициализации весов,

S – количество скрытых слоев (общее количество слоев $K = S + 2$);

$r_1, r_2, \dots, r_K$ – количество нейронов в каждом слое;

x_* – вектор тестовых признаков (одно тестовое наблюдение).

1. Инициализация начальных значений смещений слоев в соответствии с выбранным способом $LayerBiasesInitializer$.
2. Инициализация начальных значений весов в соответствии с выбранным способом $LayerWeightsInitializer$.
3. Минимизация функции средних квадратичных отклонений для каждого i из N наблюдений между выходами нейронов выходного слоя o_i^K и откликами из обучающей выборки y_i для определения значений весов W и смещений B в каждом слое нейронной сети $\eta = [W, B]$ (шаги 4-12):

$$\eta_* = \underset{\eta}{\operatorname{argmin}}(E(\eta, X)) = \underset{\eta}{\operatorname{argmin}} \left(\frac{1}{2N} \sum_{i=0}^N (o_i^K(\eta, X) - y_i)^2 \right).$$

4. Прямой ход. Определение значений на выходе каждого нейрона по текущим значениям весов W_{n-1} и смещений B_{n-1} для каждого вектора признаков из обучающей выборки x_i :
5. Инициализация входного слоя $O^0 = x_i$.
6. Обработка скрытых слоев. Для каждого k -го слоя от 1 до $K - 1$:
7. $O_k = g(B_k + O_{k-1} \cdot W_k)$
8. Обработка выходного слоя:
9. $O_K = B_K + O_{K-1} \cdot W_K$
10. Обратный ход. Обновление значений весов W_n и смещений B_n на итерации n :
11. $W_n = W_{n-1} + \lambda \frac{\partial E(\eta, X)}{\partial W_{n-1}}$
12. $B_n = B_{n-1} + \lambda \frac{\partial E(\eta, X)}{\partial B_{n-1}}$
13. По результатам обучения модели получены значения всех весов W и смещений B . С использованием данных параметров определяется прогнозное значение y_* для тестового наблюдения x_* :

14. | Выполнение шагов 5-9, в качестве x_i выступает x_* , полученное значение на выходе выходного слоя O_K равняется искомому y_* .

Выход:

y_* – прогнозное значение.

Нейронные сети прямого распространения также нашли применение в задаче прогнозирования генерации ВЭС.

Авторами исследования [211] было рассмотрено использование нейронной сети прямого распространения для прогнозирования генерации всех ВЭС в Португалии на 1 час вперед. Общая установленная мощность прогнозируемой генерации составляла 5 ГВт. Модель *FFNN* сравнивалась с наивной моделью прогноза. В зависимости от времени года *NRMSE* рассматриваемого подхода находилась в диапазоне от 1,9% до 2,7%, при этом обеспечивалось улучшение на 15-38% по сравнению с референсной моделью.

В статье [212] проанализирована эффективность использования нейронной сети прямого распространения для различных горизонтов прогнозирования от 3 ч до 36 ч. Тестирование модели *FFNN* выполнялось по данным с 9 различных ВЭС. В работе также рассматривалась эффективность использования при обучении модели метеоданных с соседних ВЭС. Результаты исследования показывают, что средняя абсолютная ошибка прогноза составила порядка 18% для горизонта 3 ч и 35% для горизонта 24 ч.

В работе [213] было предложено использование генетического алгоритма оптимизации для обучения нейронной сети прямого распространения. Данный подход был применен для краткосрочного прогнозирования генерации одной ВЭУ мощностью 2,5 МВт на сутки вперед. Предложенная методика позволила в 4 раза снизить среднюю квадратичную погрешность прогноза по сравнению с прогнозом по наивному подходу.

Исследование [214] посвящено применению модели нейронной сети прямого распространения совместно с методикой разделения исходной обучающей выборки на подвыборки с помощью метода кластеризации *k*-средних. В данной работе рассматривались горизонты прогнозирования от 1 ч до 6 ч. В зависимости от горизонта прогнозирования и количества кластеров, на которые разбивалась

обучающая выборка, предложенный подход позволял снизить среднюю квадратичную погрешность прогноза на 50-80%.

Анализ влияния на ошибку прогноза различных структур нейронных сетей прямого распространения, функций активации нейронов и способов обучения нейронной сети был проведен в работе [215]. Проверка данных моделей выполнялась для 4 различных ВЭС. Результаты исследования показывают, что в зависимости от структуры, функции активации и способа обучения средняя нормализованная квадратичная ошибка прогноза при использовании модели *FFNN* может находиться в диапазоне от 3,5% до 6%.

2.2.3. Метод опорных векторов

Идея метода опорных векторов основана на том, что восстановление регрессии должно выполняться не для всех пар наблюдение-отклик, а для их подмножества. При этом вводится понятие зоны нечувствительности, которая определяется параметром модели ε , который задается перед началом обучения. Все пары наблюдение-отклик, оказавшиеся внутри зоны нечувствительности, не попадают в минимизируемую целевую функцию. Можно сказать, что остатки регрессии для этих пар заранее принимаются равными нулю. Все остальные пары, оказавшиеся за зоной нечувствительности или на ее границах, называются опорными векторами [216, 217]. На рисунке 2.6 представлена иллюстрация к решению задачи линейной регрессии с помощью метода опорных векторов.

Решение задачи поиска аппроксимирующей гиперповерхности выполняется исходя из условий, что как можно большее количество точек должно оказаться внутри зоны нечувствительности и как можно большее количество опорных векторов оказалось на границе зоны нечувствительности. Для случаев нелинейной регрессии используются ядерные функции, которые позволяют проецировать признаки X в пространство большей размерности. Регрессионная модель принимает вид [217, 218]:

$$f(x) = \sum_{n=1}^N (\alpha_n - \alpha_n^*) G(x_n, x) + b, \quad (2.23)$$

где α_n и α_n^* – множители Лагранжа для каждого наблюдения x_n ; $G(x_n, x)$ – ядерная функция; b – свободный член (смещение).

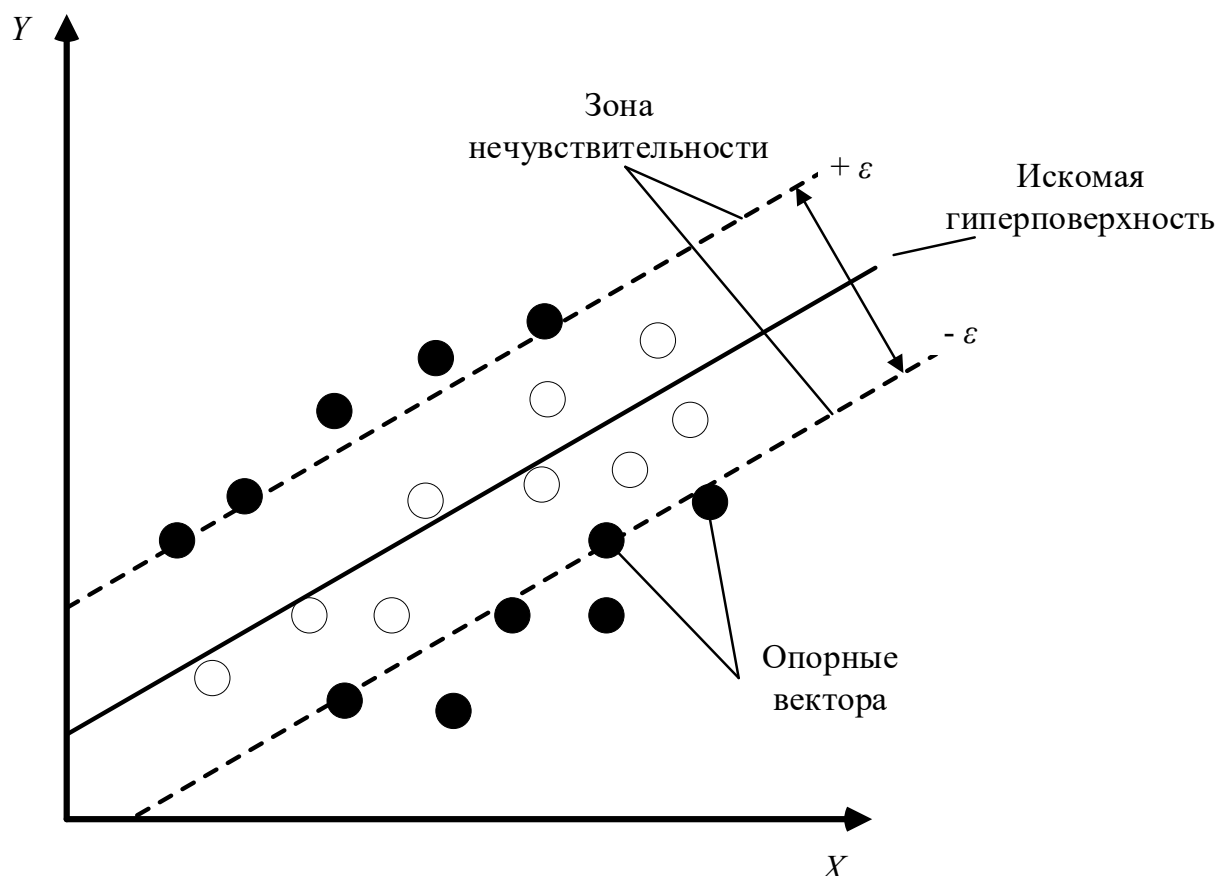


Рисунок 2.6 – Метод опорных векторов

Для наблюдений внутри зоны нечувствительности ε , не являющихся опорными векторами, α_n и α_n^* равны 0.

В данной работе рассматриваются следующие ядерные функции.

1. Линейная (*linear*):

$$G(x_i, x_j) = x_i x_j. \quad (2.24)$$

2. Функция Гаусса или радиально-базисная функция (*gaussian*):

$$G(x_i, x_j) = \exp(-\|x_i - x_j\|^2), \quad (2.25)$$

где $\|x_i - x_j\|$ – евклидово расстояние между наблюдениями x_i и x_j .

3. Полиномиальная (*polynomial*):

$$G(x_i, x_j) = (1 + x_i x_j)^q, \quad (2.26)$$

где q – порядок полинома.

В данной работе в рамках подбора гиперпараметров для модели регрессии на основе метода опорных векторов выполняется выбор следующих значений [195]:

1. *Epsilon* – половина ширины зоны нечувствительности ε , диапазон значений $[10^{-3} \cdot IQR(Y)/1,349; 10^2 \cdot IQR(Y)/1,349]$, где IQR – межквартильный размах.

2. *BoxConstraint* – параметр регуляризации C , (чем меньше параметр C , тем меньше α_n и α_n^* , тем меньше влияние наблюдений, находящихся далеко от зоны нечувствительности, на результат поиска гиперповерхности), диапазон значений $[10^{-3}; 10^3]$.

3. *KernelFunction* – вид ядерной функции, диапазон значений $[gaussian, linear, polynomial]$.

4. *KernelScale* – параметр масштабирования для ядерной функции θ , диапазон значений $[10^{-3}; 10^3]$.

5. *PolynomialOrder* порядок полинома (только для полиномиальной ядерной функции) q , диапазон значений $[2, 4]$.

Далее представлен описание алгоритма восстановления регрессии с помощью метода опорных векторов, основанное на алгоритме, представленном автором в [218], и соответствующее реализации данной модели машинного обучения в *Statistics and Machine Learning Toolbox* ПК *MATLAB* [195].

Алгоритм 3: Алгоритм обучения модели регрессии на основе метода опорных векторов (*SVM*) и получения прогнозов

Вход:

X – матрица признаков из обучающей выборки

Y – вектор откликов из обучающей выборки,

ε (*Epsilon*) – половина ширины зоны нечувствительности,

C (*BoxConstraint*) – параметр регуляризации,

KernelFunction – вид ядерной функции,

θ (*KernelScale*) – параметр масштабирования для ядерной функции,

x_* – вектор тестовых признаков (одно тестовое наблюдение).

1. Минимизация целевой функции для определения значений множителей Лагранжа α и свободного члена b (шаги 2-6):

$$[\alpha_*, b] = \operatorname{argmin}_{\alpha, b} \frac{1}{2} \sum_{i=1}^N \sum_{j=1}^N (\alpha_i - \alpha_i^*)(\alpha_j - \alpha_j^*) G(x_i x_j) + \varepsilon \sum_{i=1}^N (\alpha_i + \alpha_i^*) - \sum_{i=1}^N y_i (\alpha_i - \alpha_i^*),$$

с учетом ограничений

$$\sum_{i=1}^N (\alpha_n - \alpha_n^*) = 0,$$

$$0 \leq \alpha_n \leq C \text{ для всех } n \in N,$$

$$0 \leq \alpha_n^* \leq C \text{ для всех } n \in N,$$

$$\alpha_n (\varepsilon + \xi_n - y_n + f(x_n)) = 0 \text{ для всех } n \in N,$$

$$\alpha_n^* (\varepsilon + \xi_n^* + y_n - f(x_n)) = 0 \text{ для всех } n \in N,$$

$$\xi_n (C - \alpha_n) = 0 \text{ для всех } n \in N,$$

$$\xi_n^* (C - \alpha_n^*) = 0 \text{ для всех } n \in N.$$

2. Определение значений α и b с помощью алгоритма последовательной минимальной оптимизации (SMO) [219].
Упрощенное описание алгоритма следующее:
3. Инициализация значений α и b равными 0.
4. Выбор двух значений α_i и α_j по выбранным эвристическим правилам.
5. Оптимизация целевой функции по двум выбранным множителям Лагранжа α_i и α_j с учетом ограничений, остальные множители Лагранжа остаются неизменными.
6. Аналитическое определение величины b .
7. Повторение шагов 4-6 до тех пор, пока не будут соблюдены ограничения для всех наблюдений.
8. По результатам обучения модели получены значения всех множителей Лагранжа α и α^* , а также значение коэффициента свободного члена b . С использованием данных параметров определяется прогнозное значение y_* для тестового наблюдения x_* в соответствии с выражением:

$$y_* = \sum_{n=1}^N (\alpha_n - \alpha_n^*) G(x_*, x) + b$$

Выход:

y_* – прогнозное значение.

Использование метода опорных векторов было предложено в ряде исследований, посвященных прогнозированию генерации ВЭС.

Исследование [220] посвящено оценке применимости метода опорных векторов применительно к задаче оперативного и краткосрочного прогнозирования

генерации ВЭС. Проверка модели регрессии на основе метода опорных векторов осуществлялась по данным, полученным с ВЭС установленной мощностью 30 МВт. Средняя нормализованная ошибка прогноза при краткосрочном прогнозировании на 8 ч вперед составила 12,1%, при оперативном прогнозировании на 1 ч вперед – 1,1%. При сравнении модели *SVM* с наивной моделью было определено, что снижение погрешности составляет более 60% для оперативного прогноза и около 38% для краткосрочного.

В работе [221] было предложено использование гибридной ядерной функции, объединяющей полиномиальную функцию с функцией Гаусса, а также улучшенного алгоритма оптимизации по методу роя частиц для выбора гиперпараметров модели *SVM* и прогнозирования генерации ВЭС. Результаты валидации предложенного подхода с использованием данных с одной ВЭС мощностью порядка 1,4 ГВт показывают снижение среднеквадратичной погрешности прогноза почти в два раза по сравнению с моделью *ARMA*.

В статье [222] также проанализирована применимость биоинспирированного алгоритма (*Chicken Swarm Optimization, CSO*) и метода опорных векторов применительно к задаче прогнозирования генерации ВЭС. Проверка предложенного улучшенного алгоритма *CSO* осуществлялась по данным с одной ВЭУ мощностью 2,05 МВт. Результаты исследования показывают снижение среднеквадратичной ошибки на 25% по сравнению с классической версией *CSO*.

Авторами исследования [223] было рассмотрено совместное использование метода опорных векторов и методики кластеризации наблюдений исходной выборки. В данной работе при краткосрочном прогнозировании генерации ВЭС мощностью 45 МВт на сутки вперед повышение эффективности прогнозирования генерации ВЭС составило от 17% до 46% в зависимости от времени года по сравнению с моделью нейронной сети прямого распространения.

Анализ эффективности использования метода опорных векторов, методики отбора признаков с помощью восстановления фазового пространства и генетического алгоритма для выбора гиперпараметров модели *SVM* был проведен в работе [224]. Валидация предложенной методики выполнялась по данным одной

ВЭУ мощностью 300 кВт. Результаты показывают снижение средней нормализованной погрешности прогноза на 10% при прогнозировании на 1 ч вперед по сравнению с авторегрессионной моделью.

2.2.4. Регрессионное бинарное дерево решений

Бинарное дерево решений представляет собой древовидную структуру, в терминальных узлах которой находятся значения откликов, а во всех остальных внутренних узлах – условия перехода от данного узла к нижестоящим. Решением задачи регрессии в данном случае будет являться определение структуры дерева в соответствии с заданными гиперпараметрами дерева, условий перехода между узлами в соответствии с заданными правилами разбиения и значений откликов в терминальных узлах [225]. Тогда модель регрессионного дерева решений можно представить следующим образом [226]:

$$f(x) = \sum_{k=1}^K c_k E(x \in R_k), \quad (2.27)$$

где K – количество терминальных узлов; $R_1, R_2, \dots, R_K$ – подмножества пространства признаков, получаемые в результате разбиения деревом решений; c_k – значение отклика в терминальном узле; $E(x \in R_k)$ – единичная матрица, определяющая принадлежность признака к заданному подмножеству пространства признаков.

Значения отклика в терминальном узле c_k определяются как среднее значение откликов наблюдений y_i , попавших в подмножество R_k [226]:

$$c_k = \frac{1}{N_k} \sum_{i=1}^{N_k} (y_i | x_i \in R_k), \quad (2.28)$$

где N_k – количество наблюдений в терминальном узле; $y_i | x_i \in R_k$ – значение i -го отклика, сопоставленное наблюдению x_i , которое принадлежит подмножеству R_k .

Условие перехода между узлами можно представить в виде функции, которая представляет собой сравнение значения признака t с некоторым пороговым значением t [227]:

$$B_v(x, m, t) = [x_m \leq t]. \quad (2.29)$$

где B_v – правило разбиения во внутреннем узле v ; t – пороговое значение; x_m – значение признака m .

Если для значения признака x_m получено $B_v(x_m) = 1$, то при движении по узлам дерева выполняется переход вправо, если $B_v(x_m) = 0$, то влево.

Пример построения регрессионного дерева решений с максимальным количеством разбиений, равным 5, представлен на рисунке 2.7.

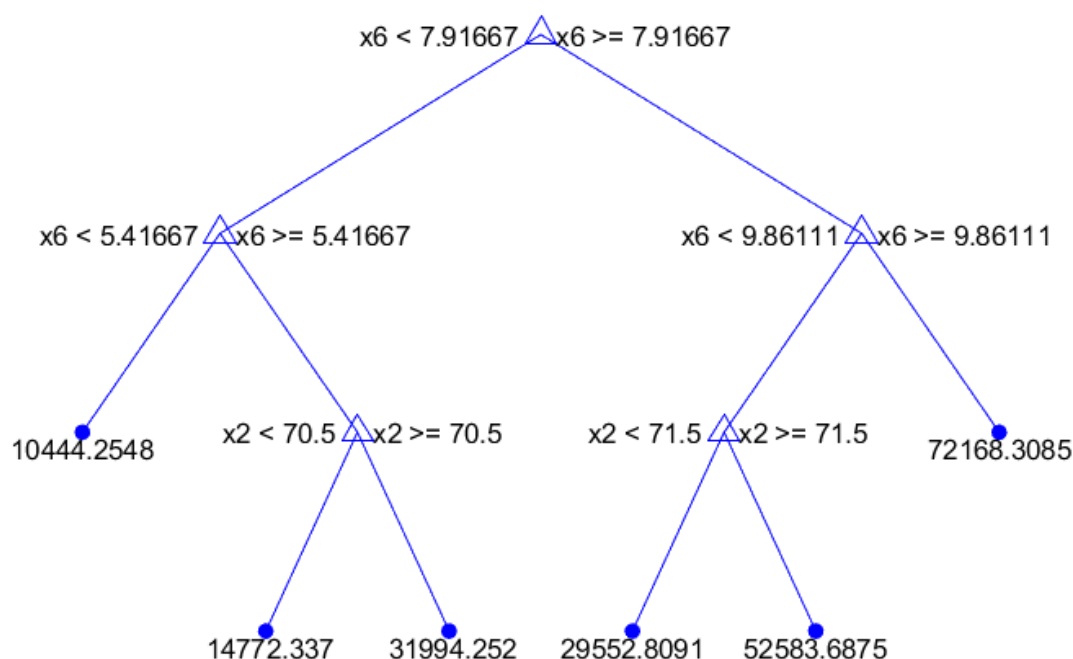


Рисунок 2.7 – Регрессионное дерево решений

В данной работе в рамках подбора гиперпараметров для модели регрессии на основе бинарного дерева решений выполняется выбор следующих значений [195]:

1. *MinLeafSize* – минимальное количество наблюдений в узле дерева, диапазон значений $[1; \max(2, \lfloor N/2 \rfloor)]$, где N – общее количество наблюдений.

2. *MaxNumSplits* – максимальное количество разбиений в дереве (количество нетерминальных узлов), диапазон значений $[1; \max(2, N - 1)]$.

3. *NumVariablesToSample* – максимальное количество случайных признаков, рассматриваемых при каждом делении M_s , диапазон значений $[1, M]$, где M – общее количество признаков.

Далее представлено описание алгоритма *CART* (*Classification and Regression Trees*) для обучения регрессионного бинарного дерева решений [225], соответствующее реализации данной модели машинного обучения в *Statistics and Machine Learning Toolbox* ПК *MATLAB* [195].

Алгоритм 4: Алгоритм обучения модели регрессионного решающего дерева (RDT) и получения прогнозов

Вход:

X – матрица признаков из обучающей выборки,

Y – вектор откликов из обучающей выборки,

$MinLeafSize$ – минимальное количество наблюдений в узле дерева,

$MaxNumSplits$ – максимальное количество разбиений в дереве,

M_s ($NumVariablesToSample$) – максимальное количество случайных признаков, рассматриваемых при каждом делении,

x_* – вектор тестовых признаков (одно тестовое наблюдение).

1. Вся матрица наблюдений X и вектор откликов Y помещается в родительский узел.
2. Разделение узла:
3. Выбирается случайный признак из M_s рассматриваемых признаков из общего набора признаков.
4. Для всех m от 1 до M_s :
5. Сортировка по значению признака всех пар признак-отклик.
6. Определение среднего значения откликов в родительском узле:

$$c_P = \overline{y_{N_P}} = \frac{1}{N_P} \sum_{i=1}^{N_P} y_i.$$
7. Определение величины дисперсии откликов ε_P в родительском узле с N_P наблюдениями:

$$\varepsilon_P = \frac{1}{N_P} \sum_{i=1}^{N_P} (y_i - \overline{y_{N_P}}).$$
8. Проверка всех возможных разбиений множества пар признак-отклик на два подмножества. Для каждого j -го разбиения:
9. Определение величины дисперсии откликов ε_L^j в левом узле-потомке с N_L^j наблюдениями:

$$\varepsilon_L^j = \frac{1}{N_L^j} \sum_{i=1}^{N_L^j} (y_i - \overline{y_{N_L^j}}).$$
10. Определение величины дисперсии откликов ε_R^j в правом узле-потомке с N_R^j наблюдениями:

- | | |
|-----|--|
| 11. | $\varepsilon_R^j = \frac{1}{N_R^j} \sum_{i=1}^{N_R^j} (y_i - \overline{y_{N_R^j}}).$ <p>Определение величины снижения дисперсии ΔI для j-го разбиения:</p> $\Delta I^j = \varepsilon_P - \frac{N_L^j}{N_P} \varepsilon_L^j - \frac{N_R^j}{N_P} \varepsilon_R^j.$ |
| 12. | <p>Определение признака m и соответствующего разбиения пар признак-отклик на два подмножества j с наибольшим значением ΔI.
Формирование правила разделения B_v.</p> |
| 13. | <p>Если критерии останова не достигнуты, разделение узла (шаги 3-12) выполняется либо для левого, либо для правого, либо для обоих узлов-потомков. Если достигнуты, процесс построения дерева окончен.
Критерии останова включают в себя:</p> |
| 14. | <p>Достигнуто максимальное количество разделений в дереве
<i>MaxNumSplits</i>;</p> |
| 15. | <p>Достигнуто минимальное количество наблюдений для узла-потомка
<i>MinLeafSize</i>;</p> |
| 16. | <p>Достигнуто допустимое отклонение дисперсии откликов в узле-потомке.</p> |
| 17. | <p>По результатам обучения модели получены значения откликов во всех терминальных узлах c_k и условия перехода между внутренними узлами B_v. С использованием данных параметров определяется прогнозное значение y_* для тестового наблюдения x_*:</p> |
| 18. | <p>Выполнение проверки условия B_v для узла v. Переход к правому узлу, если $B_v = 1$, и переход к левому, если $B_v = 0$.</p> |
| 19. | <p>Повторение шага 18 до достижения терминального узла. Значение отклика в терминальном узле c_k приравнивается прогнозному значению y_*.</p> |

Выход:

y_* – прогнозное значение.

Регрессионное дерево решений редко используется как самостоятельная модель при прогнозировании генерации ВЭС. Как правило, применяются ансамбли деревьев решений, которые рассматриваются далее. Тем не менее, существуют исследования, как например [228], в которых прогнозирование генерации ВЭС осуществляется с помощью одного регрессионного дерева решений. В данной работе *RDT* использовалось для прогнозов генерации одной ВЭУ мощностью 2 МВт. Результаты исследования показали значительное снижение средней

абсолютной погрешности прогноза, практически в 3 раза до 60 кВт, по сравнению с линейной регрессионной моделью.

2.2.5. Ансамблевые методы: случайный лес и градиентный бустинг

Суть ансамблевых методов машинного обучения сводится к комбинированию результатов, полученных от нескольких моделей. Итоговый результат определяется обобщением результатов, полученных от базовых моделей, входящих в ансамбль. Обобщение может представлять собой различные функции (среднее, сумма и т.д.) [226, 229].

В качестве элементов ансамбля могут выступать любые методы. Тем не менее, в качестве базовых моделей, как правило, применяются простые модели (например, деревья решений небольшой глубины), определение параметров которых не сопряжено со значительными затратами вычислительных ресурсов.

Сама по себе такая модель будет обладать низкой эффективностью. Но обобщенный результат большого количество таких моделей может обеспечить достаточно высокую точность. Количество простых моделей в рамках одного ансамбля может исчисляться сотнями. Параметры каждого последующего простого алгоритма могут формироваться случайно, либо определяться на основе эффективности тех базовых моделей, которые уже входят в композицию [230, 231].

В данной работе рассматривается два алгоритма формирования ансамблей: градиентный бустинг и случайный лес.

Модель регрессии на основе алгоритма случайного леса представляет собой ансамбль из регрессионных деревьев решений. Каждая базовая модель $h(x)$ в этом ансамбле обучается независимо друг от друга. При этом набор обучающих данных для каждой модели формируется из исходной обучающей выборки методом бустрепа: из исходной выборки формируется дополнительная выборка такого же размера, что и исходная, при этом элементы этой дополнительной выборки являются случайно выбранными элементами исходной выборки (могут быть выбраны даже несколько раз). Итоговый прогноз модели регрессии случайного

леса может быть определен как среднее значение прогнозов, получаемых от базовых моделей (регрессионных деревьев решений) [232, 233]:

$$f(x) = \frac{1}{K} \sum_{k=1}^K h_k(x), \quad (2.30)$$

где K – количество базовых моделей; $h_k(x)$ – базовая модель.

На рисунке 2.8 представлена иллюстрация к построению регрессионной модели с помощью алгоритма случайного леса.

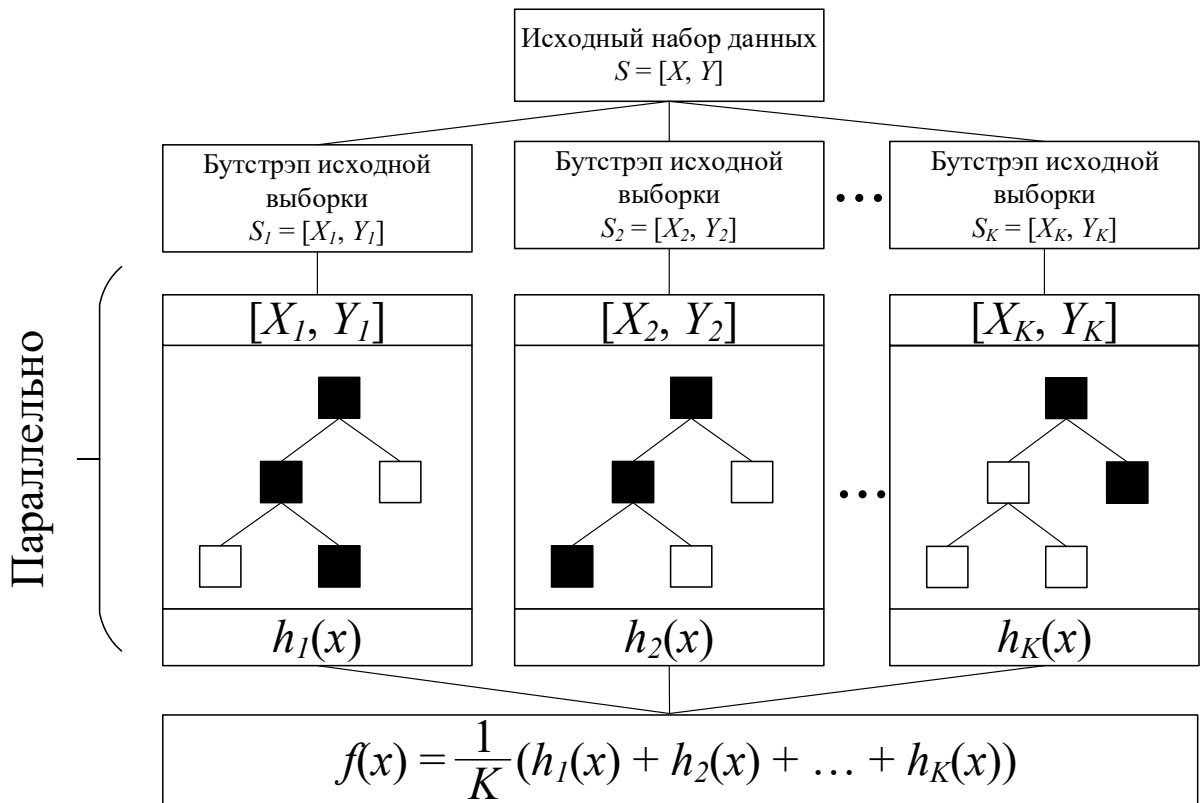


Рисунок 2.8 – Случайный лес

В данной работе в рамках подбора гиперпараметров для модели регрессии на основе алгоритма случайного леса выполняется выбор следующих значений [195]:

1. *NumLearningCycles* – количество базовых моделей в ансамбле K , диапазон значений [10, 500].

2. *MinLeafSize* – минимальное количество наблюдений в узле дерева, диапазон значений $[1; \max(2, \lfloor N/2 \rfloor)]$, где N – общее количество наблюдений.

3. *MaxNumSplits* – максимальное количество разбиений в дереве (количество нетерминальных узлов), диапазон значений $[1; \max(2, N - 1)]$.

4. *NumVariablesToSample* – максимальное количество случайных признаков, рассматриваемых при каждом делении M_s , диапазон значений $[1, M]$, где M – общее количество признаков.

Далее представлено описание алгоритма для обучения регрессионной модели случайного леса [232], соответствующее реализации данной модели машинного обучения в *Statistics and Machine Learning Toolbox* ПК *MATLAB* [195].

Алгоритм 5: Алгоритм обучения модели случайного леса (*RF*) и получения прогнозов

Вход:

X – матрица признаков из обучающей выборки,

Y – вектор откликов из обучающей выборки,

K (*NumLearningCycles*) – количество базовых моделей в ансамбле,

MinLeafSize – минимальное количество наблюдений в узле дерева,

MaxNumSplits – максимальное количество разбиений в дереве,

M_s (*NumVariablesToSample*) – максимальное количество случайных признаков,

x_* – вектор тестовых признаков (одно тестовое наблюдение).

1. Для заданного количества базовых моделей, то есть для k от 1 до K :
2. Выполняется бутстрэп. Формируются выборка $[X_k, Y_k]$.
3. Определение параметров базовой регрессионной модели $h_k(x)$ с параметрами M_s , *MinLeafSize*, *MaxNumSplits* по алгоритму 4, при этом в качестве откликов используется полученная в результате бутстрэппинга выборка, то есть в качестве обучающего набора данных выступает X_k и Y_k .
4. По результатам обучения модели получены параметры K базовых моделей $h_k(x)$. С использованием данных параметров определяется прогнозное значение y_* для тестового наблюдения x_* в соответствии с выражением:

$$y_* = \frac{1}{K} \sum_{k=1}^K h_k(x_*).$$

Выход:

y_* – прогнозное значение.

Алгоритм градиентного бустинга также, как правило, использует в качестве базовых моделей $h(x)$ регрессионные деревья решений. При этом в отличие от алгоритма случайного леса, построение ансамбля выполняется направленно: каждый новый базовый алгоритм в ансамбле дополняет уже построенные базовые модели так, чтобы уменьшить среднеквадратичное отклонение прогнозных значений, получаемых текущим ансамблем, от значений откликов обучающей

выборки [234]. Итоговый прогноз определяется как взвешенная сумма прогнозов, получаемых от базовых моделей (регрессионных деревьев решений) [235, 236]:

$$f(x) = \sum_{k=1}^K \mu_k \cdot h_k(x), \quad (2.31)$$

где K – количество базовых моделей; $h_k(x)$ – базовая модель, μ_k – весовой коэффициент, определяющий вклад базовой модели в результирующий прогноз.

На рисунке 2.9 представлена иллюстрация к построению регрессионной модели с помощью алгоритма градиентного бустинга над деревьями решений.

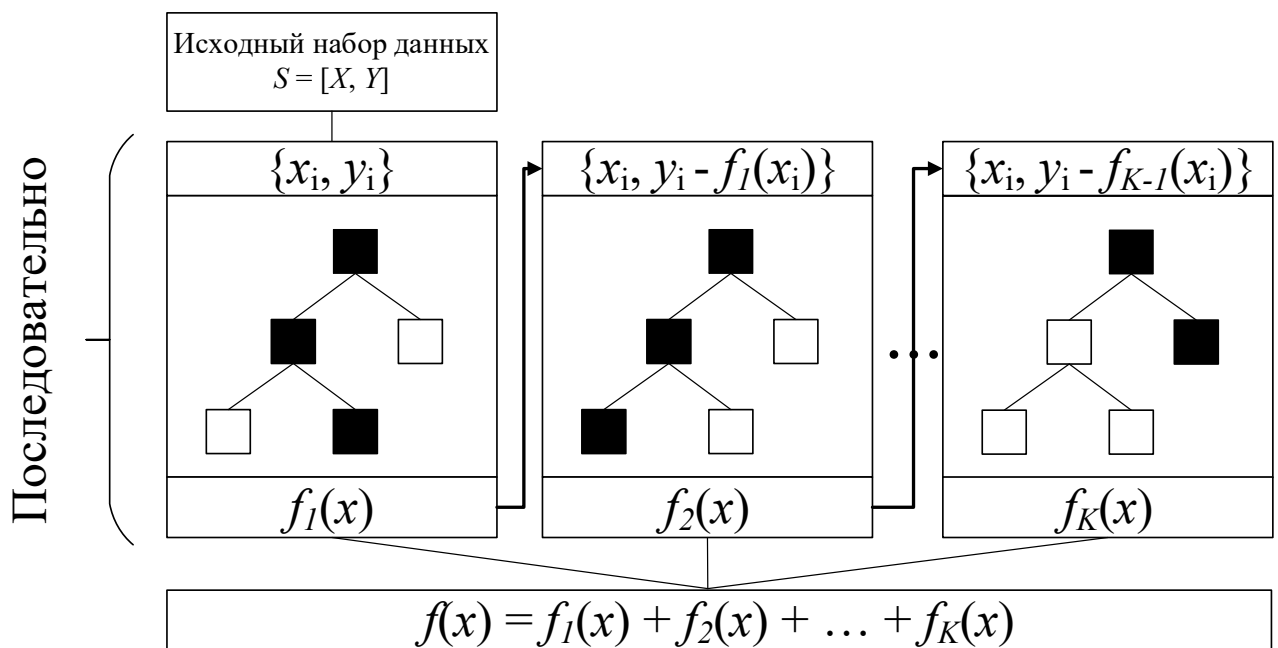


Рисунок 2.9 – Градиентный бустинг над деревьями решений

В данной работе в рамках подбора гиперпараметров для модели регрессии на основе алгоритма градиентного бустинга над деревьями решений осуществляется выбор следующих значений [195]:

1. *NumLearningCycles* – количество базовых моделей в ансамбле K , диапазон значений [10, 500].
2. *MinLeafSize* – минимальное количество наблюдений в узле дерева, диапазон значений $[1; \max(2, \lfloor N/2 \rfloor)]$, где N – общее количество наблюдений.
3. *MaxNumSplits* – максимальное количество разбиений в дереве (количество нетерминальных узлов), диапазон значений $[1; \max(2, N - 1)]$.

4. *NumVariablesToSample* – максимальное количество случайных признаков, рассматриваемых при каждом делении M_s , диапазон значений $[1, M]$, где M – общее количество признаков.

5. *LearnRate* – коэффициент скорости обучения (мера регуляризации), диапазон значений $[10^{-3}, 1]$.

Далее представлено описание алгоритма для обучения регрессионной модели градиентного бустинга над деревьями решений [235], соответствующее реализации данной модели в *Statistics and Machine Learning Toolbox* ПК MATLAB [195].

Алгоритм 6: Алгоритм обучения модели градиентного бустинга над деревьями решений (GBDT) и получения прогнозов

Вход:

X – матрица признаков из обучающей выборки,

Y – вектор откликов из обучающей выборки,

K (*NumLearningCycles*) – количество базовых моделей в ансамбле,

MinLeafSize – минимальное количество наблюдений в узле дерева,

MaxNumSplits – максимальное количество разбиений в дереве,

M_s (*NumVariablesToSample*) – максимальное количество случайных признаков,

λ (*LearnRate*) – коэффициент скорости обучения,

x_* – вектор тестовых признаков (одно тестовое наблюдение).

1. Получение первого приближения значения целевой функции как среднего значения откликов обучающей выборки Y : $f_0 = \frac{1}{N} \sum_{i=1}^N y_i$.
2. Для заданного количества базовых моделей, то есть для k от 1 до K :
3. Определение значений псевдо-остатков для каждой пары x_i и y_i для i от 1 до N как $r_{ik} = y_i - f_{k-1}(x_i)$.
4. Определение параметров базовой модели $h_k(x)$ с параметрами M_s , *MinLeafSize* и *MaxNumSplits* по алгоритму 4, при этом в качестве откликов используются значения псевдо-остатков, то есть в качестве обучающего набора данных выступает $\{x_i, r_{ik}\}_{i=1}^N$.
5. Определение величины шага: $\rho_k = \underset{\rho}{\operatorname{argmin}} \sum_{i=1}^N (r_{ik} - \rho h_k(x_i))$.
6. Обновление приближения модели: $f_k(x) = f_{k-1}(x) + \lambda \cdot \rho_k \cdot h_k(x)$.
7. По результатам обучения модели получены параметры K базовых моделей $h_k(x)$ и величина шага для каждой модели ρ_k . С использованием данных параметров определяется прогнозное значение y_* для тестового наблюдения x_* в соответствии с выражением: $y_* = f_0 + \lambda \sum_{k=1}^K \rho_k \cdot h_k(x_*)$.

Выход:

y_* – прогнозное значение.

Использованию ансамблей деревьев решений для прогнозирования генерации ВЭС посвящено значительное количество исследований. Эти методики являются распространенным инструментом для решения данной задачи.

Например, в работе [237] рассмотрена задача оперативного прогнозирования генерации ВЭС на один час вперед. В качестве исследуемого объекта использовалась ВЭС установленной мощностью 53,5 МВт. Для построения прогнозов использовалась методика, основанная на применении регрессионной модели, основанной на алгоритме случайного леса, использующей в качестве признаков как лаговые значения временного ряда мощности ВЭС, так и внешние признаки, такие, как скорость ветра и направление ветра. В качестве референсной модели выступала наивная модель прогноза. Исследование показало, что нормализованная среднеквадратичная ошибка прогноза для модели *RF* составила 3,5%, фактор улучшения составил 20% по сравнению с наивной моделью.

В статье [238] было предложено совместное использование вейвлет-преобразования для разложения временных рядов мощности и скорости ветра и алгоритма взвешенного случайного леса для прогнозирования каждой компоненты. Каждому дереву решений присваивался весовой коэффициент в итоговом прогнозе, определенный с помощью построения предварительного прогноза. Данный подход был протестирован при решении задачи оперативного прогнозирования на 5 минут вперед для двух ВЭС установленной мощностью 49,5 МВт и 53,5 МВт. При сравнении предложенной методики и классической модели регрессии на основе алгоритма случайного леса снижение среднеквадратичной ошибки прогноза составило 2,5-3 раза до величины 27-28 кВт.

Работа [239] также посвящена оперативному прогнозированию генерации ВЭС на 1 час вперед. Авторами предложена методика прогнозирования, подразумевающая совместное использование декомпозиции временного ряда мощности с помощью разложения на эмпирические моды и модели регрессии *RF*. Для проверки предложенной методики использовались данные с ВЭС установленной мощностью 16 МВт. Результаты проверки показали, что среднеквадратичная ошибка прогноза при применении предложенного подхода

составляет 1,6 МВт, при этом обеспечено снижение погрешности на 60% по сравнению с применением случайного леса без разложения на эмпирические моды.

Авторами исследования [240] было проанализировано использование градиентного бустинга над деревьями решений и различных подходов к определению обучающей выборки: использование данных только текущего месяца, только предыдущего месяца или текущего времени года. Оценка эффективности данных подходов выполнялась на данных, полученных с ВЭС установленной мощностью 70 МВт. В качестве референсной модели использовалась нейронная сеть с долгой краткосрочной памятью. В зависимости от подхода к выбору обучающей выборки среднеквадратичная ошибка прогноза для модели *RF* составила от 3,6 МВт до 4,5 МВт. Было достигнуто снижение погрешности в 1,5-2 раза по сравнению с референсной моделью.

Анализ эффективности прогнозирования суммарной генерации всех ВЭС Польши, установленная мощность которых составляет порядка 6 ГВт, был проведен в работе [241]. В статье выполнялось сравнение моделей градиентного бустинга над деревьями решений и случайного леса применительно к задаче краткосрочного прогноза на сутки вперед. Нормализованная среднеквадратичная ошибка прогноза составила 4,5% для модели *GBDT* и 5% для модели *RF*. Использование ансамблевых подходов обеспечило снижение погрешности на 14% и на 20% соответственно по сравнению с моделью нейронной сети прямого распространения с тремя скрытыми слоями.

Исследование [242] посвящено сравнению различных реализаций моделей регрессии, основанных на алгоритмах градиентного бустинга над деревьями решений и алгоритме случайного леса. Для тестирования использовались данные с одной ВЭУ мощностью 800 кВт. При выполнении краткосрочного прогнозирования на 48 часов вперед наименьшего значения нормализованной среднеквадратичной ошибки прогноза удалось получить при использовании реализации модели градиентного бустинга над деревьями решений *CatBoost* – 13%. Снижение погрешности по сравнению с референсной моделью регрессионного дерева решений *RDT* составило 30%.

В статье [243] также было выполнено сравнение различных моделей градиентного бустинга над деревьями решений, а также модели случайного леса. При прогнозировании генерации ВЭС установленной мощностью 54 МВт модели *RF* и *GBDT* позволили получить примерно одинаковые значения среднеквадратичной ошибки 0,06 МВт. При этом погрешность снизилась более, чем в 2 раза, по сравнению с референсной моделью *RDT*.

Оценка эффективности алгоритмов случайного леса и градиентного бустинга также выполнена в работе [244]. Авторами использовался набор данных, в котором значения мощности были нормализованы в диапазоне от 0 до 1. Среднеквадратичная ошибка прогноза для модели градиентного бустинга составила 0,1 о.е., почти на 40% ниже, чем у случайного леса.

Еще одно исследование, посвященное прогнозированию генерации ВЭС целой страны, выполнено в работе [245]. Авторы сравнили модели регрессии на основе случайного леса, градиентного бустинга и метода опорных векторов при прогнозировании суммарной генерации всех ВЭС Испании, а также мощности одной из этих ВЭС отдельно. Наиболее эффективной моделью при прогнозировании генерации одной ВЭС оказалась модель *RF*, нормализованная среднеквадратичная ошибка составила 7%. Фактор улучшения составил примерно 3% по сравнению с референсной моделью *SVM*. При этом при прогнозировании суммарной мощности ВЭС Испании наиболее эффективной оказались модели *SVM* и *GBDT* с нормализованной среднеквадратичной ошибкой 3,4%. Фактор улучшения по сравнению с моделью *RF* составил примерно 11%.

2.3. Подходы к оптимизации гиперпараметров и их применение для краткосрочного прогнозирования генерации ВЭС

Как было отмечено ранее, часть параметров регрессионной модели должны быть заданы перед началом обучения. Задачу определения таких параметров называют задачей оптимизации гиперпараметров. Можно выделить следующие основные подходы к определению гиперпараметров моделей машинного обучения [188, 246, 247].

Экспертный подход. Данный подход сопряжен с определением искомой модели регрессии и алгоритма ее получения, а также гиперпараметров, на основе мнения экспертов, специалистов в области машинного обучения.

Эмпирический подход. Данный подход основан на использовании эмпирических выражений для поиска гиперпараметров конкретных моделей.

Поиск по решетке или вариация параметров (Grid Search, GS). Суть данного подхода заключается в полном переборе всех возможных наборов гиперпараметров λ по заданному вручную подмножеству пространства гиперпараметров Λ [246]. На рисунке 2.10 представлена графическая интерпретация метода поиска по решетке, использованного для минимизации целевой функции (ЦФ), зависящей от двух переменных – гиперпараметр 1 и гиперпараметр 2. Всего в данном примере значение ЦФ вычислено 49 раз. Контурные линии, обозначенные оттенками фиолетового, определяют области с меньшими значениями оптимизируемой ЦФ, оттенками оранжевого – с большими. Черными звездочками обозначены наборы гиперпараметров, для которых было определено значение ЦФ. Красными метками на осях обозначены значения гиперпараметров, для которых были вычислены значения целевой функции. Как видно из рисунка, для каждого гиперпараметра было проверено лишь 7 значений. Можно сделать вывод о том, что успешность результатов поиска гиперпараметров для данного подхода в значительной степени зависит от заданных границ изменения и дискретизации каждого гиперпараметра. С расширением границ и уменьшением частоты дискретизации гиперпараметров растет эффективность данного подхода, при этом соответственно в значительной степени увеличивается суммарное время, необходимое на проверку всего рассматриваемого подмножества гиперпараметров Λ .

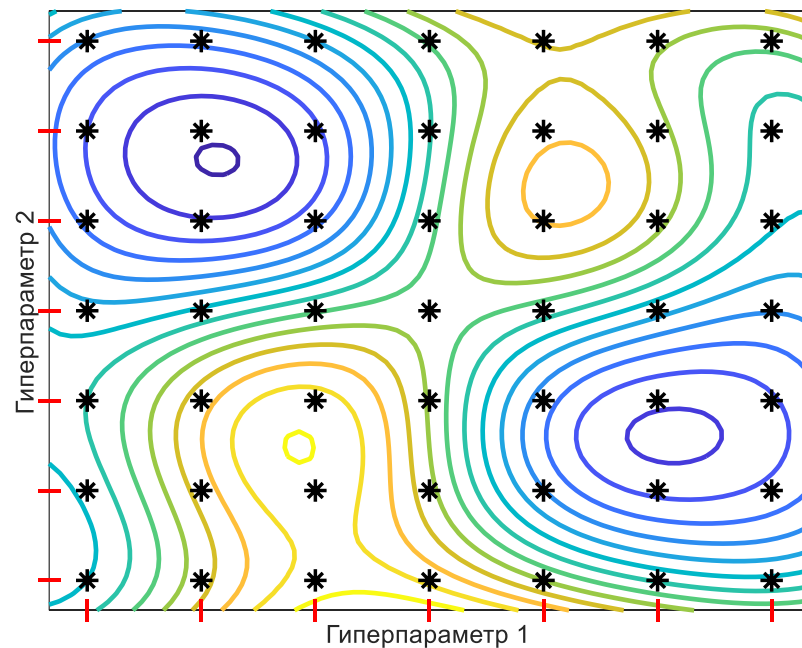


Рисунок 2.10 – Поиск по решетке

Случайный поиск (Random Search, RS). В данном подходе полный перебор заменяется на оценку эффективности некоторой выборки из случайных комбинаций гиперпараметров, принадлежащих множеству Λ . При этом частота дискретизации непрерывных величин гиперпараметров может не задаваться. В результате, можно найти точку, которая не была бы найдена при поиске по решетке. Данный метод, как правило, позволяет добиться большей эффективности при подборе гиперпараметров при меньших временных затратах [248]. На рисунке 2.11 представлена графическая интерпретация метода случайного поиска. Как и в случае с поиском по сетке, значение целевой функции было вычислено 49 раз. Красные метки на рисунке показывают, что рассматривается большее количество индивидуальных значений каждого гиперпараметра по сравнению с методом поиска по решётке. Также можно отметить, что метод случайного поиска позволил обнаружить меньшее значение целевой функции, чем метод поиска по решетке (область в правом нижнем краю).

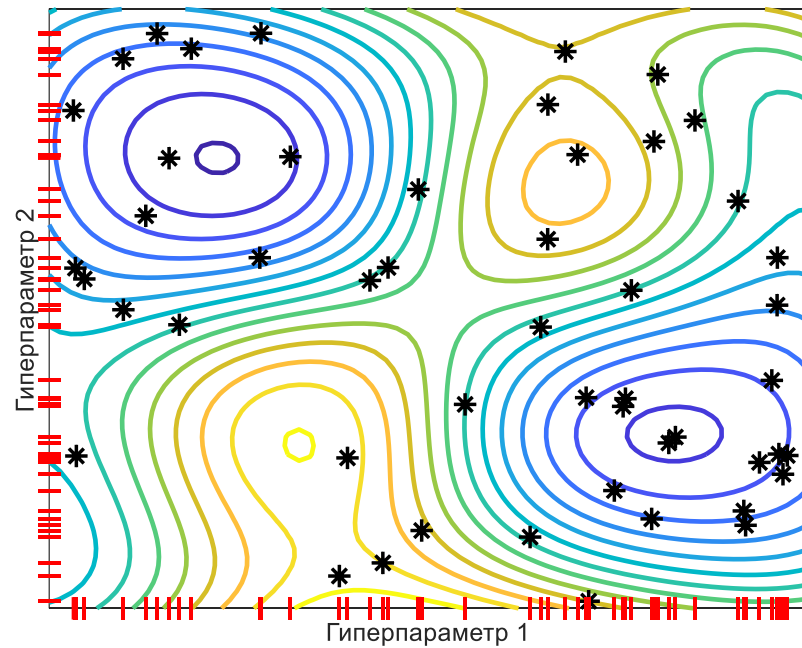


Рисунок 2.11 – Случайный поиск

Байесовская оптимизация (Bayesian Optimization, BO). Основная идея данного подхода заключается в том, что для поиска наиболее удачного набора в пространстве гиперпараметров необходимо учитывать уже рассмотренные наборы гиперпараметров и значения целевой функции, полученные для этого набора. В результате, может быть обеспечено снижение количества проверяемых наборов гиперпараметров и ускорение процесса выбора гиперпараметров в целом. Кроме того, направленный перебор позволяет достигнуть меньшего значения ожидаемой ошибки прогноза, а значит получить более качественную модель регрессии f [188, 249].

При байесовской оптимизации выполняется направленный перебор гиперпараметров для минимизации целевой функции f_B , определяющей зависимость ожидаемой ошибки прогноза модели регрессии $\mathbb{E}_{(D_{train}, D_{valid}) \sim D}(\Psi(L, A_\lambda, D_{train}, D_{valid}))$ от набора гиперпараметров λ . Оптимизация данной целевой функции, которую можно считать «черным ящиком», сопряжена с рядом сложностей, поскольку для многих моделей машинного обучения оказывается невыполнимой задачей определить аналитическое выражение и его производные для зависимости эффективности регрессионной модели от гиперпараметров. Поэтому в алгоритме байесовской оптимизации используется

суррогатная функция M (*surrogate function*), которая является вероятностным представлением зависимости f_B от λ . В данной работе в качестве суррогатной функции применяется модель регрессии на основе гауссовских процессов (*GPR*).

Важной составляющей алгоритма байесовской оптимизации является функция выбора S (*acquisition function*), определяющая следующий проверяемый набор гиперпараметров λ из пространства Λ . Данная функция должна обеспечивать баланс между исследованием новых областей, в которых стандартное отклонение суррогатной модели σ_M высоко, и уточнением тех наборов гиперпараметров, для которых среднее суррогатной модели μ_M имеет низкие значения. Одной из задач данной работы является исследование различных функций выбора при байесовской оптимизации гиперпараметров моделей машинного обучения, используемых для краткосрочного прогнозирования генерации ВЭС.

На рисунке 2.12 представлена графическая интерпретация метода байесовской оптимизации гиперпараметров. Как и в предыдущих случаях, значение целевой функции было вычислено 49 раз. Из рисунка видно, что большая часть проверяемых наборов гиперпараметров соответствует меньшим значениям целевой функции, чем при выполнении случайного поиска.

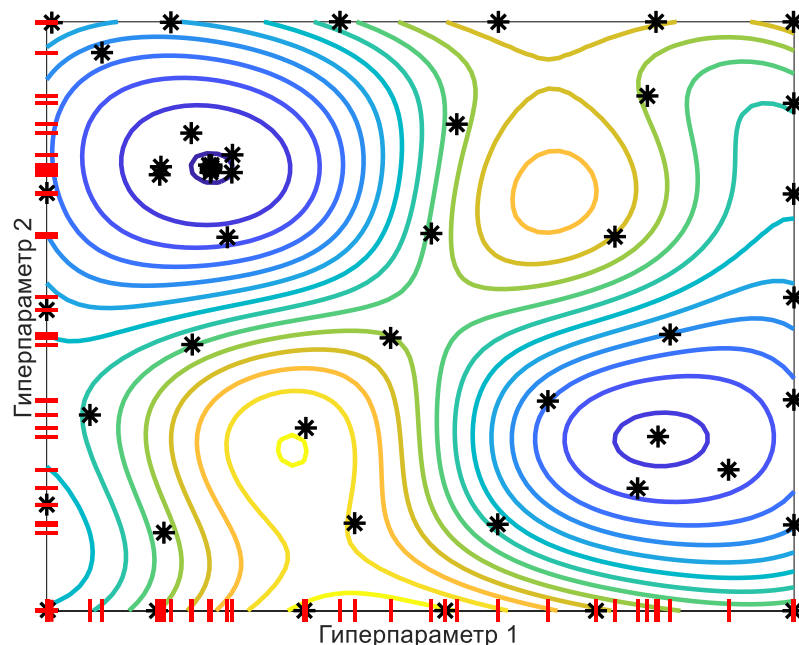


Рисунок 2.12 – Байесовская оптимизация

Далее представлено описание алгоритма байесовской оптимизации гиперпараметров [250, 251], соответствующее реализации данного алгоритма в *Statistics and Machine Learning Toolbox* ПК *MATLAB* [195].

Алгоритм 7: Алгоритм определения гиперпараметров регрессионной модели с помощью байесовской оптимизации

Вход:

f_B – целевая функция, определяющая ожидаемую ошибку прогноза регрессионной модели с учетом кроссвалидации,

K – количество итераций,

Λ – пространство гиперпараметров (количество параметров и диапазоны их изменения),

тип функции выбора S ,

R – количество случайных наборов гиперпараметров для инициализации суррогатной функции.

1. Формирование R случайных наборов гиперпараметров. Формирование обучающей выборки H для суррогатной модели.
2. Инициализация суррогатной функции M_0 по наборам гиперпараметров, полученных на шаге 1.
3. Для заданного количества итераций, то есть для i от 1 до K :
4. Определение следующего набора гиперпараметров
 $\lambda_i = \operatorname{argmax}_{\lambda} S(\lambda, M_{i-1})$
5. Определение значения целевой функции $f_B(\lambda_i)$.
6. Добавление пары λ_i и $f_B(\lambda_i)$ в обучающую выборку H .
7. Получение суррогатной функции M_i по обновленной обучающей выборке H .
8. Определение наименьшего значения целевой функции в $f_B(\lambda_i)$ среди проверенных пар H . Набор гиперпараметров, соответствующий этому значению, является искомым набором гиперпараметров λ^* .

Выход:

λ_* – искомый набор гиперпараметров.

Иллюстрация к алгоритму байесовской оптимизации представлена на рисунке 2.13. В примере рассматривается зависимость целевой функции f_B от одного гиперпараметра λ . На рисунке представлены результаты инициализации суррогатной функции, а также три первые итерации алгоритма. Для каждой итерации показаны максимум функции выбора S , полученное с помощью функции выбора значение следующего проверяемого гиперпараметра, а также суррогатная функция M , обученная с помощью предыдущих проверенных и вновь полученного значения гиперпараметра H .

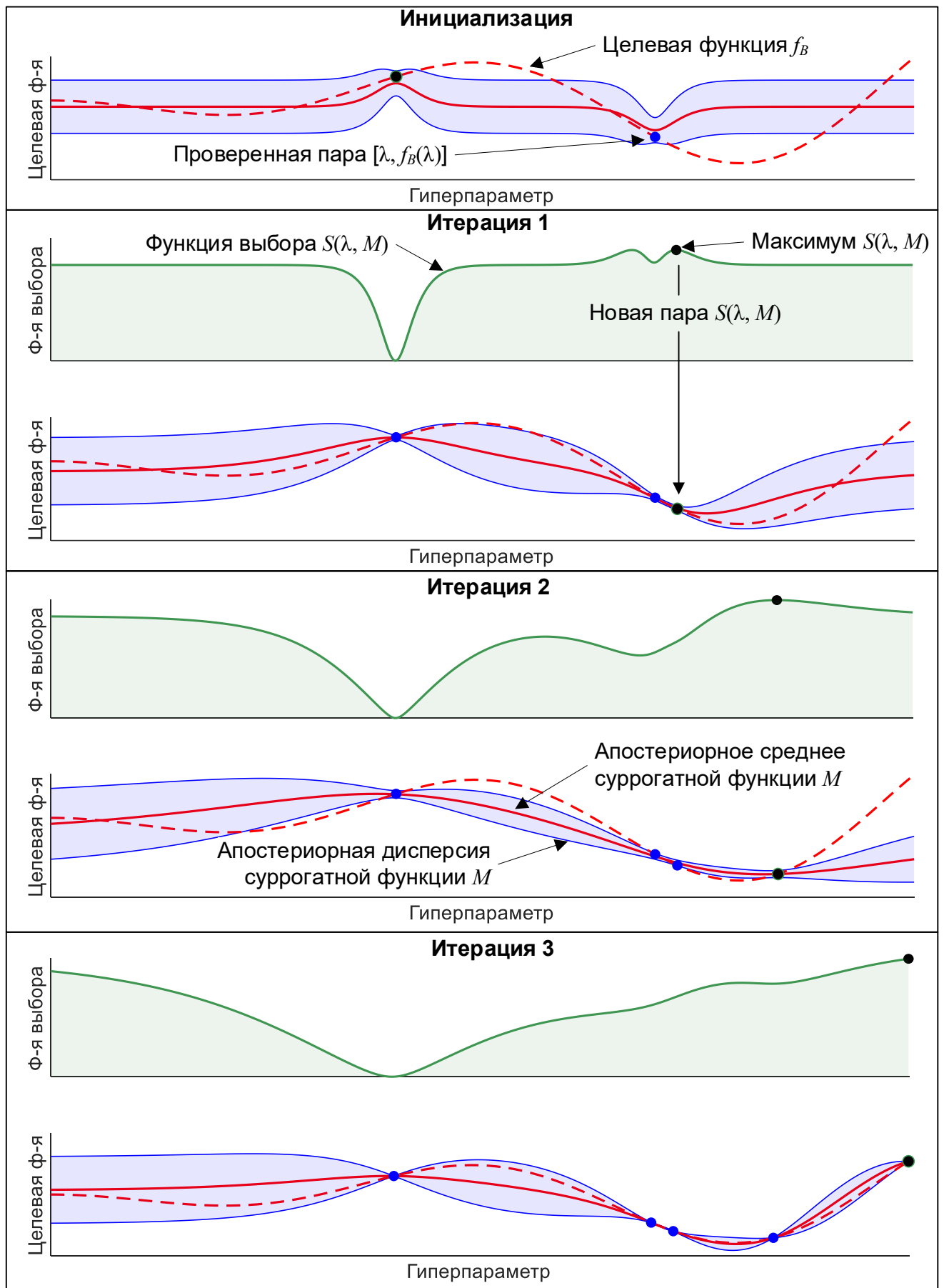


Рисунок 2.13 – Алгоритм байесовской оптимизации гиперпараметров

Наиболее широко распространенной функцией выбора при выполнении байесовской оптимизации гиперпараметров является функция «ожидаемого улучшения» (*Expected Improvement, EI*) [252]:

$$EI(\lambda, M) = \begin{cases} \mu_M(\lambda) - f_B(\lambda^+) \cdot \Phi(Z) + \sigma_M(\lambda) \cdot \phi(Z) & \text{при } \sigma_M(\lambda) > 0 \\ 0 & \text{при } \sigma_M(\lambda) = 0 \end{cases}, \quad (2.32)$$

где $\mu_M(\lambda)$ – среднее суррогатной модели для набора гиперпараметров λ ; $\sigma_M(\lambda)$ – стандартное отклонение суррогатной модели для набора гиперпараметров λ ; λ^+ – набор гиперпараметров, который на i -й итерации обеспечивает наименьшее значение целевой функции $f_B(\lambda^+)$; Φ и ϕ – функция распределения и плотность вероятности стандартного нормального распределения; параметр Z определяется в соответствии с выражением:

$$Z = \frac{\mu_M(\lambda) - f_B(\lambda^+)}{\sigma_M(\lambda)}. \quad (2.33)$$

Использованию данной функции выбора при байесовской оптимизации гиперпараметров регрессионных моделей, используемых для прогнозирования генерации ВЭС, посвящено несколько исследований. Например, в работе [253] байесовская оптимизация с функцией выбора «ожидаемое улучшение» использовалась для определения гиперпараметров модели регрессии, основанной на алгоритме градиентного бустинга над деревьями решений. Результаты исследования показали, что оптимизация гиперпараметров позволила снизить среднеквадратичную погрешность прогнозирования генерации одной ВЭУ мощностью 1,5 МВт для интервала упреждения 6 часов на 7-10%.

В статье [254] проведено аналогичное исследование, посвященное использованию байесовской оптимизации гиперпараметров модели *GBDT*, но применительно к прогнозированию генерации ВЭС установленной мощностью 200 МВт на сутки вперед. Авторами было продемонстрировано, что предложенный подход позволил снизить среднеквадратичную погрешность прогноза на 10-22% по сравнению с моделью *GBDT* без оптимизации гиперпараметров.

Авторами работы [255] рассматривалось применение методики определения гиперпараметров, основанной на байесовской оптимизации, для нескольких

различных моделей машинного обучения: модели, основанной на методе опорных векторов, модели регрессии на основе гауссовских процессов, а также регрессионном дереве решений. Кроме того, сравнивались ансамблевые модели, представляющие собой комбинации из перечисленных ранее моделей. В качестве функции выбора в данном исследовании также выступало «ожидаемое улучшение». При выполнении прогнозов генерации для одной ВЭУ мощностью 3,6 МВт на 1 час вперед получены следующие данные о снижении среднеквадратичной погрешности прогноза: для модели *GPR* – 0,5%, для модели *RDT* – 37%, для модели *SVM* – 29%. Можно сделать вывод о том, что в данном случае модель *GPR* оказалась менее чувствительна к выбору гиперпараметров, чем другие рассмотренные регрессионные модели.

Оценка эффективности применения байесовской оптимизации с функцией выбора *EI* применительно к моделям регрессии, основанных на гауссовских процессах, а также алгоритмах градиентного бустинга над деревьями решений и случайного леса, выполнена в статье [256]. Сравнение подходов выполнялось при прогнозировании генерации ВЭС для трех различных ВЭУ мощностью 2,2 МВт, 3,6 МВт и 1,7 МВт, находящихся в разных метеорологических и географических условиях. Результаты показали, что оптимизация гиперпараметров позволила снизить среднеквадратичную погрешность прогноза во всех рассматриваемых случаях, фактор улучшения достигал 11-19%.

В представленной работе впервые применительно к задаче оптимизации гиперпараметров моделей машинного обучения для краткосрочного прогнозирования генерации ВЭС предложено использование байесовской оптимизации с функциями выбора «вероятность улучшения» (*Probability of Improvement, PI*) и «нижний доверительный интервал» (*Lower Confidence Bound, LCB*).

Функции «вероятность улучшения» (*PI*) и «нижний доверительный интервал» (*LCB*) соответствуют выражениям [252]:

$$PI(\lambda, M) = \Phi \left(\frac{\mu_M(\lambda) - f_B(\lambda^+) - m}{\sigma_M(\lambda)} \right), \quad (2.34)$$

$$LCB(\lambda, M) = 2 \cdot \sigma_M(\lambda) - \mu_M(\lambda), \quad (2.35)$$

где m – параметр, контролирующий исследование новых областей.

В ряде исследований, посвященных байесовской оптимизации функций, представляющих собой «черный ящик», было выявлено, что функции выбора PI и LCB могут обеспечивать определение искомого набора гиперпараметров за меньшее количество итераций и/или обеспечить меньшее значение целевой функции. Например, в работе [257] предложено совместное использование функций выбора «ожидаемое улучшение» EI и «вероятность улучшения» PI для оптимизации тестовых функций из набора *BBOB* (*Black-Box Optimization Benchmarking*). Совместное использования функций выбора выполнялось за счет применения внутри одного процесса байесовской оптимизации разных функций выбора на разных итерациях. В статье предложено несколько стратегий, определяющих очередность использования функций выбора PI и EI . Исследование показало, что комбинирование функций выбора обеспечило достижение на 20% меньшего значения целевой функции за то же количество итераций.

Исследование [258] посвящено изучению различных функций выбора при байесовской оптимизации, выполняемой при создании материалов с улучшенными свойствами (электрической проводимостью, механической прочностью и т.д.). Результаты исследования показали, что функция выбора «нижний доверительный интервал» LCB обеспечила более быстрый поиск минимума целевой функции.

На рисунке 2.14 продемонстрировано влияние функции выбора на процесс выполнения байесовской оптимизации гиперпараметров применительно к простой целевой функции, зависящей только от одного гиперпараметра. В данной целевой функции, помимо глобального минимума, также есть локальный. Инициализация суррогатной модели выполнена таким образом, что одно из проверенных значений гиперпараметра оказалась вблизи локального минимума. Для инициализации суррогатной модели для байесовской оптимизации использованы одинаковые значения гиперпараметров.

Как видно из рисунка, при выполнении байесовской оптимизации с функцией выбора «вероятность улучшения» PI все 5 итераций рассматривались точки вблизи локального минимума. Можно сделать вывод о том, что использование данной функции в большей степени интенсифицирует уточнение уже проверенных наборов гиперпараметров, и в меньшей степени – исследование новых областей. Кроме того, при неудачной инициализации и недостаточном количестве итераций результаты оптимизации могут оказаться неудовлетворительными. Тем не менее, при удачной инициализации, когда апостериорное среднее суррогатной модели окажется вблизи глобального минимума, функция PI может обеспечить нахождение набора гиперпараметров, соответствующих меньшему значению целевой функции, чем EI и LCB .

При сравнении функций «ожидаемое улучшение» EI и «нижний доверительный интервал» LCB можно увидеть, что область, соответствующая глобальному минимуму, была обнаружена быстрее при использовании функции LCB (для LCB – на 3 итерации, для EI – на 5 итерации). Можно сделать вывод о том, что функция LCB быстрее исследует новые области и использует меньшее количество итераций для исследования локальных минимумов. Тем не менее, можно отметить, что близкое к глобальному минимуму значения целевой функции, обнаруженное с помощью функций EI и LCB на 5 итерации, практически совпадают.

Иллюстрация к процессу байесовской оптимизации, представленная на рисунке 2.14, показывает, насколько сильно функция выбора может повлиять на процесс и результаты поиска оптимального набора гиперпараметров модели регрессии для краткосрочного прогнозирования генерации ВЭС. Предварительно можно сделать вывод о том, что для более «гладких» целевых функций наиболее подходящей может оказаться функция PI , а функция LCB может оказаться лучшим вариантом для более «зашумлённых» целевых функций с большим количеством локальных минимумов.

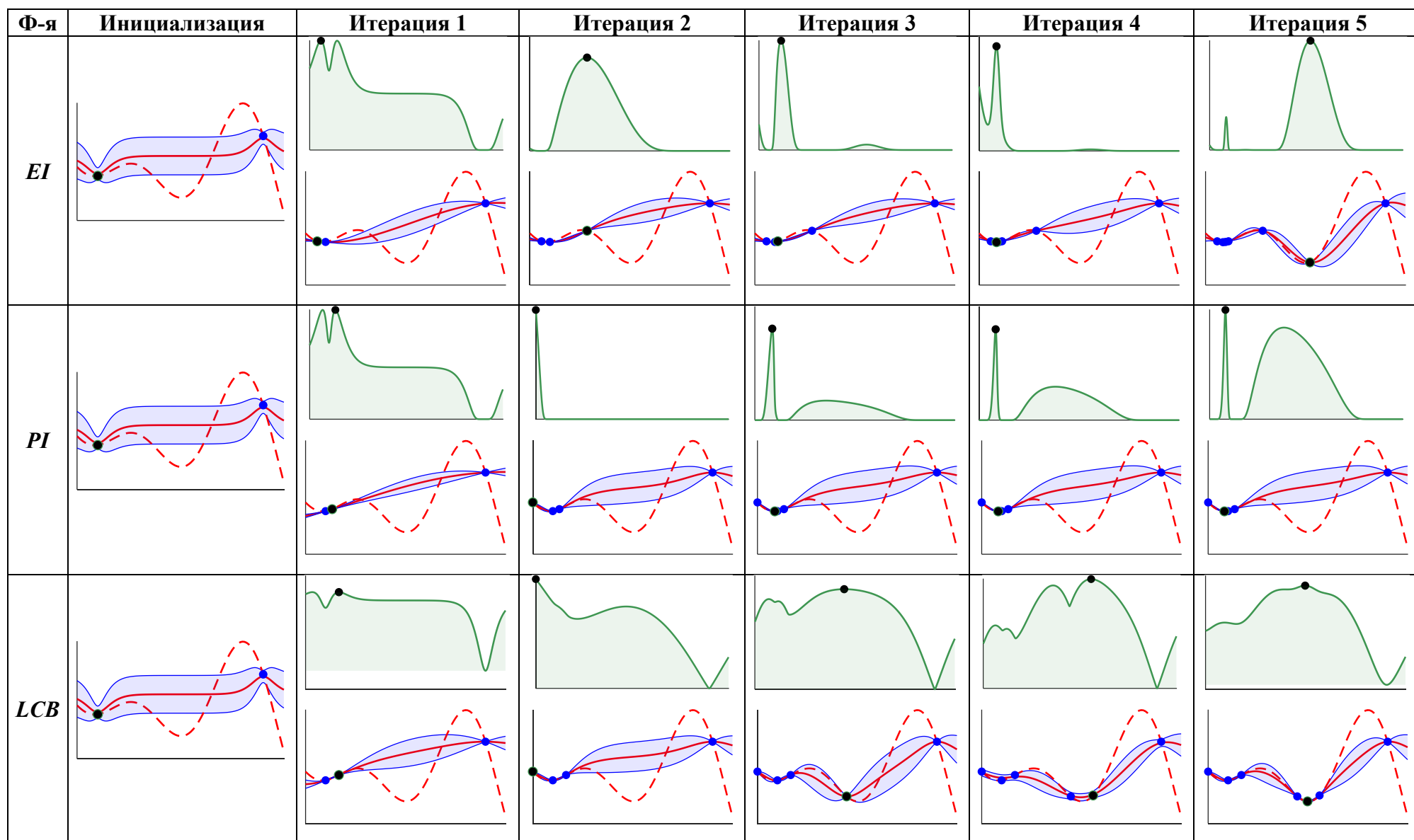


Рисунок 2.14 – Сравнение функций выбора при выполнении байесовской оптимизации гиперпараметров

2.4.Вычислительные эксперименты

В качестве объекта исследования при проведении вычислительных экспериментов рассматривалась ВЭС суммарной установленной мощностью 98,8 МВт, расположенная в Ростовской области. ВЭС состоит из 26 ВЭУ *Vestas V126-3.80*. Установленная мощность каждой турбины 3,8 МВт. Минимальная скорость ветра – 3 м/с, номинальная – 12 м/с, максимальная – 25 м/с.

Полный состав доступных исходных данных включал в себя следующие временные ряды:

- данные *SCADA* ВЭС: мощность, выдаваемая каждой турбиной (кВт), скорость ветра, измеренная на высоте гондолы (м/с);
- данные численного прогноза погоды «на сутки вперед» для места расположения ВЭС: скорость ветра (м/с), направление ветра (°), температура (°C), влажность воздуха (%), атмосферное давление (гПа).

Все временные ряды усреднены на интервале времени 1 час. Общее количество наблюдений в рассматриваемом наборе данных составляет 6343. Предварительно перед выполнением оптимизации гиперпараметров формируется тестовая выборка, не участвующая в обучении моделей. В тестовую выборку относится 20% случайных наблюдений (1268 шт.). Оставшиеся наблюдения (5075 шт.) используются для обучения и валидации моделей.

При проведении каждого вычислительного эксперимента для определения значения целевой функции при оптимизации гиперпараметров использовалась стандартная *5-fold* кросс-валидация. Данный подход основан на разделении набора данных на 5 подвыборок с одинаковым количеством наблюдений. Обучение модели выполняется 5 раз, каждый раз с разным составом обучающей выборки, состоящей из четырех подвыборок. Оставшаяся подвыборка при этом используется в качестве валидационной. Итоговое значение целевой функции для итерации определяется как среднее значение эффективности пяти обученных моделей на пяти валидационных выборках.

С использованием полученного при оптимизации набора гиперпараметров и всех пяти подвыборок, использованных при кроссвалидации, выполняется обучение итоговой модели. Далее выполняется расчет прогнозов и критериев для оценки эффективности модели на тестовой подвыборке. Иллюстрация к методике проведения вычислительного эксперимента при выполнении оптимизации гиперпараметров модели машинного обучения приведена на рисунке 2.15.

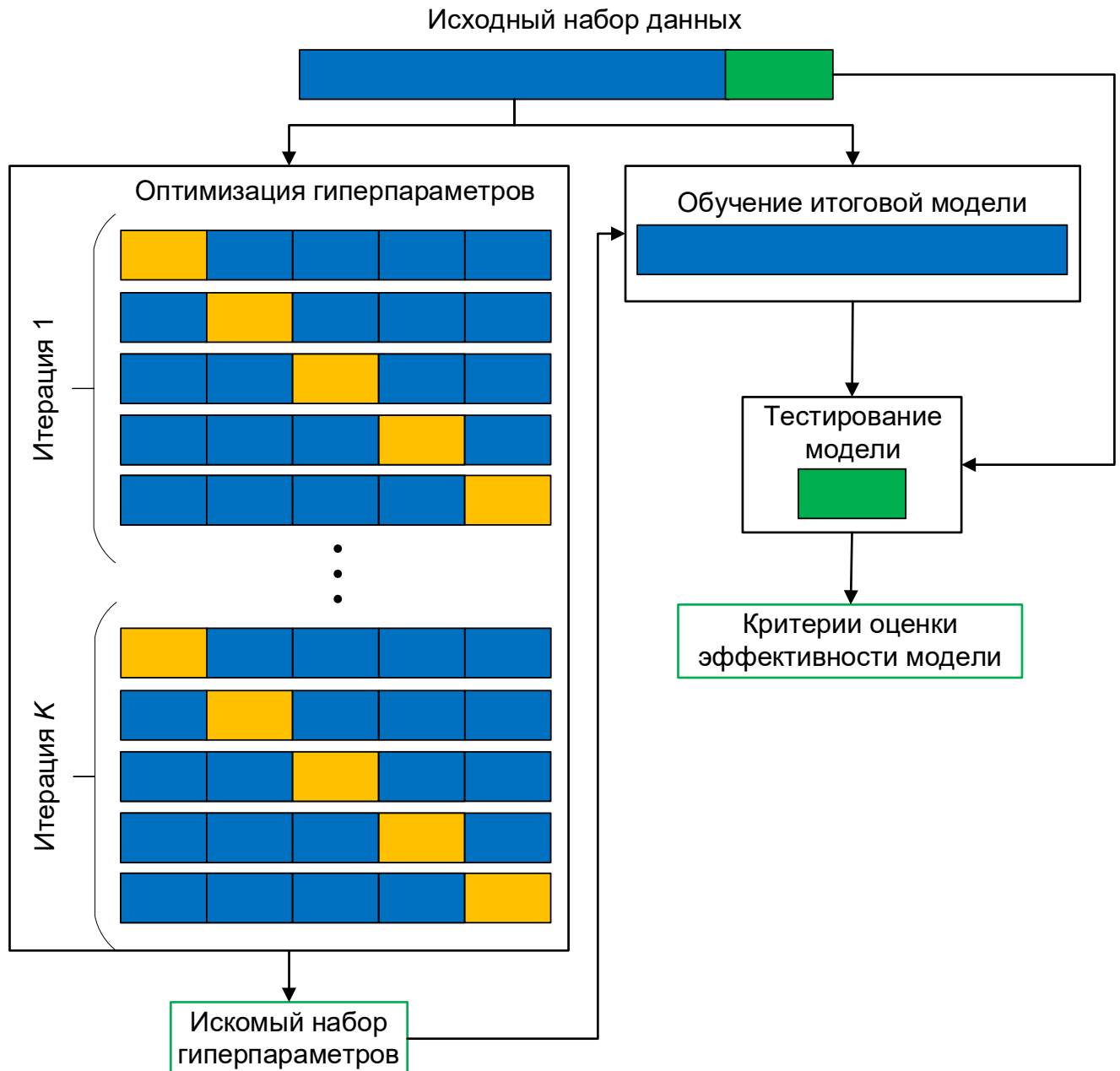


Рисунок 2.15 – Методика проведения вычислительного эксперимента при выполнении оптимизации гиперпараметров модели машинного обучения

В рамках исследования эффективности предложенных методик оптимизации гиперпараметров моделей машинного обучения выполнена серия из 31 вычислительного эксперимента. В каждом вычислительном эксперименте рассматривалась комбинация модели машинного обучения и способа оптимизации гиперпараметров (СОГ). Также проведен дополнительный вычислительный эксперимент для оценки эффективности аналитической модели.

В данной серии вычислительных экспериментов рассматривались модели машинного обучения, описанные в п 2.2:

- *RDT* – регрессионное дерево решений;
- *FFNN* – нейронная сеть прямого распространения;
- *GBDT* – градиентный бустинг над деревьями решений;
- *GPR* – регрессия на основе гауссовских процессов;
- *RF* – случайный лес;
- *SVM* – метод опорных векторов.

Также рассматривались способы оптимизации гиперпараметров из п 2.3:

- *GS* – поиск по сетке;
- *RS* – случайный поиск;
- *EI* – байесовская оптимизация с функцией выбора «ожидаемое улучшение»;
- *PI* – байесовская оптимизация с функцией выбора «вероятность улучшения»;
- *LCB* – байесовская оптимизация с функцией выбора «нижний доверительный интервал».

Количество итераций оптимизации гиперпараметров во всех экспериментах равнялось 60. Для метода поиска по сетке выбирались случайные элементы сетки пространства гиперпараметров. Результаты оптимизации гиперпараметров для рассматриваемых моделей машинного обучения, включающие общее время выполнения оптимизации, наименьшее найденное в процессе байесовской оптимизации значение целевой функции, время до нахождения минимума ЦФ, и

номер итерации минимума ЦФ, представлены в таблице 2.1. Найденные в процессе оптимизации наборы гиперпараметров приведены в приложении А в таблице А.1.

Таблица 2.1 – Результаты оптимизации гиперпараметров для рассматриваемых моделей машинного обучения

Модель	СОГ	Снижение значения ЦФ, %	Наименьшее значение ЦФ	Итерация с наименьшим значением ЦФ	Время до получения итерации с наименьшим значением ЦФ, с	Общее время, с
<i>RDT</i>	<i>GS</i>	0,0	19,72	25	12,2	16,2
	<i>RS</i>	0,9	19,71	60	7,4	8,1
	<i>EI</i>	0,1	19,72	58	11,4	13,9
	<i>PI</i>	-0,9	19,73	16	12,4	36,5
	<i>LCB</i>	-1,1	19,74	42	12,4	16,5
<i>FFNN</i>	<i>GS</i>	0,0	19,47	39	398,6	499,7
	<i>RS</i>	-5,1	19,54	34	173,2	286,6
	<i>EI</i>	-0,1	19,47	50	371,7	514,9
	<i>PI</i>	4,0	19,41	53	848,9	967,8
	<i>LCB</i>	-0,4	19,47	50	756,8	1026,6
<i>GBDT</i>	<i>GS</i>	0,0	19,34	48	13,8	17,9
	<i>RS</i>	7,5	19,23	22	7,9	17,7
	<i>EI</i>	9,1	19,20	55	16,9	19,3
	<i>PI</i>	13,7	19,13	59	41,7	42,0
	<i>LCB</i>	14,7	19,12	49	35,7	62,1
<i>GPR</i>	<i>GS</i>	0,0	19,17	5	143,1	1170,3
	<i>RS</i>	1,5	19,15	52	985,2	1108,2
	<i>EI</i>	0,8	19,16	38	786,2	1627,4
	<i>PI</i>	0,7	19,16	21	312,6	845,9
	<i>LCB</i>	0,4	19,17	47	2522,4	3935,4
<i>RF</i>	<i>GS</i>	0,0	19,32	15	7,3	22,8
	<i>RS</i>	-0,1	19,32	14	6,3	19,9
	<i>EI</i>	-1,1	19,34	43	15,9	27,4
	<i>PI</i>	2,3	19,29	59	69,3	70,7
	<i>LCB</i>	0,8	19,31	47	58,3	77,9
<i>SVM</i>	<i>GS</i>	0,0	20,53	60	151,8	151,8
	<i>RS</i>	17,2	20,28	57	183,3	191,7
	<i>EI</i>	14,8	20,31	58	314,1	318,1
	<i>PI</i>	-4,8	20,60	49	13,6	24,9
	<i>LCB</i>	4,9	20,46	59	159,6	161,1

Как видно из таблицы, в ряде случаев, общее время оптимизации гиперпараметров для одного из способов может быть значительно большим, например *LCB* для модели *GPR*. Это связано с тем, что в зависимости от значений гиперпараметров, время обучения на одной итерации может различаться. Большее общее время обучения говорит о том, что поиск гиперпараметров в основном

осуществлялся вблизи конфигурации, которая требовала больше времени на обучение.

Результаты оценки показателей погрешности и надёжности прогнозов на тестовой выборке приведены в таблице 2.2.

Таблица 2.2 – Результаты оценки погрешности и надёжности прогнозов на тестовом наборе данных для рассматриваемых моделей машинного обучения

Модель	COГ	<i>SS</i> , %	<i>NRMSE</i> , %	<i>NMAE</i> , %	<i>NBias</i> , %	<i>NSAE</i> , %	<i>EICP_20</i> , %
<i>AM</i>	-	0,0	33,0	24,1	22,5	67,5	54,3
<i>RDT</i>	<i>GS</i>	0,0	18,4	13,0	-0,1	36,2	76,3
	<i>RS</i>	-1,1	18,6	13,3	-0,8	37,1	74,4
	<i>EI</i>	-1,3	18,6	13,6	-0,7	37,9	75,4
	<i>PI</i>	0,0	18,4	13,0	-0,1	36,2	76,3
	<i>LCB</i>	-1,1	18,6	13,3	-0,8	37,1	74,4
<i>FFNN</i>	<i>GS</i>	0,00	15,6	11,4	-0,9	31,7	83,2
	<i>RS</i>	-6,9	16,7	12,4	-1,2	34,6	78,9
	<i>EI</i>	-4,8	16,3	12,3	-0,7	34,3	79,8
	<i>PI</i>	2,1	15,3	11,2	-0,9	31,3	82,1
	<i>LCB</i>	-4,0	16,2	12,2	-0,4	34,1	79,0
<i>GBDT</i>	<i>GS</i>	0,0	14,6	10,3	-0,2	28,8	85,1
	<i>RS</i>	2,8	14,2	10,2	1,1	28,6	85,0
	<i>EI</i>	4,3	14,0	9,9	-0,3	27,7	86,2
	<i>PI</i>	6,8	13,6	9,6	-0,3	26,8	86,2
	<i>LCB</i>	8,7	13,3	9,4	-0,4	26,3	86,5
<i>GPR</i>	<i>GS</i>	0,00	13,8	9,5	-0,2	26,7	86,2
	<i>RS</i>	0,0	13,8	9,5	-0,2	26,7	86,3
	<i>EI</i>	0,0	13,8	9,5	-0,2	26,7	86,2
	<i>PI</i>	-0,0	13,8	9,5	-0,2	26,7	86,3
	<i>LCB</i>	0,0	13,8	9,5	-0,2	26,7	86,2
<i>RF</i>	<i>GS</i>	0,0	15,1	11,1	-0,5	31,1	82,3
	<i>RS</i>	0,8	15,0	11,3	-0,5	31,7	82,3
	<i>EI</i>	-1,1	15,2	11,4	-0,5	32,0	81,7
	<i>PI</i>	3,4	14,6	10,8	-0,5	30,3	83,0
	<i>LCB</i>	3,2	14,6	10,9	-0,6	30,5	82,6
<i>SVM</i>	<i>GS</i>	0,0	28,1	22,7	5,3	63,4	52,2
	<i>RS</i>	15,6	23,7	18,6	4,5	51,9	62,9
	<i>EI</i>	11,6	24,9	20,6	1,8	57,7	54,2
	<i>PI</i>	-4,3	29,3	24,1	6,0	67,4	45,3
	<i>LCB</i>	-12,7	31,7	27,4	0,1	76,5	33,4

Фактор улучшения *SS* рассчитывался для каждой модели по отношению к эксперименту, в котором оптимизация гиперпараметров выполнялась с помощью метода поиска по решетке. При применении предложенных функций выбора «вероятность улучшения» *PI* и «нижний доверительный интервал» *LCB* удалось

добиться повышения эффективности прогнозирования для моделей *GBDT*, *RF* и *FFNN* по сравнению с другими рассмотренными способами оптимизации гиперпараметров. Для моделей *RDT*, *GPR* и *SVM* положительного результата добиться не удалось. При этом время оптимизации гиперпараметров для рассмотренных моделей машинного обучения для экспериментов с *PI* и *LCB*, как правило, оказывается выше.

Как видно из таблиц 2.1 и 2.2, использование метода случайного поиска для оптимизации гиперпараметров для моделей машинного обучения *GBDT* и *SVM* позволило получить меньшее значение целевой функции, а также повысить эффективность прогнозов при оценке показателей точности и надежности на тестовой выборке. Для моделей *RDT*, *FFNN*, *GPR* и *RF* заметного снижения значения ЦФ и повышения эффективности прогноза получено не было, либо результаты случайного поиска оптимальных гиперпараметров оказались хуже, чем при поиске по решетке.

На рисунке 2.16 представлено сравнение итерационного процесса поиска минимума целевой функции для модели *GBDT*, для которой было получено наименьшее значение *NRMSE*, и различных способов оптимизации гиперпараметров.

На рисунке 2.17 представлено сравнение рассмотренных моделей машинного обучения по основным критериям, используемым в данной работе для оценки точности и надежности прогноза. На рисунке 2.18 представлена диаграмма размаха («ящик с усами») для ошибок прогнозов, полученных с помощью рассматриваемых моделей машинного обучения и аналитической модели. Для каждой модели машинного обучения на рисунках 2.17 и 2.18 показаны результаты модели с гиперпараметрами, соответствующие наименьшему значению *NRMSE*.

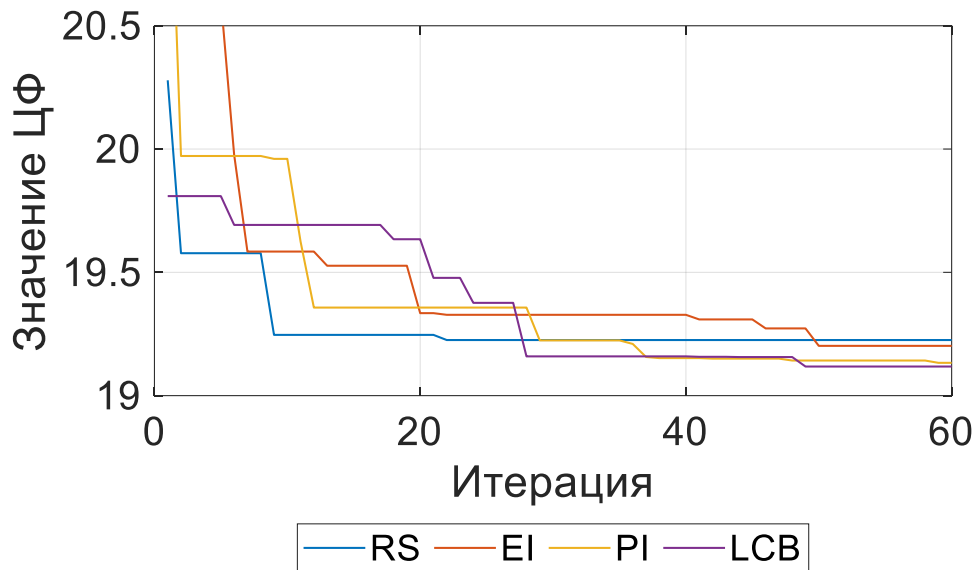


Рисунок 2.16 – Изменение значений целевой функции для различных способов оптимизации гиперпараметров: *RS* – случайный поиск; *EI* – байесовская оптимизация с функцией выбора «ожидаемое улучшение»; *PI* – «вероятность улучшения»; *LCB* – «нижний доверительный интервал»

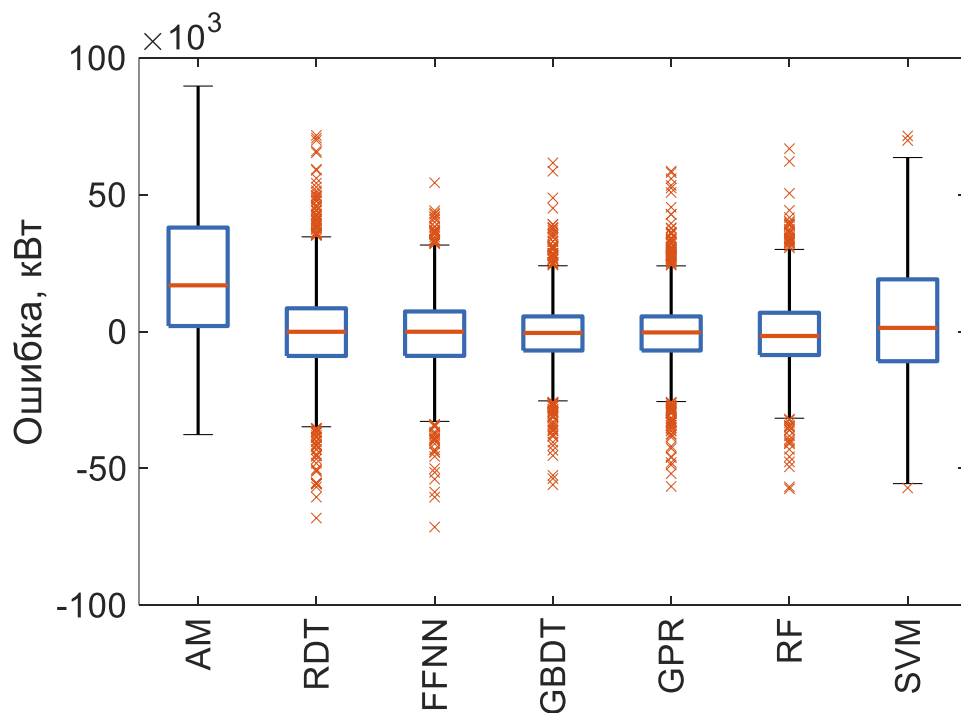


Рисунок 2.17 – Диаграмма размаха для полученных с помощью различных регрессионных моделей ошибок прогноза: *AM* – аналитическая модель; *RDT* – регрессионное дерево решений; *FFNN* – нейронная сеть прямого распространения; *GBDT* – градиентный бустинг над деревьями решений; *GPR* – регрессия на основе гауссовских процессов; *RF* – случайный лес; *SVM* – метод опорных векторов

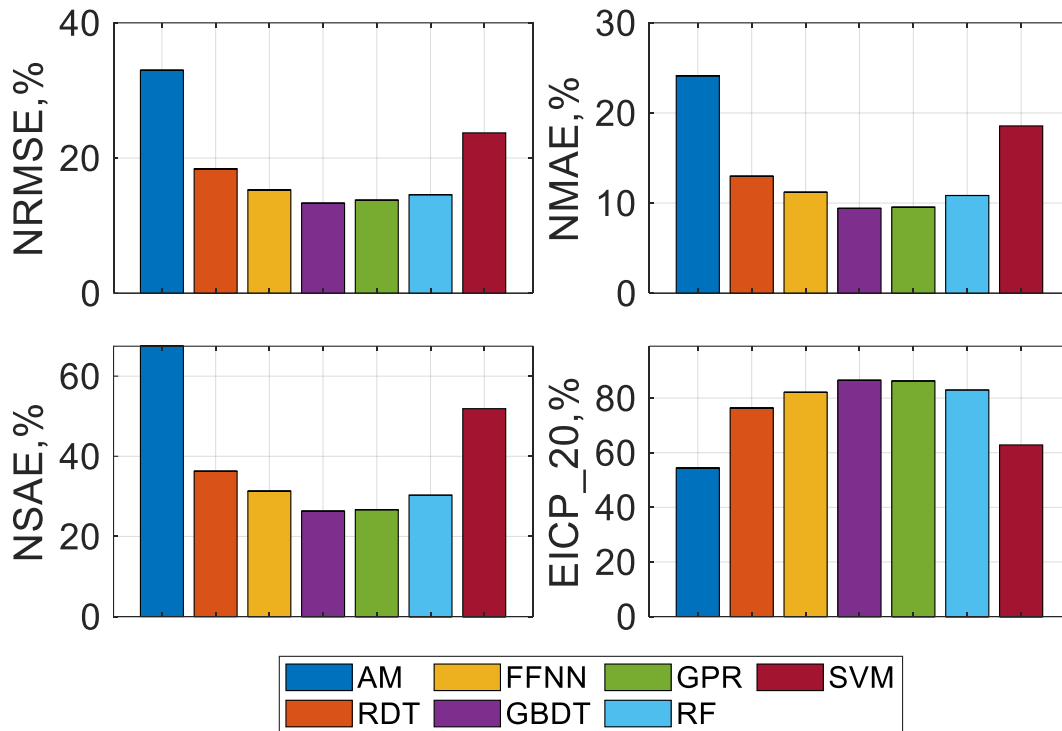


Рисунок 2.18 – Основные метрики для оценки точности и надежности прогнозов для различных регрессионных моделей: *AM* – аналитическая модель; *RDT* – регрессионное дерево решений; *FFNN* – нейронная сеть прямого распространения; *GBDT* – градиентный бустинг над деревьями решений; *GPR* – регрессия на основе гауссовских процессов; *RF* – случайный лес; *SVM* – метод опорных векторов

Среди рассмотренных моделей машинного обучения наибольшую эффективность показала модель *GBDT*. Сопоставимой эффективности удалось добиться при использовании модели *GPR*. При этом обучение модели *GPR* с оптимизацией гиперпараметров занимало в 10-60 раз большее количество времени в зависимости от способа оптимизации и проверяемых наборов гиперпараметров. При проведении последующих вычислительных экспериментов было принято решение использовать модель *GBDT*. Также по результатам проведенных вычислительных экспериментов и оценке точности результатов прогнозов можно сделать вывод, что предложенная методика оптимизации гиперпараметров, основанная на использовании байесовской оптимизации с функцией выбора «нижний доверительный интервал» *LCB*, обеспечивает наибольший фактор улучшения для модели *GBDT*. Эта методика также будет использоваться в дальнейших исследованиях.

2.5. Выводы по главе

1. Одним из способов повысить эффективность прогноза генерации ВЭС является оптимизация гиперпараметров моделей машинного обучения. Среди основных подходов для решения данной задачи выделяют: поиск по сетке (*GS*), случайный поиск (*RS*) и байесовская оптимизация (*BO*). При выполнении байесовской оптимизации, как правило, используют функцию выбора «ожидаемое улучшение» (*EI*).

2. В представленной работе впервые применительно к задаче оптимизации гиперпараметров моделей машинного обучения для краткосрочного прогнозирования выработки ВЭС предложено использование функций выбора «вероятность улучшения» (*PI*) и «нижний доверительный интервал» (*LCB*).

3. Для проверки предложенных методик и сравнения способов оптимизации гиперпараметров проведены вычислительные эксперименты с использованием моделей машинного обучения, который часто применяются для прогнозирования выработки ВЭС: регрессионное дерево решений (*RDT*), нейронная сеть прямого распространения (*FFNN*), градиентный бустинг над деревьями решений (*GBDT*), регрессия на основе гауссовских процессов (*GPR*), случайный лес (*RF*), метод опорных векторов (*SVM*).

4. Использование метода случайного поиска для оптимизации гиперпараметров для моделей машинного обучения *GBDT* и *SVM* позволило получить меньшее значение целевой функции, а также повысить эффективность прогнозов при оценке показателей точности и надежности на тестовой выборке. Для моделей *RDT*, *FFNN*, *GPR* и *RF* заметного снижения значения ЦФ и повышения эффективности прогноза получено не было, либо результаты случайного поиска оптимальных гиперпараметров оказались хуже, чем при поиске по решетке.

5. Использование байесовской оптимизации гиперпараметров с функцией ожидаемого улучшения *EI* позволило добиться меньшего значения целевой функции, чем для методов поиска по решетке и случайного поиска, лишь для модели *GBDT*. Для всех остальных рассмотренных моделей машинного обучения

использование байесовской оптимизации с данной функцией выбора приводило либо к отсутствию повышения эффективности прогноза, либо вовсе снижению. При этом, использование байесовской оптимизации, как правило, увеличивает время, необходимое на обучение.

6. При применении предложенных функций выбора «вероятность улучшения» PI и «нижний доверительный интервал» LCB для байесовской оптимизации гиперпараметров при построении регрессионных моделей для краткосрочного прогнозирования выработки ВЭС удалось добиться снижения значений целевой функции и повысить эффективность прогнозирования для моделей $GBDT$, RF и $FFNN$ по сравнению с другими рассмотренными способами оптимизации гиперпараметров. При этом время оптимизации гиперпараметров для рассмотренных моделей машинного обучения для экспериментов с PI и LCB , как правило, оказывается выше.

7. Среди рассмотренных моделей машинного обучения наибольшей эффективностью для решения поставленной задачи обладает модель $GBDT$. Сопоставимой эффективностью обладает модель GPR . При этом обучение модели GPR с оптимизацией гиперпараметров может занять в 10-60 раз большее количество времени в зависимости от способа оптимизации и проверяемых наборов гиперпараметров.

8. Для проведения дальнейших экспериментов по прогнозированию генерации ВЭС будет использоваться регрессионная модель $GBDT$ с байесовской оптимизацией гиперпараметров с помощью функции выбора LCB .

9. Предложенная методика байесовской оптимизации гиперпараметров с помощью функции «нижний доверительный интервал» LCB позволила снизить среднеквадратичную ошибку прогноза на 8,7% по отношению к методу поиска по решетке и на 6,1% по отношению к подходу, основанному на подборе гиперпараметров методом случайного поиска.

ГЛАВА 3. ОПРЕДЕЛЕНИЕ СПОСОБА АГРЕГАЦИИ МОЩНОСТИ ОТДЕЛЬНЫХ ВЕТРОУСТАНОВОК ПРИ КРАТКОСРОЧНОМ ПРОГНОЗИРОВАНИИ ГЕНЕРАЦИИ ВЕТРОЭЛЕКТРОСТАНЦИЙ

3.1. Общие подходы к агрегации мощности отдельных ветроустановок

Мощность, выдаваемая отдельными ветроустановками, расположенными на одной и той же ВЭС, может значительно отличаться. Основные факторы, приводящие к возникновению резких изменений скорости и направления ветра, а значит и мощности ВЭУ, описаны в п.1.4.4: взаимное влияние ВЭУ друг на друга, также называемое «затенением» турбин, наличие препятствий на пути воздушных масс, наклон поверхности, на которой расположены ВЭУ, и другие особенности ландшафта. Сопоставление мощности, выдаваемых турбинами за 48 часов, представлено на рисунке 3.1.

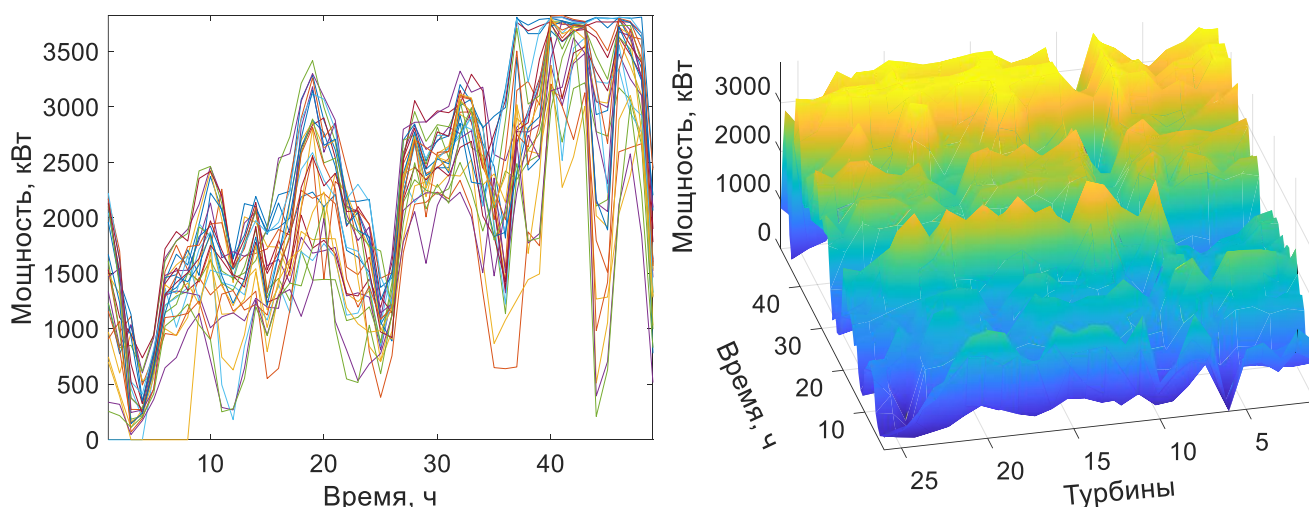


Рисунок 3.1 – Мощности турбин за 48 часов
(слева – графики, справа – поверхность)

Одним из способов учесть данный факт и повысить эффективность прогноза генерации ВЭС является изменение способа агрегации мощности отдельных ВЭУ при построении регрессионных моделей. На рисунке 3.2 представлены три основных подхода к агрегированию мощности отдельных турбин при выполнении прогнозов генерации ВЭС:

- прогноз на уровне ВЭС – суммирование мощности всех турбин ВЭС для получения отклика и прогнозирование суммарной мощности ВЭС;

- прогноз на уровне групп турбин – суммирование мощности турбин по группам, прогнозирование мощности каждой группы и суммирование прогнозов по группам для получения суммарной мощности ВЭС;
- прогноз на уровне турбин – прогнозирование мощности каждой турбины и суммирование их прогнозов для получения мощности всей ВЭС.

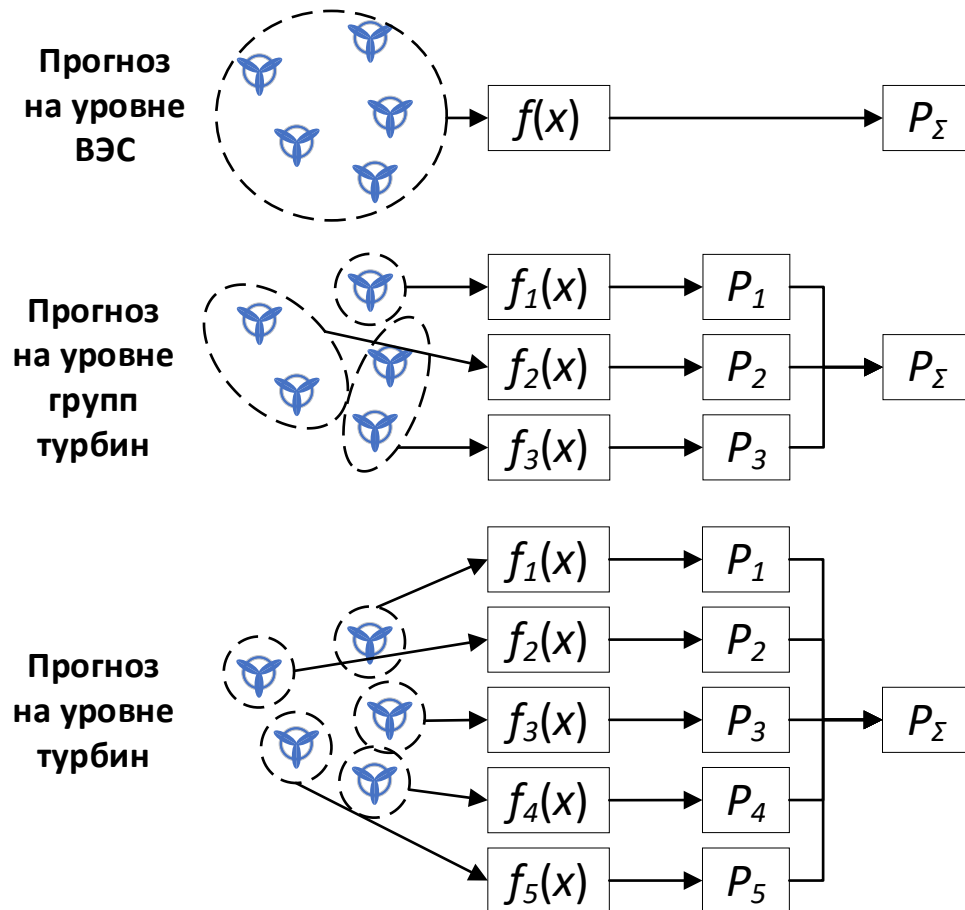


Рисунок 3.2 – Подходы к агрегации мощности отдельных ветроустановок

Изменение способа агрегации мощности отдельных ветроустановок является одним из наиболее редко используемых способов повышения эффективности прогноза генерации ВЭС. Единого мнения о том, какой подход является наиболее эффективным, пока нет. Как правило в большинстве исследований применяется методика прогноза на уровне ВЭС [131, 134, 136]. Основным преимуществом данного подхода является простота и низкие затраты вычислительных ресурсов, поскольку нет необходимости выполнять обучение нескольких моделей. Данный подход также позволяет использовать эффект пространственного сглаживания,

часто используемый при выполнении прогнозов сразу нескольких ВЭС [259, 260], и в ряде случаев позволяющий обеспечить повышение эффективности прогноза.

Тем не менее, пространственное сглаживание также может приводить к снижению информативности исходных данных из-за слишком различающихся условий, в которых могут находиться ветроустановки даже на одной ВЭС, что может снизить эффективность прогнозирования мощности на уровне ВЭС. Турбины могут находиться на значительном расстоянии друга от друга (10 и более километров), а также «в затенении» по отношению к друг другу или элементам ландшафта. Например, в исследованиях [261, 262] показано, что прогнозирование на уровне турбин позволило добиться снижения среднеквадратичной погрешности на 15% для оперативного прогноза генерации ВЭС мощностью 150 МВт на 2 часа вперед.

В ряде исследований [263–267] рассматривается возможность ограниченного использования пространственного сглаживания с помощью объединения турбин в группы турбин внутри одной ВЭС. Выполнение прогнозирования на уровне групп турбин позволяет как снизить погрешность прогнозирования генерации ВЭС, так и обеспечить более низкое потребление вычислительных ресурсов по сравнению с методом построения прогнозов на уровне отдельных турбин.

Для выполнения прогнозов на уровне групп турбин требуется решение задачи объединения ВЭУ в группы: необходимо определить количество групп, а также критерии и методику распределения ВЭУ по этим группам. Для решения данной задачи, как правило, используются различные методы машинного обучения без прецедентов (или обучения без учителя, *unsupervised learning*), главным образом методы кластерного анализа. В частности, в статье [263] рассматриваются три методики кластеризации: метод k -средних, спектральная кластеризация и самоорганизующаяся карта Кохонена. В качестве признаков при кластеризации использовались нормализованные значения средней мощности и среднеквадратичного отклонения мощности ветроустановок, а также средней скорости ветра и среднеквадратичного отклонения скорости ветра, измеренной на каждой ветроустановке. В работе исследовалась ВЭС установленной мощностью

49,5 МВт, состоящей из 33 ВЭУ по 1,5 МВт. Рассмотренные в данной статье методы кластеризации требовали предварительного определения количества кластеров, что является нетривиальной задачей. Для решения этой проблемы авторами анализировались среднеквадратичная ошибка прогноза генерации ВЭС, коэффициент силуэта кластеров и статистика Хопкинса для различного количества кластеров (от 2 до 8). По результатам анализа количество кластеров было принято равным 5 для всех трех методов кластеризации. В качестве регрессионной модели в работе использовалась нейронная сеть прямого распространения с радиально-базисной функцией активации. Результаты исследования показали, что распределение турбин по кластерам оказалось практически одинаковым для всех трех методов, среднеквадратичная погрешность прогноза генерации ВЭС на 72 часа вперед на уровне групп турбин оказалась на 30-35% ниже, чем прогноза на уровне ВЭС.

Аналогичное исследование было проведено авторами [264]. В работе также рассматривалась аналогичная ВЭС с 33 ВЭУ по 1,5 МВт и такие же признаки при кластеризации, при этом выполнялось сравнение методов кластеризации k -средних, спектральной кластеризации и иерархической агломеративной кластеризации. Оптимальное количество кластеров (в диапазоне от 2 до 9), равное 3, было выбрано по результатам анализа среднеквадратичной ошибки прогноза, коэффициента силуэта кластеров и индекса Калински-Харабаса. В этой работе также все три метода кластеризации обеспечили одинаковое распределение турбин по группам для оптимального количества кластеров. Объединение турбин в группы позволило снизить среднеквадратичную погрешность краткосрочного прогноза генерации ВЭС практически на 50%.

Алгоритм основанной на плотности пространственной кластеризации для приложений с шумами был применен в [265] для решения задачи распределения турбин по группам при оперативном прогнозировании генерации ВЭС на 6 часов вперед. В качестве признаков при кластеризации использовались данные SCADA о средней мощности и среднеквадратичных отклонениях мощности турбин, а также о средней скорости ветра. Особенностью рассматриваемого алгоритма

кластеризации являлось отсутствие необходимости выбора количества кластеров, при этом требовалось задание величины максимального радиуса соседства, которое было выбрано авторами эмпирически. Результаты проверки предложенной методики на двух ВЭС мощностью 45 МВт (17 ВЭУ) и 30 МВт (11 ВЭУ) показали, что средняя ошибка прогноза оказалась ниже на 20%, чем при прогнозировании на уровне ВЭС. Прогноз на уровне отдельных турбин обеспечил примерно такое же снижение погрешности, при этом прогноз на уровне групп турбин обладал значительно меньшими вычислительными затратами.

В работе [266] для распределения турбин по группам также предложено использование самоорганизующейся карты Кохонена. При прогнозировании генерации ВЭС мощностью 183 МВт, состоящей из 122 турбин, при использовании трех разных регрессионных моделей (метод опорных векторов, метод релевантных векторов и нейронная сеть прямого распространения) было выявлено, что кластеризация турбин позволяет снизить среднеквадратичную погрешность прогноза на 9-10%, вне зависимости от используемой регрессионной модели.

В исследовании [267] предложено использование метода нечеткой кластеризации *C*-средних для распределения турбин по группам. Кластеризация выполнялась на основе аппроксимаций характеристик мощности каждой отдельной турбины для ВЭС мощностью 49,5 МВт с 33 ВЭУ по 1,5 МВт. При выполнении краткосрочного прогноза на сутки вперед предложенный алгоритм обеспечил снижение среднеквадратичной погрешности прогноза более, чем в 2 раза.

В данной работе предлагается ряд новых методик для объединения турбин в группы и выполнения прогнозов генерации ВЭС на уровне групп турбин.

3.2. Методики объединения ветроустановок в группы

Для сравнения предлагаемых в данной работе методик с уже существующими подходами рассматривается способ объединения турбин в группы, основанный на кластеризации по методу *k*-средних, предложенный в исследованиях [263, 264].

Методика объединения ВЭУ в группы K-MS. Данная методика предложена в работах [263, 264]. Существующий подход подразумевает объединение турбин в группы по данным о средних значениях и среднеквадратичных отклонениях (СКО) измеренных мощностей и скоростей ветра каждой турбины, то есть матрица признаков принимает вид:

$$X = \begin{bmatrix} P_{1\text{ср}} & P_{1\text{ско}} & WS_{1\text{ср}} & WS_{1\text{ско}} \\ P_{2\text{ср}} & P_{2\text{ско}} & WS_{2\text{ср}} & WS_{2\text{ско}} \\ \vdots & \vdots & \vdots & \vdots \\ P_{T_n\text{ср}} & P_{T_n\text{ско}} & WS_{T_n\text{ср}} & WS_{T_n\text{ско}} \end{bmatrix} \quad (3.1)$$

где T_n – количество ВЭУ, $P_{1\text{ср}}$, $P_{2\text{ср}}$, $P_{T_n\text{ср}}$ – средние значения мощности каждой турбины в обучающей выборке; $P_{1\text{ско}}$, $P_{2\text{ско}}$, $P_{T_n\text{ско}}$ – среднеквадратичные отклонения значений мощности каждой турбин в обучающей выборке, $WS_{1\text{ср}}$, $WS_{2\text{ср}}$, $WS_{T_n\text{ср}}$ – средние значения скорости ветра, измеренной на каждой турбине, в обучающей выборке; $WS_{1\text{ско}}$, $WS_{2\text{ско}}$, $WS_{T_n\text{ско}}$ – среднеквадратичные отклонения средние значения скорости ветра, измеренной на каждой турбине, в обучающей выборке.

В работах [263, 264] при объединении турбин в группы используется алгоритм кластеризации k -средних, использующий евклидово расстояние между объектами. При этом способ определения начального приближения кластеров не указан, поэтому в данном исследовании в методике *K-MS* будет использоваться алгоритм кластеризации $k\text{-means}^{++}$.

Суть метода кластеризации k -средних заключается в итеративном повторении двух шагов: распределении объектов (в данном случае турбин) по кластерам, исходя из расстояния между объектом и центром кластера, и пересчете центров кластеров. Итерационный процесс продолжается до тех пор, пока распределение объектов по кластерам не перестанет изменяться, либо при достижении заданного количества итераций [268]. Иллюстрация к классическому алгоритму k -средних приведена на рисунке 3.3.

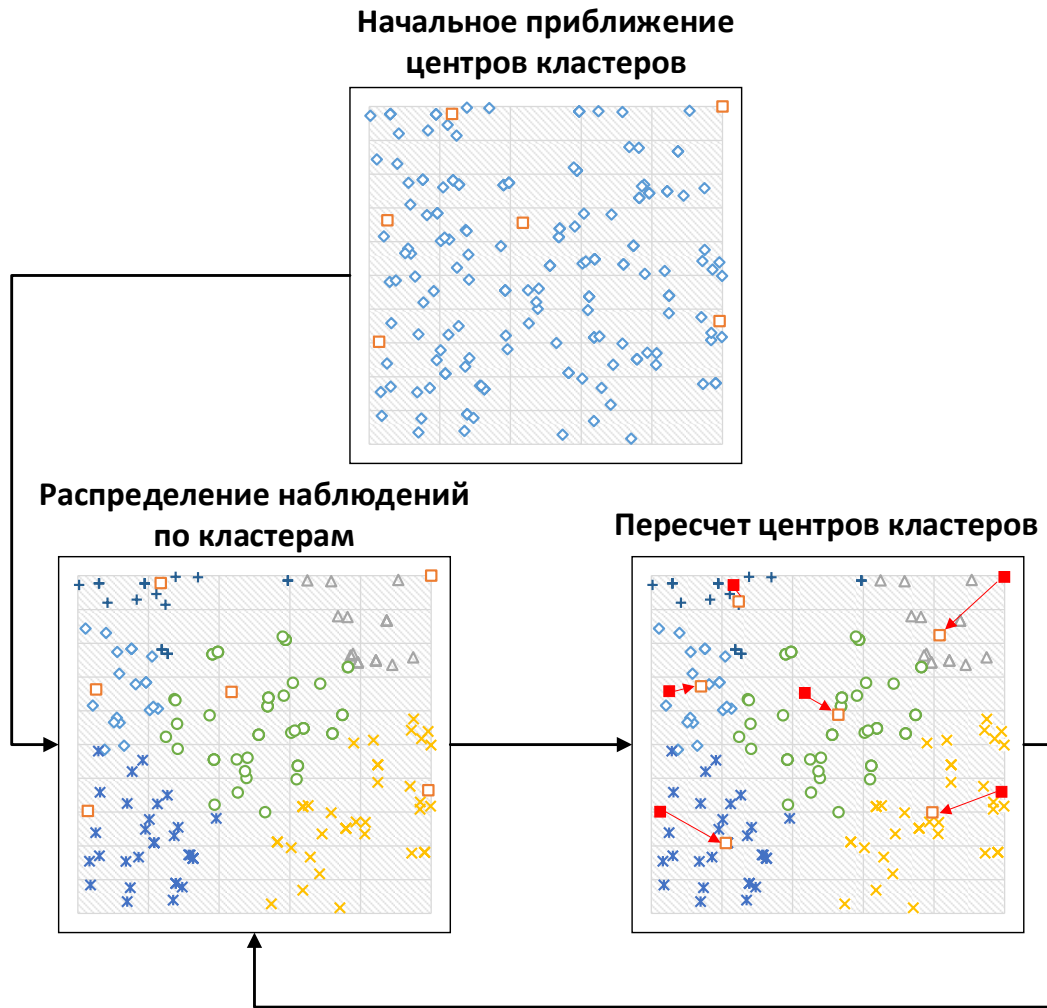


Рисунок 3.3 – Алгоритм кластеризации k -средних

Наиболее широко распространенным способом измерения расстояния между двумя объектами является евклидово расстояние:

$$d_E(\bar{x}_i, \bar{x}_j) = \|\bar{x}_i - \bar{x}_j\| = \sqrt{(\bar{x}_i - \bar{x}_j)^T (\bar{x}_i - \bar{x}_j)} = \sqrt{\sum_{m=1}^M (x_i^m - x_j^m)^2} \quad (3.2)$$

где $\bar{x}_i$ – вектор признаков объекта i ; $\bar{x}_j$ – вектор признаков объекта j ; $d_E(\bar{x}_i, \bar{x}_j)$ – евклидово расстояние между объектами i и j ; M – количество признаков; x_i^m и x_j^m – значения m -го признака для объектов i и j соответственно.

Одним из основных недостатков исходной версии алгоритма k -средних является необходимость определения первоначальных приближений центров кластеров заранее. Для решения данной проблемы был разработан алгоритм автоматического поиска начальных приближений кластеров k -means++ [269].

Далее представлено описание алгоритма кластеризации по улучшенному алгоритму k -средних с автоматическим поиском начальных приближений центров кластеров (k -means++) [269], соответствующее реализации данного алгоритма в Statistics and Machine Learning Toolbox ПК MATLAB [195].

Алгоритм 8: Алгоритм кластеризации k -means++

Вход:

X – матрица признаков для кластеризации,

K – количество кластеров,

$MaxIter$ – максимальное количество итераций,

d – способ определения расстояния между объектами (в данной работе используется евклидово расстояние (3.2)).

1. Определение начальных приближений центров кластеров
 $C^0 = \{c_1^0, c_2^0, \dots, c_K^0\}$.
2. Выбор случайного объекта (строки) из матрицы признаков X .
 Данный объект принимается как первый центроид c_1^0 .
3. Для заданного количества кластеров 1 до K .
4. | Определение квадрата расстояния $d^2(x_n, c_k^0)$ от каждого из N объектов матрицы признаков X до каждого центроида c_k^0 .
5. | Распределение объектов по кластерам. Для каждого объекта от 1 до N .
6. | | Определение расстояния $d(x_n, c_k)$ до центра каждого кластера от 1 до K .
7. | | Для каждого объекта определяется наименьшее из рассчитанных расстояний.
8. | | В вектор индексов I_c , определяющий принадлежность объектов к кластерам, для каждого объекта записывается индекс кластера, соответствующего наименьшему расстоянию.
9. | Определение суммы квадратов расстояний от каждого объекта до центра ближайшего кластера:

$$d_\Sigma = \sum_{n=1}^N d^2(x_{n \in k}, c_k).$$
10. | Выбор случайного числа R в диапазоне от 0 до N .
11. | Для каждого n , пока сумма a не окажется больше R :
12. | | $a = a + \frac{d^2(x_n, c_k)}{d_\Sigma}$
13. | | Объект с индексом n принимается новым центром кластера k .
14. Кластеризация объектов X .
15. | Распределение объектов по кластерам (аналогично шагам 5-8)

16. | Определение новых центров кластеров. Для каждого кластера от 1 до K
17. | Рассчитывается среднее значение по каждому из M признаков по всем объектам h , принадлежащих кластеру k :

$$c_k = \sum_{h=1}^H \frac{x_{h \in k}}{H}$$
18. | Повторение шагов 15-17 до тех пор, пока вектор I_c не перестанет изменяться (будет таким же, как на предыдущей итерации), или по достижению $MaxIter$

Выход:

I_c – вектор индексов, определяющий принадлежность объектов к кластерам.

Методика объединения ВЭУ в группы K-P. Первая предложенная методика заключается также в использовании метода k -средних и евклидова расстояния между объектами. Отличием от существующей методики является используемая при кластеризации матрица признаков. Предлагается использовать вместо средних значений и среднеквадратичных отклонений непосредственно наблюдения мощности. В таком случае выражение (3.2) будет использоваться для определения расстояния между временными рядами, то есть суммарного расстояния между парами измерений мощности, имеющими одинаковые метки времени (индексы). Таким образом, матрица признаков приобретает вид:

$$X = \begin{bmatrix} P_{11} & P_{12} & \dots & P_{1N} \\ P_{21} & P_{22} & \dots & P_{2N} \\ \vdots & \vdots & \vdots & \vdots \\ P_{T_n 1} & P_{T_n 2} & \dots & P_{T_n N} \end{bmatrix} \quad (3.3)$$

где T_n – количество ВЭУ; N – количество наблюдений в обучающей выборке; $P_{11}, \dots, P_{T_n N}$ – измерения мощности ВЭУ.

Методика объединения ВЭУ в группы H-PE. Вторая разработанная методика (H-PE) основана на использовании алгоритма иерархической агломеративной кластеризации (*Hierarchical Agglomerative Clustering, HAC*) и евклидова расстояния между объектами. При этом в качестве признаков при кластеризации, аналогично методике K-P, используются временные ряды мощности отдельных ВЭУ.

При выполнении иерархической агломеративной кластеризации каждый объект (в данном случае временной ряд мощности турбины) выделяется в

отдельный кластер. Далее рассчитывается расстояние между всеми объектами, например, евклидово расстояние. Два ближайших друг к другу кластера объединяются, этот процесс, как правило, продолжается вплоть до объединения всех объектов в один кластер. Результаты иерархической кластеризации принято визуализировать в виде дендрограммы, то есть бинарного дерева, в котором длина связей между объектами и кластерами соответствует расстоянию между ними [270]. Количество кластеров может быть выбрано заранее (в этом случае построение полной дендрограммы может не выполняться), либо по результатам построения дендрограммы путем выбора порогового значения расстояния и соответствующего ему распределения объектов по кластерам. Иллюстрация к алгоритму иерархической агломеративной кластеризации показана на рисунке 3.4.

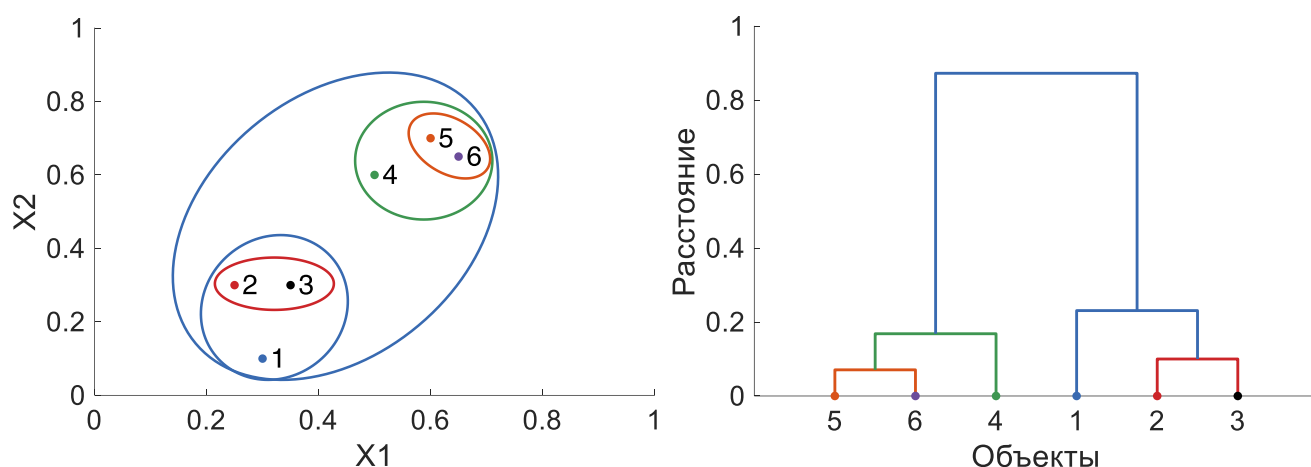


Рисунок 3.4 – Алгоритм иерархической агломеративной кластеризации: объекты и их последовательное объединение в кластеры – слева, дендрограмма – справа

Для определения меры близости между кластерами могут использоваться различные методики. Расстояние между кластерами может определяться по минимальному, максимальному или среднему расстоянию между объектами кластеров, либо по центроидам кластеров, рассчитываемых как среднее значение по каждому из признаков по всем объектам кластеров.

В данной работе для определения расстояния между кластерами применялся метод Уорда, использующий в качестве расстояния между кластерами прирост

суммы квадратов расстояний объектов до центроида, получаемого в результате объединения кластеров [271]:

$$\Delta_W^2(A, B) = \sum_{i \in A \cup B} \|x_i - c_{A \cup B}\|^2 - \sum_{i \in A} \|x_i - c_A\|^2 - \sum_{i \in B} \|x_i - c_B\|^2, \quad (3.4)$$

$$\Delta_W(A, B) = \sqrt{\frac{2 \cdot N_A \cdot N_B}{N_A + N_B} \cdot \|c_A - c_B\|}, \quad (3.5)$$

где $\Delta_W(A, B)$ – расстояние между кластерами A и B , определенное по методу Уорда; $c_A, c_B, c_{A \cup B}$ – центроиды кластеров A, B и их объединения; N_A, N_B – количество объектов в кластерах A и B . Коэффициент 2 в числителе используется для получения расстояния для кластеров из одного элемента, равного евклидову.

Далее представлено описание алгоритма иерархической агломеративной кластеризации (НАС) [271, 272], соответствующее реализации данного алгоритма в *Statistics and Machine Learning Toolbox* ПК *MATLAB* [195] (рассматривается реализация с заданным максимальным количеством кластеров).

Алгоритм 9: Алгоритм иерархической агломеративной кластеризации (НАС)

Вход:

X – матрица признаков для кластеризации,

K – максимальное количество кластеров,

d – способ определения расстояния между объектами,

Δ – способ определения расстояния между кластерами.

1. Определение расстояний между каждой парой объектов d .
2. Выделение каждого объекта в отдельный кластер. Формирование вектора I_c .
3. Объединение двух ближайших кластеров:
4. Определение наименьшего расстояния Δ между каждой парой кластеров.
5. Объединение двух кластеров с наименьшим расстоянием. Обновление вектора I_c .
6. Расчет расстояния от вновь сформированного кластера до всех оставшихся кластеров.
7. Повторение шагов 4-5 до тех пор, пока количество кластеров не достигнет K .

Выход:

I_c – вектор индексов, определяющий принадлежность объектов к кластерам.

Методика объединения ВЭУ в группы H-PD. Третья разработанная методика (H-PD) также использует иерархическую агломеративную кластеризацию и расстояние Уорда между кластерами. В качестве признаков при кластеризации, аналогично методикам K-P и H-PE, используются временные ряды мощности отдельных ВЭУ. Но для измерения расстояния между объектами применен способ, основанный на алгоритме динамической трансформации временной шкалы (*Dynamic Time Warping, DTW-расстояние*).

Евклидово расстояние не всегда позволяет корректно оценить сходство между двумя временными рядами, поскольку не позволяет учесть форму временных рядов. Для случаев, когда временные ряды имеют одинаковую форму, но смещены друг относительно друга по оси времени, результаты кластеризации с использованием евклидова расстояния могут быть искажены. Для решения данной проблемы используется *DTW-расстояние*, определяемое с помощью алгоритма динамической трансформации временной шкалы, позволяющее выполнить оптимальное сопоставление точек двух временных рядов. При этом одна и та же точка временного ряда может быть сопоставлена нескольким точкам другого временного ряда [273, 274].

Пусть даны два временных ряда $X = \{x_1, x_2, \dots, x_N\}$ и $X' = \{x'_1, x'_2, \dots, x'_L\}$ с количеством элементов N и L соответственно. Для определения *DTW-расстояния* между временными рядами выполняется расчет расстояния между всеми возможными парами элементов временных рядов $d(x_1, x'_1), \dots, d(x_N, x'_L)$. Полученные расстояния используются для определения элементов матрицы трансформации D размерностью $N \times L$. По результатам формирования D определяется оптимальный путь трансформации W из K элементов, то есть последовательность элементов матрицы трансформации D от элемента (1,1) до элемента (M, L) , которая соответствует минимальному расстоянию между временными рядами. Каждый элемент пути трансформации W_k представляет собой расстояние между парой индексов элементов временных рядов (n, l) и соответствующее расстояние между ними $d(x_n, x'_l)$. Само значение *DTW-*

расстояния оказывается равным последнему элементу матрицы трансформации [275, 276]. Допустимым считается только тот путь трансформации, который удовлетворяет:

- граничным условиям, подразумевающим, что первый элемент первого временного ряда должен быть сопоставлен первому элементу второго временного ряда и последний элемент первого временного ряда должен быть сопоставлен последнему элементу второго временного ряда [275, 277]:

$$\begin{aligned} W_1 &= (1, 1), \\ W_K &= (N, L), \end{aligned} \quad (3.6)$$

где W_1 – первый элемент пути трансформации; W_K – последний элемент пути трансформации.

- условия монотонности и непрерывности, определяющие, что сопоставленные индексы элементов временных рядов должны монотонно возрастать (если $n_2 > n_1$ являются индексами элементов первого временного ряда, то во втором временном ряде не должно быть индексов $l_2 > l_1$, таких, что n_1 сопоставлен l_2 и n_2 сопоставлен l_1 , и наоборот), а также что каждый элемент первого временного ряда должен быть сопоставлен одному или нескольким элементам второго временного ряда, и наоборот [275, 277]:

$$\begin{aligned} n_{k-1} &\leq n_k \leq n_{k-1} + 1, \\ l_{k-1} &\leq l_k \leq l_{k-1} + 1. \end{aligned} \quad (3.7)$$

где k – индекс элемента пути трансформации; n – индекс элемента временного ряда X ; l – индекс элемента временного ряда X' .

Последнее условие определяет способ расчета элементов матрицы трансформации [275, 277]:

$$D(n, l) = d(x_n, x'_l) + \min \begin{cases} D(n-1, l) \\ D(n, l-1) \\ D(n-1, l-1) \end{cases}, \quad (3.8)$$

где $D(n, l)$ – элемент матрицы трансформации с индексами n и l ; $d(x_n, x'_l)$ – расстояние между элементами временных рядов x_n и x'_l .

Сопоставление способов определения расстояния между временными рядами на примере генерации двух ВЭУ показано на рисунке 3.5.

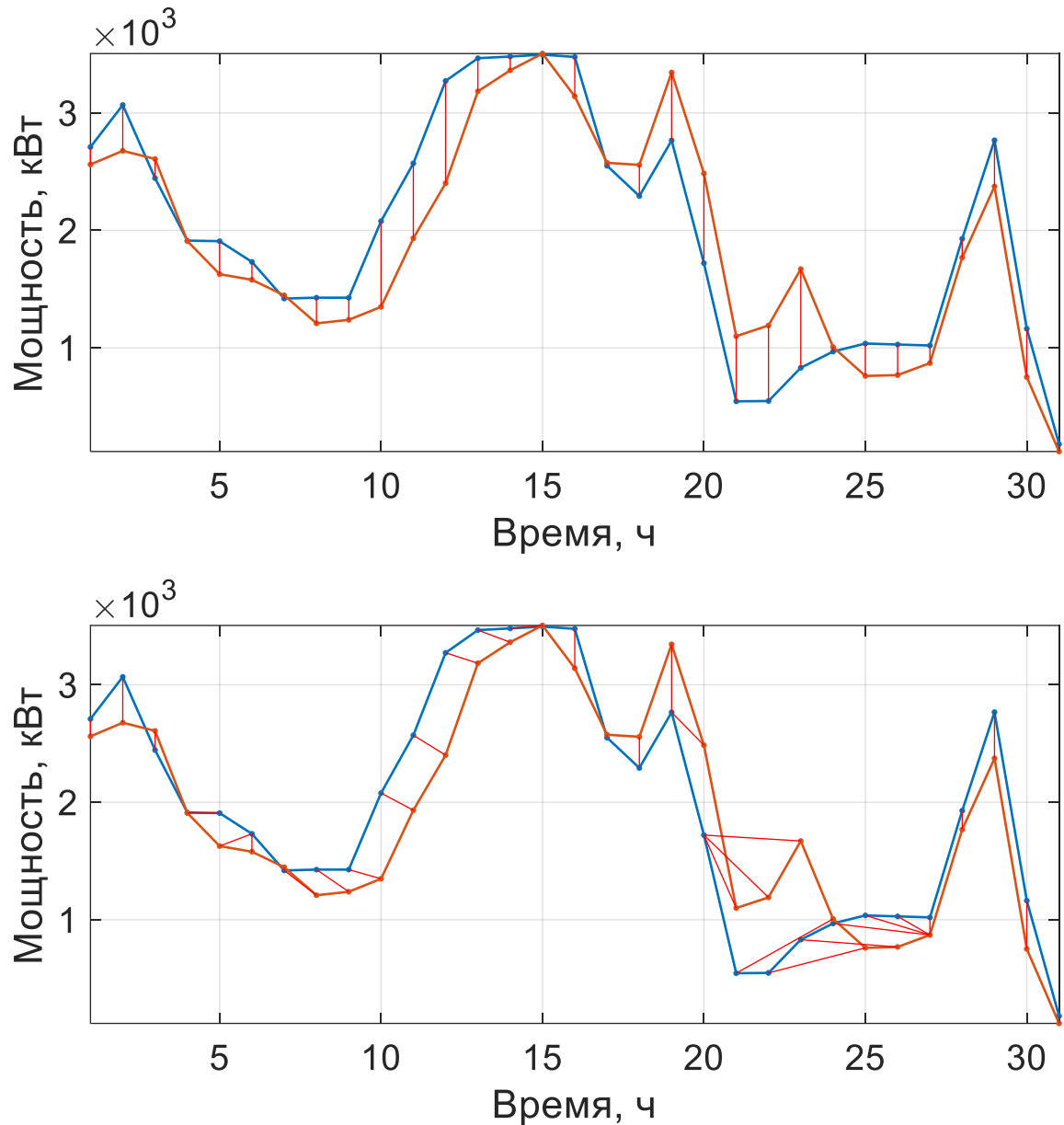


Рисунок 3.5 – Сопоставление способов определения расстояния между временными рядами мощности ВЭУ: евклидово расстояние – сверху, *DTW*-расстояние – снизу

Далее представлено описание алгоритма определения расстояния между объектами (временными рядами) с помощью динамической трансформации временной шкалы (*DTW*) [275, 277], соответствующее реализации данного алгоритма в *Signal Processing Toolbox* ПК *MATLAB*.

Алгоритм 10: Алгоритм определения расстояния между объектами с помощью динамической трансформации временной шкалы (*DTW*)

Вход:

X – первый временной ряд (вектор с количеством элементов N),

X' – второй временной ряд (вектор с количеством элементов L),

d – способ определения расстояния между элементами временных рядов (в данной работе используется евклидово расстояние (3.2)).

1. Инициализация матрицы трансформации D размерностью $N \times L$.
2. Определение расстояния $d(x_1, x'_1)$ между первыми элементами временных рядов X и X' .
3. Для всех элементов n временного ряда X от 2 до N .
4. Для всех элементов l временного ряда X' от 2 до L .
5. Определение расстояния $d(x_n, x'_l)$ между элементами временных рядов X и X' .
6. Определение значения элемента матрицы трансформации как:

$$D(n, l) = d(x_n, x'_l) + \min \begin{cases} D(n-1, l) \\ D(n, l-1) \\ D(n-1, l-1) \end{cases}$$
7. *DTW*-расстояние между двумя временными рядами определяется как последний элемент матрицы трансформации с индексами N и L :
 $d(X, X') = D(N, L)$

Выход:

$d(X, X')$ – расстояние между двумя временными рядами.

Использование кластеризации временных рядов с применением *DTW*-расстояния рассмотрено в ряде научных исследований. Например, в работе [278], посвященной прогнозированию фондовых индексов, данный способ кластеризации использовался для отбора наблюдений для обучения модели: выполнялся поиск сегментов временных рядов, близких по поведению к текущим колебаниям индекса. В статье [279] кластеризация с использованием *DTW*-расстояния использовалась при прогнозировании цен на криптовалюты. Авторами предложена методика поиска финансовых инструментов, поведение которых схоже с вновь созданными криптовалютами, ретроспективная информация по которым не имеет достаточного объема. В работах [280, 281] данная методика кластеризации использовалась при прогнозировании потребления электрической энергии. Потребители с похожим поведением распределялись по

кластерам, после чего выполнялось прогнозирование суммарного потребления каждой группы.

Рассмотренные исследования, посвященные применению *DTW*-расстояния при кластеризации временных рядов при решении различных задач прогнозирования, показывают, что применение данного подхода позволяет получить более корректные результаты кластеризации с точки зрения определения похожести временных рядов в случае, если временные ряды несколько смещены друг относительно друга по оси времени. Таким образом, использование *DTW*-расстояния может быть также оправдано и при распределении турбин на группы при прогнозировании генерации ВЭС. Эффект смещения временных рядов мощности отдельных ВЭУ может быть также связан с факторами, описанными в п.1.4.4: влиянием резко меняющихся скорости и направления ветра, «затенения» турбин друг другом и элементами ландшафта, запаздыванием в работе систем управления и т.д.

Сравнение существующей и предложенных методик объединения турбин в группы приведено в таблице 3.1.

Таблица 3.1 – Методики объединения ветроустановок в группы

Обозначение методики	Метод кластеризации	Признаки для кластеризации	Размерность матрицы признаков*	Расстояние между объектами
<i>K-MS</i> (сущ.)	<i>k</i> -средних	Средние значения и среднеквадратичные отклонения скорости ветра и мощности ВЭУ	$T_n \times 4$	евклидово
<i>K-P</i> (предл.)	<i>k</i> -средних	Измерения мощности ВЭУ	$T_n \times N$	евклидово
<i>H-PE</i> (предл.)	<i>HAC</i>	Измерения мощности ВЭУ	$T_n \times N$	евклидово
<i>H-PD</i> (предл.)	<i>HAC</i>	Измерения мощности ВЭУ	$T_n \times N$	<i>DTW</i> -расстояние

* T_n – количество ВЭУ, N – количество наблюдений.

3.3.Вычислительные эксперименты

При проведении вычислительных экспериментов рассматривался объект исследования и набор исходных данных, описанный в п.2.4. В рамках исследования эффективности предложенных методик агрегации мощности отдельных ВЭУ

выполнена серия из 99 вычислительных экспериментов. В каждом ВЭ рассматривается определенный способ агрегации мощности отдельных ветроустановок при прогнозировании генерации ВЭС. Для каждого из способов объединения турбин в группы, описанных в п.3.2, проводится по 24 эксперимента для определения количества групп. В качестве критерия, по которому определялось количество кластеров, использовался фактор улучшения SS по отношению к прогнозу по ВЭС.

Количество обученных регрессионных моделей в каждом вычислительном эксперименте равнялось количеству групп турбин. В качестве регрессионной модели использовался градиентный бустинг над деревьями решений $GBDT$. При обучении каждой модели применялась байесовская оптимизация гиперпараметров с использованием функции выбора LCB с количеством итераций 60. При выполнении оптимизации гиперпараметров и кроссвалидации использовалась методика, аналогичная описанной в п.2.4 и представленной на рисунке 2.15.

Таким образом, были проведены вычислительные эксперименты для проверки следующих методик:

- P – прогноз на уровне ВЭС (1 вычислительный эксперимент, получена 1 регрессионная модель, в рамках оптимизации гиперпараметров обучено 60 моделей);
- GF – прогноз на уровне групп турбин; объединение в 4 группы с помощью упрощённого способа распределения по группам исходя из принадлежности к фидерам (1 вычислительный эксперимент, получено 4 регрессионных модели, в рамках оптимизации гиперпараметров обучено 240 моделей);
- T – прогноз на уровне отдельных турбин (1 вычислительный эксперимент, получено 26 регрессионных моделей, в рамках оптимизации гиперпараметров обучено 1560 моделей);
- $K-MS$ – существующий подход выполнения прогноза на уровне групп турбин; объединение в группы по данным о средних значениях и среднеквадратичных отклонениях измеренных мощностей и скоростей ветра

каждой турбины методом кластеризации k -средних (24 вычислительных эксперимента, получено 324 регрессионные модели, в рамках оптимизации гиперпараметров обучено 19 440 моделей);

- $K-P$ – предложенный подход выполнения прогноза на уровне групп турбин; объединение в группы по временным рядам измерений мощности каждой турбины методом кластеризации k -средних (24 вычислительных эксперимента, получено 324 регрессионные модели, в рамках оптимизации гиперпараметров обучено 19 440 моделей);
- $H-PE$ – предложенный подход выполнения прогноза на уровне групп турбин; объединение в группы по временным рядам измерений мощности каждой турбины методом иерархической агломеративной кластеризации, для измерения расстояния между точками временных рядов используется евклидово расстояние (24 вычислительных эксперимента, получено 324 регрессионные модели, в рамках оптимизации гиперпараметров обучено 19 440 моделей);
- $H-PD$ – предложенный подход выполнения прогноза на уровне групп турбин; объединение в группы по временным рядам измерений мощности каждой турбины методом иерархической агломеративной кластеризации, для измерения расстояния между временными рядами используется алгоритм динамической трансформации временной шкалы (24 вычислительных эксперимента, получено 324 регрессионные модели, в рамках оптимизации гиперпараметров обучено 19 440 моделей).

Результаты вычислительных экспериментов по определению наиболее удачного распределения турбин по группам с точки зрения фактора улучшения по отношению к прогнозу по суммарной мощности ВЭС для методики $K-MS$ и для предложенных методик $K-P$, $H-PE$ и $H-PD$, представлены на рисунке 3.6. Количество групп, равное 1, соответствует методике P , то есть осуществлению прогноза на уровне ВЭС. Количество групп, равное 26, соответствует методике T , то есть выполнению прогноза на уровне отдельных турбин.

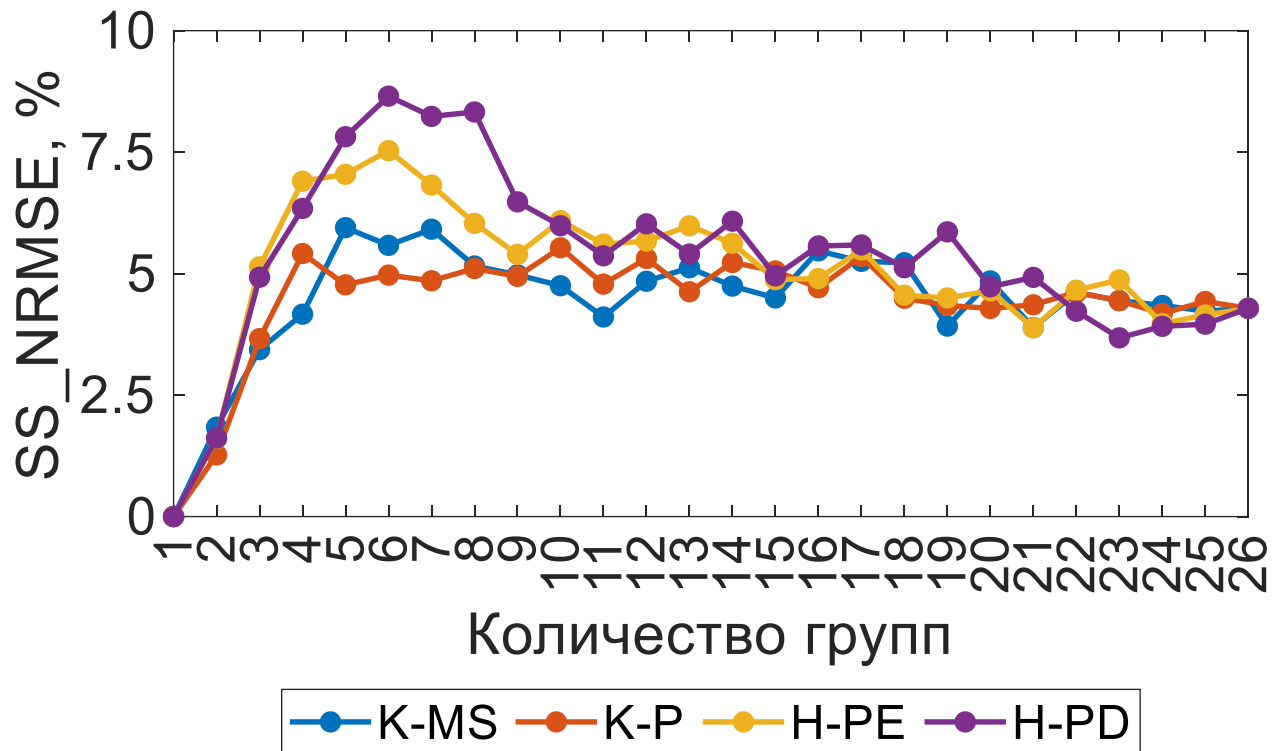


Рисунок 3.6 – Определение наиболее удачного распределения турбин по группам: *K-MS* – кластеризация *k*-средних, евклидово расстояние, среднее и СКО мощности и скорости ветра каждой турбины; *K-P* – кластеризация *k*-средних, евклидово расстояние, временные ряды мощности турбин, *H-PE* – иерархическая кластеризация, евклидово расстояние, временные ряды мощности турбин, *H-PD* – иерархическая кластеризация, *DTW*-расстояние, временные ряды мощности турбин

Из графиков на рисунке 3.6 видно, что фактор улучшения положителен для всех рассмотренных способов кластеризации ВЭУ и всех рассмотренных значений количества кластеров, в том числе для максимально возможного, когда каждая ВЭУ находится в отдельном кластере. Это позволяет сделать вывод о том, что выполнение прогнозов на уровне групп турбин и на уровне отдельных турбин позволяет снизить среднеквадратичную ошибку прогноза *NRMSE*. Для методики *K-MS* прогнозирование на уровне групп турбин имеет меньшую погрешность, чем прогнозирование на уровне отдельных турбин, для количества групп от 5 до 18, для методики *K-P* – от 4 до 19, для методики *H-PE* – от 3 до 20, для методики *H-PD* – от 3 до 21. Результаты вычислительных экспериментов показывают, что предложенный подход к выбору в качестве признаков для кластеризации непосредственно временных рядов измерений мощности ВЭУ обеспечивает

большее снижение погрешности прогноза генерации ВЭС. Наибольшего улучшения прогноза с точки зрения параметра $NRMSE$ удалось добиться при использовании одной из предложенных методик кластеризации $H-PD$ при разделении турбин на 6 групп.

Результаты распределения ВЭУ по группам при использовании различных методик кластеризации для количества групп, равного 6, представлены в таблице 3.2. Также в таблице представлено распределение ВЭУ по группам в соответствии с принадлежностью к фидерам GF .

Таблица 3.2 – Результаты распределения турбин по кластерам

Эксперимент	Группа турбин					
	1	2	3	4	5	6
GF	1, 2, 3, 4, 5, 6, 7	8, 9, 10, 11, 12, 13	14, 15, 16, 17, 18, 19, 20	21, 22, 23, 24, 25, 26	-	-
$K-MS$	1, 2, 7, 8, 14, 15, 20	3, 4, 5, 6	9, 17, 18, 22, 23	10, 24	11, 12, 13, 16, 19, 21	25, 26
$K-P$	1, 2, 8, 14, 15, 19, 20, 21	3, 4, 5, 6, 7	9, 10, 11, 12, 13	16, 17, 18, 23	22	24, 25, 26
$H-PE$	1, 2, 3, 6, 7	4, 5	8, 14, 15, 19, 20, 21	9, 10	11, 12, 13, 16, 17, 18, 22	23, 24, 25, 26
$H-PD$	1, 2, 3, 4, 5, 6, 7	8, 14, 15, 20, 21	9, 10	11, 12, 13, 16	17, 18, 19, 22	23, 24, 25, 26

Как видно из таблицы 3.2, все четыре методики распределения турбин по группам обеспечили разные результаты распределения. При этом более похожие результаты получены для методик, использующих один и тот же алгоритм кластеризации: k -средних для $K-MS$ и $K-P$, и иерархической агломеративной кластеризации для $H-PE$ и $H-PD$.

Результаты кластеризации для эксперимента с использованием методики $H-PD$ представлены на рисунках 3.7 и 3.8.

На рисунке 3.7 продемонстрировано, что в группы были объединены турбины, имеющие схожее поведение, они выделены одинаковыми цветами. На дендрограмме с результатами иерархической кластеризации на рисунке 3.8 показан процесс построения кластеров. Наиболее близкими временными рядами с точки зрения DTW -расстояния оказались мощности 1 и 2 ВЭУ.

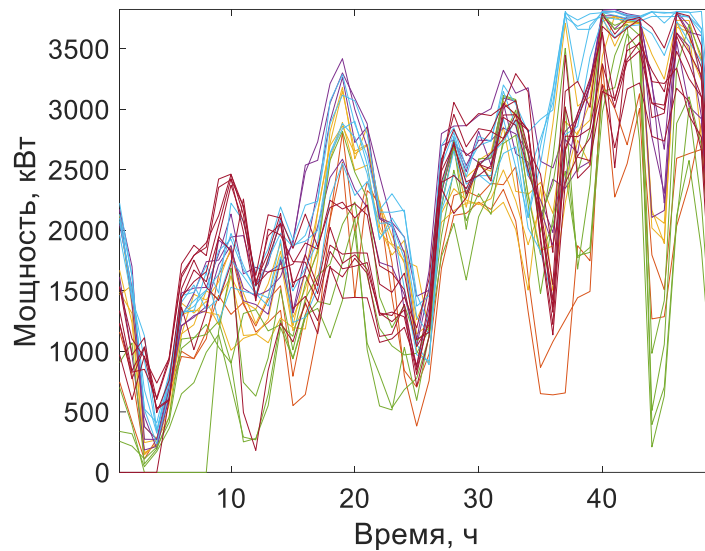


Рисунок 3.7 – Мощности турбин за 48 часов, распределенные по 6 кластерам с помощью методики *H-PD* (одинаковым цветом обозначены временные ряды в одном кластере)

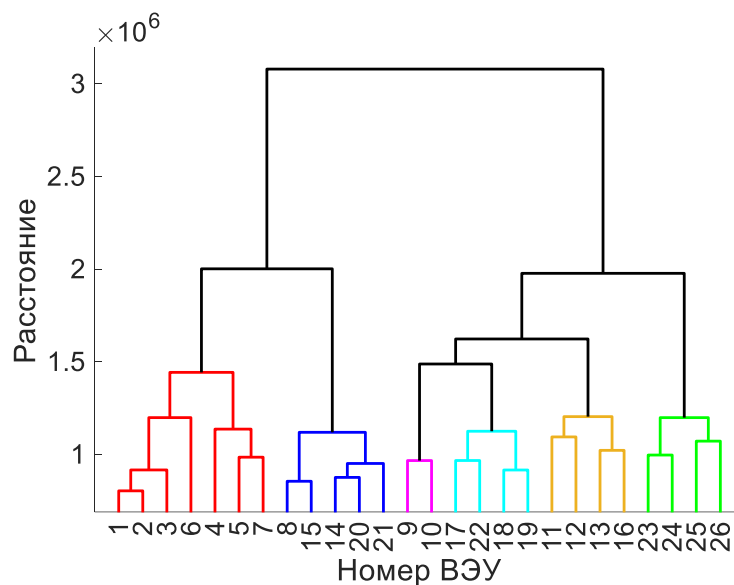


Рисунок 3.8 – Дендрограмма с результатами иерархической агломеративной кластеризации с помощью методики *H-PD*

Результаты оценки погрешности и надежности прогнозов рассматриваемых подходов к агрегации мощности отдельных установок представлены в таблице 3.3. Время обучения показано с учетом подбора гиперпараметров. Для одновременного обучения нескольких моделей в рамках одного эксперимента использовались параллельные вычисления в 12 потоках.

Результат распределения ВЭУ по группам с помощью методики *H-PD* показан на рисунке 3.9. В Приложении Б представлены результаты распределения ВЭУ по группам для методик *GF*, *K-MS*, *K-P* и *H-PE*.

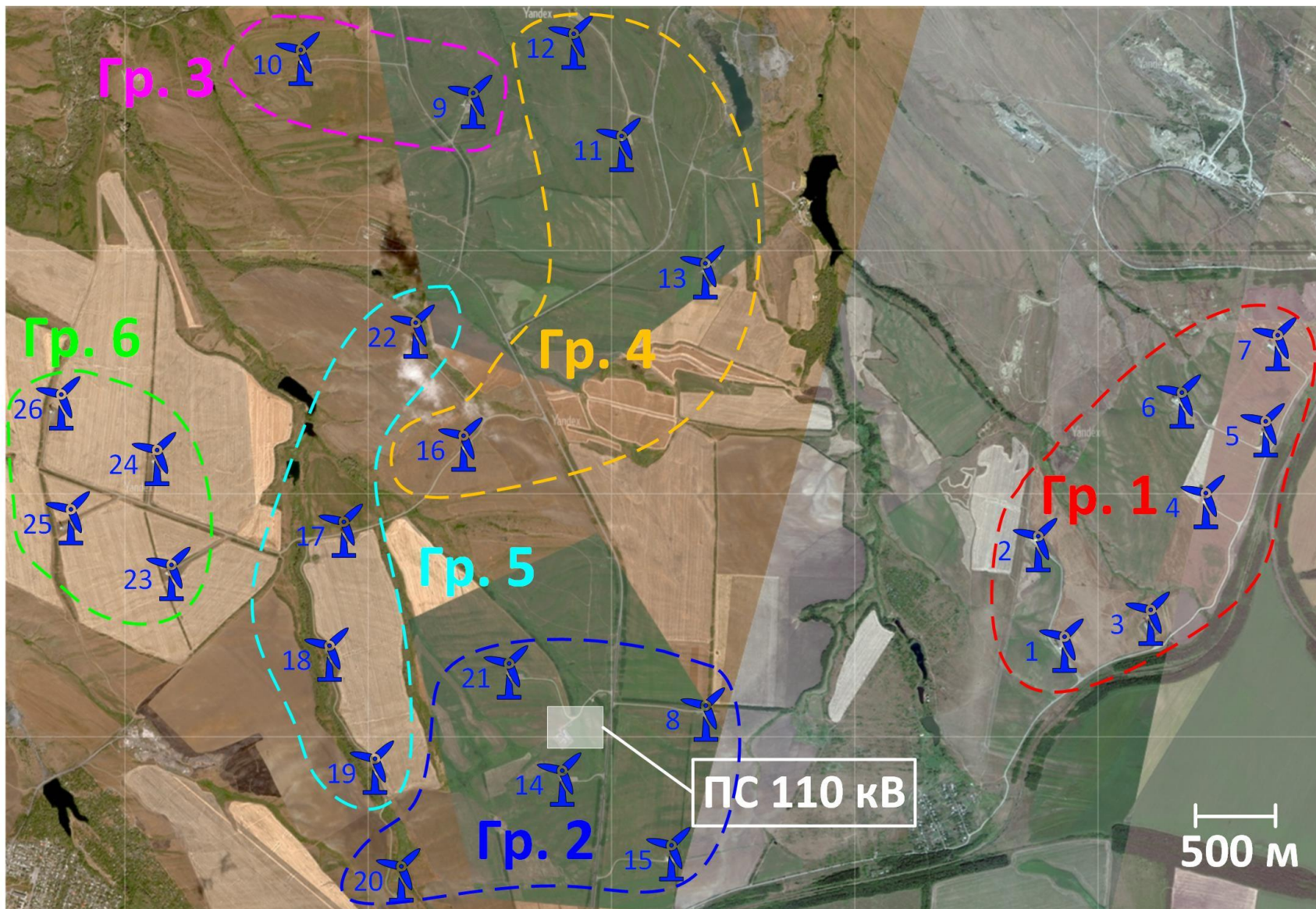


Рисунок 3.9 – Результат распределения ВЭУ по группам с помощью методики *H-PD*

Таблица 3.3 – Результаты оценки погрешности и надежности прогнозов на тестовом наборе данных для рассматриваемых подходов к агрегации мощности отдельных установок

Эксперимент	Кол- во групп	<i>SS</i> , %	<i>NRMSE</i> , %	<i>NMAE</i> , %	<i>NBias</i> , %	<i>NSAE</i> , %	<i>EICP_20</i> , %	Время обучения, с
<i>P</i>	1	0,0	13,31	9,4	-0,4	26,3	86,5	63,6
<i>GF</i>	4	5,2	12,61	9,2	-0,4	25,6	87,5	189,2
<i>T</i>	26	4,3	12,74	9,3	-0,4	26,0	87,9	757,7
<i>K-MS</i>	6	5,6	12,57	9,2	-0,3	25,6	87,8	179,3
<i>K-P</i>	6	5,0	12,65	9,3	-0,4	25,9	88,6	185,9
<i>H-PE</i>	6	7,5	12,31	9,1	-0,4	25,4	89,1	213,2
<i>H-PD</i>	6	8,7	12,16	9,0	-0,5	25,1	89,5	203,9

Как видно из таблицы 3.3, прогнозирование как на уровне отдельных турбин, так и на уровне групп турбин, даже при упрощенном разделении турбин на группы по принадлежности к фидерам, обеспечило снижение как средней, так и среднеквадратичной ошибки прогноза. При этом на обучение отдельных регрессионных моделей для каждой группы тратится большее количество времени, что особенно заметно при обучении отдельных моделей для каждой турбины. При использовании методики прогноза на уровне групп турбин с количеством групп, равном 6, время обучения моделей с учетом подбора гиперпараметров увеличилось в 3-4 раза, при прогнозировании на уровне отдельных ВЭУ – более чем в 10 раз.

Наибольший фактор улучшения был достигнут при использовании предложенной в данной работе методики *H-PD* – обеспечено снижение среднеквадратичной ошибки прогноза на 8,7 % по отношению к прогнозу по ВЭС. В дальнейших экспериментах для разделения турбин на группы будет использоваться данная методика с обозначением *GC*.

3.4. Выводы по главе

1. Мощность, выдаваемая отдельными ветроустановками, расположенными на одной и той же ВЭС, может значительно отличаться из-за их географической удаленности, «затенения ВЭУ» друг другом, наличия препятствий, наклона поверхности и других особенностей ландшафта.

2. Одним из способов повысить эффективность прогноза генерации ВЭС является определение наиболее удачного способа агрегации мощности отдельных

ВЭУ при построении регрессионных моделей. Существует три основных подхода: прогноз на уровне ВЭС, прогноз на уровне групп турбин и прогноз на уровне отдельных турбин. Единого мнения о том, какой подход является наиболее эффективным, пока нет.

3. Рассмотрена существующая методика распределения турбин по группам *K-MS*, основанная на объединении ВЭУ в группы по данным о средних значениях и среднеквадратичных отклонениях измеренных мощностей и скоростей ветра каждой турбины методом кластеризации *k*-средних

4. В представленной работе предложено три новых методики распределения ВЭУ по группам: *K-P*, *H-PE* и *H-PD*. Отличительной особенностью всех трех предложенных методик является использование в качестве признаков при кластеризации непосредственно временных рядов мощности отдельных ВЭУ. Методики *H-PE* и *H-PD* основаны на применении иерархической агломеративной кластеризации, при этом методика *H-PD* использует *DTW*-расстояние для определения схожести временных рядов.

5. Результаты вычислительных экспериментов для реальной ВЭС показали, что прогнозирование как на уровне отдельных турбин, так и на уровне групп турбин, даже при упрощенном разделении турбин на группы по принадлежности к фидерам, позволяет обеспечить снижение как средней, так и среднеквадратичной ошибки прогноза. При этом на обучение отдельных регрессионных моделей тратится большее количество времени, особенно на обучение отдельной модели для каждой турбины.

6. Выполнение прогнозирования на уровне групп турбин позволяет как снизить погрешность прогнозирования генерации ВЭС, так и обеспечить более низкое потребление вычислительных ресурсов по сравнению с методом построения прогнозов на уровне отдельных ВЭУ.

7. По результатам проведенных вычислительных экспериментов для исследуемой ВЭС по определению наиболее эффективной методики объединения ВЭУ в группы и подходящего количества кластеров было определено, что

наибольшее снижение среднеквадратичной ошибки прогноза обеспечила методика *H-PD* при разделении турбин на 6 групп.

8. Новый способ объединения турбин в группы с помощью метода иерархической агломеративной кластеризации по *DTW*-расстоянию (методика *H-PD*) позволил снизить среднеквадратичную ошибку прогноза на 8,7 % по отношению к прогнозу на уровне ВЭС.

ГЛАВА 4. ПРЕДВАРИТЕЛЬНАЯ ОБРАБОТКА ДАННЫХ ДЛЯ КРАТКОСРОЧНОГО ПРОГНОЗИРОВАНИЯ ГЕНЕРАЦИИ ВЕТРОЭЛЕКТРОСТАНЦИЙ

4.1. Идентификация выбросов на характеристиках мощности ветроустановок и ветроэлектростанций

Под понятием «выброс» (аномалия, недостоверное наблюдение, ошибочное наблюдение) обычно понимают наблюдение или объект (или подмножество наблюдений или объектов), поведение которого не согласуется с другими наблюдениями или объектами в рассматриваемом множестве [282–284]. В качестве альтернативного определения понятия «выброс» также используют следующее: «выброс – это наблюдение, которое настолько сильно отличается от других наблюдений, что вызывает подозрения, будто оно было вызвано другим механизмом» [285, 286].

Выбросы можно разделить на два типа: глобальные и локальные [287]. Глобальным выбросом можно назвать наблюдение, расположение которого в пространстве признаков значительно отличается от большинства других наблюдений в рассматриваемом множестве, выходя за пределы ожидаемых для данного множества диапазонов значений признаков. Под локальным выбросом, также называемым концептуальным, понимается наблюдение, которое значительно отличается от большинства других наблюдений в рассматриваемом множестве только в определенном контексте. Иллюстрация к двум основным типам выбросов приведена на рисунке 4.1 [287]. Наблюдения, являющиеся выбросами, обозначены красным цветом. На левом рисунке глобальными выбросами являются два наблюдения, значительно удаленных от остальных наблюдений, расположенных в плотной группе. На правом рисунке приведен пример локальных выбросов: четыре наблюдения расположены достаточно близко друг к другу, а все остальные наблюдения расположены на большем расстоянии друг от друга. При этом стоит отметить, что четыре наблюдения, являющиеся

локальными выбросами, расположены в ожидаемом диапазоне значений для наблюдений в этой выборке.

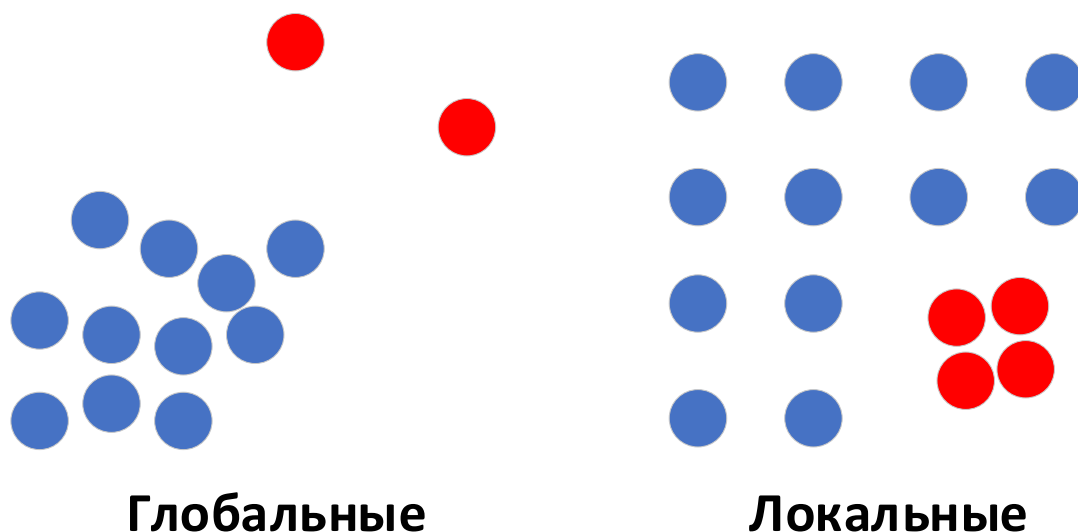


Рисунок 4.1 – Типы выбросов: глобальные – слева, локальные (концептуальные) – справа

Одним из способов повышения эффективности прогнозирования генерации ВЭС является идентификация и исключение выбросов из обучающей выборки, используемой при построении регрессионных моделей. Наиболее часто для задачи обнаружения выбросов используются характеристики мощности ВЭУ и ВЭС. В литературе при классификации типов выбросов обычно выделяют от 3 до 7 типов [288–291]. В данной работе рассматривается классификация, предложенная в [288], предполагающая наличие 5 основных типов выбросов на характеристиках мощности. Характеристика мощности ВЭУ с обозначением этих типов выбросов представлена на рисунке 4.2.

Выбросы типа 1 соответствуют области характеристики мощности с наблюдениями, имеющими близкую к нулю (иногда даже отрицательную) или равную нулю выходную мощность при значительных скоростях ветра (больше минимальной). На рисунке 4.2 выбросы данного типа представляет собой плотное скопление точек вдоль оси скорости ветра. Для значений скорости ветра ниже минимальной, выбросы данного типа неотличимы от корректных наблюдений, соответствующих отсутствию выходной мощности при недостаточной энергии воздушного потока. Среди основных причин данного типа выбросов на

характеристике мощности отдельной ВЭУ исследователи выделяют отказ ВЭУ, а также сбои в работе измерительных устройств и каналов связи. Количество выбросов данного типа на агрегированной характеристике мощности группы ВЭУ или всей ВЭС обычно значительно меньше. Основной причиной таких выбросов на этих уровнях агрегации могут стать отказы на элементах схемы выдачи мощности и последующая работа релейной защиты [288–291].

Выбросы типа 2 представляют собой наблюдения, у которых выходная мощность ВЭУ ниже номинальной, а значения мощности занижены и примерно одинаковы для широкого диапазона изменения скорости ветра. Обычно выбросы данного типа представляют собой плотную горизонтальную область под характеристикой мощности, как показано на рисунке 4.2. Основной причиной возникновения выбросов данного типа на уровне отдельных ВЭУ, групп ВЭУ и ВЭС являются искусственные ограничения выходной мощности, выполняемые с помощью системы управления инвертором или изменения угла поворота лопастей. Среди других причин также называют повреждения редуктора мультипликатора, которые не позволяют отдельной ВЭУ выйти на номинальную мощность [288–291].

Выбросы типа 3 являются наблюдениями, расположенными в области над характеристикой мощности, как показано на рисунке 4.2. Особенность данных выбросов в том, что значения выходной мощности ВЭУ могут изменяться в широком диапазоне, соответствуя практически одинаковым значениям скорости ветра, при этом скорость ветра ниже номинальной. Причины возникновения данного типа выбросов не изучены в достаточной мере. Среди наиболее вероятных причин рассматривают сбои в работе измерительных устройств и каналов связи [288]. Также возможной причиной является порывистый ветер и усреднение значений скорости ветра и мощности на часовом интервале. При усреднении ветровой поток, имеющий низкую среднюю скорость, но высокую скорость ветра на некоторых интервалах внутри часа, может увеличивать выходную мощность на данных интервалах, тем самым увеличивая также и усредненную мощность [290].

Выбросы типа 4 – это разбросанные под характеристикой мощности точки, выходная мощность, которых может быть значительно ниже номинальной, при этом скорость ветра вблизи номинальной или выше. Основной причиной возникновения таких выбросов как на уровне отдельных турбин, так и всей ВЭС, исследователи называют эффект «затенения» ВЭУ. Среди других возможных причин указывается также запаздывание в работе системы управления ориентации ВЭУ [288–291].

Выбросы типа 5 соответствуют точкам под изгибом характеристики мощности, которые имеют сниженную выходную мощность, не соответствующую ожидаемой для скорости ветра в диапазоне 0,5-1,0 о.е. от номинальной. Среди наиболее вероятных причин возникновения в литературе называют географическое расположения ВЭУ и влияние атмосферных условий, приводящее к затруднению вращения ветроколеса, например, в результате обледенения или значительного загрязнения лопастей [288, 290].

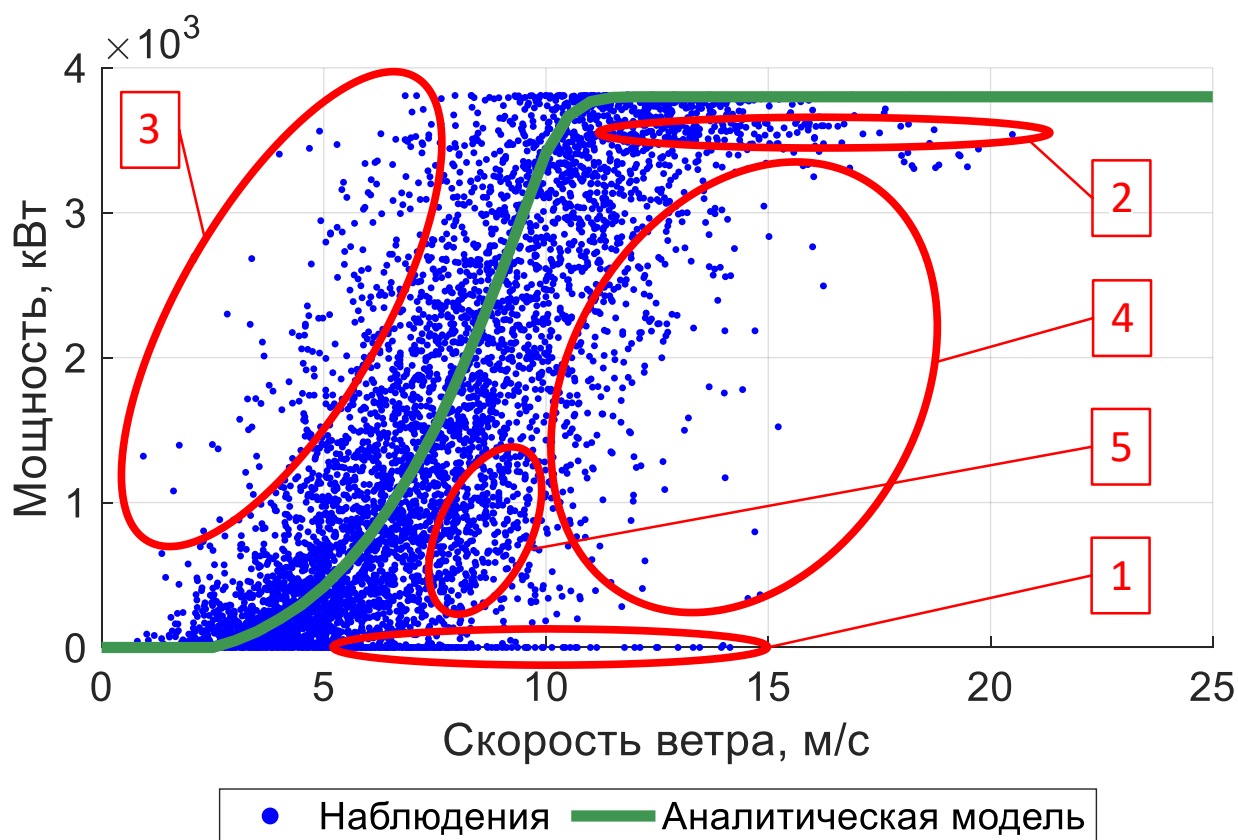


Рисунок 4.2 – Характеристика мощности ВЭУ с обозначением различных типов выбросов

Характеристики рассмотренных типов выбросов, а также их возможные причины, сведены в таблицу 4.1. Данные типы выбросов можно отнести к локальным, поскольку все наблюдения находятся в диапазоне ожидаемых значений скорости ветра (от нуля до максимальной) и мощности (от нуля до номинальной). При этом некоторые типы выбросов можно рассматривать как глобальные при добавлении некоторого контекста. Например, если выделить все наблюдения в диапазоне скорости ветра от 11 до 11,5 м/с, где ожидаются близкие к номинальным значения мощности, выбросы типа 1 окажутся глобальными.

Таблица 4.1 – Характеристики типов выбросов и их возможные причины

Типы	Характеристика	Возможные причины
Тип 1	Выходная мощность ВЭУ близка к нулю или равна нулю при значительных скоростях ветра (больше минимальной)	Отказ ВЭУ, сбой в работе измерительных устройств и каналов связи, отказы на элементах схемы выдачи мощности ВЭС
Тип 2	Выходная мощность ВЭУ ниже номинальной, значения мощности занижены и примерно одинаковы для широкого диапазона изменения скорости ветра	Искусственные ограничения выходной мощности, повреждения редуктора мультипликатора
Тип 3	Значения выходной мощности ВЭУ изменяются в широком диапазоне при отсутствии изменений скорости ветра, скорость ветра ниже номинальной	Не изучены в достаточной мере, возможные: сбой в работе измерительных устройств и каналов связи, эффект усреднения скорости ветра и выходной мощности при порывах ветра
Тип 4	Разбросанные под характеристикой мощности точки, выходная мощность ВЭУ ниже номинальной, скорость ветра вблизи номинальной или выше	Эффект «затенения» ВЭУ, запаздывание в работе системы управления ориентации ВЭУ
Тип 5	Точки под изгибом характеристики мощности, выходная мощность ВЭУ низкая, скорость ветра в диапазоне 0,5-1,0 о.е. от номинальной	Обледенение или значительное загрязнение лопастей

Среди подходов к идентификации выбросов можно выделить три группы наиболее часто используемых методов [284, 286]:

- Анализ экстремальных значений – методы, основанные на анализе распределений вероятности, в частности, хвостов распределений (например, правило трех сигм для нормального распределения, метод межквартильного размаха и т.д.).

- Оценка меры близости – методы, основанные на определении расстояний между наблюдениями в пространстве признаков, при этом мерой близости является принадлежность наблюдений к кластерам (или отсутствие принадлежности к кластеру), расстояние до ближайшего соседа или ближайшего центра кластера, количество наблюдений в заданной окрестности. Наиболее часто при поиске выбросов используются методы кластерного анализа, например, метод k -средних, основанная на плотности пространственная кластеризация для приложений с шумами (*Density Based Spatial Clustering of Applications with Noise, DBSCAN*), и т.д.
- Проецирование наблюдений в пространство меньшей размерности – методы, основанные на уменьшении размерности пространства признаков, обычно до двумерного пространства. Среди часто используемых алгоритмов можно выделить метод главных компонент (*Principal Component Analysis, PCA*) и самоорганизующиеся карты Кохонена.

Анализ существующих исследований показывает, что при решении задачи идентификации выбросов на характеристике мощности, может применяться любой из трех данных подходов или их комбинации.

Например, в исследовании [292] предложена методика идентификации выбросов типов 1, 2, 3 и 4, в качестве объекта исследования рассматривалась характеристика мощности ВЭУ 2 МВт и ВЭС 40 МВт. Для поиска выбросов типа 3 и 4, представляющих собой разбросанные вблизи характеристики мощности точки, использован метод межквартильного размаха. На первом этапе выполнялся горизонтальный поиск – все наблюдения были отсортированы по возрастанию по величине мощности, область ожидаемых значений мощности разделена на полосы шириной по 1,25% от номинальной мощности, внутри каждой полосы оценивались межквартильный размах и границы допустимых значений. Аналогично выполнялся вертикальный поиск в полосах шириной по 0,5 м/с. Для поиска выбросов типа 1 и 2, представляющих собой плотные горизонтальные области точек, использовался метод кластеризации *DBSCAN*. Кластеризация выполнялась

также внутри вертикальных полос шириной 0,5 м/с. При этом ожидалось, что выбросы будут соответствовать кластерам, расположенным в нижней части полосы. Все наблюдения, не принадлежащие кластеру с наибольшим средним значением мощности, считались выбросами. Предложенная методика позволила снизить нормализованную среднеквадратичную ошибку аппроксимации характеристики мощности более, чем в 3 раза. Похожая, но более упрощенная, методика обнаружения выбросов с помощью метода межквартильного размаха использовалась в [293]. Основным отличием являлось использование только вертикального поиска, но не в рамках полосы заданной величины скорости ветра, а в рамках наблюдений, относящихся к определенному временному промежутку (например, заданному месяцу).

Авторами статьи [294] предложен подход к идентификации выбросов на характеристике мощности, основанный исключительно на анализе экстремальных значений. Предложенный подход также основан на анализе наблюдений в ограниченных диапазонах скорости ветра (порядка 0,5 м/с), и выполнялся в два этапа. На первом этапе все наблюдения в диапазоне сортировались по значению мощности. Далее выполнялось последовательное добавление по одному наблюдению в подвыборку, начиная с первого, при этом каждый раз рассчитывалась дисперсия внутри подвыборки. По результатам расчетов обнаруживалось наблюдение, добавление которого приводило к значительному изменению дисперсии. Все последующие наблюдения после обнаруженного считались выбросами. По результатам идентификации выбросов на первом этапе выявлялись выбросы типа 1 и 2, и часть выбросов типа 4. На втором этапе применялся метод межквартильного размаха, который позволял обнаружить выбросы типа 3 и оставшиеся выбросы типа 4.

В работе [295] рассмотрено совместное применение метода кластеризации *DBSCAN* и метода кластеризации *k*-средних для идентификации выбросов на характеристике мощности двух ВЭУ номинальной мощностью 1,5 МВт и 3 МВт. При этом выбросы типа 1 идентифицировались вручную, после чего выполнялся поиск выбросов типа 3 и 4 алгоритмом *DBSCAN*, параметры алгоритма были

выбраны экспертно. Для оставшихся наблюдений выполнялась кластеризация методом k -средних, количество кластеров и начальное приближение центроидов также выбраны экспертно. Наблюдения, оказавшиеся в кластерах, центры которых соответствовали ожидаемым областям появления выбросов 2 и 4 типа, также были признаны выбросами. При определении расстояния между объектами использовалось расстояние Махаланобиса.

Идентификация выбросов на характеристике мощности отдельных ВЭУ с использованием метода кластеризации *DBSCAN* также применялась в работе [296]. Параметры алгоритма были выбраны экспертно, при этом способ определения расстояния между объектами не был указан. Проверка эффективности предложенной методики идентификации выбросов выполнялась при краткосрочном прогнозировании генерации ВЭС мощностью 50 МВт на сутки вперед. Результаты исследования показали, что алгоритм кластеризации позволил обнаружить выбросы типа 1, 3 и 4. Исключение обнаруженных выбросов привело к снижению среднеквадратичной ошибки прогноза более, чем в 1,5 раза.

В работе [297] использовался еще один метод, основанный на оценке меры близости наблюдений для идентификации выбросов – локальный уровень выбросов (*Local Outlier Factor, LOF*). Особенностью данной работы являлось предложенная мера расстояния между объектами, представляющая собой модифицированную формулу расчета евклидова расстояния, учитывающую близость наблюдения к аналитической модели характеристики мощности в виде весового коэффициента.

В статье [298] авторами рассматривался один из популярных способов идентификации выбросов на характеристике мощности ВЭУ, основанный на использовании квантильной регрессии. В качестве объекта исследования рассматривались отдельные ВЭУ на трех различных ВЭС мощностью 2 МВт и 5 МВт. С помощью квантильной регрессии осуществлялось определение характеристик мощности, соответствующих 0,05 и 0,95 квантилю. Наблюдения, находящиеся вне диапазона между характеристиками, признавались выбросами. Прогноз выполнялся по характеристике, соответствующий 0,5 квантилю. По

результатам исследования было обнаружено, что предложенный подход позволяет снизить среднеквадратичную ошибку прогноза на 5-20 %.

В данной работе предлагается ряд новых методик для идентификации выбросов на характеристике мощности всей ВЭС, отдельных ВЭУ и групп ВЭУ.

4.2. Методики идентификации выбросов по характеристике мощности

Для сравнения предлагаемых в данной работе методик с уже существующими подходами рассматривается способ идентификации выбросов по характеристике мощности, основанный на методе межквартильного размаха (*Interquartile Range, IQR*), предложенный в [292–294].

Методика идентификации выбросов по характеристике мощности *QM*. Среди существующих подходов к идентификации выбросов с помощью характеристики мощности наиболее распространенным является метод межквартильного размаха, применение которого к данной задаче рассмотрено в [292–294]. Данный подход основан на анализе экстремальных значений. Иллюстрация к методике *QM* представлена на рисунке 4.3.

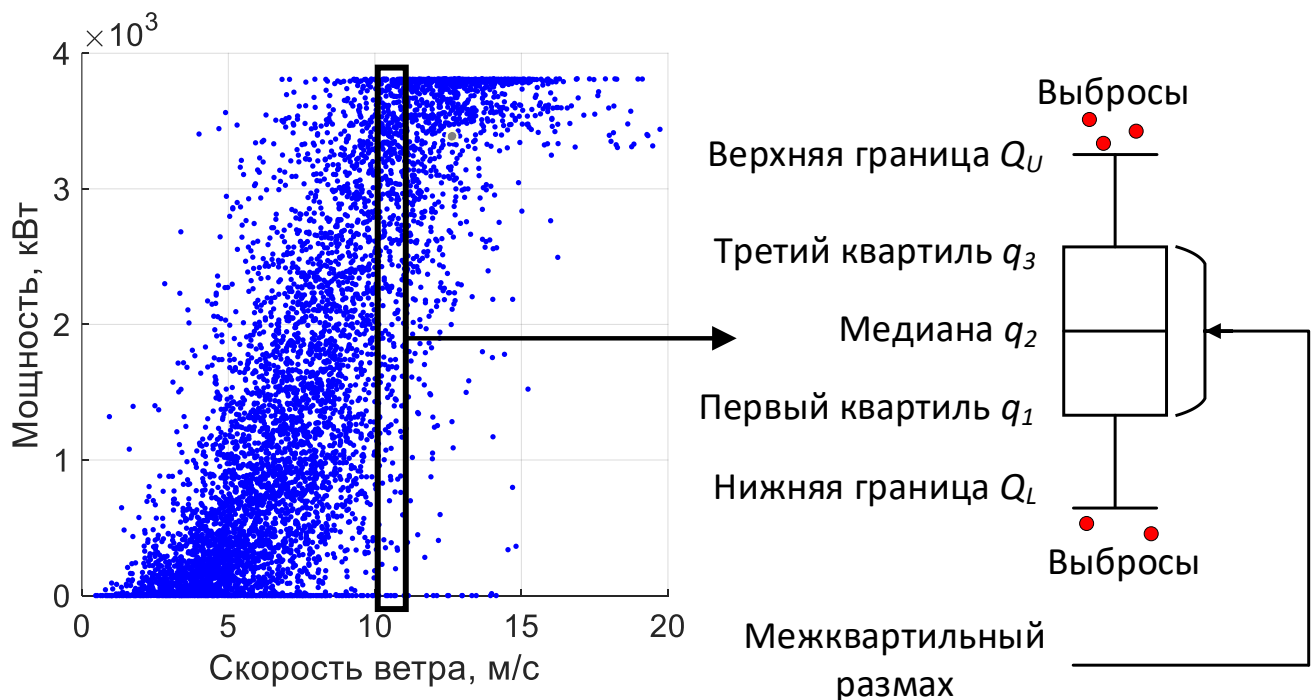


Рисунок 4.3 – Методика идентификации выбросов на характеристике мощности, основанная на методе межквартильного размаха

Методика *QM* базируется на последовательном рассмотрении подвыборок наблюдений, находящихся в ограниченных диапазонах скорости ветра. В данной работе использовалась полоса шириной 0,5 м/с. Для каждой подвыборки по отдельности используется метод межквартильного размаха. Для значений мощности наблюдений, оказавшихся в подвыборке, рассчитываются границы Q_L и Q_U в соответствии с выражениями [299]:

$$[Q_L, Q_U] = [q_1 - 1,5 \cdot Q_R, q_3 + 1,5 \cdot Q_R], \quad (4.1)$$

где q_1 – первый квартиль; q_3 – третий квартиль; Q_R – межквартильный размах, определяемый выражением $Q_R = q_3 - q_1$.

Наблюдения, для которых значения мощности оказываются вне рассчитанных диапазонов, идентифицируются как выбросы.

Методика идентификации выбросов по характеристике мощности DB.

Методика идентификации выбросов *DB* представляет собой модифицированный вариант методики, представленной в работах [295, 296]. Авторами этих исследований предложено вместо анализа экстремальных значений выполнять оценку меры близости наблюдений друг к другу с использованием алгоритма основанной на плотности пространственной кластеризации для приложений с шумами (*DBSCAN*).

При кластеризации набора точек в некотором пространстве признаков алгоритм *DBSCAN* использует меру оценки близости, основанную на плотности точек (объектов), то есть количества точек в окрестности заданного радиуса, определенного с помощью произвольной меры расстояния (например, евклидовой). Формирование кластеров осуществляется исходя из минимальной плотности точек, определяемой по двум гиперпараметрам алгоритма: минимального числа точек N_{min} , которые должны образовывать плотную область, и максимального радиуса соседства между точками (ϵ -окрестность).

Основные этапы кластеризации по методу *DBSCAN* включают [300]:

- определение для каждой точки количества точек внутри максимального радиуса соседства ϵ ;

- если их больше минимального числа точек, которые должны образовывать плотную область N_{min} , то точка считается основной;
- основные точки, находящиеся в ε -окрестности друг друга, соединяются ребром (являются напрямую достижимыми);
- выделяются компоненты связности, образующие кластеры;
- точки, в окрестности которых есть основные, но количество точек в окрестности меньше N_{min} , являются граничными точками и относятся к кластеру ближайшей основной точки (являются достижимыми);
- точки, в окрестности которых нет основных точек и в которых содержится менее N_{min} наблюдений, считаются выбросами (являются недостижимыми).

На рисунке 4.4 представлен пример результатов кластеризации по алгоритму *DBSCAN* для $N_{min} = 4$. Зеленым обозначены основные точки, оранжевым – граничные, красным – выбросы. Цифра, вписанная в точку, обозначает количество напрямую достижимых точек, включая саму рассматриваемую точку. Напрямую достижимые точки соединены ребрами и образуют кластер.

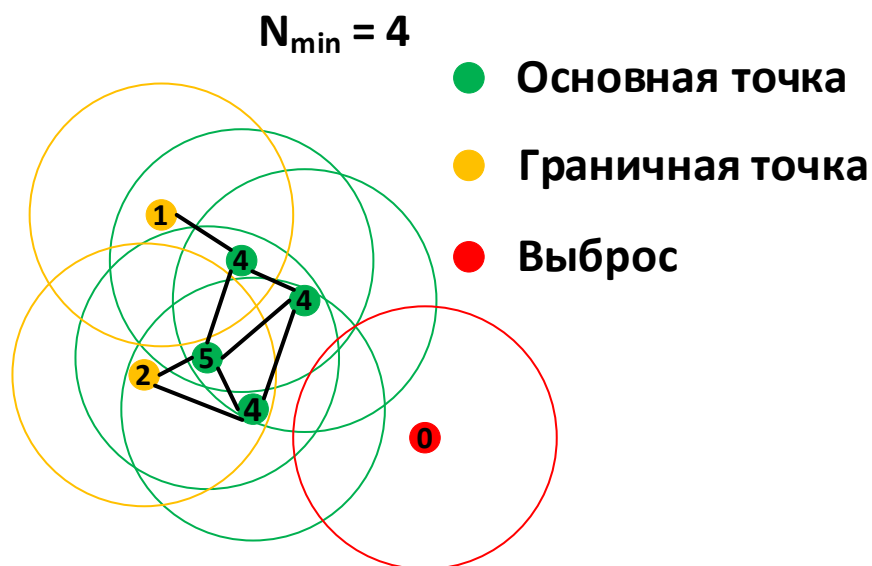


Рисунок 4.4 – Кластеризация *DBSCAN*

Далее представлено описание алгоритма основанной на плотности пространственная кластеризация для приложений с шумами *DBSCAN* [300, 301], соответствующее реализации данного алгоритма в *Statistics and Machine Learning Toolbox* ПК *MATLAB* [195].

Алгоритм 11: Алгоритм основанной на плотности пространственная кластеризация для приложений с шумами (DBSCAN)

Вход:

X – матрица признаков для кластеризации,

ε – максимальный радиус соседства,

N_{min} – минимальное количества точек в ε -окрестности,

d – способ определения расстояния между точками.

1. Формирование вектора принадлежности объектов к кластерам I_c , размер которого совпадает с количеством строк в матрице признаков X и равен N . Принадлежность к кластеру обозначается номером кластера, отсутствие принадлежности – индексом 0, выброс – индексом -1.
2. Для всех элементов n матрицы признаков X от 1 до N .
3. Если в I_c у элемента n уже обозначена принадлежность к кластеру, то выполняется переход к следующему элементу.
4. Определение расстояний d между объектом n и всеми другими объектами в матрице признаков X .
5. Определение объектов, расстояния до которых меньше, чем ε . Формирование вектора I'_c , фиксация индексов этих объектов.
6. Если размер вектора I'_c меньше N_{min} , то в векторе принадлежности объектов к кластерам I_c наблюдение обозначается как выброс. Выполняется переход к следующему элементу.
7. Если размер вектора I'_c больше или равен N_{min} , то элементы, индексы которых входят в I'_c , формируют новый кластер c .
8. Обновляется вектор индексов I_c , фиксируется принадлежность элемента n к кластеру c .
9. Для всех элементов q в векторе I'_c .
10. Если ранее элемент q был помечен как выброс, но он оказался в I'_c , то такой элемент является граничной точкой. Обновляется вектор индексов I_c , фиксируется принадлежность элемента n к кластеру c .
11. Если в I_c у элемента q уже обозначена принадлежность к кластеру, то выполняется переход к следующему элементу.
12. Определение расстояний d между объектом q и всеми другими объектами в матрице признаков X .
13. Обновляется вектор индексов I_c , фиксируется принадлежность элемента q к кластеру c .
14. Определение объектов, расстояния до которых меньше, чем ε . Обновление вектора I''_c , фиксация индексов этих объектов.
15. Если размер вектора I''_c больше или равен N_{min} , то элементы, индексы которых входят в I''_c , добавляются к элементам в I'_c .

Выход:

I_c – вектор индексов, определяющий принадлежность объектов к кластерам или выбросам.

В рассматриваемой методике, аналогично исследованиям [295, 296], для определения расстояния между точками на характеристике мощности используется расстояние Махаланобиса:

$$d_M(\bar{x}_i, \bar{x}_j) = \sqrt{(\bar{x}_i - \bar{x}_j)^T \cdot C^{-1} \cdot (\bar{x}_i - \bar{x}_j)} \quad (4.2)$$

где $\bar{x}_i$ – вектор признаков объекта i ; $\bar{x}_j$ – вектор признаков объекта j ; $d_M(\bar{x}_i, \bar{x}_j)$ – расстояние Махаланобиса между объектами i и j ; C – матрица ковариации между векторами $\bar{x}_i$ и $\bar{x}_j$.

Для определения гиперпараметров N_{min} и ε в исследованиях [295, 296] используется метод поиска по решетке. В предлагаемой модификации методики идентификации выбросов по характеристике мощности для определения N_{min} и ε используется другой подход. В качестве величины минимального количества точек N_{min} , в соответствии с рекомендациями разработчиков алгоритма *DBSCAN* [301], используется удвоенное количество признаков. В задаче идентификации выбросов по характеристике мощности количество признаков равно двум (скорость ветра и мощность), поэтому N_{min} принимается равным четырем.

Определение максимального радиуса соседства ε предлагается также выполнять в соответствии с рекомендациями разработчиков *DBSCAN*, описанными в [301]. Предлагается применять эвристический подход, основанный на построении графика, представляющего собой отсортированные по возрастанию расстояния до N ближайшего соседа, где N равняется минимальному количеству точек N_{min} , в данном случае 4. Искомое значение величины максимального радиуса соседства ε будет соответствовать точке перегиба данного графика. Точка перегиба представляет собой точку максимальной крутизны кривой. При этом определение данной точки авторами исходного алгоритма *DBSCAN* [301] предлагается выполнять визуально по результатам построения графика.

В данной работе для определения точки перегиба предлагается использовать алгоритм *KNeedle* [302]. На рисунке 4.5 представлена иллюстрация к способу определения точки перегиба. На первом этапе все значения X и Y нормализуются и оказываются в диапазоне от 0 до 1, исходная кривая показана черным цветом. Далее, как показано на рисунке слева, выполняется определение разностей величин Y исходной кривой и прямой, соответствующей уравнению $y = x$, обозначенной красным цветом. Величины разностей показаны синими стрелками. При использовании алгоритма *KNeedle* задача определения значения X , соответствующего точке перегиба, сводится к определению локального максимума кривой разностей, как показано на рисунке справа. Для каждого локального максимума определяется пороговое значение (обозначено красным пунктиром), до достижения которого не должны встречаться другие локальные максимумы, чтобы локальный максимум считался перегибом. Точка, соответствующая перегибу, обозначена оранжевым цветом.

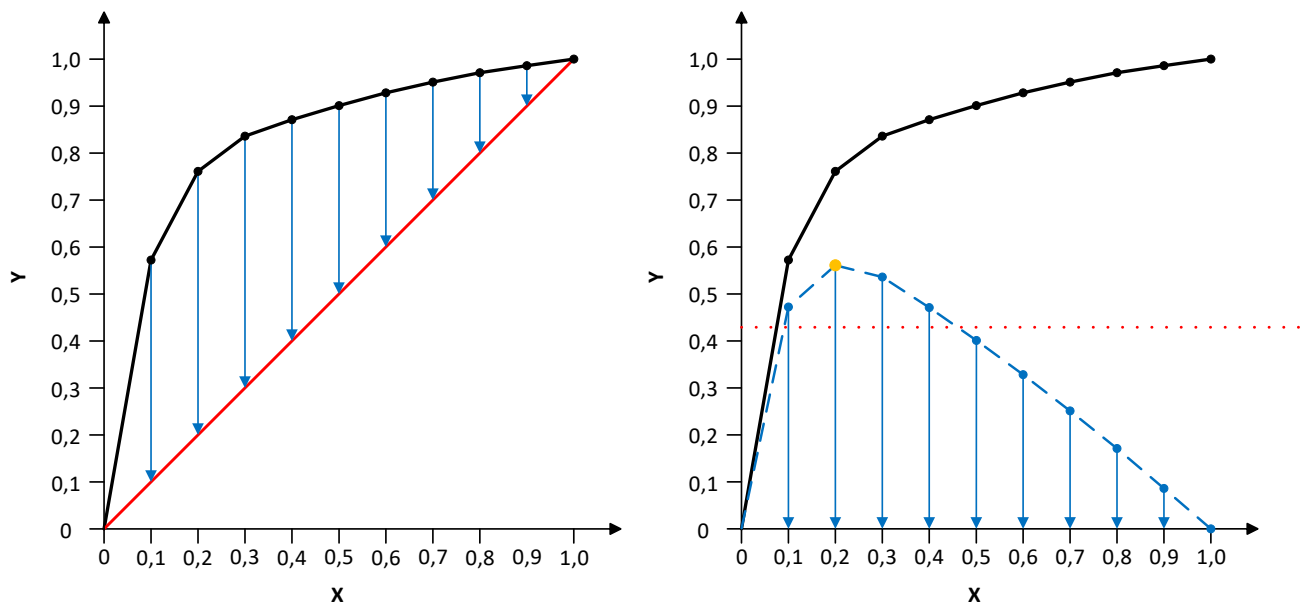


Рисунок 4.5 – Алгоритм поиска точки перегиба *KNeedle*: выбросов: определение разностей – слева, определение максимума кривой разностей – справа

Далее представлено описание алгоритма идентификации точки перегиба на графике *KNeedle* [302], соответствующее реализации данного алгоритма в библиотеке *kneed* для языка программирования *Python* [303].

Алгоритм 12: Алгоритм поиска точки перегиба на графике (*KNeedle*)**Вход:** X – вектор значений по оси абсцисс, Y – вектор значений по оси ординат, S – параметр чувствительности (по умолчанию 1).

1. Нормализация абсцисс всех значений x_i , входящих в X :

$$x_{ni} = \frac{x_i - \min X}{\max X - \min X}.$$

2. Нормализация абсцисс всех значений y_i , входящих в Y :

$$y_{ni} = \frac{y_i - \min Y}{\max Y - \min Y}.$$

3. Определение значений разностей:

$$x_{di} = x_{ni},$$

$$y_{di} = y_{ni} - x_{ni}.$$

4. Определение локальных максимумов кривой разностей, то есть значений $x_{mi} = x_{di}$ и y_{mi} , для которых $y_{di-1} > y_{di}$ и $y_{di+1} < y_{di}$.

Определение пороговых значений для каждого локального максимума:

$$T_m = y_{mi} - S \cdot \frac{\sum_{i=1}^{N-1} x_{di+1} - x_{di}}{N - 1},$$

Где N – количество элементов в X и Y .

5. Если до достижения T_m не встречаются другие локальные максимумы, то рассматриваемый локальный максимум считается точкой перегиба.

Выход: x_{mi} и y_{mi} – значения элементов X и Y , соответствующие точке перегиба.

На рисунке 4.6 представлен пример графика расстояний до 4 ближайшего соседа для точек на суммарной характеристике мощности ВЭС и результаты определения точки перегиба с помощью метода *KNeedle*.

Таким образом, среди особенностей предложенной методики *DB*, отличающей данную модификацию от исходной версии, представленной в исследованиях [295, 296], можно выделить:

- определение выбросов не только на характеристиках ВЭУ, но и групп турбин и всей ВЭС в целом;
- определение гиперпараметров алгоритма *DBSCAN*: N_{min} исходя из удвоенного количества признаков при кластеризации и ε в соответствии с графиком расстояний до ближайшего соседа с автоматическим определением точки перегиба методом *KNeedle*.

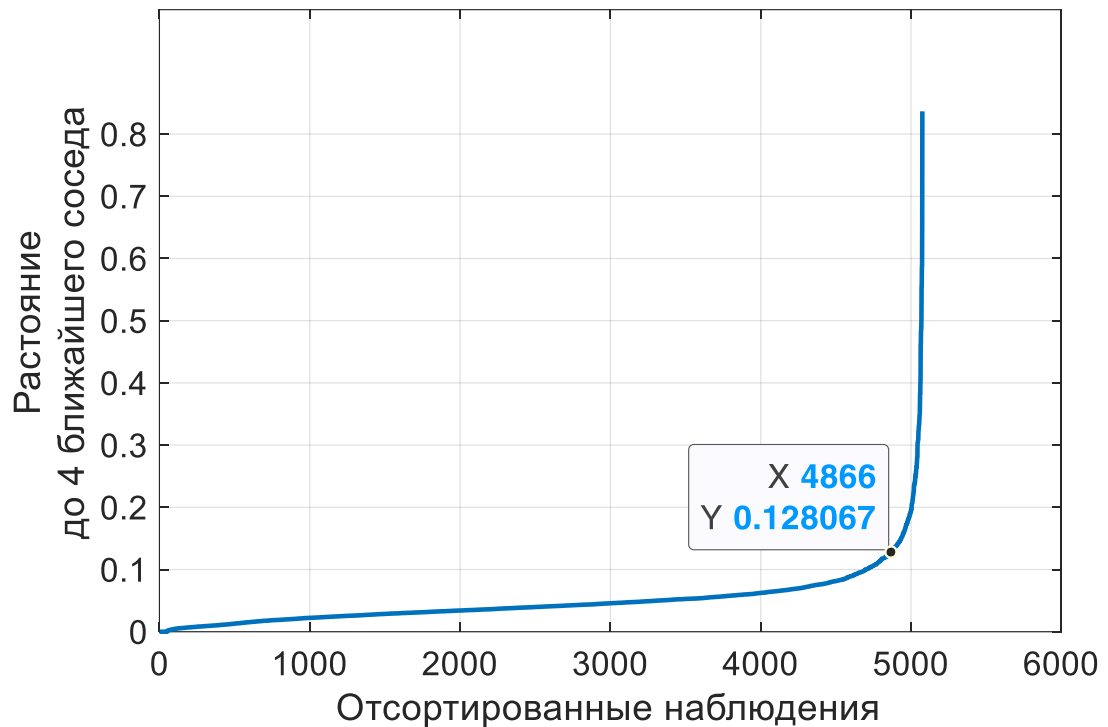


Рисунок 4.6 – График отсортированных расстояний до 4 ближайшего соседа и точка перегиба

Методика идентификации выбросов по характеристике мощности GL.

В данной работе предлагается новая методика идентификации выбросов на характеристиках мощности, основанная на алгоритме глобально-локальной оценки выбросов (*GLOSH*) по иерархической основанной на плотности пространственной кластеризации для приложений с шумами (*HDBSCAN**) (методика *GL*).

Алгоритм иерархической основанной на плотности пространственной кластеризации для приложений с шумами *HDBSCAN**, впервые предложенный в работах [304, 305], основан на алгоритме *DBSCAN**, представленный в тех же работах [304, 305]. Обозначение «*» позволяет отличить модификацию алгоритма *DBSCAN** от оригинальной версии алгоритма *DBSCAN* [301], которая описана выше. Отличие модифицированной версии от оригинальной заключается в отказе от концепции «граничных точек», которые в модифицированной версии также признаются объектами, являющимися выбросами.

Наиболее важной отличительной особенностью алгоритма *HDBSCAN** при сравнении с алгоритмом *DBSCAN* является отсутствие необходимости задания максимального радиуса соседства ε пользователем алгоритма. Для заданных

величин максимального количества объектов в кластере N_c и минимального количества точек N_{min} формируется иерархия различных реализаций кластеризаций $DBSCAN^*$, соответствующих различным значениям ε -окрестности. Полученная иерархия позволяет как выполнять поиск кластеров наблюдений с различной плотностью, так и осуществлять идентификацию локальных выбросов вблизи областей с высокой плотностью [306, 307].

Для определения различных значений величины максимального радиуса соседства ε , необходимых для построения иерархии кластеризаций, в алгоритме $HDBSCAN^*$ используется мера оценки расстояния, основанная на взаимной достижимости объектов. При этом вводится понятие основного расстояния $d_{N_{min}}(\bar{x}_n)$ для заданного значения N_{min} , рассчитываемого как расстояние $d(\bar{x}_n, \bar{x}_k)$ между объектами n и k , где $k - N_{min}$ -й (включая сам объект n) ближайший объект по отношению к объекту n . Значение расстояния взаимной достижимости $d_{mr}(\bar{x}_i, \bar{x}_j)$ между объектами i и j определяется как максимальное значение из основных расстояний по каждому из объектов i и j , а также расстояния между объектами i и j [304, 308]:

$$d_{mr}(\bar{x}_i, \bar{x}_j) = \max(d_{N_{min}}(\bar{x}_i), d_{N_{min}}(\bar{x}_j), d(\bar{x}_i, \bar{x}_j)), \quad (4.3)$$

где $\bar{x}_i$ – вектор признаков объекта i ; $\bar{x}_j$ – вектор признаков объекта j ; $d_{mr}(\bar{x}_i, \bar{x}_j)$ – расстояние взаимной достижимости между объектами i и j ; $d_{N_{min}}(\bar{x}_i)$ – основное расстояние для объекта i ; $d_{N_{min}}(\bar{x}_j)$ – основное расстояние для объекта j ; $d(\bar{x}_i, \bar{x}_j)$ – расстояние между объектами i и j , определенное по способу d .

В предлагаемой методике GL для определения расстояния между объектами, аналогично методике DB , используется расстояние Махаланобиса (4.2).

На рисунке 4.7 представлен пример определения расстояния взаимной достижимости для $N_{min} = 3$. В первом случае, при определении расстояний между точками 1 и 2, основное расстояние точек $d_3(1)$ и $d_3(2)$ оказалось меньше, чем расстояние $d(1,2)$, поэтому расстояние взаимной достижимости определяется как

расстояние между точками $d_{mr} = d(1,2)$. Во втором случае, при определении расстояний между точками 3 и 4, расстояние $d(3,4)$ и основное расстояние $d_3(3)$ меньше, чем основное расстояние $d_3(4)$, поэтому расстояние взаимной достижимости определяется как $d_{mr} = d_3(4)$.

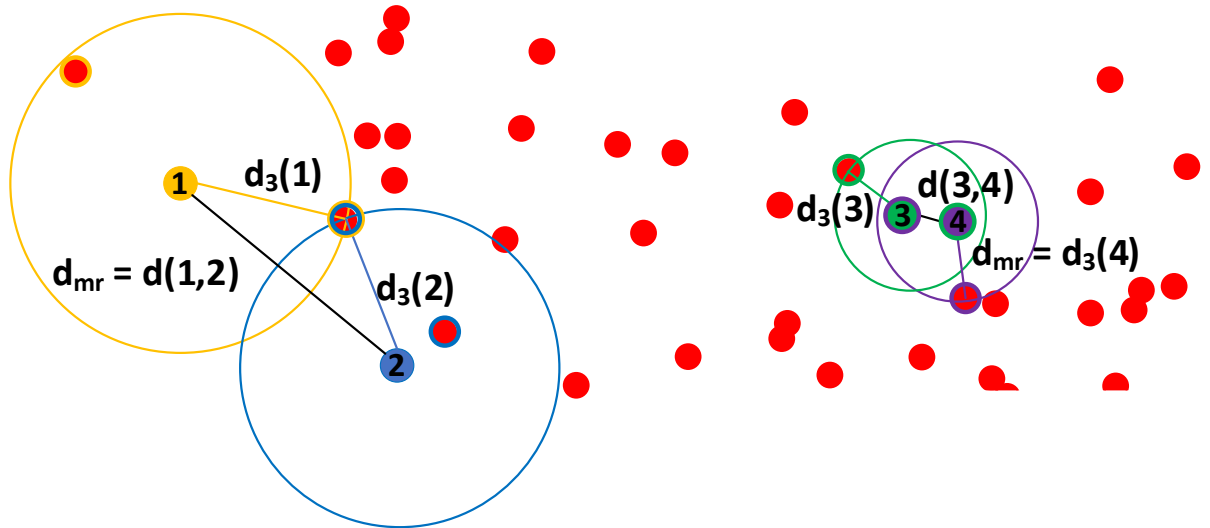


Рисунок 4.7 – Расстояние взаимной достижимости

На начальном этапе алгоритма *HDBSCAN** выполняется определение основных расстояний для всех точек. Далее, используя полученные значения, рассчитываются расстояния взаимной достижимости между каждой парой объектов. Эти расстояния образуют полный взвешенный неориентированный граф взаимной достижимости объектов $G_{N_{min}}$. Для данного графа выполняется поиск минимального остовного дерева, то есть определяется связный ациклический подграф исходного графа, содержащий все его вершины и имеющий минимальную сумму весов входящих в него ребер.

На рисунке 4.8 представлен результат построения минимального остовного дерева графа взаимной достижимости для тестового набора данных, состоящего из 60 объектов. Ребра графа, имеющие больший вес и соответствующие большим значениям расстояния взаимной достижимости, имеют зеленый оттенок, меньший вес – синий. Сами объекты обозначены красными точками.

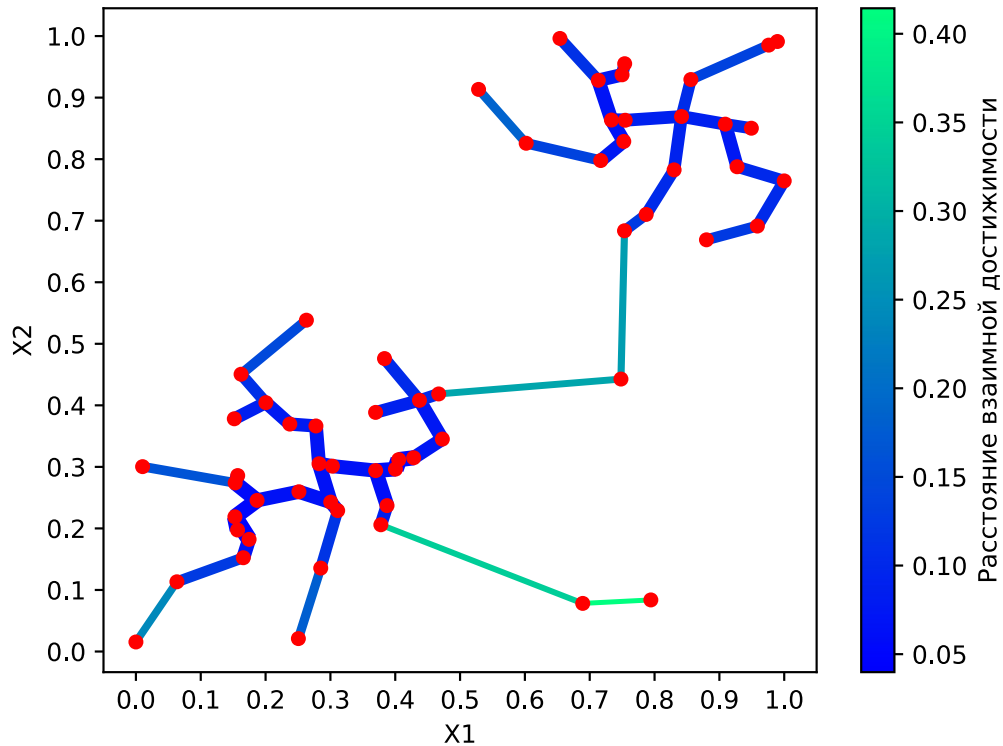


Рисунок 4.8 – Минимальное остовное дерево графа взаимной достижимости

Полученное минимальное остовное дерево расширяется с помощью добавления для каждой вершины дополнительного ребра, являющегося петлей, вес которого равен основному расстоянию $d_{N_{min}}$ для данной вершины. В результате, формируется расширенное минимальное остовное дерево MST_{ext} .

На следующем этапе, используя расширенное минимальное остовное дерево, выполняется построение иерархии различных реализаций кластеризации в виде дендрограммы. Дендрограмма представляет собой матрицу M_c размерностью $N \times L$ и вектор E размерностью $L \times 1$, где N – количество объектов, L – количество реализаций кластеризации. Каждый столбец матрицы M_c является вектором принадлежности объектов к кластерам I_c (принадлежность – номер кластера, выброс – ноль). Каждый элемент вектора E соответствует величине максимального радиуса соседства ε для данной реализации разделения объектов по кластерам.

Суть построения иерархии заключается в последовательном удалении из расширенного остовного дерева графа взаимной достижимости ребер с наибольшим весом. Процесс начинается с помещения всех объектов в единый

кластер. После удаления каждого ребра его вес фиксируется в векторе E , затем выделяются компоненты связности, определяющих принадлежность объектов к кластерам и выбросом. Полученная в результате удаления ребра реализация кластеризации фиксируется в матрице M_c . Расширение минимального остовного дерева за счет ребер, соответствующих основному расстоянию, позволяет фиксировать переход объекта, формирующего кластер из одного объекта, в разряд выбросов [304, 305].

Все компоненты связности, количество вершин в которых меньше, чем максимальное количество объектов в кластере N_c , считаются «ложными» кластерами. Вершины этих компонент отмечаются как выбросы. Если все компоненты графа данного кластера оказываются «ложными», считается, что кластер «исчез» при удалении из графа MST_{ext} ребра с весом ε .

Если только у одной из компонент связности кластера количество вершин больше или равно N_c , считается, что кластер «сократился» при удалении из графа MST_{ext} ребра с весом ε .

Если у двух и более компонент связности кластера количество вершин больше или равно N_c , считается, что кластер «разделился». Каждой из компонент присваивается новый номер кластера. Считается, что новые кластеры «возникли» при удалении из графа MST_{ext} ребра с весом ε .

После того, как из графа MST_{ext} удаляется последнее ребро, все его объекты в последней реализации кластеризации $DBSCAN^*$ оказываются помечены как выбросы. Итоговая принадлежность объектов к кластерам определяется по результатам выбора наиболее стабильных кластеров в иерархии. Мера стабильности кластеров определяется как величина λ , которая представляет собой обратную величину от веса ребра ε в расширенном минимальном остовном дереве, то есть $\lambda = 1/\varepsilon$. Для каждого возможного кластера C по результатам построения иерархии мы можем определить величину λ_b , соответствующую весу ребра ε , после удаления которого произошло «возникновение» кластера. Аналогично можно определить величину λ_d , соответствующую «исчезновению» кластера. Для

каждой точки, входящей в кластер C , также может быть определено значение λ_p , которое соответствует весу ребра ε , после удаления которого точка выходит из состава кластера (переходит в другой кластер или становится выбросом). Величина λ_p находится в диапазоне значений от λ_b до λ_d . Используя данные характеристики кластера и входящих в него объектов, стабильность можно определить как [305, 309]:

$$S_C = \sum_{p \in C} (\lambda_p - \lambda_b), \quad (4.4)$$

где S – стабильность кластера C , λ_p – обратная величина от веса ребра ε , после удаления которого объект p покинул кластер, λ_b – обратная величина от веса ребра ε , после удаления которого возник кластер C .

Определение величины стабильности для каждого кластера выполняется при обратном движении по иерархии. Сначала определяется перечень кластеров, для которых не происходило «разделение», то есть терминальных в иерархии. Далее выполняется определение величины стабильности для терминальных кластеров, данные кластеры считаются текущими выбранными. После происходит обработка более высоких в иерархии кластеров, являющихся родительскими для терминальных кластеров. Если сумма стабильности дочерних кластеров оказывается больше, чем стабильность родительского кластера, то стабильность данного кластера считается равной сумме стабильностей дочерних кластеров. Если, с другой стороны, стабильность кластера больше, чем сумма стабильности его дочерних кластеров, то данный кластер объявляется выбранным и происходит отмена выбора всех его дочерних кластеров. Процесс завершается при достижении корневого узла, когда все объекты расположены в одном кластере. Итоговая реализация кластеризации выбирается исходя из выбранных в данном процессе кластеров и рассчитанных для них значений стабильности кластеров [308, 309].

На рисунке 4.9 представлена иерархия различных реализаций кластеризации в виде дендрограммы для того же самого тестового набора данных из 60 объектов. Количество объектов в кластере обозначается цветом: большее количество

соответствует оттенкам желтого и зеленого, меньшее – оттенкам синего. На начальном этапе все объекты помещаются в один кластер (желтая область). При увеличении значения λ (и соответственно уменьшении величины ε -окрестности) происходит «возникновение», «сокращение», а также «исчезновение» кластеров. Сужение закрашенных областей соответствует уменьшению размеров кластеров, когда все большее количество наблюдений идентифицируются как выбросы. Красным цветом выделены выбранные кластеры с наибольшей стабильностью.

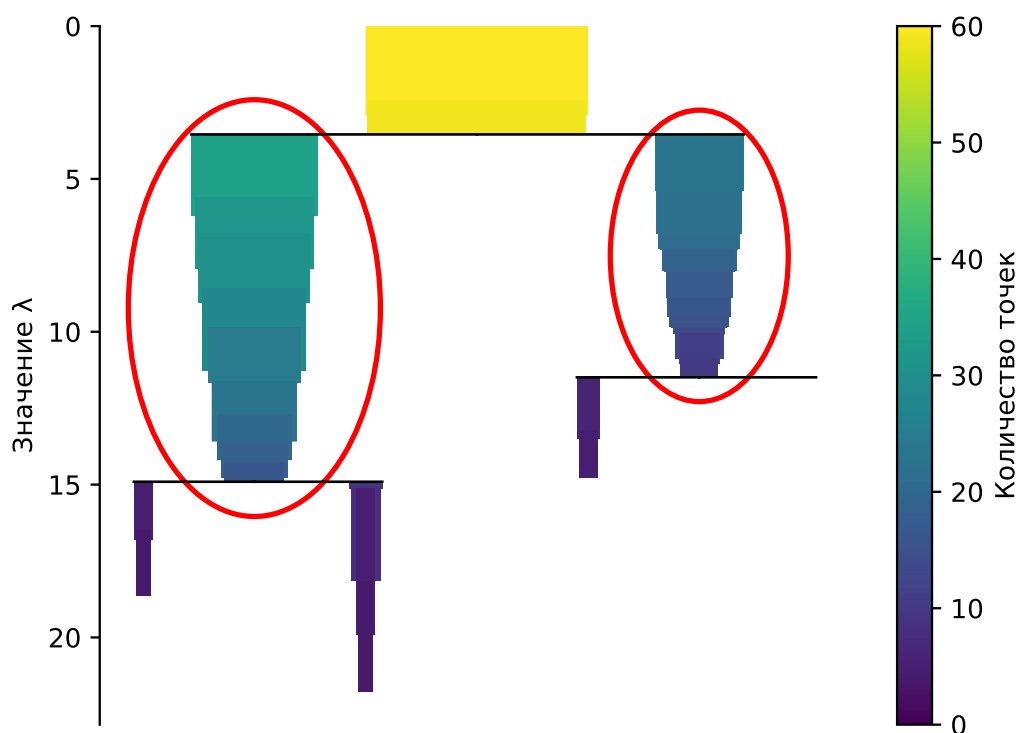


Рисунок 4.9 – Дендрограмма различных реализаций кластеризации

Полученная в результате выбора наиболее стабильных кластеров реализация кластеризации *DBSCAN** в том числе содержит результаты идентификации выбросов, но лишь для конкретного значения ε из вектора E . Более совершенным способом идентификации выбросов по полученной иерархии различных реализаций кластеризации является алгоритм глобально-локальной оценки выбросов *GLOSH*.

Основная идея глобально-локальной оценки выбросов заключается в определении некоторой величины (называемой оценкой *GLOSH*) для каждого

объекта в рассматриваемом наборе данных путем сравнения плотности точек вблизи данного объекта с наибольшим значением плотности точек для ближайшего к объекту кластера в иерархии. В зависимости от конкретной структуры иерархии и размещения объектов в кластерах, оценка *GLOSH* позволяет находить выбросы как в локальном, так и в глобальном масштабе [306, 307].

Для определения величины *GLOSH* при построении иерархий различных реализаций кластеризации по уже сформированной матрице M_c и вектору E при выполнении анализа иерархии в порядке ее построения для каждого объекта $\bar{x}_n$ из матрицы признаков X должны быть определены следующие величины [305].

- Индекс кластера C_l к которому принадлежал объект в реализации кластеризации, предшествующей идентификации объекта $\bar{x}_n$ как выброса.
- Максимальный радиус соседства $\varepsilon(\bar{x}_n)$, являющийся элементом вектора E , соответствующий реализации кластеризации, предшествующей идентификации объекта $\bar{x}_n$ как выброса.
- Максимальный радиуса соседства $\varepsilon_{max}(\bar{x}_n)$, являющийся элементом вектора E , соответствующий реализации кластеризации, предшествующей «исчезновению» кластера с индексом C_l .

Оценка *GLOSH* рассчитывается как [305]:

$$GLOSH(\bar{x}_n) = 1 - \frac{\varepsilon_{max}(\bar{x}_n)}{\varepsilon(\bar{x}_n)}, \quad (4.5)$$

где $\bar{x}_n$ – рассматриваемый объект; $\varepsilon(\bar{x}_n)$ – максимальный радиус соседства перед идентификацией $\bar{x}_n$ как выброса; $\varepsilon_{max}(\bar{x}_n)$ – максимальный радиус соседства перед «исчезновением» ближайшего к $\bar{x}_n$ кластера.

Далее представлено описание алгоритма глобально-локальной оценки выбросов (*GLOSH*) по иерархической основанной на плотности пространственной кластеризации для приложений с шумами (*HDBSCAN**) [304, 305], соответствующее реализации данного алгоритма в библиотеке *hdbscan* для языка программирования *Python* [310, 311].

Алгоритм 13: Алгоритм глобально-локальной оценки выбросов (GLOSH) по иерархической основанной на плотности пространственной кластеризации для приложений с шумами (HDBSCAN*)

Вход:

X – матрица признаков для кластеризации,

N_c – максимальное количество объектов в кластере,

N_{min} – минимальное количество точек для идентификации точки как основной,

T_{GL} – пороговое значение для глобально-локальной оценки выбросов в виде квантиля,

d – способ определения расстояния между точками.

1. Формирование вектора принадлежности объектов к выбросам I_o , размер которого совпадает с количеством строк в матрице признаков X и равен N .
2. Для всех элементов n матрицы признаков X от 1 до N .
3. | Определение для объекта n основного расстояния $d_{N_{min}}(\bar{x}_n)$.
4. Построение полного графа взаимной достижимости объектов $G_{N_{min}}$.
Вершины графа представляют собой объекты $\bar{x}_n$ из матрицы признаков X , а ребрами графа являются расстояния взаимной достижимости, рассчитываемыми между каждой парой объектов i и j :
$$d_{mr}(\bar{x}_i, \bar{x}_j) = \max(d_{N_{min}}(\bar{x}_i), d_{N_{min}}(\bar{x}_j), d(\bar{x}_i, \bar{x}_j)).$$
5. Определение минимального остовного дерева MST взвешенного неориентированного графа $G_{N_{min}}$ согласно алгоритму Прима [312].
6. Получение расширенного минимального остовного дерева MST_{ext} .
7. Получение иерархии различных реализаций кластеризации в виде дендрограммы: матрицы M_c и вектора E .
8. | Формирование исходного вектора принадлежности объектов к кластерам I_c , все объекты помещаются в один кластер. Добавление в матрицу M_c текущего вектора I_c .
9. | До тех пор, пока граф MST_{ext} имеет ребра.
10. | | Добавление величины веса ребра, имеющего наибольшее значение, в вектор весов E .
11. | | Удаление из графа MST_{ext} ребра, имеющего наибольший вес (при одинаковом весе могут быть удалены несколько ребер сразу).
12. | | Формирование текущей реализации вектора принадлежности объектов к кластерам I_c .
13. | | | Определение кластеров, в который входили вершины, которые были соединены удаленным ребром (один или два кластера).
14. | | | Индексы оставшихся кластеров остаются без изменений и вносятся в I_c .

- | | |
|-----|---|
| 15. | Для каждого из кластеров, содержащих вершину удаленного ребра. |
| 16. | Выделение компонент связности в подграфе, соответствующем данному кластеру. Для каждой компоненты связности рассчитывается количество вершин N . |
| 17. | Если у компоненты $N < N_c$, ее вершины отмечаются в I_c как выбросы. |
| 18. | Если у одной из компонент $N \geq N_c$, ее вершины вносятся в I_c с исходным номером кластера. |
| 19. | Если у двух и более компонент $N \geq N_c$, каждой из компонент присваивается новый номер кластера. Соответствующие вершины вносятся в I_c с новыми номерами кластера. |
| 20. | Выполнение для каждого элемента n матрицы признаков X глобально-локальной оценки того, что элемент может являться выбросом. |
| 21. | Для каждого объекта $\bar{x}_n$ из матрицы признаков X . |
| 22. | Определение индекса кластера C_l перед идентификацией объекта $\bar{x}_n$ как выброса. |
| 23. | Определение $\varepsilon(\bar{x}_n)$ перед идентификацией объекта $\bar{x}_n$ как выброса. |
| 24. | Определение $\varepsilon_{max}(\bar{x}_n)$ перед «исчезновением» кластера с индексом C_l . |
| 25. | Расчёт оценки $GLOSH(\bar{x}_n) = 1 - \frac{\varepsilon_{max}(\bar{x}_n)}{\varepsilon(\bar{x}_n)}$. |
| 26. | Определение величины порогового значения оценки $GLOSH_T$ в соответствии с заданным квантилем T_{GL} . |
| 27. | Для каждого объекта $\bar{x}_n$ из матрицы признаков X . |
| 28. | Если $GLOSH(\bar{x}_n) > GLOSH_T$, то объект $\bar{x}_n$ идентифицируется как выброс, в вектор I_o для данного наблюдения записывается 0. В противном случае записывается 1. |

Выход:

I_o – вектор индексов, определяющий принадлежность объектов к выбросам.

4.3. Методики учета состава работающих ветроустановок

В опубликованных исследованиях [131–137] на тему прогнозирования выработки ВЭС вопрос учета состояния турбин не освещен в должной степени. Вероятнее всего, наблюдения, соответствующие отключенному состоянию установки, исключаются из обучающих, а возможно и тестовых выборок, на этапе предварительной обработки данных при анализе выбросов с использованием характеристики мощности.

Значительное количество точек, являющихся выбросами, соответствует типу 1, то есть областям с близкими к нулю значениями мощности, возникающими вдоль всей оси скорости ветра. Эти точки могут быть связаны с отключенным состоянием ВЭУ для проведения обслуживания, в результате отказа или неправильной работой системы управления турбиной. Выявление таких выбросов возможно только при анализе характеристик мощности на уровне отдельных турбин. На уровне ВЭС или групп турбин точка с неполным составом включенных турбин будет соответствовать сниженной мощности для некоторого значения скорости ветра, то есть может оказаться как среди выбросов типа 2, 4 и 5, так и не идентифицирована как выброс вообще. В результате, такое наблюдение может оказаться включенным в обучающую выборку, что приведет к искажению регрессионной модели из-за несоответствия скорости ветра и доступной мощности турбины для данной точки.

Тем не менее, в одном из исследований [313] рассмотрен подход, заключающийся в учете состояния турбин в качестве признаков при обучении моделей (методика *A*). Этот подход будет использоваться для сравнения с разработанной методикой учета состава включенных установок.

В данной работе предлагается новая методика учета состава включенных установок (методика *RT*). Для прогноза на уровне групп турбин и на уровне ВЭС предлагается выполнять трансформацию откликов с помощью пересчета фактической измеренной мощности турбин в возможную доступную мощность, исходя из количества включенных установок. Определение доступной мощности при этом осуществляется по соотношению:

$$P_{av} = P \cdot \frac{K}{K_{av}}, \quad (4.6)$$

где P_{av} – доступная мощность, P – измеренная мощность, K – количество турбин в группы или ВЭС, K_{av} – количество включенных турбин.

Трансформация отклика по данному соотношению выполняется для обучающей выборки. Для корректировки прогнозных значений используется

обратное соотношение, получаемая прогнозная доступная мощность снижается для получения прогнозной фактической мощности группы турбин или ВЭС:

$$P_{pred} = P_{av_pred} \cdot \frac{K_{av_pred}}{K}, \quad (4.7)$$

где P_{av_pred} – прогнозная доступная мощность, P_{pred} – прогнозная мощность, K – количество турбин в группы или ВЭС, K_{av_pred} – планируемое количество включенных турбин.

Для прогноза на уровне отдельных турбин учет состава работающего оборудования в обучающей выборке, как и в работах других исследователей, выполняется за счет исключения данных точек как выбросов. При этом прогнозное значение мощности турбины приравнивается нулю в случае, если прогнозное состояние турбины – отключенное.

Использование предложенного подхода требует информации о состоянии установок, которая часто оказывается недоступной. При отсутствии этих данных можно воспользоваться эмпирическими выражениям для определения состояния установки исходя из измеренного значения скорости ветра для каждой турбины и измеренной мощности:

$$\begin{cases} V > (1 + \Delta V) \cdot V_{\min} \\ P < \Delta_o \cdot P_{\text{ном}} \cdot P_{\text{порог}} \cdot \sqrt{3\Delta I^2 + \Delta U^2 + \Delta P^2} \end{cases} \quad (4.8)$$

где V – измеренная скорость ветра; ΔV – погрешность устройств измерения скорости ветра (принимается равным 0,02); $V_{\min}$ – скорость ветра для запуска турбины; P – измеренная мощность; Δ_o – поправочный коэффициент для неучтенных погрешностей (принимается равным 1,1); $P_{\text{ном}}$ – номинальная мощность турбины; $P_{\text{порог}}$ – пороговая мощность турбины для определения отключенного состояния в о.е. (принимается равной 0,05); ΔI – погрешность трансформатора тока (принимается равной 0,05, с учетом возможного начального намагничивания при низкой загрузке увеличивается в три раза); ΔU – погрешность трансформатора напряжения (принимается равной 0,05); ΔP – погрешность измерений мощности (принимается равной 0,05).

4.4.Вычислительные эксперименты

При проведении вычислительных экспериментов для проверки предложенных методик идентификации выбросов и учета состава работающих установок рассматривался объект исследования и набор исходных данных, описанный в п 2.4.

В качестве регрессионной модели использовался градиентный бустинг над деревьями решений *GBDT*. При обучении каждой модели применялась байесовская оптимизация гиперпараметров с использованием функции выбора *LCB* с количеством итераций 60. При выполнении оптимизации гиперпараметров и кроссвалидации использовалась методика, аналогичная описанной в п.2.4 и представленной на рисунке 2.15. В качестве критерия для определения наиболее эффективного подхода использовался фактор улучшения *SS* по отношению к прогнозу по ВЭС без использования идентификации выбросов.

В рамках исследования эффективности предложенных методик идентификации выбросов на характеристиках мощности отдельных ВЭУ, групп ВЭУ и всей ВЭС выполнена серия из 12 вычислительных экспериментов. В каждом вычислительном эксперименте рассматривалась комбинация способа агрегации мощности отдельных ветроустановок при прогнозировании генерации ВЭС и способ идентификации выбросов. Для каждого из способов, описанных в п.4.2, проведено по три эксперимента для трех способов агрегации мощности ВЭУ.

Таким образом, были выполнены следующие ВЭ:

- *P* – прогноз на уровне ВЭС, идентификация выбросов не выполняется (ВЭ проведен и описан в п.3.4, результаты приводятся для сопоставления);
- *T* – прогноз на уровне отдельных ВЭУ без идентификации выбросов (ВЭ проведен и описан в п.3.4, результаты приводятся для сопоставления);
- *GC* – объединение ВЭУ в 6 групп по временным рядам измерений мощности каждой турбины методом иерархической агломеративной кластеризации, для измерения расстояния между временными рядами используется алгоритм динамической трансформации временной шкалы, идентификация выбросов

не выполняется (ВЭ проведен и описан в п.3.4 под обозначением *H-PD*, результаты приводятся для сопоставления);

- *P-QM* – прогноз на уровне ВЭС, для идентификации выбросов на характеристике мощности ВЭС используется метод межквартильного размаха (1 ВЭ, получена 1 регрессионная модель, в рамках оптимизации гиперпараметров обучено 60 моделей);
- *T-QM* – прогноз на уровне отдельных турбин, для идентификации выбросов на характеристике мощности отдельных ВЭУ используется метод межквартильного размаха (1 ВЭ, получено 26 регрессионных моделей, в рамках оптимизации гиперпараметров обучено 1560 моделей);
- *GC-QM* – объединение ВЭУ в 6 групп по временным рядам измерений мощности ВЭУ методом иерархической агломеративной кластеризации, для измерения расстояния между временными рядами используется алгоритм динамической трансформации временной шкалы, для идентификации выбросов на характеристике мощности групп ВЭУ используется метод межквартильного размаха (1 ВЭ, получено 6 регрессионных моделей, в рамках оптимизации гиперпараметров обучено 360 моделей);
- *P-DB* – прогноз на уровне ВЭС, для идентификации выбросов на характеристике мощности ВЭС используется алгоритм основанной на плотности пространственной кластеризации для приложений с шумами *DBSCAN* с автоматическим определением максимального радиуса соседства с помощью алгоритма *KNeedle* (1 ВЭ, получена 1 регрессионная модель, в рамках оптимизации гиперпараметров обучено 60 моделей);
- *T-DB* – прогноз на уровне отдельных турбин, для идентификации выбросов на характеристике мощности отдельных ВЭУ используется алгоритм основанной на плотности пространственной кластеризации для приложений с шумами *DBSCAN* с автоматическим определением максимального радиуса соседства с помощью алгоритма *KNeedle* (1 ВЭ, получено 26 регрессионных моделей, в рамках оптимизации гиперпараметров обучено 1560 моделей);

- *GC-DB* – объединение ВЭУ в 6 групп по временным рядам измерений мощности ВЭУ методом иерархической агломеративной кластеризации, для измерения расстояния между временными рядами используется алгоритм динамической трансформации временной шкалы, для идентификации выбросов на характеристике мощности групп ВЭУ используется алгоритм основанной на плотности пространственной кластеризации для приложений с шумами *DBSCAN* с автоматическим определением максимального радиуса соседства с помощью алгоритма *KNeedle* (1 ВЭ, получено 6 регрессионных моделей, в рамках оптимизации гиперпараметров обучено 360 моделей);
- *P-GL* – прогноз на уровне ВЭС, для идентификации выбросов на характеристике мощности ВЭС используется алгоритм глобально-локальной оценки выбросов *GLOSH* по кластеризации *HDBSCAN** (1 ВЭ, получена 1 регрессионная модель, в рамках оптимизации гиперпараметров обучено 60 моделей);
- *T-GL* – прогноз на уровне отдельных турбин, для идентификации выбросов на характеристике мощности отдельных ВЭУ используется алгоритм глобально-локальной оценки выбросов *GLOSH* по иерархической кластеризации *HDBSCAN** (1 ВЭ, получено 26 регрессионных моделей, в рамках оптимизации гиперпараметров обучено 1560 моделей);
- *GC-GL* – объединение ВЭУ в 6 групп по временным рядам измерений мощности ВЭУ методом иерархической агломеративной кластеризации, для измерения расстояния между временными рядами используется алгоритм динамической трансформации временной шкалы, для идентификации выбросов на характеристике мощности групп ВЭУ используется алгоритм глобально-локальной оценки выбросов *GLOSH* по иерархической кластеризации *HDBSCAN** (1 ВЭ, получено 6 регрессионных моделей, в рамках оптимизации гиперпараметров обучено 360 моделей).

Результаты оценки погрешности и надежности прогнозов для рассматриваемых подходов к поиску и исключению выбросов с использованием

характеристики мощности представлены в таблице 4.2. Время обучения показано с учетом подбора гиперпараметров. Для одновременного обучения нескольких моделей в одном ВЭ использовались параллельные вычисления в 12 потоках.

Таблица 4.2 – Результаты оценки погрешности и надежности прогнозов для рассматриваемых подходов к идентификации выбросов по характеристике мощности

Эксперимент	<i>SS</i> , %	<i>NRMSE</i> , %	<i>NMAE</i> , %	<i>NBias</i> , %	<i>NSAE</i> , %	<i>EICP_20</i> , %	Время обучения, с
<i>P</i>	0,0	13,31	9,4	-0,4	26,3	86,5	63,6
<i>T</i>	4,3	12,74	9,3	-0,4	26,0	87,9	757,7
<i>GC</i>	8,7	12,16	9,0	-0,5	25,1	89,5	203,9
<i>P-QM</i>	-0,3	13,35	9,8	-0,1	27,3	87,5	60,5
<i>T-QM</i>	4,8	12,67	9,3	-0,2	26,0	88,5	714,5
<i>GC-QM</i>	8,7	12,15	8,9	-0,1	24,9	90,3	175,6
<i>P-DB</i>	5,4	12,60	9,2	-0,5	25,6	89,0	55,9
<i>T-DB</i>	9,6	12,04	8,9	-0,4	25,0	91,2	732,4
<i>GC-DB</i>	13,7	11,49	8,4	-0,3	23,5	92,0	190,4
<i>P-GL</i>	7,9	12,26	9,0	-0,4	25,2	89,7	57,4
<i>T-GL</i>	10,7	11,88	8,8	-0,5	24,6	91,1	724,9
<i>GC-GL</i>	15,9	11,20	8,4	-0,5	23,5	92,7	172,0

Результаты, представленные в таблице 4.2, показывают, что методика идентификации выбросов, основанная на методе межквартильного размаха *QM*, обеспечивает снижение погрешности прогноза только при прогнозировании на уровне отдельных турбин. Для прогнозов на уровне ВЭС и групп турбин данный способ приводит к росту погрешности прогноза. Можно предположить, что методика *QM* ошибочно идентифицирует некоторые значимые точки как выбросы.

Методика идентификации выбросов *DB*, основанная на алгоритме *DBSCAN*, обеспечила снижение погрешности при выполнении прогнозов как на уровне ВЭС, так и уровне отдельных ВЭУ и групп ВЭУ. Аналогично, использование методики идентификации выбросов *GL*, позволило снизить погрешность при прогнозировании на всех трех уровнях. При этом, методика глобально-локальной оценки выбросов *GL* обеспечила больший фактор улучшения, чем методика *DB* для всех трех способов агрегации ВЭУ при прогнозировании генерации ВЭС: 7,9% для прогноза на уровне ВЭС, 7,2% для прогноза на уровне отдельных ВЭУ, 8,5% для прогноза на уровне групп ВЭУ.

Наибольший фактор улучшения среди проведенных экспериментов был достигнут при использовании предложенной в данной работе методики *GL* при объединении ВЭУ в группы с помощью предложенной в данной работе методики *H-PD* – обеспечено снижение среднеквадратичной ошибки прогноза на 15,9 % по отношению к прогнозу по ВЭС без идентификации выбросов.

Для всех рассматриваемых методик, в результате исключения выбросов было обеспечено уменьшение объема обучающей выборки, за счет чего было достигнуто снижение суммарного времени обучения регрессионных моделей.

На рисунке 4.10 представлено сравнение результатов идентификации выбросов на характеристике мощности ВЭС для методик *QM*, *DB* и *GL*.

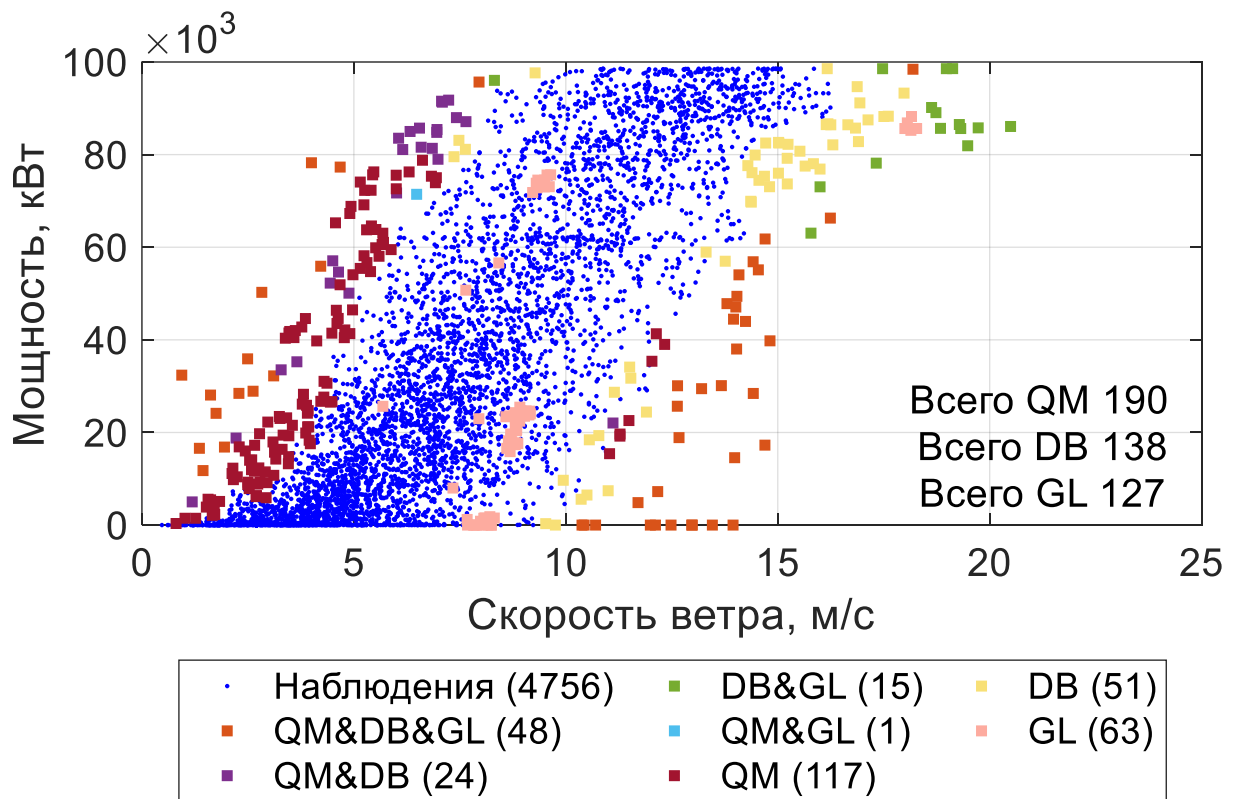


Рисунок 4.10 – Результаты идентификации выбросов: *QM* – метод межквартильного диапазона; *DB* – основанная на плотности пространственная кластеризация для приложений с шумами (*DBSCAN*); *GL* – глобально-локальная оценка выбросов (*GLOSH*) по иерархической основанной на плотности пространственной кластеризации для приложений с шумами (*HDBSCAN*); значок «&» – пересечение множеств.

Общее количество наблюдений, для которых выполнялась идентификация выбросов, равнялось 5075. Наибольшее количество выбросов было идентифицировано методом квартильного размаха QM , общее количество для данной методики составило 190. Методики идентификации DB и GL определили примерно одинаковое количество выбросов, по 138 и 127 соответственно. На рисунке также показаны выбросы, выявленные сразу несколькими методами. Как видно из рисунка, методика QM выявляет выбросы, относящиеся к типам 1 и 3. Методики DB и GL обеспечивают идентификацию выбросов типа 1, 2, 3 и 4. При этом методика GL также является единственной из рассмотренных, обеспечивающей идентификацию выбросов типа 5.

В рамках исследования эффективности предложенной методики учета состава включенных ВЭУ при прогнозировании на уровне отдельных ВЭУ, групп ВЭУ и всей ВЭС также выполнена серия из 9 вычислительных экспериментов. В каждом вычислительном эксперименте рассматривается комбинация способа агрегации мощности отдельных ветроустановок при прогнозировании генерации ВЭС и способ учета состава включенных установок.

Таким образом, были проведены следующие вычислительные эксперименты:

- P – прогноз на уровне ВЭС, учет состава включенных ВЭУ не выполняется (эксперимент проведен и описан в п.3.4, результаты приводятся повторно для сопоставления);
- T – прогноз на уровне отдельных турбин, учет состава включенных ВЭУ не выполняется (эксперимент проведен и описан в п.3.4, результаты приводятся повторно для сопоставления);
- GC – объединение ВЭУ в 6 групп по временным рядам измерений мощности каждой турбины методом иерархической агломеративной кластеризации, для измерения расстояния между временными рядами используется алгоритм динамической трансформации временной шкалы, учет состава включенных ВЭУ не выполняется (эксперимент проведен и описан в п.3.4 под обозначением $H-PD$, результаты приводятся повторно для сопоставления);

- *P-A* – прогноз на уровне ВЭС, для учета состава включенных ВЭУ при обучении регрессионной модели использовался дополнительный признак (1 ВЭ, получена 1 регрессионная модель, в рамках оптимизации гиперпараметров обучено 60 моделей);
- *T-A* – прогноз на уровне отдельных турбин, для учета состава включенных ВЭУ при обучении регрессионных моделей использовался дополнительный признак (1 ВЭ, получено 26 регрессионных моделей, в рамках оптимизации гиперпараметров обучено 1560 моделей);
- *GC-A* – объединение ВЭУ в 6 групп по временным рядам измерений мощности каждой турбины методом иерархической агломеративной кластеризации, для измерения расстояния между временными рядами используется алгоритм динамической трансформации временной шкалы, для учета состава включенных ВЭУ при обучении регрессионных моделей использовался дополнительный признак (1 ВЭ, получено 6 регрессионных моделей, в рамках оптимизации гиперпараметров обучено 360 моделей);
- *P-RT* – прогноз на уровне ВЭС, для учета состава включенных ВЭУ использовалась трансформация отклика (1 ВЭ, получена 1 регрессионная модель, в рамках оптимизации гиперпараметров обучено 60 моделей);
- *T-RT* – прогноз на уровне отдельных турбин, для учета состава включенных ВЭУ использовалась трансформация отклика (1 ВЭ, получено 26 регрессионных моделей, в рамках оптимизации гиперпараметров обучено 1560 моделей);
- *GC-RT* – объединение ВЭУ в 6 групп по временным рядам измерений мощности каждой турбины методом иерархической агломеративной кластеризации, для измерения расстояния между временными рядами используется алгоритм динамической трансформации временной шкалы, для учета состава включенных ВЭУ использовалась трансформация отклика (1 ВЭ, получено 6 регрессионных моделей, в рамках оптимизации гиперпараметров обучено 360 моделей).

Результаты оценки погрешности и надежности прогнозов для рассматриваемых подходов к учету состояния ветроустановок представлены в таблице 4.3.

Таблица 4.3 – Результаты оценки погрешности и надежности прогнозов для рассматриваемых подходов к учету состава работающих ветроустановок

Эксперимент	<i>SS</i> , %	<i>NRMSE</i> , %	<i>NMAE</i> , %	<i>NBias</i> , %	<i>NSAE</i> , %	<i>EICP_20</i> , %	Время обучения, с
<i>P</i>	0,0	13,31	9,4	-0,4	26,3	86,5	63,6
<i>T</i>	4,3	12,74	9,3	-0,4	26,0	87,9	757,7
<i>GC</i>	8,7	12,16	9,0	-0,5	25,1	89,5	203,9
<i>P-A</i>	2,9	12,93	9,2	-0,7	25,7	88,1	69,9
<i>T-A</i>	6,0	12,51	9,0	-0,4	25,1	88,1	898,8
<i>GC-A</i>	9,6	12,03	8,6	-0,5	24,0	89,1	270,1
<i>P-RT</i>	4,0	12,77	9,2	-0,2	25,7	87,9	67,8
<i>T-RT</i>	9,8	12,00	8,8	-0,5	24,5	89,1	754,7
<i>GC-RT</i>	13,5	11,51	8,5	-0,2	23,9	90,5	186,0

По результатам проведенных ВЭ можно сделать вывод о том, что учет состояния ВЭУ позволяет снизить погрешность при выполнении прогнозов генерации как на уровне отдельных ВЭУ, так и групп ВЭУ и всей ВЭС. Предложенная методика трансформации откликов оказалась более эффективна, чем учет состава включенных ВЭУ в виде признака также при прогнозировании на уровне ВЭУ, групп ВЭУ и всей ВЭС.

Наибольший фактор улучшения среди проведенных экспериментов был достигнут при использовании предложенной в данной работе методики *RT* при объединении ВЭУ в группы с помощью предложенной методики *H-PD* – обеспечено снижение среднеквадратичной ошибки прогноза на 13,5 % по отношению к прогнозу по ВЭС без учета состояния ВЭУ.

4.5.Итоговые результаты исследований

В данном разделе сведены результаты исследований способов повышения эффективности краткосрочного прогнозирования генерации ВЭС, проведенных в предыдущих главах.

Далее используются следующие обозначения проведенных экспериментов:

- *P-RS* – прогноз на уровне ВЭС, для оптимизации гиперпараметров использовался метод случайного поиска, идентификация выбросов не выполнялась, учет состава включенных ВЭУ не выполнялся (эксперимент проведен и описан в п.2.4, в таблице 2.2 модель *GBDT*, СОГ – *RS*);
- *P-LCB* – прогноз на уровне ВЭС, для оптимизации гиперпараметров использовалась байесовская оптимизация с функцией выбора «нижний доверительный интервал», идентификация выбросов не выполнялась, учет состава включенных ВЭУ не выполнялся (эксперимент проведен и описан в п.2.4, в таблице 2.2 модель *GBDT*, СОГ – *LCB*);
- *GC-LCB* – прогноз на уровне 6 групп ВЭУ, объединение ВЭУ в группы по временным рядам измерений мощности каждой турбины методом иерархической агломеративной кластеризации, для измерения расстояния между точками временных рядов использовался алгоритм динамической трансформации временной шкалы, для оптимизации гиперпараметров использовалась байесовская оптимизация с функцией выбора «нижний доверительный интервал», идентификация выбросов не выполнялась, учет состава включенных ВЭУ не выполнялся (эксперимент проведен и описан в п.3.3, в таблице 3.3 методика *H-PD*);
- *GC-LCB-GL* – прогноз на уровне 6 групп ВЭУ, объединение ВЭУ в группы по временным рядам измерений мощности каждой турбины методом иерархической агломеративной кластеризации, для измерения расстояния временными рядами использовался алгоритм динамической трансформации временной шкалы, для оптимизации гиперпараметров использовалась байесовская оптимизация с функцией выбора «нижний доверительный интервал», идентификация выбросов выполнена с помощью алгоритма глобально-локальной оценки выбросов по иерархической основанной на плотности пространственной кластеризации для приложений с шумами, учет

состава включенных ВЭУ не выполнялся (эксперимент проведен и описан в п.4.4, в таблице 4.2 методика *GC-GL*);

- *GC-LCB-RT* – прогноз на уровне 6 групп ВЭУ, объединение ВЭУ в группы по временным рядам измерений мощности каждой турбины методом иерархической агломеративной кластеризации, для измерения расстояния между точками временных рядов использовался алгоритм динамической трансформации временной шкалы, для оптимизации гиперпараметров использовалась байесовская оптимизация с функцией выбора «нижний доверительный интервал», идентификация выбросов не выполнялась, для учета состава включенных ВЭУ использовалась трансформация отклика (эксперимент проведен и описан в п.4.4, в таблице 4.3 методика *GC-RT*);
- *GC-LCB-GL-RT* – прогноз на уровне 6 групп ВЭУ, объединение ВЭУ в группы по временным рядам измерений мощности каждой турбины методом иерархической агломеративной кластеризации, для измерения расстояния между точками временных рядов использовался алгоритм динамической трансформации временной шкалы, для оптимизации гиперпараметров использовалась байесовская оптимизация с функцией выбора «нижний доверительный интервал», идентификация выбросов выполнена с помощью алгоритма глобально-локальной оценки выбросов по иерархической основанной на плотности пространственной кластеризации для приложений с шумами, для учета состава включенных ВЭУ использовалась трансформация отклика (дополнительно проведенный ВЭ для оценки эффективности совместного применения всех предложенных методик).

Итоговые результаты исследований сведены в таблицу 4.5. На рисунке 4.11 представлен график сравнения факторов улучшения и диаграмма размаха для предложенных методик повышения эффективности прогнозирования генерации ВЭС.

Таблица 4.5 – Результаты исследования по повышению эффективности краткосрочного прогнозирования выработки ВЭС

Методика	<i>SS</i> , %	<i>NRMSE</i> , %	<i>NMAE</i> , %	<i>NBias</i> , %	<i>NSAE</i> , %	<i>EICP_20</i> , %	Время обучения, с
<i>P-RS</i>	0,0	14,18	10,2	1,1	28,6	85,0	13,5
<i>P-LCB</i>	6,1	13,31	9,4	-0,4	26,3	86,5	63,6
<i>GC-LCB</i>	14,2	12,16	9,0	-0,5	25,1	89,5	203,9
<i>GC-LCB-GL</i>	21,0	11,20	8,4	-0,5	23,5	92,7	222,0
<i>GC-LCB-RT</i>	18,8	11,51	8,5	-0,2	23,9	90,5	186,0
<i>GC-LCB-GL-RT</i>	26,1	10,47	7,8	-0,2	21,9	93,1	254,6

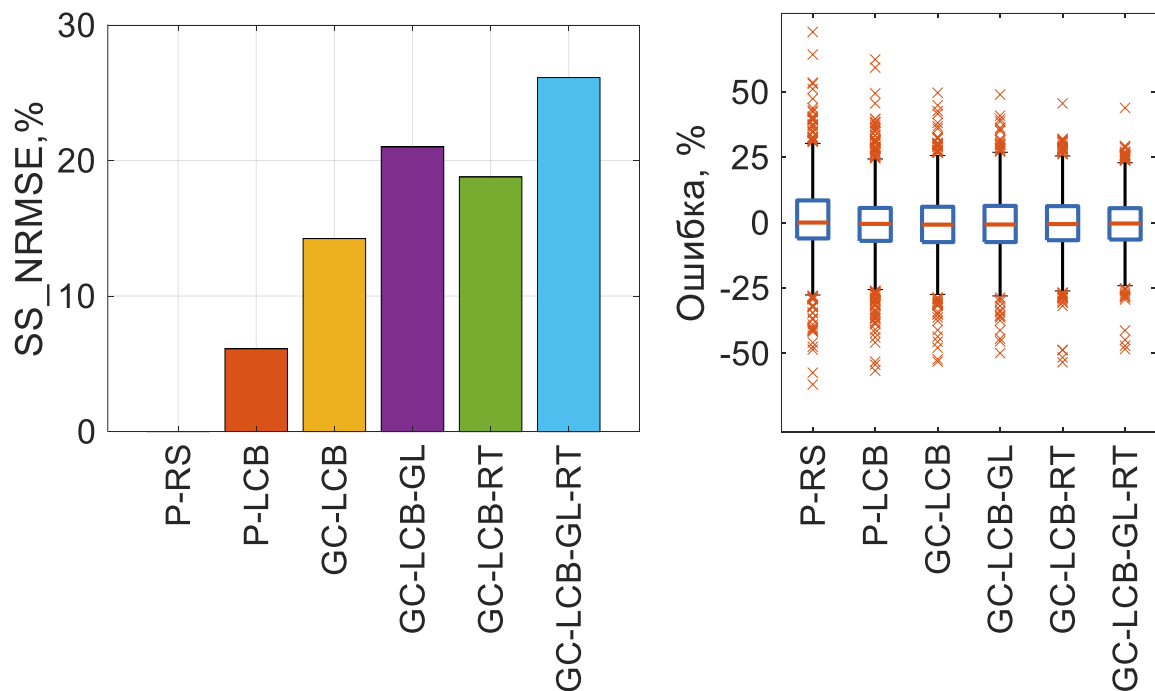


Рисунок 4.11 – Сравнение предложенных методик повышения эффективности прогнозирования генерации ВЭС (факторы улучшения – слева, диаграмма размаха – справа): *P* – прогноз на уровне ВЭС; *GC* – прогноз на уровне групп турбин, определенных иерархической кластеризацией по *DTW*-расстоянию; *RS* – оптимизация гиперпараметров методом случайного поиска; *LCB* – байесовская оптимизация гиперпараметров с функцией выбора «нижний доверительный интервал»; *GL* – глобально-локальная оценка выбросов; *RT* – учет состава работающих ВЭУ с помощью трансформации отклика.

На рисунке 4.12 представлены гистограммы ошибок прогнозов.

На рисунке 4.13 представлены графики разброса прогнозных значений генерации ВЭС относительно фактических.

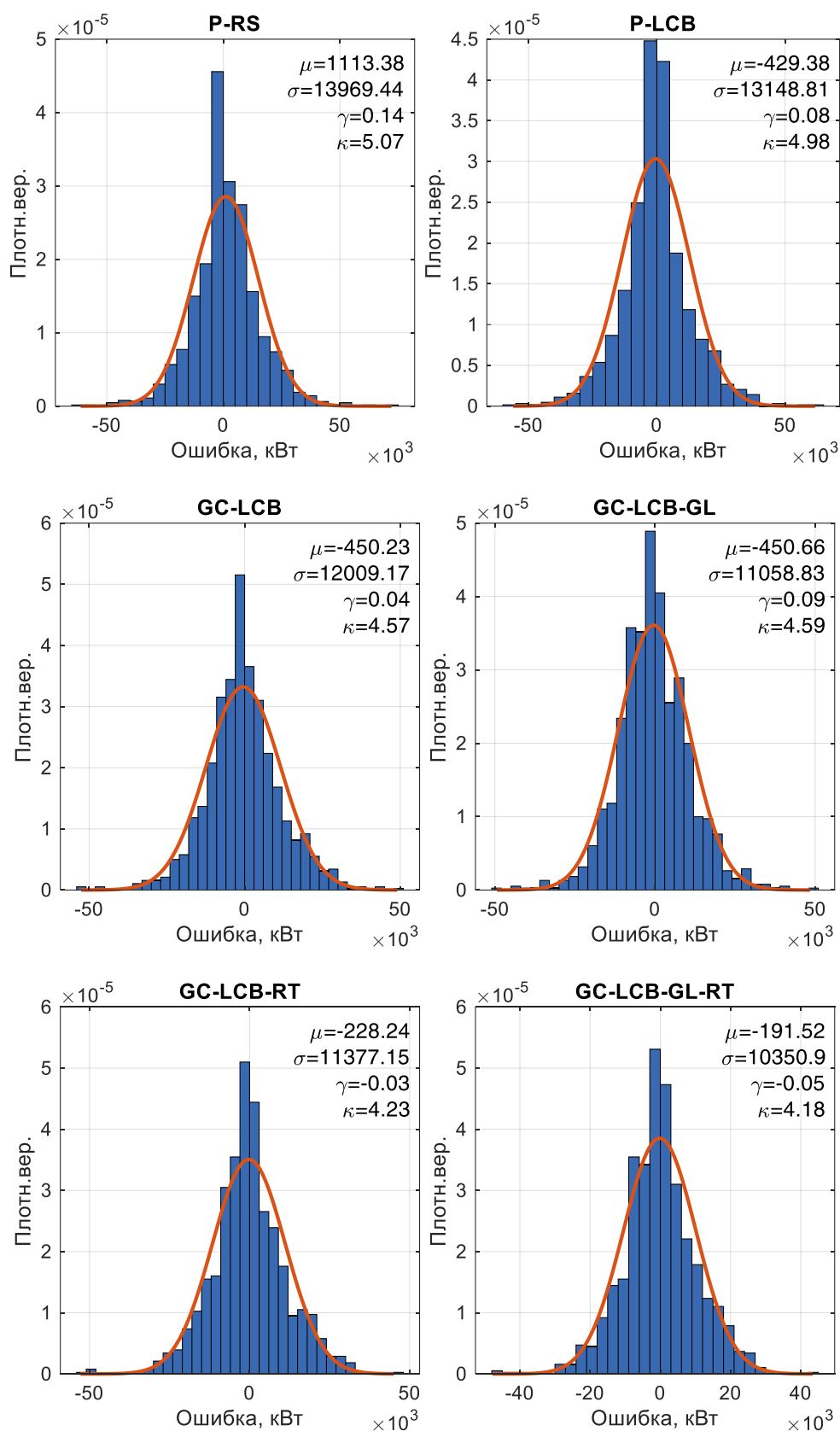


Рисунок 4.12 – Гистограммы ошибок прогноза

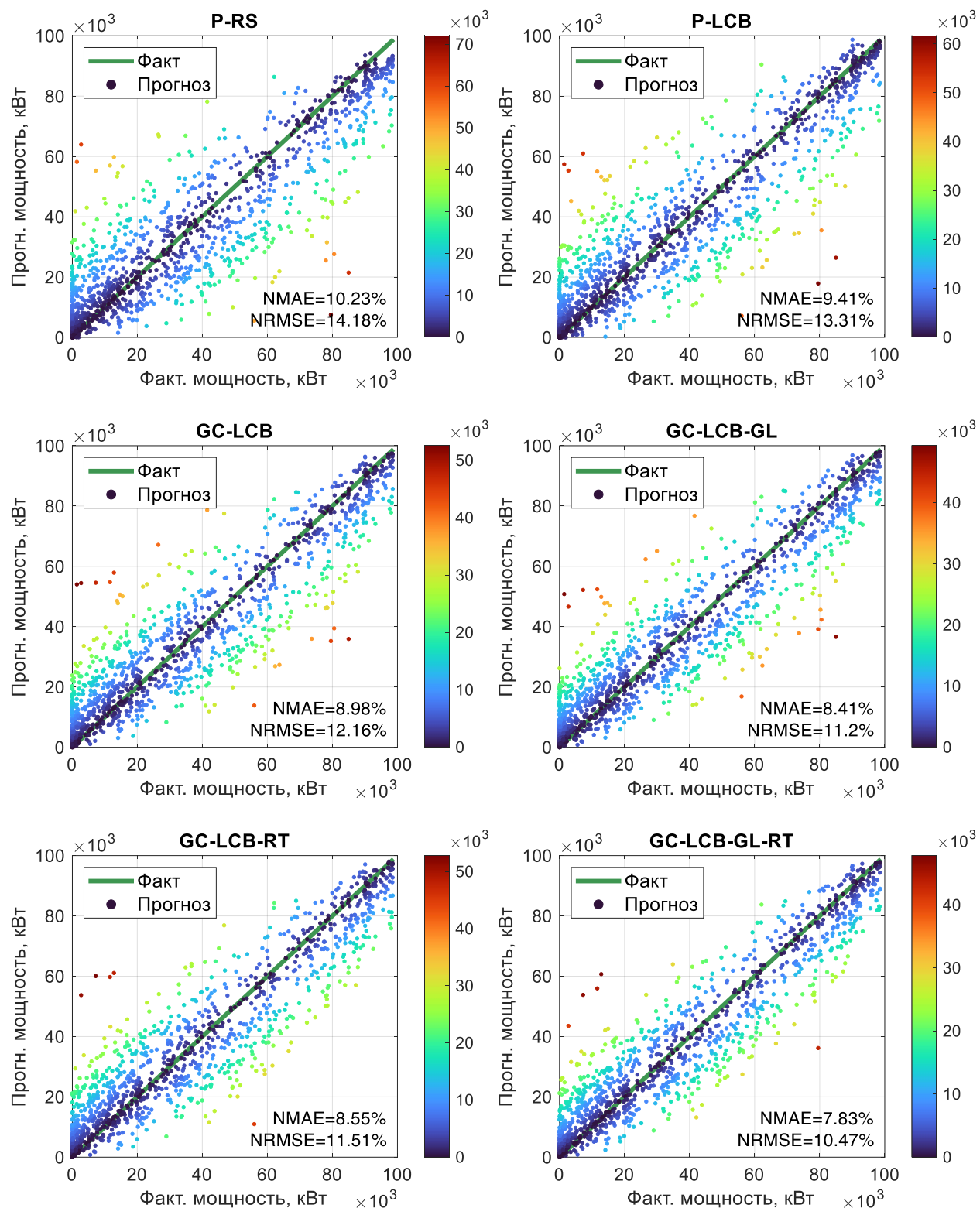


Рисунок 4.13 – Графики разброса прогнозных значений генерации ВЭС относительно фактических

4.6. Выводы по главе

1. Одним из способов повышения эффективности прогнозирования генерации ВЭС является идентификация и исключение выбросов из обучающей выборки, используемой при построении регрессионных моделей. Наиболее часто для задачи обнаружения выбросов используются характеристики мощности ВЭУ и ВЭС.

2. Как правило, выделяется 5 основных типов выбросов, каждый из которых соответствует определенной области на характеристике мощности. Идентификация этих выбросов может быть затруднительна. Все 5 типов выбросов можно отнести к локальным, поскольку все наблюдения находятся в диапазоне ожидаемых значений скорости ветра (от нуля до максимальной) и мощности (от нуля до номинальной).

3. Рассмотрена существующая методика идентификации выбросов с использованием метода межквартильного размаха, обеспечивающая поиск локальных выбросов за счет рассмотрения подвыборок наблюдений в узком диапазоне скоростей ветра.

4. Рассмотрена существующая методика идентификации выбросов, использующая алгоритм основанный на плотности пространственной кластеризации для приложений с шумами *DBSCAN*. Предложена модификация методики для автоматического определения максимального радиуса соседства с помощью алгоритма *KNeedle*.

5. Предложена новая методика идентификации и исключения выбросов с использованием алгоритма глобально-локальной оценки выбросов (*GLOSH*) по иерархической основанной на плотности пространственной кластеризации (*HDBSCAN**), которая позволила добиться наибольшего фактора улучшения среди рассмотренных способов идентификации выбросов. Предложенная методика *GL* позволила снизить среднеквадратичную ошибку прогноза на 15,9 % по отношению к прогнозу по ВЭС без исключения выбросов.

6. Использование идентификации и исключения выбросов позволяет снизить ошибку прогноза как при прогнозе на уровне ВЭС, так и на уровне отдельных турбин или групп турбин. Кроме того, за счет сокращения количества наблюдений обеспечивается снижение времени обучения регрессионных моделей.

7. Рассмотрена существующая методика учета состава работающих ВЭУ в качестве признака при обучении регрессионных моделей. Предложена новая методика для учета состава работающих ВЭУ с помощью трансформации отклика. Наибольшего фактора улучшения прогноза удалось добиться при использовании трансформации отклика совместно с предложенной методикой объединения турбин в группы *H-PD*. Предложенная методика *RT* позволила снизить среднеквадратичную ошибку прогноза на 13,5 % по отношению к прогнозу по ВЭС без учета состояния турбин.

8. Использование данных о составе работающих турбин позволило снизить ошибку прогноза для всех трех стратегий прогнозирования (на уровне ВЭС, на уровне турбин и на уровне групп турбин) как при использовании коэффициента доступности в качестве признака, так и при выполнении трансформации отклика. При этом использование дополнительного признака также увеличивает время обучения.

9. Совместное использование предложенных способов повышения эффективности краткосрочного прогнозирования выработки ВЭС позволило снизить среднеквадратичную ошибку прогноза на 26,1 %.

ЗАКЛЮЧЕНИЕ

1. Рассмотрены актуальные вопросы, касающиеся прогнозирования генерации ветроэлектрических станций. Показано, что установленная мощность ВЭС в России достигла 2 556 МВт на конец 2024 года и до 2035 года может быть введено еще 5 ГВт ВЭС. Проанализировано влияние увеличения доли установленной мощности ВИЭ с переменным характером выработки на работу энергосистем и обозначена необходимость выполнения прогнозов генерации для работы ВЭС на оптовом рынке электроэнергии и для планирования электроэнергетических режимов. Выявлена необходимость повышения эффективности краткосрочного прогнозирования генерации ВЭС на сутки вперед, что позволяет снизить убытки собственников ВЭС при отклонении фактической выработки от плановой, снизить затраты на компенсацию стоимостного небаланса для других участников оптового рынка, в том числе потребителей, и обеспечить надёжное функционирование энергосистем с ВЭС за счет более точного планирования режимов Системным оператором. Определены основные факторы, затрудняющие процесс прогнозирования генерации ВЭС.

2. Выполнен обзор литературы, исходя из которого был сделан вывод, что единого универсального способа для решения задачи краткосрочного прогнозирования генерации ВЭС на текущий момент нет. Используемые методики прогнозирования во многом носят индивидуальный характер, который определяется составом и качеством доступных исходных данных. В анализе существующих исследований продемонстрировано, что при прогнозировании на сутки вперед нормализованный корень из средней квадратичной ошибки прогноза может составлять 10 – 15%. Также определены перспективные направления повышения эффективности краткосрочного прогнозирования генерации ВЭС: оптимизация гиперпараметров моделей машинного обучения, изменение стратегии агрегации мощности отдельных ВЭУ, идентификация выбросов, учет состава работающих ВЭУ. Кроме того, был проведен анализ существующих критериев для оценки эффективности краткосрочного прогнозирования генерации ВЭС, по

результатам которого предложено два новых критерия: нормализованная суммарная абсолютная ошибка, позволяющая оценить погрешность прогноза с точки зрения вырабатываемой электроэнергии, и вероятность охвата интервала ошибки в заданном диапазоне, обеспечивающая возможность оценить надежность прогнозов.

3. Рассмотрены основные подходы к оптимизации гиперпараметров моделей машинного обучения: поиск по сетке, случайный поиск и байесовская оптимизация. Было определено, что в существующих исследованиях к задаче прогнозирования генерации ВЭС при выполнении байесовской оптимизации применялась только функция «ожидаемого улучшения». С использованием разработанного программного обеспечения выполнено сравнение эффективности данных подходов для нескольких моделей машинного обучения: регрессионное дерево решений, нейронная сеть прямого распространения, градиентный бустинг над деревьями решений, регрессия на основе гауссовских процессов, случайный лес и метод опорных векторов. Была предложена новая методика, основанная на использовании байесовской оптимизации с функцией выбора «нижний доверительный интервал», которая позволила снизить среднеквадратичную ошибку прогноза на 6,1 % по отношению к подходу, основанному на подборе гиперпараметров методом случайного поиска.

4. Показано, что мощность, выдаваемая отдельными ветроустановками, расположенными на одной и той же ВЭС, может значительно отличаться из-за их географической удаленности, «затенения» ВЭУ друг другом, наличия препятствий, наклона поверхности и других особенностей ландшафта. В качестве одного из способов учесть данные факторы используются стратегии агрегирования мощности ВЭУ при прогнозировании суммарной мощности ВЭС: прогноз на уровне всей станции, на уровне групп ВЭУ и на уровне отдельных ВЭУ. Показано, что единого мнения о том, какая стратегия является наиболее эффективной, пока нет. По результатам сравнения данных стратегий было выявлено, что выполнение прогнозирования на уровне групп ВЭУ позволяет как снизить погрешность прогноза, так и обеспечить более низкое потребление вычислительных ресурсов по

сравнению с прогнозом на уровне отдельных ВЭУ. Разработан новый способ объединения турбин в группы с помощью метода иерархической агломеративной кластеризации по *DTW*-расстоянию, который позволил снизить среднеквадратичную ошибку прогноза на 8,7 % по отношению к прогнозу по ВЭС.

5. Проанализированы вопросы предварительной обработки данных для краткосрочного прогнозирования генерации ВЭС, в том числе идентификации и исключения выбросов. Было обнаружено, что наиболее часто для задачи обнаружения выбросов используются характеристики мощности ВЭУ и ВЭС. При этом в существующих исследованиях выделяется 5 основных типов выбросов, каждый из которых соответствует определенной области на характеристике мощности. Идентификация этих выбросов может быть затруднительна. Все 5 типов выбросов можно отнести к локальным, поскольку все наблюдения находятся в диапазоне ожидаемых значений скорости ветра (от нуля до максимальной) и мощности (от нуля до номинальной). Были рассмотрены существующие методики идентификации, заключающиеся в применении метода межквартильного размаха и основанной на плотности пространственной кластеризации для приложений с шумами. Предложена новая методика идентификации выбросов по характеристике мощности с использованием алгоритма глобально-локальной оценки выбросов по иерархической кластеризации, которая позволила снизить среднеквадратичную ошибку прогноза на 15,9 % по отношению к прогнозу по ВЭС без исключения выбросов. Кроме того, результаты исследований показали, что идентификация и исключение выбросов позволяет снизить ошибку прогноза как при прогнозе на уровне ВЭС, так и на уровне отдельных ВЭУ или групп ВЭУ.

6. Рассмотрен один из наименее освещенных в литературе вопросов, связанный с учетом состава работающего оборудования при краткосрочном прогнозировании генерации ВЭС. Была проверена существующая методика, заключающаяся в использовании состояния ВЭУ в качестве признака при обучении регрессионных моделей. Разработана новая методика учета состава работающих ветроустановок с помощью трансформации отклика, которая позволила снизить среднеквадратичную ошибку прогноза на 13,5 % по отношению к прогнозу без

учета состояния турбин. Было выявлено, что информация о состоянии ВЭУ позволяет снизить ошибку прогноза для всех трех стратегий прогнозирования. При этом применение существующей методики и использование дополнительного признака также увеличивает время обучения.

7. Совместное использование предложенных методик (байесовская оптимизация гиперпараметров с функцией выбора «нижний доверительный интервал»; прогнозирование на уровне групп турбин, определенных иерархической кластеризацией по *DTW*-расстоянию; глобально-локальная оценка выбросов; учет состава работающих установок с помощью трансформации отклика) позволило снизить среднеквадратичную ошибку прогноза генерации ВЭС на 26,1 %.

8. Практическая ценность исследования подтверждается актом внедрения. Результаты работы были использованы в ООО «Прософт-Системы» при разработке модуля краткосрочного прогнозирования выработки ветроэлектрических станций для программного комплекса «Энергосфера».

Рекомендации по развитию и разработке темы

Среди направлений будущих исследований по данной теме можно рассматривать разработку новых методик и повышение эффективности прогнозов при долгосрочном планировании энергетических режимов, краткосрочном прогнозировании на несколько суток вперед для задач ВСВГО и внутрисуточном прогнозировании для задач балансирования системы и диспетчерского управления.

Повышение эффективности долгосрочных прогнозов на месяц, год и несколько лет вперед позволит более точно осуществлять планирование балансов мощности и электроэнергии, а значит снизить риск принятия неправильных решений при планировании ремонтов, расхода топлива электростанций, режимов работы ГЭС и водохранилищ и т.д.

Повышение эффективности краткосрочных прогнозов на несколько суток вперед обеспечит более точный учет генерации ВЭС при расчете ВСВГО и при определении объемов резерва третичного регулирования. В результате, может быть

достигнуто снижение объема и более оптимальное распределение резерва третичного регулирования, а значит снижение затрат на его обеспечение.

Повышение эффективности внутрисуточных прогнозов позволит формировать собственникам ВЭС более точные оперативные ценопринимающие заявки на балансирующем рынке. В результате, будет обеспечено снижение затрат собственников ВЭС на компенсацию стоимости отклонений.

ПЕРЕЧЕНЬ СОКРАЩЕНИЙ И УСЛОВНЫХ ОБОЗНАЧЕНИЙ

АО «СО ЕЭС»	-	Акционерное общество «Системный оператор Единой энергетической системы»;
АРВЭ	-	Ассоциация развития возобновляемой энергетики;
БР	-	балансирующий рынок;
ВАК	-	Высшая аттестационная комиссия;
ВИЭ	-	возобновляемые источники энергии;
ВСВГО	-	выбор состав включенного генерирующего оборудования;
ВЭ	-	вычислительный эксперимент;
ВЭС	-	ветроэлектрическая станция;
ВЭУ	-	ветроэнергетическая установка;
ДПМ	-	договор о предоставлении мощности;
ЕЭС	-	единая энергетическая система;
ЛЭП	-	линия электропередач;
МДП	-	максимально допустимый переток;
МЭА	-	международное энергетическое агентство (IEA);
НТД	-	нормативно-техническая документация;
ООН	-	организация объединенных наций;
ОРЭМ	-	оптовый рынок электроэнергии и мощности;
ОЭС	-	объединенная энергосистема;
ПАК	-	программно-аппаратный комплекс;
ПБР	-	план балансирующего рынка;
ПДГ	-	прогнозный диспетчерский график;
ПК	-	программный комплекс;
ППБР	-	предварительный план балансирующего рынка;
ПЭР	-	предварительный электроэнергетический режим;
Росстат	-	Федеральная служба государственной статистики;
РСВ	-	оптовый рынок электроэнергии «на сутки вперед»;
РФ	-	Российская Федерация;

СиПР	-	схема и программа развития;
СКО	-	среднеквадратичные отклонения;
СССР	-	Союз Советских Социалистических Республик;
СОГ	-	способ оптимизации гиперпараметров;
США	-	Соединенные штаты Америки;
ЭВМ	-	электронная вычислительная машина;
ЦАГИ	-	Центральный аэрогидродинамический институт;
ЦФ	-	Целевая функция;
ЧПП	-	численный прогноз погоды;
ЭЭ	-	электроэнергия;
ЭЭС	-	электроэнергетическая энергосистема;
ЮАР	-	Южно-Африканская Республика;
<i>A</i>	-	предложенная методика учета состава включенных ВЭУ при обучении регрессионной модели в виде дополнительного признака;
<i>ACO</i>	-	<i>Ant Colony Optimization</i> (алгоритм оптимизации подражанием муравьиной колонии);
<i>ANFIS</i>	-	<i>Adaptive Network-based Fuzzy Inference System</i> (адаптивная сеть на основе системы нечеткого вывода);
<i>AR</i>	-	<i>AutoRegressive model</i> (модель авторегрессии);
<i>ARMA</i>	-	<i>AutoRegressive Moving Average model</i> (модель авторегрессии скользящего среднего);
<i>ARIMA</i>	-	<i>AutoRegressive Integrated Moving Average model</i> (модель авторегрессии интегрированного скользящего среднего);
<i>ARIMA-GARCH</i>	-	<i>AutoRegressive Integrated Moving Average General AutoRegressive Conditional Heteroskedasticity model</i> (модель авторегрессии интегрированного скользящего среднего с условной гетероскедастичностью);
<i>ARX</i>	-	<i>AutoRegressive model with eXogenous input</i> (модель авторегрессии с внешним входом);
<i>BBOB</i>	-	<i>Black-Box Optimization Benchmarking</i> (набор тестовых функций для задач, представляющих собой «черный ящик»);

- CART* - *Classification and Regression Trees* (деревья решений для задач классификации и регрессии);
- CNN* - *Convolutional Neural Network* (сверточная нейронная сеть);
- CNN-LSTM* - *Convolutional Neural Network with a Long Short-Term Memory Layer* (сверточная нейронная сеть с долгой краткосрочной памятью);
- CMI* - *Conditional Mutual Information* (оценка взаимной информации);
- CSO* - *Chicken Swarm Optimization* (алгоритм оптимизации подражанием группы кур);
- DB* - предложенный способ идентификации выбросов на характеристике мощности с применением основанной на плотности пространственной кластеризации для приложений с шумами *DBSCAN* с автоматическим определением максимального радиуса соседства с помощью алгоритма *KNeedle*;
- DBN* - *Deep Belief Network* (глубокая сеть доверия);
- DBSCAN* - *Density Based Spatial Clustering of Applications with Noise* (основанная на плотности пространственная кластеризация для приложений с шумами);
- DTW* - *Dynamic Time Warping* (алгоритм динамической трансформации временной шкалы);
- DWT* - *Discrete Wavelet Transform* (дискретное-вейвлет преобразование);
- EMD* - *Ensemble Empirical Mode Decomposition* (ансамблевое разложение на эмпирические моды);
- EICP* - *Error Interval Coverage Probability* (вероятность охвата интервала ошибки);
- EMD* - *Empirical Mode Decomposition* (разложение на эмпирические моды);
- EI* - *Expected Improvement* (ожидаемое улучшение);
- FFNN* - *Feed Forward Neural Network* (нейронная сеть прямого распространения);
- GC* - то же, что и *H-PD*;

- GBDT* - *Gradient Boosting on Decision Trees* (градиентный бустинг над деревьями решений);
- GF* - прогноз на уровне групп турбин, объединение в группы с помощью упрощённого способа распределения исходя из принадлежности к фидерам;
- GL* - предложенный способ идентификации выбросов на характеристике мощности с применением алгоритма глобально-локальной оценки выбросов *GLOSH* по иерархической основанной на плотности пространственной кластеризации для приложений с шумами *HDBSCAN**;
- GLOSH* - *Global-Local Outlier Scores based on Hierarchies* (глобально-локальной оценки выбросов по иерархии *HDBSCAN**);
- GPR* - *Gaussian Process Regression* (регрессия на основе гауссовских процессов);
- GS* - *Grid Search* (поиск по решетке);
- H-PD* - предложенный подход выполнения прогноза на уровне групп турбин; объединение в группы по временным рядам измерений мощности каждой турбины методом иерархической агломеративной кластеризации, для измерения расстояния между временными рядами используется алгоритм динамической трансформации временной шкалы;
- H-PE* - предложенный подход выполнения прогноза на уровне групп турбин; объединение в группы по временным рядам измерений мощности каждой турбины методом иерархической агломеративной кластеризации, для измерения расстояния между точками временных рядов используется евклидово расстояние;
- HAC* - *Hierarchical Agglomerative Clustering* (иерархическая агломеративная кластеризация);
- HDBSCAN**- *Hierarchical Density Based Spatial Clustering of Applications with Noise* (иерархическая основанная на плотности пространственная кластеризация для приложений с шумами);
- HHT* - *Hilbert–Huang Transform* (преобразование Гильберта-Хуанга);
- IEA* - *International Energy Agency* (МЭА);

- Inf-FS* - *Infinite Feature Selection* (графовый метод отбора признаков);
- IQR* - *InterQuartile Range* (межквартильный размах);
- IWT* - *Improved Wavelet Transform* (улучшенное вейвлет преобразование);
- K-MS* - существующий подход выполнения прогноза на уровне групп турбин; объединение в группы по данным о средних значениях и среднеквадратичных отклонениях измеренных мощностей и скоростей ветра каждой турбины методом кластеризации *k*-средних;
- K-P* - предложенный подход выполнения прогноза на уровне групп турбин; объединение в группы по временным рядам измерений мощности каждой турбины методом кластеризации *k*-средних;
- LCB* - *Lower Confidence Bound* (нижний доверительный интервал);
- LOF* - *Local Outlier Factor* (локальный уровень выбросов);
- LSTM* - *Long Short-Term Memory* (нейронная сеть с долгой краткосрочной памятью);
- LVRT* - *Low Voltage Ride Through* (поддержание непрерывности электроснабжения при низком напряжении);
- MRMR* - *Minimum-Redundancy-Maximum-Relevance* (критерий минимальной избыточности – максимальной релевантности);
- NBias* - *Normalized Bias* (нормализованное смещение);
- NMAE* - *Normalized Mean Absolute Error* (нормализованная средняя абсолютная ошибка);
- NRMSE* - *Normalized Root Mean Squared Error* (нормализованный квадратный корень из средней квадратичной ошибки);
- NSAE* - *Normalized Sum of Absolute Errors* (нормализованная сумма абсолютных ошибок);
- P* - прогноз на уровне ВЭС;
- PCA* - *Principal Component Analysis* (метод главных компонент);
- PI* - *Probability of Improvement* (вероятность улучшения);

- PSR* - *Phase Space Reconstruction* (реконструкция фазового пространства);
- QM* - существующий способ идентификации выбросов на характеристике мощности с помощью метода межквартильного размаха;
- QR* - *Quantile Regression* (квантильная регрессия);
- RBFNN* - *Radial Basis Functions Neural Network* (нейронная сеть радиально базисных функций);
- RDT* - *Regression Decision Tree* (регрессионное дерево решений);
- RF* - *Random Forest* (случайный лес);
- RFE* - *Recursive Feature Elimination* (рекурсивное исключение признаков);
- RNN-GRU* - *Recurrent Neural Network with Gated Recurrent Units* (рекуррентная нейронная сеть с управляемым рекуррентным блоком);
- RNN-LSTM* - *Recurrent Neural Network with a Long Short-Term Memory Layer* (рекуррентная нейронная сеть с долгой краткосрочной памятью);
- RS* - *Random Search* (случайный поиск);
- RT* - предложенная методика учета состава включенных ВЭУ при обучении регрессионной модели с помощью трансформации отклика;
- SCADA* - *Supervisory Control and Data Acquisition* (диспетчерское управление и сбор данных);
- SDAE* - *Stacked Denoising Autoencoder* (нейронная сеть с архитектурой вложенных автокодировщиков);
- SFFS* - *Sequential Forward Feature Selection* (последовательный прямой плавающий отбор);
- SMO* - *Sequential Minimal Optimization* (последовательная минимальная оптимизация);
- SS* - *Skill Score* (фактор улучшения);
- SSA* - *Singular Spectrum Analysis* (анализа сингулярного спектра);
- SVM* - *Support Vector Machine* (метод опорных векторов);
- T* - прогноз на уровне отдельных турбин;

- VIM* - *Variable Importance Measure* (индекс важности переменных);
- VMD* - *Variational Mode Decomposition* (разложение на вариационные моды);
- VPD* - *Wavelet Packet Decomposition* (вейвлет-пакетная декомпозиция);
- WoS* - *Web of Science*;
- WTD* - *Wavelet Threshold Denoising* (пороговая обработка сигнала с помощью вейвлет-разложения);

СПИСОК ЛИТЕРАТУРЫ

1. Electricity, 2025. Analysis and forecast to 2027 / Intern. Energy Agency // IEA : Offic. site. – 2025. – February. – P. 200. – URL: <https://www.iea.org/reports/electricity-2025> (date of access: 05.08.2025).
2. Renewables, 2024. Analysis and forecasts to 2030 / Intern. Energy Agency // IEA : Offic. site. – 2024. – October. – P. 177. – URL: <https://www.iea.org/reports/renewables-2024> (date of access: 05.08.2025).
3. Rohde, R. Global Temperature Report, 2024 / R. Rohde // Berkeley Earth : Offic. site. – 2025. – 10 Jan. – URL: <https://berkeleyearth.org/global-temperature-report-for-2024/> (date of access: 01.08.2025).
4. Gipe, P. An overview of the history of wind turbine development. Pt. I. The early wind turbines until the 1960s / P. Gipe, E. Möllerström. – DOI 10.1177/0309524X221117825 // Wind Engineering. – 2022. – Vol. 46, Iss. 6. – P. 1973–2004. – URL: <https://journals.sagepub.com/doi/full/10.1177/0309524X221117825> (date of access: 11.10.2024).
5. Beurskens, J. The history of wind energy. Chap. 1 / J. Beurskens. – DOI 10.1002/9781118701492.ch1 // Understanding wind power technology: theory, deployment and optimisation / ed. by A. Schaffarczyk. – Wiley, 2014. – P. 1–44. – URL: <https://onlinelibrary.wiley.com/doi/abs/10.1002/9781118701492.ch1> (date of access: 11.10.2024).
6. Pasqualetti, M. History of wind energy / M. Pasqualetti, R. Richter, P. Gipe // Encyclopedia of energy / ed. by C. Cleveland. – [Boston], 2004. – Vol. 6. – P. 419–433.
7. Renewable Energy Progress Tracker / Intern. Energy Agency // IEA. – 2024. – 9 Oct. – URL: <https://www.iea.org/data-and-statistics/data-tools/renewable-energy-progress-tracker> (date of access: 05.08.2025).
8. Global Wind Report, 2025 / Global wind energy council : Overview // GWEC : Offic. site. – 2025. – URL: <https://www.gwec.net/reports/globalwindreport> (date of access: 05.08.2025).
9. Wind Power Generation Capacity, 2010–2022 // Incorrrys : Offic. site. – 2023. – 31 July. – URL: <https://incorrrys.com/power-generation/wind-power-generation/wind-power-generation-capacity-2010-2022/> (date of access: 14.08.2024).
10. Wind Power Capacity Additions, 2022–2030 // Incorrrys : Offic. site. – 2023. – 2 Aug. – URL: <https://incorrrys.com/power-generation/wind-power-generation/wind-power-capacity-additions-2022-2030/> (date of access: 14.08.2024).
11. Renewable Capacity Statistics, 2025 / Intern. renewable energy agency. – Abu Dhabi : [s. n.], 2025. – P. 288. – URL: <https://www.irena.org/Publications/2025/Jul/Renewable-energy-statistics-2025> (date of access: 05.08.2025). – ISBN: 978-92-9260-675-6.

12. Бутузов, В. А. В. Сто лет развития ветроэнергетики в России. Ч. 1 / В. А. Бутузов, П. П. Безруких, С. В. Грибков // Окружающая среда и энерговедение. – 2021. – № 3. – С. 8–24.
13. Бутузов, В. А. Сто лет развития ветроэнергетики в России. Ч. 2 / В. А. Бутузов, П. П. Безруких, С. В. Грибков // Окружающая среда и энерговедение. – 2021. – № 4. – С. 18–34.
14. Об основных направлениях государственной политики в сфере повышения энергетической эффективности электроэнергетики на основе использования возобновляемых источников энергии на период до 2035 года : распоряжение Правительства Рос. Федерации № 1-р от 08.01.2009 (ред. от 03.05.2024). – URL: <https://docs.cntd.ru/document/902137809> (дата обращения: 21.08.2024).
15. О механизме стимулирования использования возобновляемых источников энергии на оптовом рынке электрической энергии и мощности : постановление Правительства Рос. Федерации № 449 от 28.05.2013 (ред. от 03.05.2024) // Кодекс : электрон. фонд правовых и нормат.-техню док. – URL: <https://docs.cntd.ru/document/499023293> (дата обращения: 21.08.2024).
16. Статистика ВИЭ / Ассоц. развития возобновляемой энергетики // АРВЭ : офиц. сайт. – 2025. – 9 июля. – URL: <https://rreda.ru/industry/statistics/#graph3> (дата обращения: 05.08.2024).
17. Выработка электрической энергии, всего / Единая межведомств. информ.-статист. система // ЕМИСС. Государственная статистика. – URL: <https://fedstat.ru/indicator/62001> (дата обращения: 05.08.2025).
18. Возобновляемая энергетика в России и мире / Рос. энергет. Агентство. – Москва : РЭА, 2022. – URL: <https://rosenergo.gov.ru/upload/iblock/e04/3xtm87iv99x76b23c6wjul3as5pzz8zj.pdf?ysclid=m661a3xx45274276204> (дата обращения: 14.08.2024).
19. Объекты ВИЭ-генерации в регионах России / Ассоц. развития возобновляемой энергетики // АРВЭ : сайт. – 2025. – URL: <https://rreda.ru/industry/imap/> (дата обращения: 05.08.2025).
20. Реестр квалифицированных генерирующих объектов, функционирующих на основе использования возобновляемых источников энергии // Центр энергосертификации : офиц. сайт. – 2025. – URL: <https://green-e-track.ru/doc/rqf.xls> (дата обращения: 05.08.2025).
21. Результаты отборов проектов ВИЭ / Администратор торговой системы // АТС : офиц. сайт. – 2025. – URL: <https://www.atsenergo.ru/vie/proresults> (дата обращения: 05.08.2025).
22. Конкурсные отборы инвестиционных проектов ВИЭ / Ассоц. развития возобновляемой энергетики // АРВЭ : офиц. сайт. – 2025. – URL: <https://rreda.ru/industry/competitive-selection/> (дата обращения: 05.08.2025).
23. Отчеты о функционировании ЕЭС России за 2017–2024 гг. // Системный оператор Единой энергетической системы : офиц. сайт. – URL: <https://www.sops.ru/functioning/tech-disc/tech-disc-ups/> (дата обращения: 05.08.2025).

24. Илюшин, П. В. Интеграция электростанций на основе возобновляемых источников энергии в Единую энергетическую систему России: обзор проблемных вопросов и подходов к их решению / П. В. Илюшин. – DOI 10.24160/1993-6982-2022-4-98-107 // Вестник Московского энергетического института. – 2022. – № 4. – С. 98–107.
25. Offshore wind farm-grid integration: A review on infrastructure, challenges, and grid solutions / S. W. Ali, M. Sadiq, Y. Terriche [et al.] – DOI 10.1109/ACCESS.2021.3098705 // IEEE Access. – 2021. – Vol. 9. – P. 102811–10282. – URL: <https://ieeexplore.ieee.org/abstract/document/9492079> (date of access: 12.12.2022).
26. Potential challenges of integrating large-scale wind energy into the power grid: A review / G. M. Shafiullah, M. T. Oo Amanullah, A. B. M. Shawkat Ali, P. Wolfs. – DOI 10.1016/j.rser.2012.11.057 // Renewable and sustainable energy reviews. – 2013. – Vol. 20. – P. 306–321. – URL: <https://www.sciencedirect.com/science/article/abs/pii/S1364032112006715?via%3Dihub> (date of access: 14.12.2022).
27. Challenges of grid integration of wind power on power system grid integrity: A review / T. R. Ayodele, A. Jimoh, J. L. Munda, A. J. Tehile // International journal of renewable energy research. – 2012. – Vol. 2, Iss. 4. – P. 618–626. – URL: <https://dergipark.org.tr/en/pub/ijrer/issue/16081/168306> (date of access: 14.12.2022).
28. Grid integration challenges of wind energy: A review / S. D. Ahmed, F. S. M. Al-Ismail, Md Shafiullah [et al.]. – DOI 10.1109/ACCESS.2020.2964896 // IEEE Access. – 2020. – Vol. 8. – P. 10857–10878. – URL: <https://ieeexplore.ieee.org/abstract/document/8952713> (date of access: 14.12.2022).
29. Халле, Ф. В. Ветроэнергетика: цифры и факты. Ч. 2. Сетевая интеграция / Ф. В. Халле, П. Гарднер ; пер. и коммент. Е. Карташев // Силовая электроника. – 2014. – Т. 6, № 51. – С. 116–119.
30. Integrating Wind. Developing Europe's power market for the large-scale integration of wind power : Final report / F. V. Hulle, J. O. Tande, K. Uhlen [et al.] ; European Wind Energy Association (EWEA). – Nederland : EWEA, 2009. – 53 p.
31. Getting Wind and Sun onto the Grid. A Manual for Policy Makers / Intern. Energy Agency // IEA : Off. site. – 2017. – March. – P. 65. – URL: https://iea.blob.core.windows.net/assets/1b5de86e-4499-4ae7-84b6-948bc2ca2759/Getting_Wind_and_Sun.pdf (date of access: 05.11.2022).
32. Крамской, Ю. Г. Интеграция возобновляемых источников электроэнергии в электрические сети с применением силовой электроники / Ю. Г. Крамской // Энергия единой сети. – 2017. – Т. 30, №. 1. – С. 53–66. – URL: <https://www.elibrary.ru/item.asp?id=28800667> (дата обращения: 07.12.2022).
33. Жданеев, О. В. Развитие ВИЭ и формирование новой энергополитики России / О. В. Жданеев, С. С. Зуев. – DOI 10.46920/2409-5516_2020_2144_8 // Энергетическая политика. – 2020. – № 2 (144). – С. 84–95. – URL: <https://www.elibrary.ru/item.asp?id=42801927> (дата обращения: 07.12.2022).

34. Дворкин, Д. В. Нормативные проблемы разработки схем выдачи мощности электростанций на базе возобновляемых источников энергии / Д. В. Дворкин, А. А. Антонов, И. Ю. Кошкин // Электроэнергия. Передача и распределение. – 2021. – № 6 (69). – С. 66–73.
35. Дворкин, Д. В. Средства и способы снижения влияния ВИЭ на снижение качества электроэнергии и надежности электроснабжения при переходе к зеленой энергетике / Д. В. Дворкин, Д. А. Халезин // Электроэнергия. Передача и распределение. – 2023. – № 1 (76). – С. 62–68.
36. Power quality aspects in a wind power plant / E. Muljadi, C. P. Butterfield; J. Chacon [et al.]. – DOI 10.1109/PES.2006.1709244 // 2006 IEEE Power Engineering Soc. Gen. Meet : conf. proc. (Montreal, 18–22 June 2006). – New York, 2006. – P. 18–22. – URL: <https://ieeexplore.ieee.org/abstract/document/1709244> (date of access: 21.12.2022).
37. Saqib, M. A. Power-quality issues and the need for reactive-power compensation in the grid integration of wind power / M. A. Saqib, A. Z. Saleem. – DOI 10.1016/j.rser.2014.11.035 // Renewable and Sustainable Energy Reviews. – 2015. – Vol. 43. – P. 51–64. – URL: <https://www.sciencedirect.com/science/article/abs/pii/S1364032114009666> (date of access: 21.12.2022).
38. Илюшин, П. В. Обеспечение устойчивой работы ветроэнергетических установок при кратковременных провалах напряжения / П. В. Илюшин, А. В. Симонов // Методические вопросы исследования надежности больших систем энергетики : материалы конф. (Волжский, 13–17 сент. 2021 г.) / отв. ред. Н. И. Воропай. – Иркутск, 2021. – Вып. 72, кн. 1. – С. 166–175. – URL: <https://www.elibrary.ru/item.asp?id=47150880> (дата обращения: 23.12.2022).
39. Симонов, А. В. Способы обеспечения устойчивой работы ветроэнергетических установок ветровых электростанций в составе отечественных распределительных электрических сетей / А. В. Симонов, П. В. Илюшин // Электроэнергия. Передача и распределение. – 2022. – № 3 (72). – С. 60–70.
40. Comprehensive overview of low voltage ride through methods of grid integrated wind generator / O. P. Mahela, N. Gupta, M. Khosravy, N. Patel. – DOI 10.1109/ACCESS.2019.2930413 // IEEE Access. – 2019. – Vol. 7. – P. 99299–99326. – URL: <https://ieeexplore.ieee.org/abstract/document/8768384> (date of access: 07.12.2022).
41. Nasiri, M. A review of low-voltage ride-through enhancement methods for permanent magnet synchronous generator based wind turbines / M. Nasiri, J. Milimonfared, S. H. Fathi. – DOI 10.1016/j.rser.2015.03.079 // Renewable and sustainable energy reviews. – 2015. – Vol. 47. – P. 399–415. – URL: <https://www.sciencedirect.com/science/article/abs/pii/S1364032115002324> (date of access: 03.12.2022).
42. Синтез и тестирование типовых структур систем автоматического управления на основе виртуального синхронного генератора для генерирующих установок с силовым преобразователем / А. А. Суворов, А. Б. Аскаров, В. Е. Рудник [и

- др.] // Электрические станции. – 2022. – № 3 (1088). – С. 43–57. – URL: <https://www.elibrary.ru/item.asp?id=48119573> (дата обращения: 11.06.2023).
43. Улучшение демпфирующих свойств виртуального синхронного генератора с помощью корректирующего согласо-параллельного регулятора / А. Б. Аскарлов, А. А. Суворов, Н. Ю. Рубан, П. П. Радько // Электричество. – 2024. – № 1. – С. 18–36. – URL: <https://www.elibrary.ru/item.asp?id=47150880> (дата обращения: 15.09.2024).
 44. Астапов, В. Ю. Мультиагентное управление виртуальной инерцией ветроустановок. Анализ актуальности и эффективности / В. Ю. Астапов, А. А. Ачитаев, А. В. Домышев // Электроэнергия. Передача и распределение. – 2024. – № 4(85). – С. 22–27. – URL: <https://www.elibrary.ru/item.asp?id=68635939> (дата обращения: 15.02.2025).
 45. Implementing virtual inertia in DFIG-based wind power generation / M. F. M. Arani, E. F. El-Saadany // IEEE Transactions on Power Systems. – 2012. – Vol. 28, Iss. 2. – P. 1373–1384. – URL: <https://ieeexplore.ieee.org/abstract/document/6289404> (date of access: 07.12.2022).
 46. Virtual inertia: Current trends and future directions / U. Tamrakar, D. Shrestha, M. Maharjan, B. P. Bhattarai. – DOI 10.3390/app7070654 // Applied sciences. – 2017. – Vol. 7, Iss. 7. – P. 654. – URL: <https://www.mdpi.com/2076-3417/7/7/654> (date of access: 23.11.2022).
 47. Особенности выполнения защит линий при наличии ветровых электростанций / В. А. Ефремов, А. В. Ефремов, М. Ю. Петрушков, Е. В. Широкина // Проблемы и перспективы развития энергетики, электротехники и энергоэффективности : материалы III Междунар. науч.-техн. конф. (Чебоксары, 14–16 нояб. 2019 г.) / гл. ред. В. Г. Ковалев. – Чебоксары, 2019. – С. 193–199. – URL: <https://www.elibrary.ru/item.asp?id=41868774&pf=1> (дата обращения: 23.11.2022).
 48. Исследование влияния возобновляемых источников энергии на функционирование релейной защиты / М. В. Андреев, А. А. Суворов, А. Б. Аскарлов [и др.]. – DOI 10.14529/power190404 // Вестник Южно-Уральского государственного университета. Серия «Энергетика». – 2019. – Т. 19, № 4. – С. 34–40. – URL: <https://www.elibrary.ru/item.asp?id=41561942> (дата обращения: 23.11.2022).
 49. Оводов, И. Ю. Оценка влияния ветряных электростанций на работу релейной защиты в электроэнергетической системе / И. Ю. Оводов // Вестник науки. – 2024. – Т. 1, № 11(80). – С. 1199–1206. – URL: <https://www.elibrary.ru/item.asp?id=74483973> (дата обращения: 15.02.2025).
 50. Eissa M. M. Protection techniques with renewable resources and smart grids—A survey / M. M. Eissa. – DOI 10.1016/j.rser.2015.08.031 // Renewable and sustainable energy reviews. – 2015. – Vol. 52. – P. 1645–1667. – URL: <https://www.sciencedirect.com/science/article/abs/pii/S1364032115008771?via%3Dihub> (date of access: 23.11.2022).

51. Distance protection of AC grid with HVDC-connected offshore wind generators / L. He, Ch.-Ch. Liu, A. Pitto, D. Cirio. – DOI 10.1109/TPWRD.2013.2271761 // IEEE Transactions on Power Delivery. – 2014. – Vol. 29, Iss. 2. – P. 493–501. – URL: <https://ieeexplore.ieee.org/document/6563170> (date of access: 07.12.2022).
52. System Integration of Renewables: An update on Best Practice / Intern. Energy Agency // IEA : Offic. site. – 2018. – January. – P. 154. – URL: <https://www.iea.org/reports/system-integration-of-renewables> (date of access: 07.12.2022).
53. The power of transformation. Wind, Sun and the Economics of Flexible Power Systems / Intern. Energy Agency // IEA : Offic. site. – 2014. – 236 P. – URL: https://iea.blob.core.windows.net/assets/b6a02e69-35c6-4367-b342-2acf14fc9b77/The_power_of_Transformation.pdf (date of access: 07.12.2022).
54. Renewables 2017 Global Status Report / Renewable Energy Policy Network for the 21st Cent. // UN CC:Learn : Offic. site. – Paris, 2017. – 302 P. – URL: https://www.uncclearn.org/wp-content/uploads/library/17-8399_gsr_2017_full_report_0621_opt.pdf (date of access: 01.12.2022).
55. Annual variable renewable energy share and corresponding system integration phase in selected countries/regions / Intern. Energy Agency // IEA : Offic. site. – 2022. – URL: <https://www.iea.org/data-and-statistics/charts/annual-variable-renewable-energy-share-and-corresponding-system-integration-phase-in-selected-countries-regions-2022> (date of access: 01.12.2022).
56. Установленная электрическая мощность электростанций / Единая межведомств. информ.-стат. система // ЕМИС : офиц. сайт. – URL: <https://fedstat.ru/indicator/43486> (дата обращения: 05.08.2025).
57. Об утверждении Правил технологического функционирования электроэнергетических систем и о внесении изменений в некоторые акты Правительства Российской Федерации : постановление Правительства Рос. Федерации от № 937 13.08.2018 (ред. от 24.07.2024) // Кодекс. Электронный фонд правовых нормативно-технических документов : офиц. сайт. – URL: <https://docs.cntd.ru/document/550919677> (дата обращения: 21.08.2024).
58. Об утверждении Правил разработки и утверждения документов перспективного развития электроэнергетики, изменении и признании утратившими силу некоторых актов и отдельных положений некоторых актов Правительства Российской Федерации : постановление Правительства Рос. Федерации № 2556 от 30 декабря 2022 г. (ред. от 24.07.2024) // Кодекс. Электронный фонд правовых нормативно-технических документов : офиц. сайт. – URL: <https://docs.cntd.ru/document/1300668311> (дата обращения: 21.08.2024).
59. Об утверждении Методических указаний по проектированию развития энергосистем и о внесении изменений в приказ Минэнерго России № 1195 от 28 декабря 2020 г. : приказ Министерства энергетики Рос. Федерации № 1286 от 6 декабря 2022 г. // Кодекс. Электронный фонд правовых нормативно-технических документов : офиц. сайт. – URL: <https://docs.cntd.ru/document/1300344098> (дата обращения: 21.08.2024).

60. Айзенберг, Н. И. Среднесрочное планирование режимов электроэнергетических систем в условиях оптового рынка электроэнергии / Н. И. Айзенберг, С. И. Паламарчук // Известия Российской академии наук. Энергетика. – 2020. – № 6. – С. 17–30.
61. Об утверждении Правил оптового рынка электрической энергии и мощности и о внесении изменений в некоторые акты Правительства Российской Федерации по вопросам организации функционирования оптового рынка электрической энергии и мощности : постановление Правительства Рос. Федерации № 117 от 27.12.2010 (ред. от 24.07.2024) // Кодекс. Электронный фонд правовых нормативно-технических документов : офиц. сайт. – URL: <https://docs.cntd.ru/document/90227061> (дата обращения: 21.08.2024).
62. Регламент внесения изменений в расчетную модель электроэнергетической системы. Приложение № 2 к Договору о присоединении к торговой системе оптового рынка : с изм. от 27 окт. 2023 г. // НП Совет рынка : офиц. сайт. – 2024. – URL: https://lk.np-sr.ru/sites/default/files/sr_regulation/reglaments/r2_01112023_27102023.pdf (дата обращения: 21.08.2024).
63. Регламент актуализации расчетной модели. Приложение № 3 к Договору о присоединении к торговой системе оптового рынка : с изм. от 24 июня 2024 г. // НП Совет рынка : офиц. сайт. – 2024. – URL: https://www.np-sr.ru/sites/default/files/sr_regulation/reglaments/r3_01102024_24092024.docx (date of access: 21.08.2024).
64. Регламент проведения расчетов выбора состава генерирующего оборудования. Приложение № 3.1 к Договору о присоединении к торговой системе оптового рынка : с изм. от 25 янв. 2022 г. // НП Совет рынка : офиц. сайт. – 2024. – URL: https://www.np-sr.ru/sites/default/files/sr_regulation/reglaments/r3_1_01022025_24092024.docx (date of access: 21.08.2024).
65. Регламент подачи ценовых заявок участниками оптового рынка. Приложение № 5 к Договору о присоединении к торговой системе оптового рынка : с изм. от 24 июля 2024 г. // НП Совет рынка : офиц. сайт. – 2024. – URL: <https://www.np-sr.ru/ru/regulation/joining/reglaments/1962> (дата обращения: 21.08.2024).
66. Регламент проведения конкурентного отбора ценовых заявок на сутки вперед Приложение № 7 к Договору о присоединении к торговой системе оптового рынка : с изм. от 22 июля 2024 г. // НП Совет рынка : офиц. сайт. – URL: https://www.np-sr.ru/sites/default/files/sr_regulation/reglaments/r7_01012026_26112024.docx (дата обращения: 21.08.2024).
67. Регламент расчета плановых объемов производства и потребления и расчета стоимости электроэнергии на сутки вперед. Приложение № 8 к Договору о присоединении к торговой системе оптового рынка : с изм. от 24 июня 2024 г. // НП Совет рынка : офиц. сайт. – 2024. – URL: <https://www.np-sr.ru>

- sr.ru/sites/default/files/sr_regulation/reglaments/r8_01012021_24112020.pdf (date of access: 21.08.2024).
68. Регламент оперативного диспетчерского управления электроэнергетическим режимом объектов управления ЕЭС России. Приложение № 9 к Договору о присоединении к торговой системе оптового рынка : с изм. от 24 июня 2024 г. // НП Совет рынка : офиц. сайт. – 2024. – URL: https://lk.np-sr.ru/sites/default/files/sr_regulation/reglaments/r9_01082021_22072021.pdf (date of access: 21.08.2024).
 69. Регламент проведения конкурентного отбора заявок для балансирования системы. Приложение № 10 к Договору о присоединении к торговой системе оптового рынка : с изм. от 27 марта 2024 г. // НП Совет рынка : офиц. сайт. – 2024. – URL: https://www.np-sr.ru/sites/default/files/sr_regulation/reglaments/r10_01092024_27032024.pdf (date of access: 21.08.2024).
 70. Регламент определения объемов, инициатив и стоимости отклонений. Приложение № 12 к Договору о присоединении к торговой системе оптового рынка : с изм. от 22.10.2024 // Кодекс. Электронный фонд правовых нормативно-технических документов : офиц. сайт. – URL: <https://docs.cntd.ru/document/1200120235> (дата обращения: 21.08.2024).
 71. Давидсон, М. Р. Математическая модель расчета ценовых индикаторов в задаче выбора состава генерирующего оборудования в условиях конкурентного рынка электроэнергии в России / М. Р. Давидсон, А. В. Селезнев // Известия Российской академии наук. Теория и системы управления. – 2014. – №. 3. – С. 61.
 72. Архипов, А. Два года новой модели ВСВГО / А. Архипова, М. Долматова // ЭнергоРынок. – 2016. – № 2 (137). – С. 52–58.
 73. Добрынина, Ю. Э. Анализ учета критериев системной надёжности в моделях выбора состава включенного генерирующего оборудования в электроэнергетических системах / Ю. Э. Добрынин, Д. С. Крупенёв. – DOI 10.21285/1814-3520-2024-2-273-28 // iPolytech Journal. – 2024. – Т. 28, №. 2. – С. 273–289. – URL: <https://ipolytech.elpub.ru/jour/article/view/826> (дата обращения: 07.09.2024).
 74. Макоклюев, Б. И. Методы и программные средства планирования электропотребления / Б. И. Макоклюев // Энергия единой сети. – 2022. – № 2 (63). – С. 58–69.
 75. Макоклюев, Б. И. Краткосрочное прогнозирование электропотребления в энергосистемах России / Б. И. Макоклюев, А. С. Полижаров, С. В. Локтионов // Электрические станции. – 2018. – № 4 (1041). – С. 24–35.
 76. Обеспечение надёжности расчётов по планированию и оперативной коррекции графиков потребления электрической мощности / А. С. Полижаров, Б. И. Макоклюев, А. В. Антонов // Энергетик. – 2021 – № 2. – С. 28–30.

77. International Best Practices in Solar and Wind Power Forecasting // GET.transform : Offic. site. – 2023. – 20 Nov. – URL: <https://www.get-transform.eu/international-best-practices-in-solar-and-wind-power-forecasting/> (date of access: 07.09.2024).
78. Grid Integration Serie: Variable Renewable Energy Forecasting. Scaling Up Renewable Energy Project / United States Agency for Intern. Development // Usaid : Off. site. – 2020. – July. – 41 P. – URL: https://www.usaid.gov/sites/default/files/2022-05/USAID_SURE_Variable-Renewable-Energy-Forecasting.pdf (date of access: 07.12.2022).
79. Сергеев, Н. Н. Обзор международного опыта в прогнозировании генерации возобновляемых источников энергии с помощью методов машинного обучения / Н. Н. Сергеев, П. В. Матренин. – DOI 10.21285/1814-3520-2023-2-354-36 // iPolytech Journal. – 2023. – Т. 27, № 2. – С. 354–369. – URL: <https://elibrary.ru/item.asp?id=54124532> (дата обращения: 05.07.2024).
80. A review of wind power and wind speed forecasting methods with different time horizons / S. S. Soman, H. Zareipour, O. Malik, P. Manda / DOI 10.1109/NAPS.2010.5619586 // North American power symposium, 2010 : conf. proc. (Arlington, 26–28 Sept 2010). – [Piscataway], 2010. – P. 1–8. – URL: <https://ieeexplore.ieee.org/abstract/document/5619586> (date of access: 07.12.2022).
81. A control strategy based on power forecasting micro-grid systems / A. Elmouatamid, R. Ouladsine, M. Bakhouya [et al.] // The Fifth IEEE Annual International Smart Cities Conference (ISC2, 2019) : conf. proc. (Casablanca, 14–17 Oct 2019) / Inst. of Electrical and Electronics Eng. (IEEE). – [Piscataway], 2019. – P. 735–740. – URL: <https://ieeexplore.ieee.org/document/9071722> (date of access: 07.12.2022).
82. Gensler, A. Wind Power Ensemble Forecasting: Performance Measures and Ensemble Architectures for Deterministic and Probabilistic Forecasts / A. Gensler. – Kassel : Kassel Univ. Press, 2019. – 214 p. – (Intelligent Embedded Systems ; Vol. 12). – ISBN 978-3737606363. – ISBN 3737606366. – URL: <https://www.uni-kassel.de/upress/online/OpenAccess/978-3-7376-0636-3.OpenAccess.pdf> (date of access: 15.03.2023).
83. Review of meta-heuristic algorithms for wind power prediction: Methodologies, applications and challenges / P. Lui, L. Ye, Y. Zhao [et al.]. – DOI 10.1016/j.apenergy.2021.11744 // Applied Energy. – 2021. – Vol. 301. – Art. 11744. – URL: <https://www.sciencedirect.com/science/article/abs/pii/S0306261921008369> (date of access: 15.03.2023).
84. Pu, Z. Numerical weather prediction basics: Models, numerical methods, and data assimilation / Z. Pu, E. Kalnay // Handbook of hydrometeorological ensemble forecasting / eds.: Q. Duan [et al.]. – Berlin : Heidelberg, 2018. – P. 1–31. – URL: https://www.inscc.utah.edu/~pu/6500_sp12/Pu-Kalnay2018_NWP_basics.pdf (date of access: 05.03.2023).
85. Белов, П. Н. Численные методы прогноза погоды : учебник / П. Н. Белов, Е. П. Борисенков, Б. Д. Панин. – Ленинград : Гидрометеоиздат, 1989. – 376 с. – ISBN 5-286-00148-3.

86. Гандин, Л. С. Численные методы краткосрочного прогноза погоды : учебник / Л. С. Гандин, А. С. Дубов. – Ленинград : Гидрометиздат. – 1968. – 427 с.
87. Al-Yahyai, S. Review of the use of Numerical Weather Prediction (NWP) Models for wind energy assessment / S. Al-Yahyai, Y. Charabi, A. Gastli. – DOI 10.1016/j.rser.2010.07.001 // Renewable and Sustainable Energy Rev. – 2010. – Vol. 14, Iss. 9. – P. 3192–3198. – URL: <https://www.sciencedirect.com/science/article/abs/pii/S1364032110001814> (date of access: 05.03.2023).
88. Orlanski, I. A rational subdivision of scales for atmospheric processes / I. Orlanski // Bulletin of the American Meteorological Society. – 1975. – Vol. 56, Iss. 5. – P. 527–530. – URL: <https://www.jstor.org/stable/26216020> (date of access: 12.03.2023).
89. Цепелев, В. Ю. Ансамблевый прогноз: составление, использование, интерпретация : учеб. пособие / В. Ю. Цепелев, О. Г. Анискина. – Санкт-Петербург : Издат.-полиграф. ассоц. высш. учеб. заведений, 2022. – 110 с. – ISBN 978-5-91155-150-6.
90. Yang, D. Summarizing ensemble NWP forecasts for grid operators: Consistency, elicibility, and economic value / D. Yang, J. Kleissl. – DOI 10.1016/j.ijforecast.2022.08.002 // International Journal of Forecasting. – 2023. – Vol. 39, Iss. 4. – P. 1640–1654. – URL: <https://www.sciencedirect.com/science/article/abs/pii/S016920702200111X> (date of access: 12.03.2023).
91. González-Sopeña, J. M., An overview of performance evaluation metrics for short-term statistical wind power forecasting / J. M. González-Sopeña, V. Pakrashi, B. Ghosh. – DOI 10.1016/j.rser.2020.110515 // Renewable and Sustainable Energy Reviews. – 2021. – Vol. 138. – Art. 110515. – URL: <https://www.sciencedirect.com/science/article/pii/S1364032120308005> (date of access: 05.01.2023).
92. Evaluation metrics for wind power forecasts: A comprehensive review and statistical analysis of errors / P. Piotrowski, I. Rutyna, D. Baczyński, M. Kopyt. – DOI 10.3390/en15249657 // Energies. – 2022. – Vol. 15, Iss. 24. – Art. 9657. – URL: <https://www.mdpi.com/1996-1073/15/24/9657> (date of access: 05.01.2023).
93. A protocol for standardising the performance evaluation of short-term wind power prediction models / H. Madsen, Pierre Pinson, G. Kariniotakis [et al.]. – DOI 10.1260/030952405776234599 // Wind Engineering. – 2005. – Vol. 29, Iss. 6. – P. 475–489. – URL: <https://minesparis-psl.hal.science/hal-00527248/document> (date of access: 05.01.2023).
94. An improved residual-based convolutional neural network for very short-term wind power forecasting / C. Yildiz, H. Acikgoz, D. Korkmaz, U. Budak. – DOI 10.1016/j.enconman.2020.113731 // Energy Conversion and Management. – 2021. – Vol. 228. – Art. 113731. – URL: <https://www.sciencedirect.com/science/article/abs/pii/S0196890420312553> (date of access: 12.03.2023).

95. Short term wind power forecasting using hybrid variational mode decomposition and multi-kernel regularized pseudo inverse neural network / J. Naik, S. Dash, P. K. Dash, R. Bisoi. – DOI 10.1016/j.renene.2017.10.111 // Renewable Energy. – 2018. – Vol. 118. – P. 180–212. – URL: <https://www.sciencedirect.com/science/article/abs/pii/S0960148117310893> (date of access: 12.03.2023).
96. A novel hybrid model for short-term wind power forecasting / P. Du, J. Wang, W. Yang, T. Niu. – DOI 10.1016/j.asoc.2019.03.035 // Applied Soft Computing. – 2019. – Vol. 80. – P. 93–196. – URL: <https://www.sciencedirect.com/science/article/abs/pii/S1568494619301619> (date of access: 12.03.2023).
97. Deterministic and probabilistic wind power forecasting using a variational Bayesian-based adaptive robust multi-kernel regression model / Y. Wang, Q. Hu, D. Meng, P. Zhu. – DOI 10.1016/j.apenergy.2017.09.043 // Applied energy. – 2017. – Vol. 208. – P. 1097–1112. – URL: <https://www.sciencedirect.com/science/article/abs/pii/S0306261917313259> (date of access: 17.03.2023).
98. A novel bidirectional mechanism based on time series model for wind power forecasting / Y. Zhao, L. Ye, Z. Li [et al.]. – DOI 10.1016/j.apenergy.2016.03.096 // Applied energy. – 2016. – Vol. 177. – P. 793–803. – URL: <https://www.sciencedirect.com/science/article/abs/pii/S0306261916304263> (date of access: 17.03.2023).
99. Lee D. Short-term wind power ensemble prediction based on Gaussian processes and neural networks / D. Lee, R. Baldick. – DOI 10.1109/TSG.2013.2280649 // IEEE Transactions on Smart Grid. – 2013. – Vol. 5, Iss. 1. – P. 501–510. – URL: <https://ieeexplore.ieee.org/abstract/document/6606922> (date of access: 21.03.2023).
100. Advanced ensemble methods using machine learning and deep learning for one-day-ahead forecasts of electric energy production in wind farms / P. Piotrowski, D. Baczyński, M. Kopyt, T. Gulczyński. – DOI 10.3390/en15041252 // Energies. – 2022. – Vol. 15, Iss. 4. – Art. 1252. – URL: <https://www.mdpi.com/1996-1073/15/4/1252> (date of access: 21.03.2023).
101. A nonparametric Bayesian framework for short-term wind power probabilistic forecast / W. Xie, P. Zhang, R. Chen, Z. Zhou / DOI 10.1109/TPWRS.2018.2858265 // IEEE Transactions on Power Systems. – 2018. – Vol. 34, Iss. 1. – P. 371–379. – URL: <https://ieeexplore.ieee.org/document/8417899> (date of access: 21.03.2023).
102. Безруких, П. П. Ветроэнергетика : справ.-метод. изд. / П. П. Безруких, П. П. Безруких (мл.), С. В. Грибков ; под общ. ред. П. П. Безруких. – Москва : Интехэнерго-Издат, 2014. – 299 с. – (Библиотека энергоэффективности и энергосбережения. Серия 3. Возобновляемая энергетика ; т. 8). – ISBN 978-5-98385-014-9.
103. Обухов, С. Г. Системы генерирования электрической энергии с использованием возобновляемых энергоресурсов: учеб. пособие / С. Г. Обухов. – Томск : Изд-во Томс. политехн. ун-та. – 2008. – 140 с.

104. Wind turbine technology: fundamental concepts of wind turbine engineering / ed by D. A. Spera. – New York : ASME Press, 1994. – 638 p. – ISBN 978-0791812051. – ISBN 0791812057.
105. Wind energy handbook / T. Burton, N. Jenkins, D. Sharpe, E. Bossanyi. – Hoboken : John Wiley & Sons, 2011. – 642 p. – ISBN 9780470699751.
106. A wind turbine on top of a hill : photo / photographer D. Vives // Unsplash : site. – URL: <https://unsplash.com/photos/a-wind-turbine-on-top-of-a-hill-SMwPMQrFM70> (date of access: 21.10.2024).
107. El-Ahmar M. H. Evaluation of factors affecting wind turbine output power / M. H. El-Ahmar, A. H. M. El-Sayed, A. M. Hemeida. – DOI 10.1109/MEPCON.2017.8301377 // 2017 Nineteenth International Middle East Power Systems Conference (MEPCON) : conf. proc. (Nas, 19–21 Dec 2017) / Inst. of Electrical and Electronics Eng. (IEEE). – [S. l.], 2017. – P. 1471–1476. – URL: <https://ieeexplore.ieee.org/abstract/document/8301377> (date of access: 09.06.2023).
108. Grogg, K. Harvesting the wind: The physics of wind turbines. Carleton College / K. Grogg // Physics and Astronomy Department. – 2005. – Vol. 4, Iss. 13. – URL: <http://large.stanford.edu/courses/2017/ph240/perry2/docs/grogg-13apr05.pdf> (date of access: 09.06.2023).
109. Hau, E. Wind turbines: fundamentals, technologies, application, economics. Vol. 2 / E. Hau, H. Renouard. – 2nd ed. – Berlin : Springer, 2006. – 801 p. – ISBN 3540242406. – ISBN 978-3540242406.
110. Understanding wind power technology: theory, deployment and optimization / ed. by A. Schaffarczyk. – Hoboken : John Wiley & Sons, 2014. – 488 p. – ISBN 978-1-118-64751-6.
111. ГОСТ Р 54418.1–2012. Возобновляемая энергетика. Ветроэнергетика. Установки ветроэнергетические. Часть 1. Технические требования : нац. стандарт Рос. Федерации : утв. Приказом Федер. агентства по техн. регулированию и метрологии от 20 сент. 2012 г. № 378-ст : дата введ. 2014-01-01 // Кодекс. Электронный фонд правовых нормативно-технических документов : офиц. сайт. – URL: <https://docs.cntd.ru/document/1200104895> (дата обращения: 09.06.2023).
112. ГОСТ Р 54418.12.1–2012. Возобновляемая энергетика. Ветроэнергетика. Установки ветроэнергетические. Часть 12-1. Измерение мощности, вырабатываемой ветроэлектрическими установками : нац. стандарт Рос. Федерации : утв. Приказом Федер. агентства по техн. регулированию и метрологии от 28 сент. 2011 г. № 379-ст : дата введ. 2012-07-01 // Кодекс. Электронный фонд правовых нормативно-технических документов : офиц. сайт. – URL: <https://docs.cntd.ru/document/1200091404> (дата обращения: 09.06.2023).
113. ГОСТ Р 54418.12.2–2012. Возобновляемая энергетика. Ветроэнергетика. Установки ветроэнергетические. Часть 12-2. Измерение показателей мощности ветроэнергетической установки с использованием анемометра, установленного на гондоле : нац. стандарт Рос. Федерации : утв. Приказом

- Федер. агентства по техн. регулированию и метрологии от 6 сент. 2013 г. № 1053-ст : дата введ. 2015-07-01. // Кодекс. Электронный фонд правовых нормативно-технических документов : офиц. сайт. – URL: <https://docs.cntd.ru/document/1200107582> (дата обращения: 09.06.2023)..
114. ГОСТ Р 54418.12.3–2012. Возобновляемая энергетика. Ветроэнергетика. Установки ветроэнергетические. Часть 12-3. Методы испытаний для определения количества вырабатываемой электроэнергии : нац. стандарт Рос. Федерации : утв. Приказом Федер. агентства по техн. регулированию и метрологии от 20 сент. 2012 г. № 380-ст. : дата введ. 2014-01-01 // Кодекс. Электронный фонд правовых нормативно-технических документов : офиц. сайт. – URL: <https://docs.cntd.ru/document/1200104415> (дата обращения: 09.06.2023).
 115. Heier, S. Grid integration of wind energy: onshore and offshore conversion systems / S. Heier. – 3rd ed. – Hoboken: John Wiley & Sons, 2014. – 528 p. – ISBN 9781119962946.
 116. Wagner, H. J. Introduction to wind energy systems. Basics, technology and operation / H. J. Wagner, J. Mathur. – Berlin : Springer, 2009. – 103 p. – ISBN 3642020224. – ISBN 978-3642020223.
 117. Salih, S. M. Performance analysis of wind turbine systems under different parameters effect / S. M. Salih, M. Q. Taha, M. K. Alawsaj // International Journal of Energy and Environment. – 2012. – Vol. 3, Iss. 6. – P. 895–904. – URL: https://www.researchgate.net/profile/Salih-Salih-4/publication/257827106_Performance_Analysis_of_Wind_Turbine_Systems_under_Different_Parameters_Effect/links/004635260de15cf93c000000/Performance-Analysis-of-Wind-Turbine-Systems-under-Different-Parameters-Effect.pdf (date of access: 12.06.2023).
 118. Ackermann, T. Wind power in power systems / T. Ackermann. – 2nd ed.– Hoboken : John Wiley & Son, 2012. – 1064 p. – ISBN 9780470974162.
 119. Shelquist R. An introduction to Air Density and Density Altitude Calculations / R. Shelquist. – 2010. – 2 Aug. // Internet Archive : site. – URL: https://web.archive.org/web/20101130082435/http://wahiduddin.net/calc/density_altitude.htm (date of access: 12.06.2023).
 120. Янченко, Г. А. Инженерные методы: расчёта плотности влажного воздуха / Г. А. Янченко, И. В. Кириченко // Горный информационно-аналитический бюллетень (научно-технический журнал). – 2008. – № 11. – С. 112–115. – URL: <https://www.elibrary.ru/item.asp?id=12363861> (дата обращения: 12.06.2023).
 121. Uncovering wind power forecasting uncertainty sources and their propagation through the whole modelling chain / J. Yan, C. Möhrlen, T. Göçmen [et al.]. – DOI 10.1016/j.rser.2022.112519 // Renewable and Sustainable Energy Reviews. – 2022. – № 165. – Art. 112519. – URL: <https://www.sciencedirect.com/science/article/pii/S1364032122004221> (date of access: 12.06.2023).

122. Manwell, J. F. Wind energy explained: theory, design and application / J. F. Manwell, J. G. McGowan, A. L. Rogers. – Hoboken : John Wiley & Sons, 2009. – 704 p. – ISBN 978-0-470-01500-1.
123. Pinson, P. Estimation and evaluation of prediction intervals of wind power, Estimation of the uncertainty in wind power forecasting : Ph.D. diss. / P. Pinson ; École Nationale Supérieure des Mines de Paris. – Paris, 2006. – URL: <https://pastel.hal.science/pastel-00002187/document> (date of access: 12.06.2023).
124. Bianchi, F. D. Wind turbine control systems: Principles, Modelling and Gain Scheduling Design / F. D. Bianchi, H. De Battista, R. J. Mantz. – London : Springer, 2007. – 228 p. – ISBN 1846284929 – ISBN 978-1846284922.
125. Wind power plants: Fundamentals, Design, Construction and Operation : textbook / eds. R. Gasch, J. Tvele. – 2nd. – Berlin : Springer Science & Business Media, 2011. – XVIII, 548 p. – ISBN 978-3-642-22937-4.
126. Renewable electricity and the grid. The challenge of variability / ed. G. Boyle. – [S. l.] : Routledge, 2009. – 244 p. – ISBN 9781844077892.
127. Brown, B. G. Time series models to simulate and forecast wind speed and wind power / B. G. Brown, R. W. Katz, A. H. Murphy. – DOI 10.1175/1520-0450(1984)023<1184:TSMTSA>2.0.CO;2 // Journal of Applied Meteorology and Climatology. – 1984. – Vol. 23, Iss. 8. – P. 1184–1195. – URL: https://journals.ametsoc.org/view/journals/apme/23/8/1520-0450_1984_023_1184_tsmtsa_2_0_co_2.xml (date of access: 15.06.2023).
128. Wegley, H. L. Test application of a semi-objective approach to wind forecasting for wind energy applications : № PNL-4403 / H. L. Wegley, W. J. Formica. – Richland : Pacific Northwest National Laboratory, 1983. – URL: <https://digital.library.unt.edu/ark:/67531/metadc1103217/m1/102/> (date of access: 15.06.2023).
129. Bossanyi, E. A. Short-term wind prediction using Kalman filters / E. A. Bossanyi // Wind Engineering. – 1985. – Vol. 9, Iss. 1. – P. 1–8. – URL: <https://www.jstor.org/stable/43749025> (date of access: 15.06.2023).
130. A review on the young history of the wind power short-term prediction / A. Costa, A. Crespo, J. Navarro [et al.]. – DOI 10.1016/j.rser.2007.01.015 // Renewable and Sustainable Energy Reviews. – 2008. – Vol. 12, Iss. 6. – P. 1725–1744. – URL: <https://www.sciencedirect.com/science/article/abs/pii/S1364032107000354> (date of access: 20.09.2023).
131. A critical review of wind power forecasting methods – past, present and future / S. Hanifi, X. Liu, Z. Lin, S. Lotfian. – DOI 10.3390/en13153764 // Energies. – 2020. – Vol. 13, Iss. 15. – Art. 3764. – URL: <https://www.mdpi.com/1996-1073/13/15/3764> (date of access: 20.09.2023).
132. Liu, H. Data processing strategies in wind energy forecasting models and applications: A comprehensive review / H. Liu, C. Chen. – DOI 10.1016/j.apenergy.2019.04.188 // Applied Energy. – 2019. – Vol. 249. – P. 392–408. – URL:

- <https://www.sciencedirect.com/science/article/abs/pii/S0306261919308517> (date of access: 20.09.2023).
133. Data Enrichment as a Method of Data Preprocessing to Enhance Short-Term Wind Power Forecasting / Y. Zhou, L. Ma, W. Ni, C. Yu. – DOI 10.3390/en16052094 // *Energies*. – 2023. – Vol. 16, Iss. 5. – Art. 2094. – URL: <https://www.mdpi.com/1996-1073/16/5/2094> (date of access: 20.09.2023).
 134. Giebel, G. Wind power forecasting – a review of the state of the art / G. Giebel, G. Kariniotakis // *Renewable energy forecasting. From Models to Applications* / ed. G. Kariniotakis. – Cambridge, 2017. – P. 59–109.
 135. Maldonado-Correa, J. Wind power forecasting: A systematic literature review / J. Maldonado-Correa, J. C. Solano, M. Rojas-Moncayo. – DOI 10.1177/0309524X198916 // *Wind Engineering*. – 2021. – Vol. 45, Iss. 2. – P. 413–426. – URL: <https://journals.sagepub.com/doi/full/10.1177/0309524X19891672> (date of access: 20.09.2023).
 136. Bazionis, I. K. Review of deterministic and probabilistic wind power forecasting: Models, methods, and future research / I. K. Bazionis, P. S. Georgilakis. – DOI 10.3390/electricity2010002 // *Electricity*. – 2021. – Vol. 2, Iss. 1. – P. 13–47. – URL: <https://www.mdpi.com/2673-4826/2/1/2> (date of access: 05.04.2023).
 137. Zhang, Y. Review on probabilistic forecasting of wind power generation / Y. Zhang, J. Wang, X. Wang. – DOI 10.1016/j.rser.2014.01.033 // *Renewable and Sustainable Energy Reviews*. – 2014. – Vol. 32. – P. 255–270. – URL: <https://www.sciencedirect.com/science/article/abs/pii/S1364032114000446?via%3Dihub> (date of access: 05.04.2023).
 138. Artipoli, G., Durante F. Physical modeling in wind energy forecasting. / G. Artipoli, F. Durante // *DEWI Magazin*. – 2014. – Vol. 44. – P. 10–15. – URL: https://www.researchgate.net/profile/Gianluca-Artipoli/publication/312159717_Physical_Modeling_in_Wind_Energy_Forecasting/links/587345f208ae8fce49238d23/Physical-Modeling-in-Wind-Energy-Forecasting.pdf (date of access: 10.09.2023).
 139. Bhaskar, K. A WNN-assisted wind power forecasting using feed-forward neural network / K. Bhaskar, S. N. Singh. – DOI 10.1109/TSTE.2011.2182215 // *IEEE transactions on sustainable energy*. – 2012. – Vol. 3, Iss. 2. – P. 306–315. – URL: <https://ieeexplore.ieee.org/abstract/document/6170987> (date of access: 10.09.2023).
 140. Mandzhieva, R. Data-driven applications for wind energy analysis and prediction: The case of «La Haute Borne» wind farm / R. Mandzhieva, R. Subhankulova. – DOI 10.1016/j.dche.2022.100048 // *Digital Chemical Engineering*. – 2022. – Vol. 4. – Art. 100048. – URL: <https://www.sciencedirect.com/science/article/pii/S2772508122000382> (date of access: 10.09.2023).

141. Kisvari, A. Wind power forecasting – A data-driven method along with gated recurrent neural network / A. Kisvari, Z. Lin, X. Liu. – DOI 10.1016/j.renene.2020.10.119 // Renewable Energy. – 2021. – Vol. 163. – P. 1895–1909. – URL: <https://www.sciencedirect.com/science/article/abs/pii/S0960148120316918> (date of access: 10.09.2023).
142. Huang, B. Wind power forecasting using attention-based recurrent neural networks: a comparative study / B. Huang, Y. Liang, X. Qiu. – DOI 10.1109/ACCESS.2021.3065502 // IEEE Access. – 2021. – Vol. 9. – P. 40432–40444. – URL: <https://ieeexplore.ieee.org/abstract/document/9374915> (date of access: 10.09.2023).
143. Short term wind power prediction using ANFIS / Y. Kassa, J. H. Zhang, D. H. Zheng, D. Wei. – DOI 10.1109/ICPRE.2016.7871238 // 2016 IEEE international conference on power and renewable energy (ICPRE) : conf. proc. (Shanghai, 21–23 Oct 2016) / Inst. of Electrical and Electronics Eng. (IEEE). – [S. l.], 2016. – P. 388–393. – URL: <https://ieeexplore.ieee.org/abstract/document/7871238> (date of access: 10.09.2023).
144. Amroune, M. Support vector regression-bald eagle search optimizer-based hybrid approach for short-term wind power forecasting / M. Amroune. – DOI 10.1186/s44147-022-00161-w // Journal of Engineering and Applied Science. – 2022. – Vol. 69, Iss. 1. – Art. 107. – URL: <https://link.springer.com/article/10.1186/s44147-022-00161-w#citeas> (date of access: 11.09.2023).
145. Hao, J. Wind power short-term forecasting model based on the hierarchical output power and poisson re-sampling random forest algorithm / J. Hao, C. Zhu, X. Guo. – DOI 10.1109/ACCESS.2020.3048382 // IEEE Access. – 2020. – Vol. 9. – P. 6478–6487. – URL: <https://ieeexplore.ieee.org/abstract/document/9311608> (date of access: 11.09.2023).
146. Wind Power Forecasting Based on ARIMA-LGARCH Model / S. Tian, Y. Fu, P. Ling [et al.]. – DOI 10.1109/POWERCON.2018.8601740 // 2018 International Conference on Power System Technology (POWERCON) : conf. proc. (Guangzhou, 6–8 Nov 2018) / Inst. of Electrical and Electronics Eng. (IEEE). – [S. l.], 2018. – P. 1285–1289. – URL: <https://ieeexplore.ieee.org/document/8601740> (date of access: 11.09.2023).
147. Pessanha, J. F. M. Short-Term Wind Power Forecasting Based On Quantile Regression / J. F. M. Pessanha, V. L. O. Castellani, V. A. Andrade // Brazil Windpower – Conference & Exhibition : conf. proc. (Rio de Janeiro 2017, 29–31 Aug 2017) / Departamento de Otimização Energética e Meio Ambiente. – Rio de Janeiro, 2017. – URL: <https://www.abeeolica.org.br/wp-content/uploads/2017/07/Short-Term-Wind-Power-Forecasting-Based-On-Quantile-Regression.pdf> (date of access: 11.09.2023).
148. An application of ensemble/multi model approach for wind power production forecasting / S. Alessandrini, P. Pinson, R. Hagedorn [et al.]. – DOI 10.5194/asr-6-

- 35-2011 // *Advances in Science and Research*. – 2011. – Vol. 6, Iss. 1. – P. 35–37. – URL: <https://asr.copernicus.org/articles/6/35/2011/> (date of access: 15.09.2023).
149. A hybrid wind power forecasting approach based on Bayesian model averaging and ensemble learning / G. Wang, R. Jia, J. Liu, H. Zhang. – DOI 10.1016/j.renene.2019.07.166 // *Renewable energy*. – 2020. – Vol. 145. – P. 2426–2434. – URL: <https://www.sciencedirect.com/science/article/abs/pii/S0960148119311863> (date of access: 15.09.2023).
 150. A combined model for short-term wind power forecasting based on the analysis of numerical weather prediction data / B. He, R. Jia, J. Liu, H. Zhang // *Energy Reports*. – 2022. – Vol. 8. – P. 929–939. – URL: <https://www.sciencedirect.com/science/article/abs/pii/S0960148119311863> (date of access: 15.09.2023).
 151. Kumar, D. An ensemble model for short-term wind power forecasting using deep learning and gradient boosting algorithms / D. Kumar, R. Abhinav, N. Pindoriya. – DOI 10.1109/NPSC49263.2020.9331902 // 21st National Power Systems Conference (NPSC) : conf. proc. (Gandhinagar, 17–19 Dec 2020) / Inst. of Electrical and Electronics Eng. (IEEE). – [S. l.], 2020. – P. 1–6. – URL: <https://ieeexplore.ieee.org/abstract/document/9331902> (date of access: 15.09.2023).
 152. Li, S. Wind power forecasting using neural network ensembles with feature selection / S. Li, P. Wang, L. Goel. – DOI 10.1109/TSTE.2015.2441747 // *IEEE Transactions on sustainable energy* : conf. proc. (Piscataway, 10 July 2015) / Inst. of Electrical and Electronics Eng. (IEEE). – [S. l.], 2015. – Vol. 6, Iss. 4. – P. 1447–1456. – URL: <https://ieeexplore.ieee.org/abstract/document/7154499> (date of access: 15.09.2023).
 153. A three-stage ensemble short-term wind power prediction method based on VMD-WT transform and SDAE deep learning / P. Xiaosheng, Z. Zuowei, X. Qiyu [et al.]. – DOI 10.1109/ICPSAsia48933.2020.9208460 // *IEEE/IAS industrial and commercial power system asia (I&CPS asia)* : conf. proc. (Weihai, 13–15 July 2020) / Inst. of Electrical and Electronics Eng. (IEEE). – [S. l.], 2020. – P. 1350–1356. – URL: <https://ieeexplore.ieee.org/abstract/document/9208460> (date of access: 15.09.2023).
 154. New hybrid deep neural architectural search-based ensemble reinforcement learning strategy for wind power forecasting / S. M. J. Jalali, G. J. Osório, S. Ahmadian [et al.]. – DOI 10.1109/TIA.2021.3126272 // *IEEE Transactions on Industry Applications*. – 2021. – Vol. 58, Iss. 1. – P. 15–27. – URL: <https://ieeexplore.ieee.org/abstract/document/9609674> (date of access: 16.09.2023).
 155. Harrou, F. Wind power prediction using bootstrap aggregating trees approach to enabling sustainable wind power integration in a smart grid / F. Harrou, A. Saidi, Y. Sun. – DOI 10.1016/j.enconman.2019.112077 // *Energy Conversion and Management*. – 2019. – Vol. 201. – Art. 112077. – URL: <https://doi.org/10.1016/j.enconman.2019.112077> (date of access: 16.09.2023).
 156. A novel hybrid model based on nonlinear weighted combination for short-term wind power forecasting / J. Duan, P. Wang, W. Ma [et al.]. – DOI 10.1016/j.ijepes.2021.107452 // *International Journal of Electrical Power & Energy*

- Systems. – 2022. – Vol. 134. – Art. 107452. – URL: <https://www.sciencedirect.com/science/article/abs/pii/S0142061521006918> (date of access: 16.09.2023).
157. Wang, H. K. A hybrid forecasting model based on CNN and informer for short-term wind power / H. K. Wang, K. Song, Y. Cheng. – DOI 10.3389/fenrg.2021.788320 // *Frontiers in Energy Research*. – 2022. – T. 9. – Art. 788320. – URL: <https://www.frontiersin.org/journals/energy-research/articles/10.3389/fenrg.2021.788320/full> (date of access: 16.09.2023).
 158. Tasdemir, S. A. The effect on the wind power performance of different normalization methods by using multilayer feed-forward backpropagation neural network / S. Tasdemir, B. Yaniktepe, A. Burak Guher. – DOI 10.31593/ijeat.464210 // *International Journal of Energy Applications and Technologies*. – 2018. – Vol. 5, Iss. 3. – P. 131–139. – URL: <https://dergipark.org.tr/en/pub/ijeat/article/464210> (date of access: 20.09.2023).
 159. Ultra-Short-Term Wind Power Prediction Based on the ZS-DT-PatchTST Combined Model / Y. Gao, F. Xing, L. Kang [et al.]. – DOI 10.3390/en17174332 // *Energies*. – 2024. – Vol. 17, Iss. 17. – Art. 4332. – URL: <https://www.mdpi.com/1996-1073/17/17/4332> (date of access: 11.07.2024).
 160. Makhloufi, S. Wind speed and wind power forecasting using wavelet denoising-GMDH neural network / S. Makhloufi, G. G. Pillai. – DOI 10.1109/ICEE-B.2017.8192155 // 2017 5th International Conference on Electrical Engineering-Boumerdes (ICEE-B) : conf. proc. (Boumerdes, 29–31 Oct. 2017) / Inst. of Electrical and Electronics Eng. (IEEE). – [S. l.], 2017. – P. 1–5. – URL: <https://ieeexplore.ieee.org/abstract/document/8192155> (date of access: 20.09.2023).
 161. Kahraman, A. Wind power forecasting using LSTM incorporating fourier transformation based denoising technique / A. Kahraman, G. Yang, P. Hou. – DOI 10.1049/icp.2021.2604 // 20th International Workshop on Large-Scale Integration of Wind Power into Power Systems as well as on Transmission Networks for Offshore Wind Power Plants (WIW 2021) : conf. proc. (Hybrid, 29–30 Sept 2021) / Institution of Engineering and Technology. – [S. l.], 2021. – P. 94–98. – URL: <https://ieeexplore.ieee.org/abstract/document/9770652> (date of access: 20.09.2023).
 162. Han, L. Wind power forecasting based on principle component phase space reconstruction / L. Han, C. E. Romero, Z. Yao. – DOI 10.1016/j.renene.2015.03.037 // *Renewable Energy*. – 2015. – Vol. 81. – P. 737–744. – URL: <https://www.sciencedirect.com/science/article/abs/pii/S0960148115002190> (date of access: 20.09.2023).
 163. A new wind power prediction method based on ridgelet transforms, hybrid feature selection and closed-loop forecasting / H. Leng, X. Li, J. Zhu [et al.]. – DOI 10.1016/j.aei.2018.02.006 // *Advanced Engineering Informatics*. – 2018. – Vol. 36. – P. 20–30. – URL: <https://www.sciencedirect.com/science/article/abs/pii/S1474034617303154> (date of access: 21.09.2023).

164. Wind power short-term prediction based on LSTM and discrete wavelet transform / Y. Liu, L. Guan, C. Hou [et al.]. – DOI 10.3390/app9061108 // Applied Sciences. – 2019. – Vol. 9, Iss. 6. – Art. 1108. – <https://www.mdpi.com/2076-3417/9/6/1108> (date of access: 21.09.2023).
165. Wang, C. Wind power forecasting based on singular spectrum analysis and a new hybrid Laguerre neural network / C. Wang, H. Zhang, P. Ma. – DOI 10.1016/j.apenergy.2019.114139 // Applied Energy. – 2020. – Vol. 259. – Art. 114139. – URL: <https://www.sciencedirect.com/science/article/abs/pii/S0306261919318264> (date of access: 21.09.2023).
166. Liu, H. Smart multi-step deep learning model for wind speed forecasting based on variational mode decomposition, singular spectrum analysis, LSTM network and ELM / H. Liu, X. Mi, Y. Li. – DOI 10.1016/j.enconman.2018.01.010 // Energy Conversion and Management. – 2018. – Vol. 159. – P. 54–64. – URL: <https://www.sciencedirect.com/science/article/abs/pii/S0196890418300104> (date of access: 21.09.2023).
167. Wind power short-term forecasting hybrid model based on CEEMD-SE method / K. Wang, D. Niu, L. Sun [et al.]. – DOI 10.3390/pr7110843 // Processes. – 2019. – Vol. 7, Iss. 11. – P. 843. – URL: <https://www.mdpi.com/2227-9717/7/11/843> (date of access: 23.09.2023).
168. Short term wind power forecasting using Hilbert-Huang Transform and artificial neural network / J. Shi, W. Lee, Y. Liu [et al.]. – DOI 10.1109/DRPT.2011.5993881 // 2011 4th International Conference on Electric Utility Deregulation and Restructuring and Power Technologies (DRPT) : conf. proc. (Weihai, 6–9 July 2011) / Inst. of Electrical and Electronics Eng. (IEEE). – [S. l.], 2011. – P. 162–167. – URL: <https://ieeexplore.ieee.org/abstract/document/5993881> (date of access: 23.09.2023).
169. A new combinatory approach for wind power forecasting / O. Abedinia, M. Bagheri, M. S. Naderi, N. Ghadimi. – DOI 10.1109/JSYST.2019.2961172 // IEEE Systems Journal. – 2020. – Vol. 14, Iss. 3. – P. 4614–4625. – URL: <https://ieeexplore.ieee.org/abstract/document/8963864> (date of access: 23.09.2023).
170. Short-term wind power forecasting using empirical mode decomposition and RBFNN / Z. W. Zheng, W.-Y. Yuan-Yi Chen, X.-W. Zhou, M.-M. Huo. – DOI 10.12720/sgce.2.2.192-199 // International Journal of Smart Grid and Clean Energy. – 2013. – Vol. 2, Iss. 2. – P. 192–99. – URL: <https://www.ijsgce.com/index.php?m=content&c=index&a=show&catid=29&id=99> (date of access: 23.09.2023).
171. The hybridization of ensemble empirical mode decomposition with forecasting models: Application of short-term wind speed and power modeling / N. Bokde, A. Feijóo, N. Al-Ansari [et al.]. – DOI 10.3390/en13071666 // Energies. – 2020. – Vol. 13, Iss. 7. – Art. 1666. – URL: <https://www.mdpi.com/1996-1073/13/7/1666> (date of access: 23.09.2023).
172. Wind power forecast based on variational mode decomposition and long short term memory attention network / X. Zhou, C. Liu, Y. Luo, [et al.]. – DOI 10.1016/j.egy.2022.08.159 // Energy Reports. – 2022. – Vol. 8. –

- P. 922–931. – URL: <https://www.sciencedirect.com/science/article/pii/S235248472201602X> (date of access: 23.09.2023).
173. Qin, L. Weather division-based wind power forecasting model with feature selection / L. Qin, Y. Xiong, K. Liu. – DOI 10.1049/iet-rpg.2019.0263 // IET Renewable Power Generation. – 2019. – Vol. 13, Iss. 16. – P. 3050–3060. – URL: <https://ietresearch.onlinelibrary.wiley.com/doi/full/10.1049/iet-rpg.2019.0263> (date of access: 26.09.2023).
 174. Novel forecasting model based on improved wavelet transform, informative feature selection, and hybrid support vector machine on wind power forecasting / Z. Liu, M. Hajiali, A. Torabi, [et al.] // Journal of Ambient Intelligence and Humanized Computing. – 2018. – Vol. 9. – P. 1919–1931. – URL: <https://link.springer.com/article/10.1007/s12652-018-0886-0> (date of access: 26.09.2023).
 175. Cali U. Short-term wind power forecasting using long-short term memory based recurrent neural network model and variable selection / U. Cali, V. Sharma. – DOI 10.12720/sgce.8.2.103-110 // International Journal of Smart Grid and Clean Energy. – 2019. – Vol. 8, Iss. 2. – P. 103–110. – URL: <https://www.ijsgce.com/index.php?m=content&c=index&a=show&catid=78&id=418> (date of access: 26.09.2023).
 176. Wind power prediction based on variational mode decomposition and feature selection / G. Zhang, B. Xu, H. Liu [et al.]. – DOI 10.35833/MPCE.2020.000205 // Journal of Modern Power Systems and Clean Energy. – 2020. – Vol. 9, Iss. 6. – P. 1520–1529. – URL: <https://ieeexplore.ieee.org/abstract/document/9205720> (date of access: 26.09.2023).
 177. An improved random forest model of short-term wind-power forecasting to enhance accuracy, efficiency, and robustness / K. Shi, Y. Qiao, W. Zhao [et al.]. – DOI 10.1002/we.2261 // Wind energy. – 2018. – Vol. 21, Iss. 12. – P. 1383–1394. – URL: <https://onlinelibrary.wiley.com/doi/full/10.1002/we.2261> (date of access: 26.09.2023).
 178. Hybrid wind power forecasting based on extreme learning machine and improved TLBO algorithm / W. Xue, C. Wang, J. Tian, K. Li. – DOI 10.1063/5.0020759 // Journal of renewable and sustainable energy. – 2020. – Vol. 12, Iss. 5. – URL: <https://pubs.aip.org/aip/jrse/article/12/5/053309/1017500/Hybrid-wind-power-forecasting-based-on-extreme> (date of access: 28.09.2023).
 179. Short-term wind power prediction for wind farm clusters based on SFFS feature selection and BLSTM deep learning / X. Peng, K. Cheng, J. Lang // Energies. – 2021. – Vol. 14, Iss. 7. – Art. 1894. – URL: <https://www.mdpi.com/1996-1073/14/7/1894> (date of access: 28.09.2023).
 180. Wind power production forecasting using ant colony optimization and extreme learning machines / M. Carrillo, J. Del Ser, M. N. Bilbao [et al.] // Intelligent Distributed Computing. / eds.: M. Ivanović [et al.]. – [S. l.], 2018. – Vol. 11. – P.

- 175–184. – URL: <https://link.springer.com/content/pdf/10.1007/978-3-319-66379-1.pdf> (date of access: 28.09.2023).
181. Wu, Q. Wind power generation forecasting using least squares support vector machine combined with ensemble empirical mode decomposition, principal component analysis and a bat algorithm / Q. Wu, C. Peng. – DOI 10.3390/en9040261 // *Energies*. – 2016. – Vol. 9, Iss. 4. – P. 261–280. – URL: <https://www.mdpi.com/1996-1073/9/4/261> (date of access: 28.09.2023).
 182. Shao, H. A novel deep learning approach for short-term wind power forecasting based on infinite feature selection and recurrent neural network / H. Shao, X. Deng, Y. Jiang. – DOI 10.1063/1.5024297 // *Journal of renewable and sustainable energy*. – 2018. – Vol. 10, Iss. 4. – Art. 043303. – URL: <https://pubs.aip.org/aip/jrse/article/10/4/043303/383869/A-novel-deep-learning-approach-for-short-term-wind> (date of access: 28.09.2023).
 183. Zheng, H. A XGBoost model with weather similarity analysis and feature engineering for short-term wind power forecasting / H. Zheng, Y. Wu. – DOI 10.3390/app9153019 // *Applied Sciences*. – 2019. – Vol. 9, Iss. 15. – P. 3019–3031. – URL: <https://www.mdpi.com/2076-3417/9/15/3019> (date of access: 01.10.2023).
 184. Wind power prediction for wind farm clusters based on the multi-feature similarity matching method / X. Peng, Yu. Chen, K. Cheng [et al.]. – DOI 10.1109/TIA.2020.3010776 // 2019 IEEE Industry Applications Society Annual Meeting. – Baltimore, 2019. – P. 1–11. – URL: <https://ieeexplore.ieee.org/document/8912320> (date of access: 01.10.2023).
 185. Short-term wind power forecasts by a synthetical similar time series data mining method / G. Sun, Ch. Jiang, P. Cheng [et al.]. – DOI 10.1016/j.renene.2017.08.071 // *Renewable energy*. – 2018. – Vol. 115. – P. 575–584. – URL: <https://www.sci-hub.ru/10.1016/j.renene.2017.08.071> (date of access: 01.10.2023).
 186. Cao, Y. Multi-step wind power forecasting model using LSTM networks, similar time series and LightGBM / Y. Cao, L. Gui. – DOI 10.1109/ICSAI.2018.8599498 // 2018 5th International Conference on Systems and Informatics (ICSAI). – IEEE, 2018. – P. 192–197. – URL: <https://www.sci-hub.ru/10.1109/ICSAI.2018.8599498?ysclid=m67nbsf4i7467168366> (date of access: 01.10.2023).
 187. Ultra-short-term prediction of wind power based on sample similarity analysis / C. Miao, H. Li, X. Wang [et al.]. – DOI 10.1109/ACCESS.2021.3080140 // *IEEE Access*. – 2021. – Vol. 9. – P. 72730–72742. – URL: <https://ieeexplore.ieee.org/document/9430545> (date of access: 01.10.2023).
 188. Feurer, M. Hyperparameter optimization / M. Feurer, F. Hutter. – DOI 10.1007/978-3-030-05318-5_1 // *Automated Machine Learning*. – Springer, 2019. – P. 3–33. – (The Springer Series on Challenges in Machine Learning (SSCML)). – URL: https://link.springer.com/chapter/10.1007/978-3-030-05318-5_1 (date of access: 18.02.2023).
 189. Zaytsev, A. A. Sparse gaussian processes regression for the data fusion problem / A. A. Zaytsev, E. V. Burnaev, Yu A. Yanovich // *Intellectualization of information*

- processing IOI : 9th international conference (Budva, Sept 16–22, 2012) : collection of papers. – Moscow, 2012. – P. 101–104.
190. Бурнаев Е. В. Регрессия на основе нестационарных гауссовских процессов с байесовской регуляризацией / Е. В. Бурнаев, М. Е. Панов, А. А. Зайцев // Информационные процессы. – 2015. – Т. 15, №. 3. – С. 298–313.
 191. Wang J. An intuitive tutorial to Gaussian processes regression / J. Wang. – DOI 10.1109/mcse.2023.3342149 // Computing in Science & Engineering.– Computing in Science and Engineering. – 2023. – Vol 25, Iss. 4. – P. 4–11. – URL: <https://ieeexplore.ieee.org/document/10360364> (date of access: 15.09.2022).
 192. Williams C. K. Gaussian Processes for Machine Learning / C. K. Williams, C. E. Rasmussen. – London : MIT Press, 2006. – 248 p. – URL: <https://gaussianprocess.org/gpml/chapters/RW.pdf> (date of access: 17.09.2022). – ISBN 026218253X.
 193. Duvenaud D. Automatic model construction with Gaussian processes : diss. ... D-r of Philosophy / D. Duvenaud ; University of Cambridge. – London, 2014. – 144 p. – DOI 10.17863/CAM.14087.
 194. Machine Learning: A First Course for Engineers and Scientists / A. Lindholm, N. Wahlström, F. Lindsten, T. B. Schön ; Cambridge University Press. – New York, 2022. – 338 p. – ISBN 110-88-436-03. – ISBN 9-781-10-8843-607.
 195. Statistics and Machine Learning Toolbox: Analyze and model data using statistics and machine learning // MathWorks : site. – 2024. – URL: <https://www.mathworks.com/help/stats/index.html> (date of access: 25.11.2024).
 196. Short-Term Wind Power Forecasting Using Gaussian Processes / Chen N., Zh. Qian, X. Meng, I. T. Nabney // Proceedings of the Twenty-Third International Joint Conference on Artificial Intelligence (IJCAI). – 2013. – P. 2790–2796. – URL: <https://www.ijcai.org/Proceedings/13/Papers/411.pdf> (date of access: 17.09.2022).
 197. Wind power forecasts using Gaussian processes and numerical weather prediction / N. Chen, Zh. Qian, I. T. Nabney, X. Meng. – DOI 10.1109/TPWRS.2013.2282366 // IEEE Transactions on Power Systems. – 2013. – Vol. 29, Iss. 2. – P. 656–665. – URL: <https://ieeexplore.ieee.org/document/6617679> (date of access: 19.09.2022).
 198. Kou, P. Sparse online warped Gaussian process for wind power probabilistic forecasting / P. Kou, F. Gao, X. Guan. – DOI 10.1016/j.apenergy.2013.03.038 // Applied energy. – 2013. – Vol. 108. – P. 410–428. – URL: <https://www.sciencedirect.com/science/article/abs/pii/S0306261913002341> (date of access: 19.09.2022).
 199. Probabilistic wind power forecasting with online model selection and warped gaussian process / P. Kou, D. Liang, F. Gao, L. Gao. – DOI 10.1016/j.enconman.2014.04.051 // Energy conversion and Management. – 2014. – Vol. 84, Iss. 2. – P. 649–663. – URL: <https://eprints.whiterose.ac.uk/142979/1/07274766.pdf> (date of access: 19.09.2022).
 200. Hybrid probabilistic wind power forecasting using temporally local Gaussian process / J. Yan, K. Li, E-W Bai [et al.]. – DOI 10.1109/TSTE.2015.2472963 // IEEE Transactions on sustainable energy. – 2015. – Vol. 7, Iss. 1. – P. 87–95. – URL:

- https://www.researchgate.net/publication/282468913_Hybrid_Probabilistic_Wind_Power_Forecasting_Using_Temporally_Local_Gaussian_Process (date of access: 21.09.2022).
201. Time series wind power forecasting based on variant Gaussian Process and TLBO / J. Yan, K. Lia, E. Bai [et al.]. – DOI 10.1016/j.neucom.2015.12.081 // *Neurocomputing*. – 2016. – Vol. 189. – P. 135–144. – URL: <https://eprints.whiterose.ac.uk/135733/7/Variant%20Gaussian%20Process%20and%20TLBO.pdf> (date of access: 21.09.2022).
 202. Hoolohan, V. R. The use of Gaussian process regression for wind forecasting in the UK : diss. ... D-r of Philosophy / V. R. Hoolohan ; University of Leeds [et al.]. – University of Leeds, 2018. – 235 p. – URL: <https://etheses.whiterose.ac.uk/21544/1/VH%20Thesis%20-%20corrected.pdf> (date of access: 21.09.2022).
 203. Liu, T. Wind power prediction with missing data using Gaussian process regression and multiple imputation / T. Liu, H. Wei, K. Zhang. – DOI 10.1016/j.asoc.2018.07.027 // *Applied Soft Computing*. – 2018. – Vol. 71. – P. 905–916. – URL: <https://www.sciencedirect.com/science/article/abs/pii/S1568494618304150?via%3Dihub> (date of access: 21.09.2022).
 204. Ahmadpour, A. Gaussian models for probabilistic and deterministic Wind Power Prediction: Wind farm and regional / A. Ahmadpour, S. G. Farkoush. – DOI 10.1016/j.ijhydene.2020.07.081 // *International Journal of Hydrogen Energy*. – 2020. – Vol. 45, Iss. 51. – P. 27779–27791. – URL: <https://www.sciencedirect.com/science/article/abs/pii/S0360319920326562?via%3Dihub> (date of access: 21.09.2022).
 205. Probabilistic wind power forecasting using selective ensemble of finite mixture Gaussian process regression models / H. Jin, L. Shi, X.Chen [et al.]. – DOI 10.1016/j.renene.2021.04.028 // *Renewable Energy*. – 2021. – Vol. 174. – P. 1–18. – URL: <https://www.sci-hub.ru/10.1016/j.renene.2021.04.028?ysclid=m67sv2tc68925122052> (date of access: 21.09.2022).
 206. Bishop, C. M. *Neural networks for pattern recognition* / C. M. Bishop. – New York : Oxford University Press, 1995. – XVII, 482 p. – ISBN 0198538499.
 207. Lee, C. W. Training feedforward neural networks: An algorithm giving improved generalization / C. W. Lee. – DOI 10.1016/s0893-6080(96)00071-8 // *Neural Networks*. – 1997. – Vol. 10, Iss. 1. – P. 61–68. – URL: <https://www.sciencedirect.com/science/article/abs/pii/S0893608096000718?via%3Dihub> (date of access: 25.09.2022).
 208. Sazlı, M. H. A brief review of feed-forward neural networks / M. H. Sazlı. – DOI 10.1501/0003168 // *Communications Faculty of Sciences University of Ankara Series A2-A3*. – 2006. – Vol. 50, Iss. 1. – P. 11–17. – URL: https://www.researchgate.net/publication/228394623_A_brief_review_of_feed-forward_neural_networks (date of access: 25.09.2022).

209. Aldakheel, F. Feed-forward neural networks for failure mechanics problems / F. Aldakheel, R. Satari, P. Wriggers. – DOI 10.3390/app11146483 // Applied Sciences. – 2021. – Vol. 11, Iss. 14. – P. 6483. – URL: <https://www.mdpi.com/2076-3417/11/14/6483> (date of access: 25.09.2022).
210. McGonagle, J. Feedforward Neural Networks / J. McGonagle, J. A. García, S. Mollick // Brilliant : site. – URL: <https://brilliant.org/wiki/feedforward-neural-networks/> (date of access: 25.09.2022).
211. Godinho, M. Comparative performance of AI methods for wind power forecast in Portugal / M. Godinho, R. Castro. – DOI 10.1002/we.2556 // Wind Energy. – 2021. – Vol. 24, Iss. 1. – P. 39–53. – URL: <https://docs.wind-watch.org/wind-power-forecasting.pdf> (date of access: 25.09.2022).
212. Medina, S. V. Performance improvement of artificial neural network model in short-term forecasting of wind farm power output / S. V. Medina, U. P. Ajenjo. – DOI 10.35833/MPCE.2018.000792 // Journal of Modern Power Systems and Clean Energy. – 2020. – Vol. 8, Iss. 3. – P. 484–490. – URL: <https://ieeexplore.ieee.org/document/9082068> (date of access: 25.09.2022).
213. A GA-BP hybrid algorithm based ANN model for wind power prediction / Y. Kassa, J. Zhang, D. Zheng, Dan Wei. – DOI 10.1109/SEGE.2016.7589518 // 2016 IEEE Smart Energy Grid Engineering (SEGE). – [S. l.], 2016. – P. 158–163.
214. Short-term wind power forecast based on cluster analysis and artificial neural networks / J. Lorenzo, J. Méndez, M. Castrillón, D. Hernández // Advances in Computational Intelligence : 11th International Work-Conference on Artificial Neural Networks : IWANN–2011 (Torremolinos-Málaga, June 8–10, 2011) : Proceedings. – Berlin, 2011. – Pt. 1. – P. 191–198.
215. Wind turbine power output prediction model design based on artificial neural networks and climatic spatiotemporal data / B. Bilal, M. Ndong, K. H. Adjallah [et al.]. – DOI 10.1109/ICIT.2018.8352329 // 2018 IEEE International Conference on Industrial Technology (ICIT) (20–22 Feb 2018). – [S. l.], 2018. – P. 1085–1092. – URL: <https://ieeexplore.ieee.org/document/8352329> (date of access: 25.09.2022).
216. Вапник, В. Н. Восстановление зависимостей по эмпирическим данным / В. Н. Вапник. – Москва : Наука, 1979. – 447 с.
217. Вьюгин, В. В. Математические основы теории машинного обучения и прогнозирования / В. В. Вьюгин ; Москов. физ.-техн. ин-т, Лаб. структур. методов анализа данных в предсказат. моделировании. – Москва : Изд-во МЦНМО, 2013. – 304 с. – ISBN 978-5-4439-0111-4.
218. Vapnik, V. N. The nature of statistical learning theory / V. N. Vapnik. – New York : Springer, 2013. – XX, 314 p. – DOI 10.1007/978-1-4757-3264-1. – ISBN 978-0-387-98780-4.
219. Platt, J. C. Sequential Minimal Optimization: A Fast Algorithm for Training Support Vector Machines / J. C. Platt // Microsoft Research Technical Report. – 1998. – URL: <http://luthuli.cs.uiuc.edu/~daf/courses/optimization/Papers/smoTR.pdf> (date of access: 01.10.2022).
220. Zeng, J. Support vector machine-based short-term wind power forecasting / J. Zeng, W. Qiao. – DOI 10.1109/PSCE.2011.5772573 // 2011 IEEE/PES power systems

- conference and exposition (20–23 March 2011). – [S. l.], 2011. – P. 1–8. – URL: <https://ieeexplore.ieee.org/document/5772573> (date of access: 01.10.2022).
221. Wind power prediction method based on hybrid kernel function support vector machine / Z. Tia, S. Li, Y. Wang, X. Wang. – DOI 10.1177/0309524X17737337 // Wind Engineering. – 2018. – Vol. 42, Iss. 3. – P. 252–264. – URL: <https://journals.sagepub.com/doi/epub/10.1177/0309524X17737337> (date of access: 01.10.2022).
 222. Short-term wind power prediction based on improved chicken algorithm optimization support vector machine / C. Fu, K.-P. Lin, H.-J. Zhang. – DOI 10.3390/su11020512 // Sustainability. – 2019. – Vol. 11, Iss. 2. – P. 512.
 223. Short term wind power prediction based on data regression and enhanced support vector machine / Chia-Sheng Tu, Chih-Ming Hong, Hsi-Shan Huang, Chiung-Hsing Chen. – DOI 10.3390/en13236319 // Energies. – 2020. – Vol. 13, Iss. 23. – P. 6319.
 224. Santamaría-Bonfil, G. Wind speed forecasting for wind farms: A method based on support vector regression / G. Santamaría-Bonfil, A. Reyes-Ballesteros, C. Gershenson // Renewable Energy. – 2016. – Vol. 85. – P. 790–809.
 225. Classification and Regression Trees / L. Breiman, J. H. Friedman, R. A. Olshen, C. J. Stone. – New York : Chapman & Hall, 1984. – 368 p. – DOI 10.1201/9781315139470. – ISBN 9781315139470.
 226. Hastie, T. The elements of statistical learning: data mining, inference, and prediction / T. Hastie, R. Tibshirani, J. Friedman. – DOI 10.1007/BF02985802. – New York : Springer, 2009. – 745 p. – URL: <https://codowd.com/bigdata/misc/ESLII.pdf> (date of access: 01.10.2022).
 227. Учебник по машинному обучению. / М. Артемьев, А. Ашуха, М. Буркина [и др.] // Яндекс.Образование. – Разд. 2.3 : Решающие деревья. – URL: <https://education.yandex.ru/handbook/ml/article/reshayushchiye-derevya> (дата обращения: 03.09.2024).
 228. Performance comparison of different machine learning algorithms on the prediction of wind turbine power generation / O. Eyecioglu, B. Hangan, K. Kayisli, M. – DOI 10.1109/ICRERA47325.2019.8996541 // 2019 8th International conference on renewable energy research and applications (ICRERA). – Brasov, 2019. – P. 922–926.
 229. Breiman, L. Bagging predictors / L. Breiman // Machine learning. – 1996. – Vol. 24. – P. 123–140.
 230. Чистяков, С. П. Случайные леса : обзор / С. П. Чистяков // Труды Карельского научного центра Российской академии наук. – 2013. – № 1. – С. 117–136.
 231. Учебник по машинному обучению / М. Артемьев, А. Ашуха, М. Буркина [и др.] // Яндекс.Образование. – Разд. 2.4 : Ансамбли в машинном обучении. – URL: <https://education.yandex.ru/handbook/ml/article/ansambli-v-mashinnom-obuchenii> (дата обращения: 03.09.2024).
 232. Breiman, L. Random forests / L. Breiman // Machine learning. – 2001. – Vol. 45. – P. 5–32.
 233. Biau, G. A random forest guided tour / G. Biau, E. Scornet // Test. – 2016. – Vol. 25. – P. 197–227.

234. Учебник по машинному обучению / М. Артемьев, А. Ашуха, М. Буркина [и др.] // Яндекс.Образование. – Рзд. 2.5 : Градиентный бустинг. – URL: <https://education.yandex.ru/handbook/ml/article/gradientnyj-busting> (дата обращения: 03.09.2024).
235. Friedman, J. H. Greedy function approximation: a gradient boosting machine / J. H. Friedman. – DOI 10.1214/aos/1013203451 // Annals of statistics. – 2001. – P. 1189–1232. – URL: <https://projecteuclid.org/journals/annals-of-statistics/volume-29/issue-5/Greedy-function-approximation-A-gradient-boosting-machine/10.1214/aos/1013203451.full> (date of access: 15.02.2023).
236. Moudiki, T. LSBoost, gradient boosted penalized nonlinear least squares / T. Moudiki. – DOI 10.13140/RG.2.2.26553.80485 // ResearchGate. – 2020. – P. 1–7. – URL: https://www.researchgate.net/publication/346059361_LSBoost_gradient_boosted_penalized_nonlinear_least_squares (date of access: 01.10.2022).
237. Lahouar, A. H. Hour-ahead wind power forecast based on random forests / A. Lahouar, J. B. Slama. – DOI 10.1016/j.renene.2017.03.064 // Renewable energy. – 2017. – Vol. 109. – P. 529–541. – URL: <https://www.sciencedirect.com/science/article/abs/pii/S0960148117302550> (date of access: 01.10.2022).
238. Niu, D. Ultra-short-term wind-power forecasting based on the weighted random forest optimized by the niche immune lion algorithm / D. Niu, D. Pu, S. Dai. – DOI 10.3390/en11051098 // Energies. – 2018. – Vol. 11, Iss. 5. – P. 1098. – URL: <https://www.mdpi.com/1996-1073/11/5/1098> (date of access: 01.10.2022).
239. Shen, W. An EMD-RF based short-term wind power forecasting method / W. Shen, N. Jiang, N. Li. – DOI 10.1109/DDCLS.2018.8515901 // 2018 IEEE 7th Data Driven Control and Learning Systems Conference (DDCLS). – IEEE, 2018. – P. 283–288.
240. A short-term forecasting of wind power outputs based on gradient boosting regression tree algorithms / S. Park, S. Jung, J. Lee, J. Hur. – DOI 10.3390/en16031132 // Energies. – 2023. – Vol. 16, Iss. 3. – P. 1132. – URL: <https://typeset.io/papers/a-short-term-forecasting-of-wind-power-outputs-based-on-8uv8aosr?ysclid=m68zly8k2x381353028> (date of access: 13.09.2024).
241. Day-ahead wind power forecasting in Poland based on numerical weather prediction / B. Bochenek, J. Jurasz, A. Jaczewski [et al.]. – DOI 10.3390/en14082164 // Energies. – 2021. – Vol. 14, Iss. 8. – P. 2164. – URL: <https://www.mdpi.com/1996-1073/14/8/2164> (date of access: 06.10.2022).
242. Sobolewski, R. A. Gradient boosting-based approach for short-and medium-term wind turbine output power prediction / R. A. Sobolewski, M. Tchakorom, R. Couturier. – DOI 10.1016/j.renene.2022.12.040 // Renewable Energy. – 2023. – Vol. 203. – P. 142–160. – URL: <https://hal.science/hal-04224796/document> (date of access: 13.09.2024).
243. A machine learning-based gradient boosting regression approach for wind power production forecasting: A step towards smart grid environments / U. Singh, M. Rizwan, M. Alaraj, I. Alsaidan. – DOI 10.3390/en14165196 // Energies. – 2021. –

- Vol. 14, Iss. 16. – Art. 5196. – URL: <https://www.mdpi.com/1996-1073/14/16/5196> (date of access: 06.10.2022).
244. Comparative assessment of regression techniques for wind power forecasting / R. Pathak, A. Wadhwa, P. Khetarpal, N. Kumar. – DOI 10.1080/03772063.2020.1869591 // IETE Journal of Research. – 2023. – Vol. 69, Iss. 3. – P. 1393–1402. – URL: <https://www.sci-hub.ru/10.1080/03772063.2020.1869591?ysclid=m690b2ufqo671002383> (date of access: 14.09.2024).
 245. Alonso, Á. Random forests and gradient boosting for wind energy prediction / Á. Alonso, A. Torres, J. R. Dorronsoro. – DOI 10.1007/978-3-319-19644-2-3 // Hybrid Artificial Intelligent Systems : 10th International Conference : HAIS 2015 (Bilbao, June 22–24, 2015). – [S. n.], 2015. – P. 26–37.
 246. Algorithms for hyper-parameter optimization // Advances in neural information processing systems / J. Bergstra, R. Bardenet, Y. Bengio, B. Kégl // NIPS'11 : 25th International Conference on Neural Information Processing Systems. – [S. n.], 2011. – Vol. 24. – P. 2546–2554. – URL: https://www.researchgate.net/publication/216816964_Algorithms_for_Hyper-Parameter_Optimization (date of access: 04.05.2023).
 247. Оптимизация гиперпараметров // Википедия. Свободная энциклопедия. – Сан-Франциско, 2024. – URL: <https://ru.wikipedia.org/?curid=7631179&oldid=137301666> (дата обращения: 14.09.2024).
 248. Bergstra, J. Random search for hyper-parameter optimization / J. Bergstra, Y. Bengio // Journal of machine learning research. – 2012. – Vol. 13, Iss. 2. – P. 281–305.
 249. Оптимизация гиперпараметров на основе объединения априорных и апостериорных знаний о задаче классификации / В. С. Смирнова, В. В. Шаламов, В. А. Ефимова, А. А. Фильченков. – DOI 10.17586/2226-1494-2020-20-6-828-834 // Научно-технический вестник информационных технологий, механики и оптики. – 2020. – Т. 20, № 6. – С. 828–834.
 250. Frazier, P. I. A tutorial on Bayesian optimization / P. I. Frazier. – DOI 10.48550/arXiv.1807.02811 // arXiv. – 2018. – 8 Jul. – URL: <https://arxiv.org/pdf/1807.02811> (date of access: 09.10.2022).
 251. Brochu, E. A tutorial on Bayesian optimization of expensive cost functions, with application to active user modeling and hierarchical reinforcement learning / E. Brochu, V. M. Cora, N. De Freitas. – DOI 10.48550/arXiv.1012.2599 // arXiv. – 2010. – 12 Dec. – URL: <https://www.cs.ubc.ca/~nando/papers/bayopt.pdf> (date of access: 09.10.2022).
 252. Garnett, R. Bayesian optimization / R. Garnett. – Cambridge : Cambridge University Press, 2023. – 358 p. – DOI 10.1017/9781108348973. – ISBN 978-1108425780.
 253. Feature selection and hyper parameters optimization for short-term wind power forecast / H. Huang, R. Jia, X. Shi [et al.]. – DOI 10.1007/s10489-021-02191-y // Applied Intelligence. – 2021. – Vol. 51. – P. 6752–6770. – URL:

- <https://link.springer.com/article/10.1007/s10489-021-02191-y> (date of access: 21.03.2023).
254. A short-term wind power forecast method via XGBoost hyper-parameters optimization / X. Xiong, X. Guo, P. Zeng [et al.]. – DOI 10.3389/fenrg.2022.905155 // *Frontiers in energy research*. – 2022. – Vol. 10. – Art. 905155. – URL: <https://www.frontiersin.org/journals/energy-research/articles/10.3389/fenrg.2022.905155/full> (date of access: 07.06.2023).
 255. Yazıcı, K. Short-Term Wind Power Prediction Approach Based On Bayesian Optimization and Ensemble Learning / K. Yazıcı, S. Boran. – DOI 10.38016/jista.889991 // *Journal of Intelligent Systems: Theory and Applications*. – 2021. – Vol. 4, Iss. 2. – P. 142–154. – URL: <https://dergipark.org.tr/en/download/article-file/1613731> (date of access: 10.10.2022).
 256. Alkesaiberi, A. Efficient wind power prediction using machine learning methods: A comparative study / A. Alkesaiberi, F. Harrou, Y. Sun. – DOI 10.3390/en15072327 // *Energies*. – 2022. – Vol. 15, Iss. 7. – Art. 2327. – URL: <https://www.mdpi.com/1996-1073/15/7/2327> (date of access: 21.03.2023).
 257. PI is back! switching acquisition functions in bayesian optimization / C. Benjamins, E. Raponi, A. Jankovic [et al.]. – DOI 10.48550/arXiv.2211.01455 // *arXiv*. – 2022. – 2 Nov. – URL: <https://hal.sorbonne-universite.fr/hal-03861120v1/file/PIisBack.pdf> (date of access: 23.03.2023).
 258. Benchmarking the performance of Bayesian optimization across multiple experimental materials science domains/ Q. Liang, A. E. Gongora, Z. Ren [et al.]. – DOI 10.1038/s41524-021-00656-9 // *npj / Computational Materials*. – 2021. – Vol. 7, Iss. 1. – P. 1–10.
 259. Short-term prediction of the aggregated power output of wind farms – a statistical analysis of the reduction of the prediction error by spatial smoothing effects / U. Focken, M. Lange, K. Monnich [et al.]. – DOI 10.1016/S0167-6105(01)00222-7 // *Journal of Wind Engineering and Industrial Aerodynamics*. – 2002. – Vol. 90, Iss. 3. – P. 231–246. – URL: [https://www.sci-hub.ru/10.1016/s0167-6105\(01\)00222-7](https://www.sci-hub.ru/10.1016/s0167-6105(01)00222-7) (date of access: 10.10.2022).
 260. Lobo, M. G. Regional wind power forecasting based on smoothing techniques, with application to the Spanish peninsular system / M. G. Lobo, I. Sanchez // *IEEE Transactions on Power Systems*. – 2012. – Vol. 27, Iss. 4. – P. 1990–1997.
 261. Yakoub, G. Intelligent estimation of wind farm performance with direct and indirect ‘point’ forecasting approaches integrating several NWP models / G. Yakoub, S. Mathew, J. Leal. – DOI 10.1016/j.energy.2022.125893 // *Energy*. – 2023. – Vol. 263. – Art. 125893. – URL: <https://www.sciencedirect.com/science/article/pii/S0360544222027797> (date of access: 01.05.2024).
 262. Yakoub, G. Direct and indirect short-term aggregated turbine-and farm-level wind power forecasts integrating several NWP sources / G. Yakoub, S. Mathew, J. Leal. – DOI 10.1016/j.heliyon.2023.e21479 // *Heliyon*. – 2023. – Vol. 9, Iss. 11. – Art. e21479. – URL:

- <https://www.sciencedirect.com/science/article/pii/S2405844023086875> (date of access: 01.05.2024).
263. Clustering methods of wind turbines and its application in short-term wind power forecasts / Y. Liu, X. Gao, Jie Yan [et al.]. – DOI 10.1063/1.4898361 // Journal of Renewable and Sustainable Energy. – 2014. – Vol. 6, Iss. 5. – Art. 053119. – URL: <https://pubs.aip.org/aip/jrse/article/6/5/053119/901754/Clustering-methods-of-wind-turbines-and-its> (date of access: 15.12.2023).
 264. Short-term wind power forecasting based on clustering pre-calculated CFD method / Y. Wang, Y. Liu, Li Li [et al.]. – DOI 10.3390/en11040854 // Energies. – 2018. – Vol. 11, Iss. 4. – P. 854. – URL: <https://www.mdpi.com/1996-1073/11/4/854> (date of access: 15.12.2023).
 265. Turbine-level clustering for improved short-term wind power forecasting / J. M. G. Sopena, C. Maury, V. Pakrashi, B. Ghosh. – DOI 10.1088/1742-6596/2265/2/022052 // Journal of Physics: Conference Series. – 2022. – Vol. 2265, Iss. 2. – Art. 022052. – URL: <https://www.mdpi.com/1996-1073/11/4/854> (date of access: 15.12.2023).
 266. Wind power grouping forecasts and its uncertainty analysis using optimized relevance vector machine / J. Yan, Y. Liu, S. Han, M. Qiu. – DOI 10.1016/j.rser.2013.07.026 // Renewable and sustainable energy reviews. – 2013. – Vol. 27. – P. 613–621.
 267. Yang, M. Day-ahead wind power forecasting based on the clustering of equivalent power curves / M. Yang, C. Shi, H. Liu. – DOI 10.1016/j.energy.2020.119515 // Energy. – 2021. – Vol. 218. – Art. 119515. – URL: <https://www.sciencedirect.com/science/article/abs/pii/S0360544220326220> (date of access: 15.12.2023).
 268. Lloyd S. Least squares quantization in PCM / S. Lloyd // IEEE transactions on information theory. – 1982. – Vol. 28, Iss. 2. – P. 129–137.
 269. Arthur, D. k-means++: The advantages of careful seeding / D. Arthur, S. Vassilvitskii. – DOI 10.1145/1283383.1283494 // Eighteenth Annual ACM-SIAM Symposium on Discrete Algorithms, SODA 2007 (New Orleans, Louisiana, Jan. 7–9, 2007). – Stanford, 2006. – P. 1–11. – URL: <https://theory.stanford.edu/~sergei/papers/kMeansPP-soda.pdf> (date of access: 17.12.2023).
 270. Учебник по машинному обучению / М. Артемьев, А. Ашуха, М. Буркина [и др.] // Яндекс.Образование. – Разд. 10.2 : Кластеризация. – URL: <https://education.yandex.ru/handbook/ml/article/klasterizaciya> (дата обращения: 03.09.2024).
 271. Ward, Jr J. H. Hierarchical grouping to optimize an objective function / Jr J. H. Ward // Journal of the American statistical association. – 1963. – Vol. 58, Iss. 301. – P. 236–244.
 272. Ward method of hierarchical clustering for non-Euclidean similarity measures / S. Miyamoto, R. Abe, Y. Endo, Jun-ichi Takeshita. – DOI 10.1109/SOCPAR.2015.7492784 // 7th International Conference of Soft Computing

- and Pattern Recognition (SoCPar). – Fukuoka, 2015. – P. 60–63. – URL: <https://ieeexplore.ieee.org/document/7492784> (date of access: 23.07.2024).
273. Мастицкий, С. Э. Анализ временных рядов с помощью R / С. Э. Мастицкий. – [Б. г.] : [б. и.], 2020. – Электрон. версия. – URL: <https://ranalytics.github.io/tsa-with-r> (дата обращения: 12.04.2022).
 274. Алгоритм динамической трансформации временной шкалы // Википедия. Свободная энциклопедия. – Сан-Франциско, 2022. – URL: <https://ru.wikipedia.org/?curid=5327577&oldid=122890693> (дата обращения: 12.09.2024).
 275. Sakoe, H. Dynamic programming algorithm optimization for spoken word recognition / H. Sakoe, S. Chiba // IEEE transactions on acoustics, speech, and signal processing. – 1978. – Vol. 26, Iss. 1. – P. 43–49.
 276. Paliwal, K. K. A modification over Sakoe and Chiba's dynamic time warping algorithm for isolated word recognition / K. K. Paliwal, A. Agarwal, S. S. Sinha // Signal Processing. – 1982. – Vol. 4, Iss. 4. – P. 329–333.
 277. Dynamic time warping under limited warping path length / Z. Zhang, R. Tavenard, A. Bailly [et al.]. – DOI 10.1016/j.ins.2017.02.018 // Information Sciences. – 2017. – Vol. 393. – P. 91–107. – URL: https://hal.science/hal-01470554/file/LDTW_author_version.pdf (date of access: 18.01.2024).
 278. Nakagawa, K. Stock price prediction using k-medoids clustering with indexing dynamic time warping / K. Nakagawa, M. Imamura, K. Yoshida. – DOI 10.1541/ieejieiss.138.986 // Electronics and Communications in Japan. – 2019. – Vol. 138, Iss. 8. – P. 986–991. – URL: <https://onlinelibrary.wiley.com/doi/full/10.1002/ecj.12140> (date of access: 16.01.2024).
 279. Clustering and Modelling of the Top 30 Cryptocurrency Prices Using Dynamic Time Warping and Machine Learning Methods / T. Šťastný, J. Koudelka, D. Bílková, L. Marek [et al.]. – DOI 10.3390/math10193672 // Mathematics. – 2022. – Vol. 10, Iss. 19. – P. 3672. – URL: <https://www.mdpi.com/2227-7390/10/19/3672> (date of access: 16.01.2024).
 280. Simmhan, Y. Scalable prediction of energy consumption using incremental time series clustering / Y. Simmhan, M. U. Noor. – DOI 10.1109/BigData.2013.6691774 // 2013 IEEE International Conference on Big Data. – Silicon Valley, 2013. – P. 29–36. – URL: <https://ieeexplore.ieee.org/document/6691774> (date of access: 16.01.2024).
 281. Time series clustering based on dynamic time warping / W. Wang, G. Lyu, Y. Shi, X. Liang. – DOI 10.1109/ICSESS.2018.8663857 // 2018 IEEE 9th international conference on software engineering and service science (ICSESS). – Beijing, 2018. – P. 487–490. – URL: <https://ieeexplore.ieee.org/document/8663857> (date of access: 21.01.2024).
 282. Grubbs, F. E. Procedures for detecting outlying observations in samples / F. E. Grubbs // Technometrics. – 1969. – Vol. 11, Iss. 1. – P. 1–21.
 283. Barnett, V. Outliers in statistical data / V. Barnett, T. Lewis. – 3rd Ed. – New York : Wiley, 1994. – 608 p. – ISBN 978-0-471-93094-5.

284. Hodge, V. A survey of outlier detection methodologies / V. Hodge, J. Austin. – DOI 10.1023/B:AIRE.0000045502.10941.a9 // Artificial intelligence review. – 2004. – Vol. 22. – P. 85–126. – URL: <https://link.springer.com/article/10.1007/s10462-004-4304-y> (date of access: 16.10.2024).
285. Hawkins, D. Identification of outliers / D. Hawkins. – London : Chapman and Hall, 1980. – 188 p. – ISBN 041221900X. – ISBN 9780412219009.
286. Aggarwal, C. C. An introduction to outlier analysis / C. C. Aggarwal. – DOI 10.1007/978-3-319-47578-3_1 // Outlier Analysis / Ch. C. Aggarwal. – [S. l.], 2017. – P. 1–34. – URL: https://link.springer.com/chapter/10.1007/978-3-319-47578-3_1 (date of access: 16.10.2024).
287. Smiti, A. A critical overview of outlier detection methods / A. Smiti. – DOI 10.1016/j.cosrev.2020.100306 // Computer Science Review. – 2020. – Vol. 38. – Art. 100306. – URL: <https://www.sciencedirect.com/science/article/abs/pii/S1574013720304068> (date of access: 26.10.2024).
288. Applications and modeling techniques of wind turbine power curve for wind farms—A review / F. Bilendo, A. Meyer, H. Badihi [et al.]. – DOI 10.3390/en16010180 // Energies. – 2022. – Vol. 16, Iss. 1. – P. 180. – URL: <https://www.mdpi.com/1996-1073/16/1/180> (date of access: 17.03.2024).
289. Morrison, R. Anomaly detection in wind turbine SCADA data for power curve cleaning / R. Morrison, X. Liu, Z. Lin. – DOI 10.1016/j.renene.2021.11.118 // Renewable Energy. – 2022. – Vol. 184. – P. 473–486. – URL: <https://www.sciencedirect.com/science/article/abs/pii/S0960148121017134> (date of access: 17.03.2024).
290. Zou, M. A review of approaches for the detection and treatment of outliers in processing wind turbine and wind farm measurements / M. Zou, S. Z. Djokic. – DOI 10.3390/en13164228 // Energies. – 2020. – Vol. 13, Iss. 16. – Art. 4228. – URL: <https://www.mdpi.com/1996-1073/13/16/4228> (date of access: 17.03.2024).
291. Identification and correction of outliers in wind farm time series power data / X. Ye, Z. Lu, Y. Qiao [et al.]. – DOI 10.1109/TPWRS.2015.2512843 // IEEE Transactions on power systems. – 2016. – Vol. 31, Iss. 6. – P. 4197–4205. – URL: <https://ieeexplore.ieee.org/stamp/stamp.jsp?tp=&arnumber=7384769> (date of access: 17.03.2024).
292. Data-driven correction approach to refine power curve of wind farm under wind curtailment / Y. Zhao, Lin Ye, W. Wang [et al.]. – DOI 10.1109/TSTE.2017.2717021 // IEEE Transactions on Sustainable Energy. – 2017. – Vol. 9, Iss. 1. – P. 95–105. – URL: <https://ieeexplore.ieee.org/document/7953532> (date of access: 17.03.2024).
293. Uncertainty analysis of wind power probability density forecasting based on cubic spline interpolation and support vector quantile regression / Y. He, H. Li, S. Wang, Xin Yao. – DOI 10.1016/j.neucom.2020.10.093 // Neurocomputing. – 2021. – Vol. 430. – P. 121–137. – URL: <https://www.sciencedirect.com/science/article/abs/pii/S0925231220317069> (date of access: 18.03.2024).

294. Shen, X. A combined algorithm for cleaning abnormal data of wind turbine power curve based on change point grouping algorithm and quartile algorithm / X. Shen, X. Fu, C. Zhou. – DOI 10.1109/TSTE.2018.2822682 // IEEE Transactions on Sustainable Energy. – 2018. – Vol. 10, Iss. 1. – P. 46–54. – URL: <https://ieeexplore.ieee.org/stamp/stamp.jsp?tp=&arnumber=8330024> (date of access: 18.03.2024).
295. Paik, C. Power Curve Modeling of Wind Turbines through Clustering-Based Outlier Elimination / C. Paik, Y. Chung, Y. J. Kim. – DOI 10.3390/asi6020041 // Applied System Innovation. – 2023. – Vol. 6, Iss. 2. – P. 41.
296. Short-term wind power prediction using GA-BP neural network based on DBSCAN algorithm outlier identification / P. Zhang, Y. Wang, L. Liang [et al.]. – DOI 10.3390/pr8020157 // Processes. – 2020. – Vol. 8, Iss. 2. – P. 157.
297. Wind turbine power curve modelling with logistic functions based on quantile regression / B. Jing, Z. Qian, H. Zareipour [et al.]. – DOI 10.3390/app11073048 // Applied Sciences. – 2021. – Vol. 11, Iss. 7. – P. 3048.
298. Zheng, L. Raw wind data preprocessing: A data-mining approach / L. Zheng, W. Hu, Y. Min. – DOI 10.1109/TSTE.2014.2355837 // IEEE Transactions on Sustainable Energy. – 2014. – Vol. 6, Iss. 1. – P. 11–19.
299. Tukey J. W. Exploratory data analysis / J. W. Tukey. – Reading ; Mass ; London : Addison-Wesley, 1977. – XVI, 688 p.
300. DBSCAN revisited, revisited: why and how you should (still) use DBSCAN / E. Schubert, J. Sander, M. Ester [et al.] // ACM Transactions on Database Systems (TODS). – 2017. – Vol. 42, Iss. 3. – P. 19-1–19-21. – URL: https://www.khoury.northeastern.edu/home/vip/teach/DMcourse/2_cluster_EM_mixed/notes_slides/revisitotherevisitedDBSCAN.pdf (date of access: 16.10.2024).
301. A density-based algorithm for discovering clusters in large spatial databases with noise / M. Ester, H.-P. Kriegel, J. Sander, X. Xu // KDD'96: Proceedings of the Second International Conference on Knowledge Discovery and Data Mining. – 1996. – Vol. 96, Iss. 34. – P. 226–231. – URL: <https://www2.cs.uh.edu/~ceick/DM/WS-P1.pdf> (date of access: 04.04.2023).
302. Finding a "kneedle" in a haystack: Detecting knee points in system behavior / V. Satopaa, J. Albrecht, D. Irwin [et al.]. – DOI 10.1109/ICDCSW.2011.20 // ICDCSW '11: Proceedings of the 2011 31st International Conference on Distributed Computing Systems Workshops. – Washington, 2011. – P. 166–171.
303. Welcome to kneed's documentation! : (knead Python package documentation). – 2020. – URL: <https://knead.readthedocs.io/en/stable/index.html> (date of access: 04.05.2023).
304. Campello, R. J. G. B. Density-based clustering based on hierarchical density estimates / R. J. G. B. Campello, D. Moulavi, J. Sander // Advances in Knowledge Discovery and Data Mining : 17th Pacific-Asia Conference : PAKDD 2013 (Gold Coast, April 14-17, 2013) : Proceedings. – Berlin, 2013. – Part 2. – P. 160–172.
305. Hierarchical density estimates for data clustering, visualization, and outlier detection / R. Campello, D. Moulavi, Arthur Zimek, J. Sander. – DOI 10.1145/27333 // ACM

- Transactions on Knowledge Discovery from Data (TKDD). – 2015. – Vol. 10, Iss. 1. – P. 1–51.
306. Swersky, L. A study of unsupervised outlier detection for one-class classification : Thesis / L. Swersky ; Department of Computing Science. – Lorne Swersky, 2018. – 69 p. – DOI 10.7939/R3GB1XZ93.
 307. Ghosh, K. Beyond Defaults: Parameter Free Outlier Detection using GLOSH / K. Ghosh. – Kushankur Ghosh, 2024. – 87 p. – Access mode: ERA (Education and Research Archive). – URL: https://era.library.ualberta.ca/items/988e813a-06d7-4434-935d-655d34ba7be0/view/c38349cb-5fd7-4ea0-ad1b-27f55c6245b9/Ghosh_Kushankur_202401_MSc.pdf (date of access: 18.10.2024). – DOI 10.7939/r3-zs51-gq83.
 308. Model-based clustering with HDBSCAN / M. Strobl, J. Sander, R. J. G. B. Campello, O. Zaïane // Machine Learning and Knowledge Discovery in Databases : European Conference : ECML PKDD 2020 (Ghent, Sept. 14–18, 2020). – Ghent, 2021. – Part 2. – P. 364–379. – URL: <https://webdocs.cs.ualberta.ca/~zaiane/postscript/PKDD2020-hdbscan.pdf> (date of access: 18.10.2024).
 309. Stewart, G. An implementation of the HDBSCAN* clustering algorithm / G. Stewart, M. Al-Khassaweneh. – DOI 10.3390/app12052405 // Applied Sciences. – 2022. – Vol. 12, Iss. 5. – P. 2405. – URL: <https://www.mdpi.com/2076-3417/12/5/2405> (date of access: 19.10.2024).
 310. hdbscan: Hierarchical density based clustering / L. McInnes, J. Healy, S. Astels. – DOI 10.21105/joss.00205 // Journal of Open Source Software. – 2017. – Vol. 2, Iss. 11. – P. 205-1–205-3. – URL: https://www.researchgate.net/publication/315508524_hdbscan_Hierarchical_density_based_clustering (date of access: 21.10.2024).
 311. McInnes, L. The hdbscan Clustering Library / L. McInnes, J. Healy, S. Astels // hdbscan : Python package documentation. – 2016. – URL: <https://hdbscan.readthedocs.io/en/latest/index.html> (date of access: 21.10.2024).
 312. Prim, R. C. Shortest connection networks and some generalizations / R. C. Prim // The Bell System Technical Journal. – 1957. – Vol. 36, Iss. 6. – P. 1389–1401.
 313. The ultra-short term power prediction of wind farm considering operational condition of wind turbines / R. Fang, Y. Wang, R. Shang [et al.]. – DOI 10.1016/j.ijhydene.2016.03.173 // International Journal of Hydrogen Energy. – 2016. – Vol. 41, Iss. 35. – P. 15733–15739. – URL: <https://www.sciencedirect.com/science/article/abs/pii/S0360319916308941?via%3Dihub> (date of access: 22.10.2024).

ПРИЛОЖЕНИЕ А

ОПТИМИЗАЦИЯ ГИПЕРПАРАМЕТРОВ

Таблица А.1 – Результаты вычислительных экспериментов по определению гиперпараметров моделей машинного обучения

Модель	COГ	Гиперпараметры
RDT	GS	MinLeafSize = 6; MaxNumSplits = 762
	RS	MinLeafSize = 8; MaxNumSplits = 629
	EI	MinLeafSize = 10; MaxNumSplits = 1206
	PI	MinLeafSize = 6; MaxNumSplits = 1501
	LCB	MinLeafSize = 8; MaxNumSplits = 799
FFNN	GS	NumLayers = 2; Activations - relu; Lambda = 4,2452e-06; LayerWeightsInitializer - gloriot; LayerBiasesInitializer - zeros; Layer_1_Size = 300; Layer_2_Size = 13
	RS	NumLayers = 2; Activations - relu; Lambda = 1,5794e-06; LayerWeightsInitializer - gloriot; LayerBiasesInitializer - zeros; Layer_1_Size = 82; Layer_2_Size = 41
	EI	NumLayers = 1; Activations - relu; Lambda = 6,3761; LayerWeightsInitializer - he; LayerBiasesInitializer - zeros; Layer_1_Size = 267
	PI	NumLayers = 2; Activations - relu; Lambda = 2,9884; LayerWeightsInitializer - gloriot; LayerBiasesInitializer - zeros; Layer_1_Size = 269; Layer_2_Size = 148
	LCB	NumLayers = 2; Activations - relu; Lambda = 0,001481; LayerWeightsInitializer - gloriot; LayerBiasesInitializer - zeros; Layer_1_Size = 294; Layer_2_Size = 8
GBDT	GS	NumLearningCycles = 136; LearnRate = 0,46416; MinLeafSize = 14; MaxNumSplits = 762; NumVariablesToSample = 4
	RS	NumLearningCycles = 88; LearnRate = 0,034638; MinLeafSize = 4; MaxNumSplits = 2579; NumVariablesToSample = 3
	EI	NumLearningCycles = 193; LearnRate = 0,17337; MinLeafSize = 7; MaxNumSplits = 3876; NumVariablesToSample = 2
	PI	NumLearningCycles = 324; LearnRate = 0,05301; MinLeafSize = 3; MaxNumSplits = 1106; NumVariablesToSample = 2
	LCB	NumLearningCycles = 488; LearnRate = 0,035108; MinLeafSize = 4; MaxNumSplits = 1821; NumVariablesToSample = 2
GPR	GS	Sigma = 1,6664; BasisFunction - pureQuadratic; KernelFunction - exponential; KernelScale = 190,3051
	RS	Sigma = 1444,0957; BasisFunction - constant; KernelFunction - exponential; KernelScale = 1257,8536
	EI	Sigma = 0,033109; BasisFunction - pureQuadratic; KernelFunction - exponential; KernelScale = 4846,3625
	PI	Sigma = 0,31945; BasisFunction - pureQuadratic; KernelFunction - exponential; KernelScale = 81,861
	LCB	Sigma = 0,18095; BasisFunction - pureQuadratic; KernelFunction - exponential; KernelScale = 230,2639

Продолжение таблицы А.1

RF	GS	NumLearningCycles = 37; MinLeafSize = 2; MaxNumSplits = 762; NumVariablesToSample = 5
	RS	NumLearningCycles = 51; MinLeafSize = 1; MaxNumSplits = 754; NumVariablesToSample = 2
	EI	NumLearningCycles = 104; MinLeafSize = 3; MaxNumSplits = 576; NumVariablesToSample = 3
	PI	NumLearningCycles = 419; MinLeafSize = 1; MaxNumSplits = 1394; NumVariablesToSample = 3
	LCB	NumLearningCycles = 419; MinLeafSize = 1; MaxNumSplits = 1895; NumVariablesToSample = 3
SVM	GS	BoxConstraint = 0,46416; Epsilon = 6796,6503; KernelFunction - linear
	RS	BoxConstraint = 2,5208; Epsilon = 75,6296; KernelFunction - linear
	EI	BoxConstraint = 1,7353; Epsilon = 19506,0839; KernelFunction - linear
	PI	BoxConstraint = 842,4693; KernelScale = 22,7652; Epsilon = 11914,6056; KernelFunction - gaussian
	LCB	BoxConstraint = 4,7071; Epsilon = 34566,2687; KernelFunction - linear

ПРИЛОЖЕНИЕ Б
РАСПРЕДЕЛЕНИЕ ВЕТРОУСТАНОВОК ПО ГРУППАМ

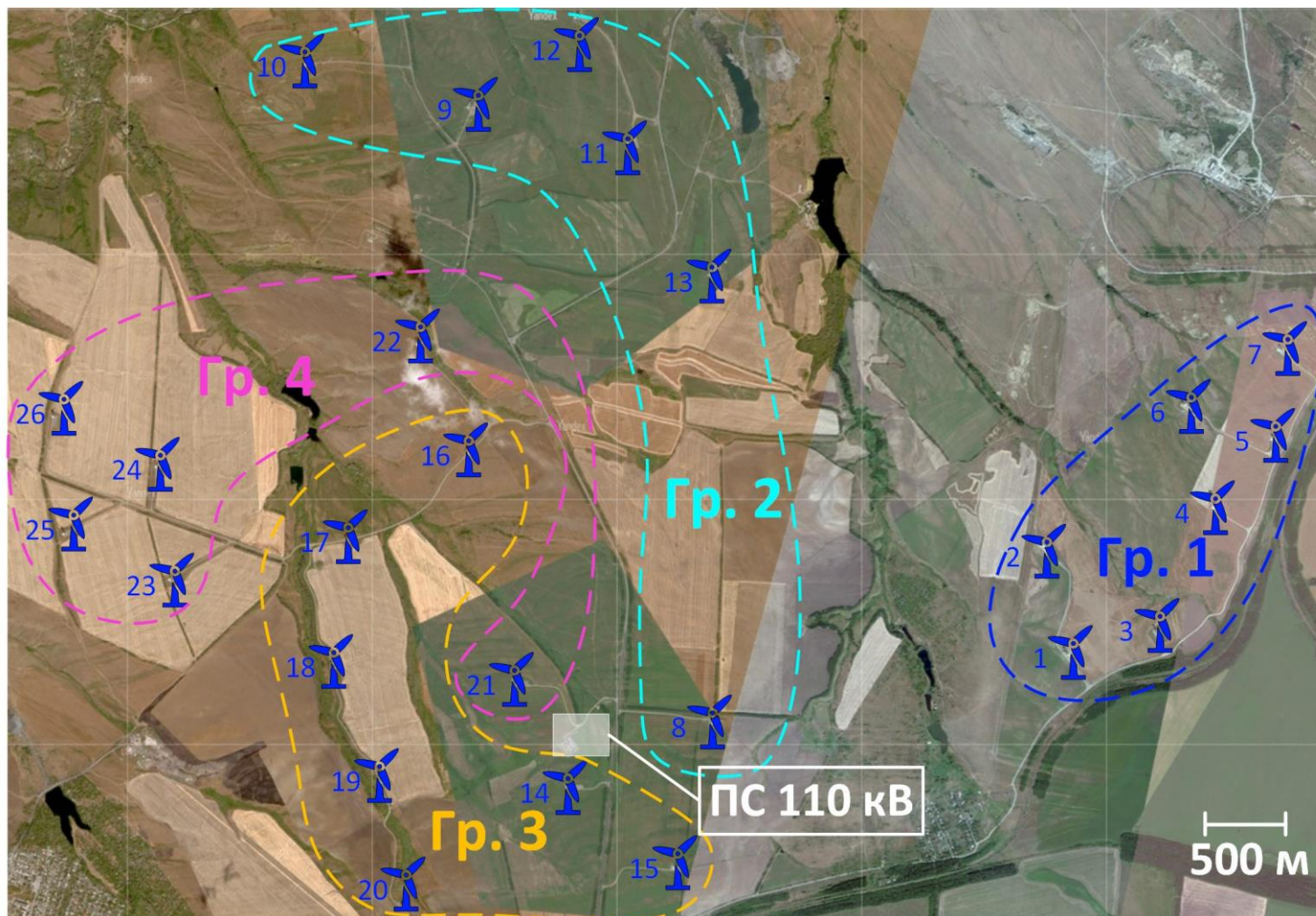


Рисунок Б.1 –Результат распределения ВЭУ по группам с помощью методики *GF*

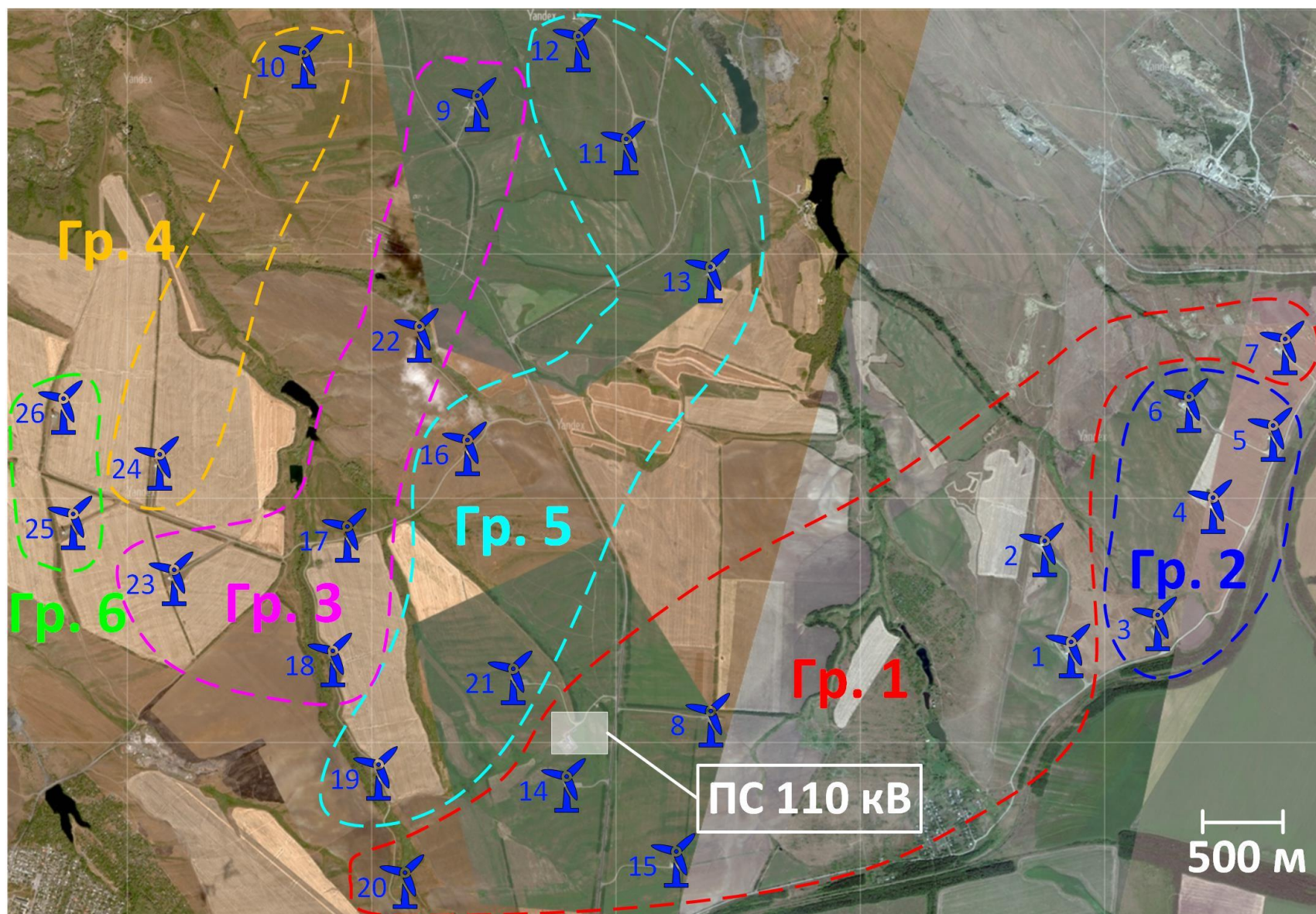


Рисунок Б.2 –Результат распределения ВЭУ по группам с помощью методики *K-MS*

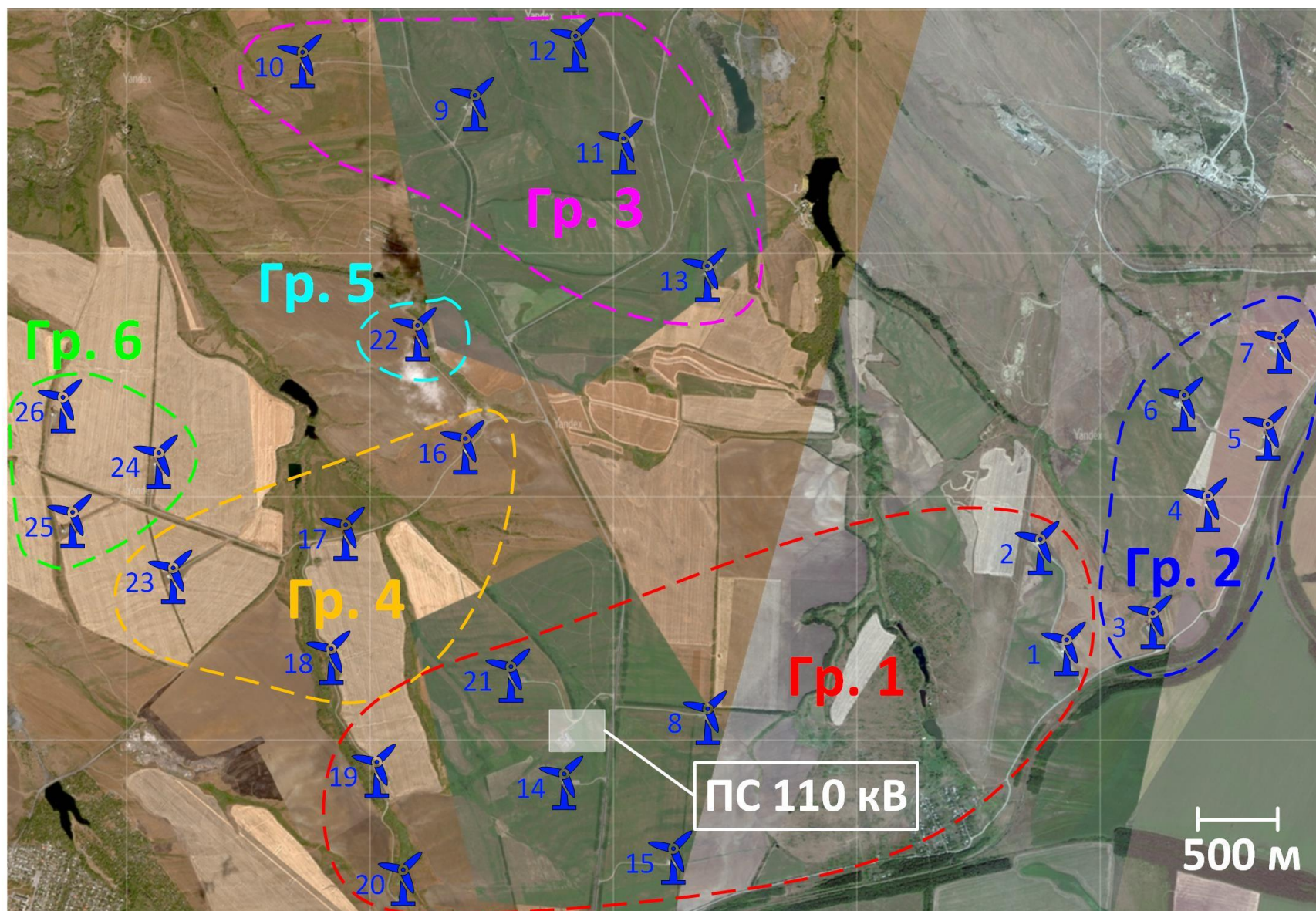


Рисунок Б.3 –Результат распределения ВЭУ по группам с помощью методики *K-P*

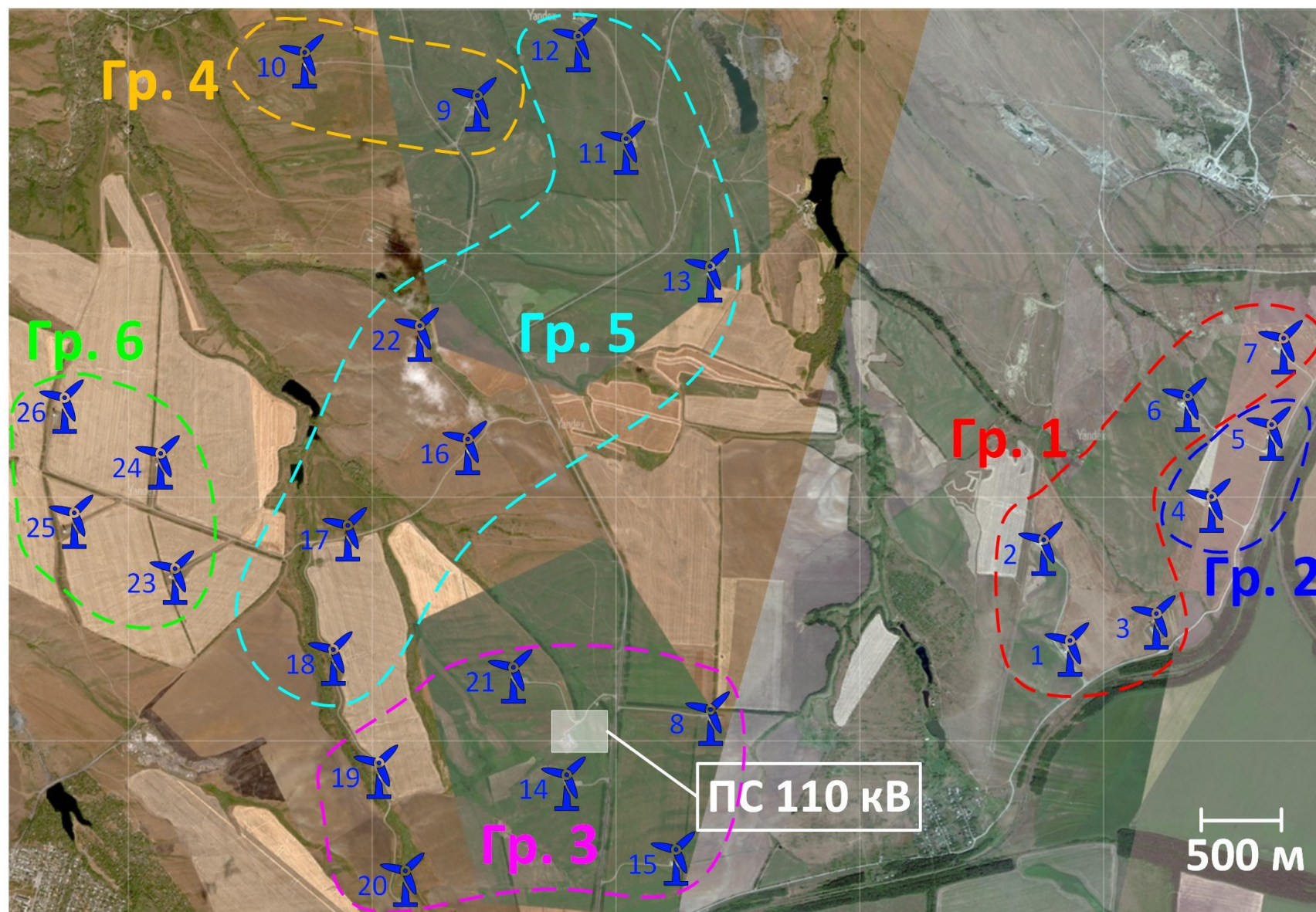


Рисунок Б.4 –Результат распределения ВЭУ по группам с помощью методики *H-PE*

ПРИЛОЖЕНИЕ В

АКТ О ВНЕДРЕНИИ РЕЗУЛЬТАТОВ ДИССЕРТАЦИОННОЙ РАБОТЫ



ООО «Прософт-Системы»
620102, г. Екатеринбург, ул. Волгоградская, 194а
Тел.: +7 (343) 356-51-11, info@prosoftsystems.ru
www.prosoftsystems.ru

р/с 40702810700280007084, к/с 30101810145250000411
Филиал «Центральный» Банка ВТБ (ПАО)
БИК 044525411, ИНН 6660149600, КПП 665801001
ОКВЭД 72.19, 27.12; 62.01; 46.69.5, ОКПО 55181848

УТВЕРЖДАЮ
Заместитель генерального директора
по автоматизации энергосистем
общества с ограниченной ответственностью
«Прософт-Системы»
Дымшаков Алексей Викторович
_____ 2024г.

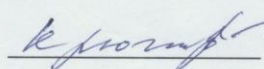


АКТ
о внедрении результатов диссертационной работы
на соискание ученой степени кандидата технических наук
Снегирева Дениса Алексеевича
«Повышение эффективности
краткосрочного прогнозирования генерации ветроэлектрических станций»

Настоящим актом подтверждается, что результаты диссертационного исследования Снегирева Д.А. использованы при разработке программного обеспечения ООО «Прософт-Системы». Совместная научно-исследовательская работа ФГАОУ ВО «УрФУ имени первого Президента России Б.Н. Ельцина» и ООО «Прософт-Системы», посвященная разработке методик прогнозирования генерации ветроэлектрических станций, осуществлялась в рамках договоров № 21-53810 от 26 января 2021 г. и № 22-59892 от 01 апреля 2022 г.

Разработанная Снегиревым Д.А. методика краткосрочного прогнозирования генерации ветроэлектрических станций, основанная на методах машинного обучения, внедрена в модуле краткосрочного прогнозирования выработки ветроэлектрических станций для программного комплекса «Энергосфера». Эффективность предложенной методики подтверждается результатами краткосрочного прогнозирования генерации реальной ветроэлектрической станции «Гуковская ВЭС».

Начальник отдела,
Департамент автоматизации электрических систем,
общество с ограниченной ответственностью
«Прософт-Системы»,
доцент, кандидат технических наук

 Крючков Павел Анатольевич



Система менеджмента качества соответствует ISO 9001:2015
Система экологического менеджмента соответствует ГОСТ Р ИСО 14001-2016 (ISO 14001:2015)
Система менеджмента безопасности труда и охраны здоровья соответствует ГОСТ Р ИСО 45001-2020 (ISO 45001:2018)

ПРИЛОЖЕНИЕ Г
СВИДЕТЕЛЬСТВА О ГОСУДАРСТВЕННОЙ РЕГИСТРАЦИИ
ПРОГРАММ ДЛЯ ЭВМ

РОССИЙСКАЯ ФЕДЕРАЦИЯ



СВИДЕТЕЛЬСТВО

о государственной регистрации программы для ЭВМ

№ 2022616750

WindPowerForecast

Правообладатель: **Снегирев Денис Алексеевич (RU)**

Авторы: **Снегирев Денис Алексеевич (RU), Паздерин Андрей Владимирович (RU), Самойленко Владислав Олегович (RU)**



Заявка № **2022615338**

Дата поступления **29 марта 2022 г.**

Дата государственной регистрации

в Реестре программ для ЭВМ **15 апреля 2022 г.**

Руководитель Федеральной службы
по интеллектуальной собственности

Ю.С. Зубов

РОССИЙСКАЯ ФЕДЕРАЦИЯ



СВИДЕТЕЛЬСТВО

о государственной регистрации программы для ЭВМ

№ 2023614895

WindPowerForecastML

Правообладатель: *Снегирев Денис Алексеевич (RU)*Авторы: *Снегирев Денис Алексеевич (RU), Паздерин Андрей Владимирович (RU), Самойленко Владислав Олегович (RU)*

Заявка № 2023613489

Дата поступления **22 февраля 2023 г.**

Дата государственной регистрации

в Реестре программ для ЭВМ **07 марта 2023 г.**Руководитель Федеральной службы
по интеллектуальной собственности

Ю.С. Зубов