

**Федеральное государственное бюджетное учреждение
«Гидрометеорологический научно-исследовательский центр Российской Федерации»**

На правах рукописи

Быков Филипп Леонидович

**ПОСТПРОЦЕССИНГ ЧИСЛЕННЫХ ПРОГНОЗОВ
ПРИЗЕМНЫХ МЕТЕОРОЛОГИЧЕСКИХ ПАРАМЕТРОВ
НА ОСНОВЕ НЕЙРОСЕТЕВЫХ МЕТОДОВ**

Специальность

25.00.29 – физика атмосферы и гидросферы

ДИССЕРТАЦИЯ

на соискание ученой степени

кандидата физико-математических наук

Научный руководитель:
доктор физико-математических наук,
В.А.Гордин

Москва – 2021

Оглавление

ВВЕДЕНИЕ	3
ГЛАВА 1. Используемые методы и данные	16
1.1. Обзор технологии искусственных нейронных сетей	17
1.2. Используемые наборы данных.....	46
1.3. Предпосылки и перспективы применения нейронных сетей в геофизике.....	52
1.4. Технические подробности реализации предлагаемого постпроцессинга	58
Выводы из главы 1	62
ГЛАВА 2. Систематическая коррекция и комплексификация прогнозов	64
2.1. Систематическая коррекция.....	64
2.2. Комплексный (мультимодельный) прогноз.....	80
Выводы из главы 2	96
ГЛАВА 3. Нейросетевая коррекция	97
3.1. Используемые предикторы	97
3.2. Оптимизация нейронной сети	100
3.3 Учет локальных параметров с помощью Embedding слоёв	109
3.4 Оценка важности различных предикторов	112
Выводы из главы 3	115
ГЛАВА 4. Интерполяция поправок по горизонтальным координатам.....	117
4.1. Корреляционные функции и оптимальная интерполяция	117
4.2. Квазилинейная неоднородная оптимальная интерполяция инкрементов	139
Выводы из главы 4	161
ГЛАВА 5. Сравнительные оценки качества прогнозов.....	163
5.1. Сравнительные оценки качества прогнозов температуры воздуха и точки росы.....	163
5.2. Сравнительные оценки качества прогнозов скорости ветра.....	171
5.3. Сравнительные оценки качества прогнозов давления	176
5.4. Сравнительные оценки качества прогнозов экстремальных температур воздуха.....	179
Выводы из главы 5	182
Заключение.....	183
Список сокращений и условных обозначений.....	188
Список литературы	190
Приложение А. Оценки прогнозов от начального срока 12:00 ВСВ	210
Приложение Б. Примеры продукции, использующей предлагаемый постпроцессинг.....	221
Приложение В. Письма руководителей и специалистов региональных управлений Росгидромета	231

ВВЕДЕНИЕ

Несмотря на все успехи применения численного моделирования атмосферы для задачи прогноза погоды, пользователи метеопрогноза могут быть неудовлетворены прогнозом численной модели прогноза погоды (далее ЧМПП) по двум основным причинам: а) если в прогнозе не представлена (или представлена в неудовлетворительной форме) интересующая пользователя информация о погоде; б) если ошибки предоставленных прогнозов велики.

Актуальность темы исследования

Уменьшить эти недостатки можно, применяя методы статистической интерпретации (постпроцессинга) ЧМПП. Задачи по интерпретации ЧМПП условно можно разделить на две группы согласно устраняемым недостаткам ЧМПП: а) задачи, целью которых является прогноз метеорологических параметров, отсутствующих в прогнозе ЧМПП; б) задачи коррекции, целью которых является улучшение качества прогноза ЧМПП.

Задачи первой группы могут решаться в двух концепциях [244]: либо в так называемой концепции точного прогноза (perfect prognosis), когда параметры постпроцессинга настраиваются по данным наблюдений или по данным из полей анализа, а при оперативном использовании вместо этих данных на вход ему подаются прогностические данные; либо в концепции model output statistic (MOS), когда параметры метода постпроцессинга сразу настраиваются по данным прогноза ЧМПП.

Цель задачи коррекции – вычислить поправки, улучшающие качество прогностической продукции, выдаваемой ЧМПП. Задача коррекции может решаться только в концепции MOS с использованием статистических методов, в том числе методами машинного обучения (МО). Для её решения необходимо иметь соответствующий архив данных.

Задача коррекции была и остается актуальной, поскольку в любой ЧМПП используется множество физических и вычислительных допущений, без которых было бы невозможно рассчитать прогноз к требуемому времени. Тем не менее, эти допущения могут существенно повлиять на качество прогноза для конечного пользователя. Для постановки математической задачи интерпретации (постпроцессинга) прогноза ЧМПП необходим архив данных наблюдений, также разумный и строго сформулированный заданный критерий качества.

Концепция точного прогноза имеет существенный недостаток: прогноз ЧМПП может иметь статистическую структуру, отличающуюся от статистической структуры используемых

при настройке параметров данных наблюдений или анализа. Поэтому предварительное применение коррекции может улучшить прогноз, разработанный в концепции точного прогноза.

Любой метод статистической интерпретации ЧМПП и принимает на вход некоторый набор предикторов, в том числе параметры, прогнозируемые ЧМПП (direct model output) и применяет к ним некоторую линейную или нелинейную функцию (модель МО).

Применение линейных моделей МО не гарантирует достижения наилучшего возможного решения задачи интерпретации, но они сравнительно просты, общеприняты и наиболее разработаны. Существенно улучшить качество прогнозов линейных моделей возможно лишь, конструируя новые предикторы (так называемая задача feature engineering), что требует ручной работы специалиста. Конструирование новых предикторов часто улучшает качество прогноза и при применении нелинейных моделей МО [227].

Линейные модели МО в метеорологии применяются традиционно [16], [67], [46], а также разрабатываются традиционно различные нелинейные предикторы (индексы) поэтому вопрос о преимуществе новых нелинейных моделей МО над линейными остается актуальным.

Объект и предмет исследования

Объект исследования – погрешности прогностических полей ЧМПП.

Предмет исследования – моделирование поправок к полям ЧМПП нейронными сетями.

Цель исследования

Основной целью диссертационной работы является разработка основанного на применении нейросетевых методов нового автоматизированного комплекса постпроцессинга выходной продукции одной или нескольких численных прогностических моделей с учетом последних наблюдений (за период длительностью T) приземных метеорологических величин на синоптических станциях. Необходимо, чтобы разработанный комплекс:

- был применим в случаях частичного отсутствия прогностической информации и в случаях отсутствия (отбраковки алгоритмом контроля качества) части данных наблюдений за период T ;
- обеспечивал получение прогностических данных в местах расположения синоптических станций и в узлах модельной сетки;
- уточнял результаты численных прогнозов погоды;
- мог работать в оперативном режиме на доступных вычислительных системах.

Задачи исследования

Были сформулированы и решены с применением нейросетевых методов следующие задачи:

- систематическая коррекция (комплексификация) прогнозов приземных метеорологических параметров в местах расположения синоптических станций; поправка вычисляется как результат применения к известным погрешностям (за период длительностью T) прогнозов модели (моделей) нелинейного оператора, коэффициенты которого вычисляются нейронной сетью;
- определение поправок к прогнозам ЧМПП, учитывающий сдвинутые по начальному сроку и заблаговременности прогнозы;
- реализована квазилинейная неоднородная оптимальная интерполяция (КНОИ) по горизонтальным координатам для вычисления поправок к прогностическим полям приземных метеорологических параметров в точках модельной сетки; КНОИ использует нейронные сети для учёта неоднородностей полей поправок;
- оптимизация параметров нейронных сетей, использованных в предложенном постпроцессинге приземных прогностических полей одной и нескольких ЧМПП;
- сравнение различных конфигураций предлагаемого постпроцессинга;

Предложенный постпроцессинг протестирован в непрерывном режиме счета и на большом статистически независимом архиве. Получены сравнительные оценки качества прогнозов, составляемых путем применения предлагаемого постпроцессинга и других известных методов.

Причины выбора именно квазилинейных методов пояснены в первой главе, параграф 1.3.4.

Методология и методы исследования

Для решения сформулированных в диссертационной работе задач использованы теоретические результаты и методы линейной алгебры, математического анализа, функционального анализа, интерполяции, оптимизации, математической статистики и теории вероятностей, теории случайных полей, статистической обработки больших данных, нейронных сетей и машинного обучения, а также методы структурного и объектно-ориентированного программирования и различные методики оценки качества метеорологических прогнозов.

Использованы архивы прогнозов различных ЧМПП в точках синоптических станций и архивов наблюдений на синоптических станциях (используемые архивы описаны во второй части главы 2).

Сформулируем основную задачу машинного обучения (МО) – подробнее задача и её решение методом нейронных сетей рассмотрены в главе 1. Применение МО возможно при наличии достаточно большого архива данных из M реализаций вектора (тензора) предикторов $\vec{X}_{input,j}, j=1,...,M$ размерности N_0 каждый, которые будут известны в момент составления прогноза и соответствующие им результаты наблюдений $\vec{X}_{fact,j}$.

Допустим, что заданы:

- а) параметрическое семейство моделей МО $\vec{X}_{pred} = F(\vec{X}_{input}, \vec{\theta})$, где $\vec{X}_{input}$ – аргументы (предикторы); $\vec{\theta} \in \mathbb{R}^{n_\theta}$ – параметры, которые будем оптимизировать; $\vec{X}_{pred}$ – предсказание модели;
- б) функция потерь $e(\vec{X}_{fact}, \vec{X}_{pred})$ – функция, выпуклая вверх по второй переменной;

тогда имеет смысл задача минимизации штрафного функционала $L(\vec{\theta})$, вычисляемого как среднее значение функции потерь e по архиву:

$$L(\vec{\theta}) = \frac{1}{M} \sum_{j=1}^M e(\vec{X}_{fact,j}, F(\vec{X}_{input,j}, \vec{\theta})) \rightarrow \min_{\vec{\theta}}, \quad (1.1)$$

Оптимизационная задача (1.1) называется задачей машинного обучения с учителем или обучением модели $F(\vec{X}_{input}, \vec{\theta})$. Выделим четыре основных подхода к построению моделей МО:

1. Линейный подход: модель – линейная [244];
2. Решающие деревья: модель – кусочно-постоянная [206], [114];
3. Нейронные сети: модель – нейронная сеть [215], [51], [85], [108];
4. Генетическое программирование: модель – композиция некоторых заранее заданных «элементарных» функций [173], [174].

В линейном подходе функция $F(\vec{X}_{input}, \vec{\theta})$ линейна по обоим переменным: $F(\vec{X}_{input}, \vec{\theta}) = \vec{\theta}^T \vec{X}_{input}$. Рассмотрим в качестве функции потерь квадрат евклидова расстояния между предсказанными и фактическими значениями $e = \|\vec{X}_{fact} - \vec{X}_{pred}\|_{L_2}^2$, тогда задача (1.1) является классической задачей линейной регрессии, которая может быть решена методом наименьших квадратов.

Некоторые модели МО допускают наличие пропусков некоторых компонент $\vec{X}_{input}$ в некоторых реализациях. Например, применим метод наименьших квадратов: для его применения достаточно оценить положительно определенную ковариационную матрицу, а при отсутствии части $\vec{X}_{input}$ решать систему линейных алгебраических уравнений (СЛАУ) на коэффициенты линейной регрессии с соответствующим минором ковариационной матрицы. Отметим, что оценка ковариационной матрицы по неполным данным может приводить к вырожденной или плохо обусловленной ковариационной матрице [45].

При постпроцессинге применяются решающие деревья [90], [91], делаются (причем, не всегда удачные [139], [194]) попытки применения ансамблей решающих деревьев. В последнее время все более применение находят и нейронные сети [180], [209], [28], [235].

Чтобы обучать модели с большим количеством параметров (большой размерностью вектора параметров n_θ) на больших архивах данных необходимо применять эффективные методы обучения, например методы градиентного спуска [68], [37], [63]. Если выполнены условия:

1. модель МО $F(\vec{X}_{input}, \vec{\theta})$ непрерывна по обеим переменным и кусочно-гладка по параметрам $\vec{\theta}$;
2. градиент $\nabla_{\vec{\theta}} F(\vec{X}_{input}, \vec{\theta})$ равномерно ограничен во всей рассматриваемой области в пространстве параметров;
3. функция потерь $e(\vec{X}_{fact}, \vec{X}_{pred})$ непрерывна и кусочно-гладка по второму аргументу,

то для оптимизации функционала $L(\vec{\theta})$ можно использовать различные варианты метода градиентного спуска (градиентный метод МО). При нелинейном градиентном МО наиболее эффективными оказываются модели МО $F(\vec{X}_{input}, \vec{\theta})$:

- а) основанные на ансамблях деревьев решений [114], которые кусочно-постоянны по первому аргументу;
- б) использующие нейронные сети [85], [108], которые кусочно-гладки по обоим аргументам.

Опишем некоторые преимущества и недостатки различных способов построения моделей МО и тем самым обоснуем наш выбор нейронных сетей. Более подробно выбор нейронных сетей обоснован в 3 части первой главы.

Решающие деревья широко применяются при прогнозировании явлений погоды. Например, в [90] при прогнозе метеорологических условий для авиации для каждого аэропорта строится своё решающее дерево.

В последние годы методика решающих деревьев была существенно усовершенствована и стали рассматриваться ансамбли решающих деревьев, например, случайный лес и градиентный бустинг над решающими деревьями (gradient boosting decision trees) [114], [184]. При градиентном бустинге обучение деревьев происходит поэтапно, а именно каждое следующее дерево обучается на погрешностях ансамбля деревьев, построенного на предыдущем шаге алгоритма. Существенным недостатком метода решающих деревьев для его применения в метеорологии является отсутствие достаточно проработанных подходов к обработке временных рядов и полей.

И нейронные сети, и градиентный бустинг над решающими деревьями, как правило, обучаются методом градиентного спуска, и функционал качества $L(\vec{\theta})$ в процессе обучения убывает достаточно быстро.

При генетическом программировании модель МО строится как композиция некоторых заранее заданных «элементарных» функций. Преимущества генетического программирования: быстрое применение найденной модели; отсутствие переобучения; отсутствие ограничений на оптимизируемый функционал и рассматриваемые элементарные функции. Последнее означает, что возможна оптимизация разрывных (кусочно-постоянных) функционалов качества и моделей. Недостатки: обучение происходит генетическим алгоритмом [173], который является алгоритмом перебора с применением различных эвристик, что делает процесс обучения чрезвычайно затратным с вычислительной точки зрения (хотя он очень хорошо распараллеливается).

Поскольку существующие ЧМПП основаны на решении дифференциальных уравнений в частных производных в предположении гладкости метеорологических полей, то применение разрывной модели решающих деревьев зачастую возможно лишь при постпроцессинге. Из нелинейных подходов наибольшей популярностью в метеорологии пользуются нейронные сети [180], [235], которые обеспечивают непрерывность (гладкость) модели МО и хорошо подходят для обработки больших объемов информации. Кроме того, нейронные сети позволяют легко обучать квазилинейные методы, применимые в том числе и при отсутствии информации о некоторых из предикторов (параграф 1.3.4).

Учитывая вышесказанное (более подробное обоснование приведено в третьей части главы 1), по мнению диссертанта, наиболее перспективным из подходов к построению моделей машинного обучения в метеорологии является подход нейронных сетей.

Научная новизна

В диссертационной работе получены следующие новые научные результаты:

- Использование нейронных сетей для вычисления коэффициентов систематической коррекции по сравнению систематической коррекцией с экспоненциально убывающими весами уменьшает среднюю погрешность прогнозов приземной температуры воздуха на величину, эквивалентную уменьшению заблаговременности на 1 сутки (глава 2);
- Разработан автоматизированный метод коррекции с помощью нейронных сетей (КНС, глава 3) для расчета нелинейных поправок к прогностическим полям. КНС учитывает индивидуальные параметры для каждой синоптической станции и сдвинутые по начальному сроку и заблаговременности прогнозы. Определены оптимальные гиперпараметры метода КНС для различных корректируемых метеорологических параметров, прогнозируемым по различным методам;
- Предложен метод двумерной квазилинейной неоднородной анизотропной оптимальной интерполяции (КНОИ, глава 4). КНОИ применяет к известным значениям интерполируемого поля квазилинейный оператор, коэффициенты которого вычисляются методом оптимальной интерполяции [36] после вычисляемого нейронной сетью вложения рассматриваемой двумерной области на поверхности Земли в пространство большей размерности. КНОИ по сравнению с обычной однородной оптимальной интерполяцией позволило объяснить (в 1.5 и более раз) большую часть дисперсии найденных ранее поправок в точках синоптических станций.

Положения, выносимые на защиту

1. Новая алгоритм систематической коррекции прогнозов приземных метеорологических параметров для вычисления поправок путем применения оператора с коэффициентами, вычисляемыми нейронной сетью, к известным погрешностям численной модели.

2. Новый алгоритм комплексификации прогнозов нескольких численных моделей, основанный на методе нейронных сетей.

3. Новый алгоритм квазилинейной неоднородной оптимальной интерполяции (КНОИ) инкрементов приземных метеорологических параметров. Алгоритм КНОИ реализует метод оптимальной интерполяции с корреляционной функцией, зависящей от расстояния не в геометрическом пространстве, а в линейном пространстве большей размерности. Вложение в

пространство большей размерности вычисляется нейронной сетью и учитывает предикторы неоднородности (в том числе поля первого приближения). Этот учет предикторов неоднородности позволяет объяснить большую часть дисперсии интерполируемого поля.

Практическая значимость работы

Предлагаемый автоматизированный комплекс постпроцессинга применяется в повседневной практике и результаты публикуются на интернет-сайтах (примеры продукции, использующей прогнозы, рассчитанные по предлагаемым постпроцессингом, представлены в приложении Б):

- методического кабинета ФГБУ «Гидрометцентр России» (далее Гидрометцентр России) <http://method.meteorf.ru/ansambl/ansambl.html> для населенных пунктов России и Беларуси. Прогнозы с заблаговременностью до 96 ч доступны в виде карт по каждому из регионов РФ, а для некоторых населенных пунктов – в виде метеограмм с заблаговременностью до 144 ч;
- проекта всемирной метеорологической организации (ВМО) по прогнозированию суровых погодных условий в Центральной Азии <http://swfdp-ca.meteoinfo.ru/prognozy/mmforecasts> для населенных пунктов Центральной Азии в виде метеограмм;
- Численных прогнозов погоды для метеорологического обеспечения <http://u2019.meteoinfo.ru/services> на метеограммах прогнозов по модели COSMO-Ru в конфигурациях COSMO-Ru6ENA и COSMO-RuBy демонстрируются графики поправленных прогнозов температуры и точки росы на высоте 2 м;
- «О погоде для специалистов» Гидрометцентра России <https://special.meteoinfo.ru/> для населенных пунктов России в разделах «метеограммы» (дублируются метеограммы с сайта методического кабинета и с сайта численных прогнозов погоды для метеорологического обеспечения) и «прогнозы по пунктам».

Качество этих прогнозов оперативно оценивается, и их результаты оперативно публикуются на сайте методического кабинета Гидрометцентра России. Полученные оценки качества полученных прогнозов максимальной дневной и минимальной ночной температур превосходят соответствующие оценки качества прогнозов синоптиков УГМС.

Имеются отзывы, положительно отмечающие полезность и качество представляемых на сайте методического кабинета Гидрометцентра России прогнозов от синоптиков Приволжского УГМС, Саратовского ЦГМС, Читинского ГМЦ, Ульяновского ЦГМС, Дальневосточного УГМС, Мурманского УГМС, Владимирского ЦГМС (приложение В).

Личный вклад автора

Все представленные в диссертации результаты получены лично автором. Автор лично предложил и реализовал в программном виде предлагаемые методы и провел их отладку, обучение, валидацию, тестирование и организовал работу в автоматизированном режиме.

Апробация работы

Основные положения и результаты диссертации докладывались и обсуждались на международных и российских научно-технических конференциях и семинарах:

1. «М.А.Петросянц и отечественная метеорология» в 2009 и 2019 годах.
2. Научная сессия Совета РАН по нелинейной динамике в 2009, 2011, 2014, 2016, 2019 годах.
3. XXVIII International Conference on Mathematical Geophysics “Modelling Earth Dynamics: Complexity, Uncertainty and Validation” в 2010 году.
4. European Geophysical Union General Assembly в 2011 году.
5. XX Всероссийская конференция «Теоретические основы и конструирование численных алгоритмов решения задач математической физики» в 2014 году.
6. XVI Всероссийская открытая конференция «Современные проблемы дистанционного зондирования земли из космоса» в 2016 году.
7. The China-Russia Conference on Numerical Algebra with Applications (CRCNAA) в 2017 году.
8. Вторая научно-практическая конференция «Современные информационные технологии в гидрометеорологии и смежных с ней областях» в 2017 году.
9. V Международная научная конференция «Региональные проблемы дистанционного зондирования Земли» в 2018 году.
10. Вторая Всероссийская научная конференция с международным участием «Применение средств дистанционного зондирования Земли в сельском хозяйстве» в 2018 году.
11. Семинар в главной геофизической обсерватории им. А.И.Воейкова в 2018 году.
12. Межвузовская научно-техническая конференция студентов, аспирантов и молодых специалистов им. Е.В. Арменского в 2019 и 2021 годах.
13. COSMO General Meeting в 2020 году.
14. International Conference on Computer Simulation in Physics and beyond в 2020 году.
15. ICON/COSMO/CLM/ART User Seminar (ICCARUS) в 2020 и 2021 годах.

16. Семинар лаборатории вычислительного интеллекта Сколковского института науки и технологий в 2021 году.

Публикации

Основные результаты по теме диссертации изложены в 16 научных работах, 6 из которых изданы в периодических научных журналах, рекомендованных ВАК и индексируемых в базе данных Scopus; 1 — в материалах конференций; 5 — в сборниках тезисов конференций.

Разработки по теме исследования

Методы машинного обучения в общем (и, в частности, нейронные сети) рассматривались многими авторами для задач постпроцессинга полей приземных метеорологических параметров.

Статистическая структура метеорологических полей изучалась при помощи разложений метеорологических полей на естественные ортогональные составляющие (ЕОС) [16], [65], [55], [59], [57]. Так как коэффициенты разложения на первые ЕОС более устойчивы к шумам, чем исходные метеорологические поля, то получающиеся на их основе статистические прогнозы метеорологических величин в некоторых случаях оказываются точнее [244], [67], [46].

Хотя использование ЕОС не позволяет выйти из класса линейных моделей МО, устойчивость метода разложения на ЕОС позволяет искать параметры, индивидуальные для отдельных групп случаев, например, для отдельных регионов, временных периодов или синоптических ситуаций.

Опишем сначала преимущества использования именно нейронных сетей с теоретической точки зрения, а затем перейдём к описанию конкретных результатов.

Нейронные сети являются гладкой (кусочно-гладкой) моделью МО. Любую непрерывную функцию на n -мерном кубе можно приблизить нейронной сетью в метрике C [128], [254], параграф 1.1.3. Нейронные сети с вычислительной точки зрения хорошо подходят для обработки большого объёма данных и не предполагают, что значения исследуемых параметров распределены согласно какому-то вероятностному распределению.

Нейронные сети типа автокодировщик (autoencoder) [113], [208] являются нелинейным обобщением идеи ЕОС. Последовательное рассмотрение нескольких автокодировщиков приводит к так называемым глубоким нейронным сетям [156], [109].

Большой обзор применения МО в задачах прогноза погоды сделан в книге [180], а в задачах постпроцессинга ЧМПП – в недавнем обзоре [235]. Приведем некоторые результаты по коррекции прогнозов ЧМПП.

В [209] предлагается для коррекции среднего и разброса ансамблевых прогнозов использовать нейронные сети с индивидуальными параметрами для каждой синоптической станции (так называемые Embeddings). Использовались данные с 537 синоптических станций на территории Германии. Показано, что добавление Embedding улучшает качество ансамблевых прогнозов приземной температуре воздуха с заблаговременностью 48 ч в метрике continuous ranked probability score (CRPS) на 10%: с 0.9 °C до 0.82 °C.

В [150] предложено использовать свёрточные нейронные сети (convolution neural network, CNN) специальной конфигурации U-net [214] для коррекции среднего и разброса ансамблевых прогнозов с заблаговременностью 48 ч. Прогнозы сравнивались с полями анализа. При количестве членов ансамбля от 1 до 9 применение коррекции эквивалентно уменьшению количества членов ансамбля на 1-2, например, качество откорректированного прогноза с 8 членами близко к качеству исходных прогнозов с 10 членами ансамбля. При 10 членах ансамбля эффект, выраженный в процентном уменьшении среднеквадратической ошибки RMS, при коррекции поля T850 (RMS уменьшилась на 7.9% при использовании CNN против 4.8% при использовании линейного метода) существенно больше эффекта при коррекции поля H500 (RMS уменьшилась на 2.6% при использовании CNN против 2.1% при использовании линейного метода).

В [225] для коррекции ансамблевых прогнозов температуры и количества осадков на территории Западной Европы применяется метод, основанный на ансамблях решающих деревьев (так называемых случайных лесах), использующий в том числе, данные об орографии. Использовались данные с 2000 синоптических станций. Продemonстрировано улучшение качества ансамблевых прогнозов приземной температуры воздуха в метрике CRPS с 1.3-1.5°C у исходных прогнозов до 0.7-1.2 °C у откорректированных (в зависимости от заблаговременности прогноза, которая не превосходит 90 ч).

В [191] предложено использовать нейронные сети для поиска корреляционных связей прогнозов на срок от 2 недель до 2 месяцев. Из более ранних работ [86], [80], [189] (не использующих нелинейное машинное обучение) известно, что при прогнозировании на такие сроки оказывают влияние нелокальные статистические связи, задача выявления которых не тривиальна, особенно при использовании нелинейных статистических моделей.

Во всех приведенных работах данные наблюдений используются только при обучении модели МО. Использование данных наблюдений так же и при применении модели МО может

заметно уменьшить погрешности откорректированного прогноза, но тогда возникает проблема применимости коррекции вне мест расположения синоптических станций.

Перейдем к обзору результатов прогнозирования методами машинного обучения параметров и явлений, которые рассматриваемая численная модель не прогнозирует.

В статье [139] проведено сравнение различных подходов к задаче постпроцессинга ансамблевых прогнозов осадков на территории США. Прогнозы сравнивались с данными композита радарных наблюдений. Показано, что качество прогнозов с заблаговременностью 30 ч, полученных линейной логистической регрессией, в среднем немного лучше прогнозов, полученных методом случайного леса.

В работе [192] в задаче прогноза скорости порывов ветра на 50 аэродромах и озерах в Швейцарии сравнивались результаты постпроцессингов, основанных на генетическом программировании и на традиционном методе логистической регрессии. Показано, что среднее качество прогнозов, полученных этими методами весьма близко (процент попаданий 80-81% против 55% у исходных прогнозов модели COSMO (Consortium for Small-scale Modeling) [122] в конфигурации COSMO-EE-2 при близком количестве ложных тревог), но метод, основанный на генетическом программировании, имеет меньший разброс качества для различных прогностических пунктов. В [192] продемонстрировано, что предупреждения, выдаваемые синоптиками, имеют существенно меньший процент попаданий.

В статье [194] производилось сравнение постпроцессингов, основанных на случайном лесе, и некоторых линейных методов по территории США для прогноза различных редких явлений. Например, при прогнозе града размером более 25мм качество прогнозов, полученных с помощью леса решающих деревьев и лучшим из рассмотренных линейных методов, совпадает. При прогнозе турбулентности на высоте полета самолетов метод, основанный на решающих деревьях, имеет преимущество.

В работе [251] предложена методика прогноза вероятности конвективных явлений с заблаговременностью до 48 ч, основанный на свёрточных нейронных сетях (CNN) с $1,65 \times 10^6$ параметров. Для прогноза вероятности конвективных явлений CNN принимает на вход 144 поля метеорологических параметров на 1° сетке в окрестности размера $7^\circ \times 7^\circ$ вокруг данной точки. Для обучения использованы архивы наблюдений на 20000 автоматических и 2420 синоптических станций на территории Китая. Демонстрировано, что прогнозы, вычисленные CNN, превосходят по качеству прогнозы, полученные другими методами (решающие деревья, линейные методы). Также на архиве за 2015-2017 года прогнозы CNN сравнивались с независимыми прогнозами синоптиков. Прогнозы, сделанные синоптиками, имели большее количество ложных тревог при прогнозировании грозы при аналогичной предупреденности, а

при прогнозировании сильного дождя, града и конвективных явлений имели примерно в 2 раза меньшую предупреденность чем прогнозы CNN при аналогичной доле ложных тревог.

Суммируем результаты обзора применений машинного обучения в постпроцессинге численных прогнозов. Оценки качества прогнозов, полученных методом решающих деревьев, не всегда демонстрируют превосходство над линейными методами. Наиболее перспективным подходом к постпроцессингу среди нелинейных методов является подход нейронных сетей [235]. При прогнозировании методами машинного обучения параметров и явлений, которые рассматриваемая ЧМПП не прогнозирует, не во всех задачах наилучший результат дают новые, нелинейные методы. Наиболее интересные результаты демонстрируют реализации методов, при обучении которых использованы большие архивы фактической и соответствующей прогностической информации.

Благодарности

Автор благодарит научного руководителя доктора физико-математических наук В.А.Гордина за полезные обсуждения, ссылки на литературу и критику текста диссертации, которые помогли сделать текст лучше.

Автор выражает благодарность сотрудникам ФГБУ «Гидрометцентр России»: А.Н.Багрову за помощь в работе с архивами наблюдений на синоптических станциях и прогнозов иностранных и отечественных ЧМПП, а так же за возможность использовать программные коды разработанной им системы оценок качества прогнозов приземных метеорологических параметров; Г.С.Ривину, без поддержки и наставничества которого работа могла не состояться; Д.В.Блинову за техническую помощь в работе с архивами прогнозов модели COSMO-Ru и за визуализацию данных; А.Ю.Бундель, А.В. Муравьеву, И.А.Розинкиной, А.В.Романову, Л.Л.Тарасовой и М.Д.Цырульникову за полезные обсуждения и ссылки на литературу; Ю.А.Степанову за помощь в настройке и обеспечении совместимости используемых прикладных программ; И.И.Жабиной, И.И.Кулаковой и А.Ю.Недачиной за организацию и техническую поддержку используемых баз данных «Прогноз» Гидрометцентра России; Н.А.Светлову, Е.Н.Шакотько и И.А.Уманскую за поддержку сайта комплексного прогноза и проведенные сравнения прогнозов, размещенных на некоторых интернет-сайтах.

Автор благодарит директора ФГБУ «ГВЦ Росгидромета» С.В.Лубова за помощь с установкой используемых Python пакетов на суперкомпьютерные системы Cray XC40-LC и T-Platforms V6000 и предоставление доступа к этим вычислительным комплексам.

ГЛАВА 1. Используемые методы и данные

В данной работе приняты следующие обозначения: t – исходный срок прогнозов численной модели прогноза погоды (ЧМПП), τ – заблаговременность прогноза, i – номер ЧМПП, результаты которой используются в качестве одного из предикторов. Пусть вектор-функция $\vec{\Psi}_i(t, \tau, \vec{x})$ – прогноз $k = \dim \Psi_i$ приземных метеорологических параметров в точке $\vec{x}$ области Ω на поверхности Земли от срока t на момент времени $t + \tau$ согласно i -й ЧМПП.

Рассматривалась вектор-функция $\vec{\Psi}_i(t, \tau, \vec{x})$, состоящая из $k = 7$ компонент:

- 1) температура ($^{\circ}\text{C}$) воздуха на высоте 2 м $T_i(t, \tau, \vec{x})$;
- 2) точка росы ($^{\circ}\text{C}$) на высоте 2 м Td_i ;
- 3) давление (гПа), приведенное к уровню моря P_i ;
- 4) зональная компонента скорости ветра (м/с) на высоте 10 м U_i ;
- 5) меридиональная компонента скорости ветра (м/с) на высоте 10 м V_i ;
- 6) модуль скорости ветра (м/с) на высоте 10 м $S_i = |\vec{U}_i|$;
- 7) скорость порывов ветра (м/с) на высоте 10 м G_i .

Для некоторых из рассмотренных ЧМПП некоторые из указанных полей в Гидрометцентре России недоступны (вторая часть главы) – в этом случае размерность вектор-функции $\vec{\Psi}_i(t, \tau, \vec{x})$ для i -ой модели была меньше.

Фактически измеренные в момент времени t на синоптической станции, расположенной в точке $\vec{x}$, значения обозначим:

- 8) «срочная» температура воздуха ($^{\circ}\text{C}$) на высоте 2 м T_{fact} ;
- 9) экстремальная (минимальная ночная $T \min_{fact}(t) = \min_{t_1 \in [t-6ч, t+6ч]} T(t_1)$ и максимальная дневная $T \max_{fact} = \max_{t_1 \in [t-6ч, t+6ч]} T(t_1)$) температура воздуха ($^{\circ}\text{C}$) на высоте 2 м TE_{fact} ;
- 10) точка росы ($^{\circ}\text{C}$) на высоте 2 м Td_{fact} ;
- 11) давление (гПа), приведенное к уровню моря P_{fact} ;
- 12) скорость ветра (м/с) на высоте 10 м $\vec{U}_{fact}$;
- 13) модуль скорости ветра (м/с) на высоте 10 м $S_{fact} = |\vec{U}_{fact}|$;
- 14) скорость (м/с) порывов ветра на высоте 10 м G_{fact} ;

15) количество (мм) осадков за предыдущие 12 ч $B_{12, fact}$.

1.1. Обзор технологии искусственных нейронных сетей

1.1.1 Одновременная оптимизация прогноза нескольких метеорологических параметров

Будем для каждого из следующих трех наборов прогнозируемых метеорологических параметров: а) $X_{fact} = P$; б) $\vec{X}_{fact} = (T, Td, TE)$; в) $\vec{X}_{fact} = (U, V, G)$ строить свою нейронную сеть $\vec{X}_{NN} = F(\vec{X}_{input}, \vec{\theta})$, где $\vec{X}_{NN}$ – вектор-столбец, позволяющий однозначно вычислить $\vec{X}_{pred}$ согласно формулам (1.3) далее. Таким образом, размерности вектора наблюдений $\vec{X}_{fact, j}$ и вектора соответствующего прогноза $\vec{X}_{pred, j}$ были равны 3 (б, в) или 1 (а). Параметры из одного набора имеют одинаковые физические размерности.

Прогнозы $\vec{X}_{pred}$ из одного набора должны удовлетворять очевидным неравенствам:

$$T_{pred} \geq Td_{pred}, T_{pred}(\tau + \Delta\tau) \geq T \min_{pred}(\tau), T_{pred}(\tau + \Delta\tau) \leq T \max_{pred}(\tau), \quad (1.2)$$

$$U_{pred}^2 + V_{pred}^2 = S_{pred}^2 \leq G_{pred}^2$$

при всех $\Delta\tau \in [-6ч, 6ч]$.

Нейронные сети, предназначенные для прогнозирования скорости ветра, возвращали 4 параметра: $\vec{X}_{NN} = (U_{NN}, V_{NN}, S_{NN}, G_{NN})$, а все остальные нейронные сети возвращали столько же параметров, сколько в векторе наблюдений $\vec{X}_{fact}$. Чтобы обеспечить выполнение системы неравенств (1.2) перед вычислением штрафного функционала будем к $\vec{X}_{NN}$ применять процедуру согласования, то есть по $\vec{X}_{NN}$ вычислять $\vec{X}_{pred}$ согласно формулам:

$$\begin{aligned} P_{pred} &= P_{NN}, T_{pred} = T_{NN}, Td_{pred} = \min(Td_{NN}, T_{NN}), \\ Tmin_{pred} &= \min\left(Tmin_{NN}, \min_{\Delta t = -6ч, 6ч} T_{NN}(t + \Delta t)\right), \\ Tmax_{pred} &= \max\left(Tmax_{NN}, \max_{\Delta t = -6ч, 6ч} T_{NN}(t + \Delta t)\right), \\ |\vec{U}_{pred}| &= \max(0, S_{NN}), U_{pred} = \frac{U_{NN}}{|\vec{U}_{NN}|} |\vec{U}_{pred}|, V_{pred} = \frac{V_{NN}}{|\vec{U}_{NN}|} |\vec{U}_{pred}|, \\ G_{pred} &= \min\left(\max(G_{NN}, g_0 |\vec{U}_{pred}|), g_1 |\vec{U}_{pred}|\right) \end{aligned} \quad (1.3)$$

где $g_1 > g_0 \geq 1$ – дополнительные оптимизируемые (одновременно с параметрами $\vec{\theta}$ нейронной сети F) константы. Параметры g_1, g_0 оптимизировались в смысле минимизации G_{loss} (параграф 1.1.2, формула (1.16)) – одного из слагаемых функции потерь при прогнозе ветра. Перед началом минимизации значения констант инициализировались следующим образом: $g_0 = 1.1, g_1 = 3$. Далее применение формул (1.3) будем обозначать как применение оператора *adapt*, например $(T_{pred}, Td_{pred}) = adapt(T_{NN}, Td_{NN})$.

Для того, чтобы гарантировать выполнение неравенств (1.2), необходима совместная оптимизация прогнозов соответствующих метеорологических параметров, а при совместной оптимизации прогнозов нескольких различных параметров необходим совместный критерий качества (совместные функции потерь).

В данной работе качестве функции потерь e при прогнозировании метеопараметров $X = T, D, \vec{U}, P$ использовалась функция потерь Хьюбера [162] – кусочно-аналитическая комбинация среднеквадратичной ошибки при малых погрешностях и средней абсолютной ошибки при больших (её график приведен на рисунке 1.1):

$$e_{huber}(X_{fact}, X_{pred}) = \begin{cases} c^{-1}(X_{pred} - X_{fact})^2 & \Leftarrow |X_{pred} - X_{fact}| < c; \\ 2|X_{pred} - X_{fact}| - c & \Leftarrow |X_{pred} - X_{fact}| \geq c, \end{cases} \quad (1.4)$$

где $c = 2$ °C при прогнозировании температуры и точки росы ($X = T, D$); $c = 2$ м/с при прогнозировании скорости ветра ($X = \vec{U}$); $c = 2$ гПа при прогнозировании давления ($X = P$), приведенного к уровню моря. Эксперименты показали, что выбор константы c не оказывает существенного влияния на получаемые результаты.

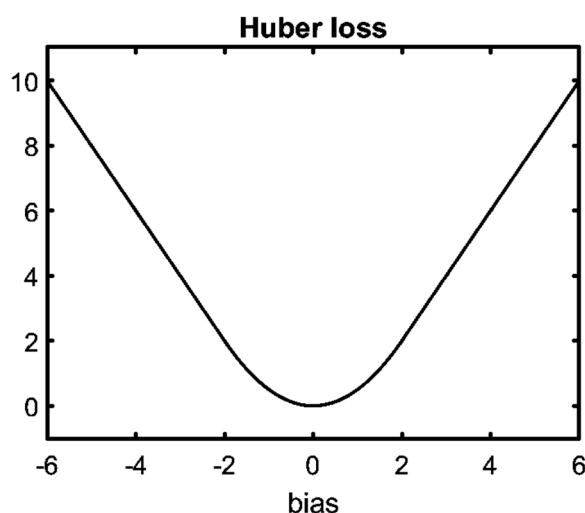


Рисунок 1.1 – Функция потерь Хьюбера (1.4) в зависимости от погрешности прогноза

В качестве совместных функций потерь для каждого из трех наборов прогнозируемых величин будем использовать функции потерь, являющиеся линейной комбинацией некоторых функций потерь от компонент вектора $\vec{X}_{pred}$, а именно:

$$\begin{aligned} e\left((T_{fact}, Td_{fact}, TE_{fact}), (T_{pred}, Td_{pred}, TE_{pred})\right) &= c_T e_{huber}(T_{fact}, T_{pred}) + \\ &+ c_{Td} e_{huber}(Td_{fact}, Td_{pred}) + c_{TE} e_{huber}(TE_{fact}, TE_{pred}), \\ e\left((T_{fact}, Td_{fact}), (T_{pred}, Td_{pred})\right) &= c_T e_{huber}(T_{fact}, T_{pred}) + c_{Td} e_{huber}(Td_{fact}, Td_{pred}), \\ e\left((U_{fact}, V_{fact}, G_{fact}), (U_{pred}, V_{pred}, G_{pred})\right) &= c_U e_{huber}(U_{fact}, U_{pred}) + \\ &+ c_V e_{huber}(V_{fact}, V_{pred}) + c_S e_{huber}(|U_{fact}|, |U_{pred}|) + c_G G_{loss}(\tilde{G}_{fact}, G_{pred}), \end{aligned} \quad (1.5)$$

где константы выбраны эмпирически:

$$c_T = 4, c_{Td} = 3, c_{TE} = 20, c_U = c_V = 1, c_S = 3, c_G = 30 \text{ м/с}.$$

Вообще говоря, выбор этих констант должен выполняться с учётом технических требований пользователя прогнозов к качеству прогнозов. По результатам численных экспериментов, эти константы существенно влияют на оптимальный прогноз в тех случаях, когда оператор *adapt* (формулы (1.3)) не тождественен $\vec{X}_{NN} \neq \vec{X}_{pred} = \text{adapt}(\vec{X}_{NN})$.

Функция потерь, используемая при оптимизации прогноза порывов ветра, G_{loss} будет описана ниже в параграфе 1.1.2, а величина $\tilde{G}_{fact}$ вычислялась по формуле:

$$\tilde{G}_{fact} = \begin{cases} G_{fact} & \Leftarrow \text{синоптическая станция передала } G_{fact} \\ \min(1.9 \cdot |\vec{U}_{fact}|, 10.5 \text{ м/с}) & \Leftarrow \text{иначе} \end{cases} \quad (1.6)$$

Различие между величинами G_{fact} и $\tilde{G}_{fact}$ обусловлено тем, что, согласно коду КН-01 [84], значение G_{fact} может не передаваться (а может и передаваться), если измеренный максимальный ветер не превосходит 10 м/с. Выбранная нами константа 1.9 в формуле (1.6) соответствует среднему отношению между $|\vec{U}_{fact}|$ и G_{fact} . Как правило, станции, привязанные к портам и аэропортам, передают значения G_{fact} в том числе и меньшие 10 м/с.

Если предсказываемая величина $X_{fact,j} \in \{0,1\}$, где $j = 1 \dots M$ – номер реализации интерпретируется как бинарное вероятностное событие, а $X_{pred,j} \in [0,1]$ – как вероятность этого события, то задача (1.1) является задачей прогноза вероятности. Предположим, что события $X_{fact,j}$ и $X_{fact,k}$ независимы при $k \neq j$. В такой ситуации часто используют логарифмическую функцию потерь:

$$e(\vec{X}_{fact}, \vec{X}_{pred}) = -\ln p(\vec{X}_{fact} | \vec{X}_{pred}), \quad (1.7)$$

При использовании функции потерь (1.7) задача МО (1.1) эквивалентна максимизации логарифмической функции правдоподобия [138]. Действительно, подставляя формулу (1.7) в формулу (1.1) и используя независимость событий при $k \neq j$, получим:

$$L(\vec{\theta}) = -\frac{1}{M} \ln \prod_{j=1}^M p(\vec{X}_{fact,j} | \vec{X}_{pred,j}) = -\frac{1}{M} \sum_{j=1}^M \ln p(\{\vec{X}_{fact,j}\} | \{\vec{X}_{pred,j}\}).$$

Таким образом, при минимизации функционала $L(\vec{\theta})$ максимизируется $p(X_{fact} | X_{pred})$, то есть правдоподобие – вероятность того, что вероятностный прогноз X_{pred} события X_{fact} верен.

Если предсказываемая величина интерпретируется как детерминированная величина, то задачу (1.1) называют задачей регрессии. Как правило, в задаче регрессии используют функцию потерь, такую, что её градиент по прогностическому значению $\nabla_{\vec{X}_{pred}} e(\vec{X}_{fact}, \vec{X}_{pred})$ кососимметричен.

Чтобы обеспечить несмещённость оценок, то есть выполнение условия $\overline{\vec{X}_{fact} - \vec{X}_{pred}} = \vec{0}$, нужно использовать функцию потерь с кососимметричным градиентом $\nabla_{\vec{X}_{pred}} e(\vec{X}_{fact}, \vec{X}_{pred})$. Действительно, если функция двух векторных аргументов $\nabla_{\vec{X}_{pred}} e$ не кососимметрична, то есть существуют $\vec{X}_1, \vec{X}_2$ такие, что $\nabla_{\vec{X}_2} e(\vec{X}_1, \vec{X}_2) + \nabla_{\vec{X}_1} e(\vec{X}_2, \vec{X}_1) \neq 0$, тогда (в силу определения градиента) существует вектор $\vec{\varepsilon} \neq 0$, такой что

$$e(\vec{X}_1, \vec{X}_2 + \vec{\varepsilon}) + e(\vec{X}_2, \vec{X}_1 + \vec{\varepsilon}) < e(\vec{X}_1, \vec{X}_2) + e(\vec{X}_2, \vec{X}_1),$$

то есть прогноз со смещением равным $\vec{\varepsilon} \neq 0$ имеет меньшее значение штрафного функционала (1.1). Отметим, что градиент логарифмической функции потерь (1.7) кососимметричен в силу определения условной вероятности.

При оптимизации функционала (1.1) вычисляется градиент функции потерь $\nabla_{X_{pred}} e$ по прогностическому значению. Для функции потерь Хьюбера (1.4) этот градиент равен:

$$\nabla_{X_{pred}} e_{huber}(X_{fact}, X_{pred}) = \begin{cases} 2, & X_{pred} - X_{fact} \geq 2; \\ X_{pred} - X_{fact}, & -2 < X_{pred} - X_{fact} < 2; \\ -2, & -2 \geq X_{pred} - X_{fact}. \end{cases} \quad (1.8)$$

Таким образом, функция потерь Хьюбера (1.4) имеет ограниченный градиент $\left| \nabla_{X_{pred}} e_{huber} \right| \leq 2$, что гарантирует сходимость метода градиентного спуска. Градиент более традиционной среднеквадратичной функции потерь

$$MSE(X_{fact}, X_{pred}) = (X_{pred} - X_{fact})^2$$

вообще говоря, не ограничен:

$$\nabla_{X_{pred}} MSE = 2(X_{pred} - X_{fact}).$$

Поэтому использование функции Хьюбера (1.4) в функционале (1.1) позволяет (по сравнению со среднеквадратичной метрикой MSE) исключить излишнюю подстройку параметров $\vec{\theta}$ функции $F(\vec{X}_{input}, \vec{\theta})$ под прогнозы с очень большой ошибкой: влияние прогнозов с ошибкой больше константы c на получившиеся в результате оптимизации параметры $\vec{\theta}$ одинаково.

Хотя градиент MSE не ограничен, это не мешает сходимости градиентного спуска. Действительно, ограничимся рассмотрением шара в пространстве параметров

$$S = \{ \vec{\theta} : |\vec{\theta} - \vec{\theta}_*| \leq |\vec{\theta}_0 - \vec{\theta}_*| \},$$

где $\vec{\theta}_0$ – начальное значение вектора параметров, $\vec{\theta}_*$ – реализующее минимум функционала L значение вектора параметров. Тогда в этом шаре градиент MSE будет ограничен.

1.1.2. Задача прогноза порывов ветра и порядковая регрессия

Согласно наставлению [60], [61] прогноз скорости порывов ветра считается оправдавшимся, если он попадает в ту же численную градацию, что и измеренное значение максимального ветра. Градации определены в наставлении так: 0-12, 12-18, 18-24, более 24 м/с.

Задача порядковой регрессия (ordinal regression, [193], [185], [202]) заключается в максимизации оправдываемости, то есть количества попаданий прогнозируемой величины G_{pred} в ту же заранее заданную градацию, $\xi_i, i = 1 \dots n_g$, где находится G_{fact} :

$$U_{loss}(G_{fact}, G_{pred}) = \sum_{i=1}^{n_g} v_i I_{\xi_i}(G_{pred}) I_{\xi_i}(G_{fact}) \rightarrow \max_{G_{pred}}, \quad (1.9)$$

$v_i > 0$ – безразмерные весовые коэффициенты, описывающие важность соответствующей градации (в наставлении [61] все эти весовые коэффициенты равны 100%). Для решения задачи порядковой регрессии (максимизации оправдываемости (1.9)) методами машинного обучения

необходимо её формализовать в виде задачи (1.1). Максимизировать оправдываемость (1.9) нельзя методом градиентного спуска, поскольку факт попадания в определенный интервал $I_{\xi_i}(G_{pred})$ не является непрерывной функцией от G_{pred} , а тем более кусочно-гладкой функцией. Приблизим сначала разрывную функцию потерь (1.9) семейством гладких функций потерь $U_{loss,\delta}$, при $\delta \rightarrow +\infty$ почти во всех точках стремящихся к U_{loss} .

Чтобы получить функции потерь $U_{loss,\delta}$, перейдем от детерминированной логики к нечеткой, а затем используем функцию потерь (1.7) метода максимального правдоподобия. Обозначим как $p_\delta(G > g)$ вероятность того, что прогностическая (при $G = G_{pred}$) или фактическая (при $G = G_{fact}$) скорость порывов ветра больше порога g :

$$p_\delta(G > g) = \sigma(\delta \cdot (G - g)),$$

где $\sigma(x)$ – логистическая функция, а константа $\delta > 0$ имеет физическую размерность, обратную к размерности скорости G (то есть с/м). Величина δ^{-1} имеет смысл допустимой погрешности. При $\delta \rightarrow +\infty$ выражение для $p_\delta(G > g)$ при $G \neq g$ будет стремиться к функции Хевисайда $I_{G>g}$:

$$\lim_{\delta \rightarrow +\infty} p_\delta(G > g) = I_{G>g}.$$

Логистическая функция удовлетворяет соотношению $\sigma(x) + \sigma(-x) = 1$, следовательно, выражение $\sigma[-\delta(G - g)]$ равно вероятности того, что $G \leq g$.

Рассмотрим вероятностный прогноз «скорость порывов ветра превзойдет константу g с вероятностью $p_\delta(G_{pred} > g) = \sigma[\delta(G_{pred} - g)]$ ». Логарифм функции правдоподобия этого прогноза будет равен:

$$\ln p_\delta(\tilde{G}_{fact} > g | G_{pred} > g) = \sum_{\delta_k = -\delta, \delta} \sigma[\delta_k(\tilde{G}_{fact} - g)] \ln \sigma[\delta_k(G_{pred} - g)]. \quad (1.10)$$

При оптимизации оправдываемости прогноза скорости порывов ветра будем использовать функцию потерь, равную взвешенной сумме нескольких логарифмов правдоподобия (1.10):

$$U_{loss,\delta}(\tilde{G}_{fact}, G_{pred}) = \sum_{i=1}^{n_g} v_i \ln p_{\delta_i}(\tilde{G}_{fact} > g_i | G_{pred} > g_i), \quad (1.11)$$

где $g_i = 12, 18, 24$ м/с – заданные в наставлении [61] уровни отсечения; $n_g = 3$ – количество градаций; v_i – безразмерные весовые коэффициенты, описывающие важность прогноза события $\tilde{G}_{fact} > g_i$. Для всех градаций использовались одинаковые значения параметров $v_i = 1$, $\delta_i^{-1} = 1$ м/с.

Знак минус убран в (1.11) поскольку функцию правдоподобия (1.10) необходимо максимизировать, а при оптимизации (1.1) минимизируется функция потерь (1.11).

При оптимизации (1.1) методами градиентного спуска вычисляют частную производную $\partial_{G_{pred}} e$, которая для функции потерь (1.11) кососимметрична и равна:

$$\frac{\partial U_{loss,\delta}}{\partial G_{pred}} = \sum_{i=1}^{n_g} \delta_i v_i \left(\sigma \left[\delta_i (G_{pred} - g_i) \right] - \sigma \left[\delta_i (\tilde{G}_{fact} - g_i) \right] \right). \quad (1.12)$$

Таким образом, получена гладкая функция потерь, приближающая кусочно-постоянную функцию оправдываемости U , то есть формализована задача порядковой регрессии для прогноза скорости порывов ветра. Градиент с компонентами, вычисленными по формуле (1.12), является гладкой ограниченной функцией, а именно по модулю не превосходит $\sum_{i=1}^{n_g} \delta_i v_i$, а значит, метод градиентного спуска будет сходиться.

Если в формуле (1.10) устремить параметр сглаживания $\delta \rightarrow +\infty$, то логарифм правдоподобия (1.10) будет стремиться к 0, если G_{pred} и $\tilde{G}_{fact}$ лежат по одну сторону от g и к $-\infty$, если по разные стороны. То есть, если все значения δ_i велики, то оптимизация функции потерь (1.11) эквивалентна оптимизации общей оправдываемости попадания прогностических и фактических значений в один и тот же интервал среди набора интервалов $(-\infty; g_1), (g_1; g_2), \dots, (g_{n_g}; +\infty)$, где $g_1 < g_2 < \dots < g_{n_g}$.

Если же устремить все значения δ_i к 0, то градиент (1.12) будет эквивалентен следующему выражению:

$$\frac{\partial U_{loss,\delta}}{\partial G_{pred}} = \frac{1}{4} \sum_{i=1}^{n_g} \delta_i^2 v_i (G_{pred} - \tilde{G}_{fact}) + O(\delta_i^4) = \frac{1}{2} \sum_{i=1}^{n_g} \delta_i^2 v_i \frac{\partial}{\partial G_{pred}} (G_{pred} - \tilde{G}_{fact})^2 + O(\delta_i^4). \quad (1.13)$$

Из формулы (1.13) следует, что если все значения δ_i малы, то градиент функции потерь (1.11) пропорционален градиенту среднеквадратической ошибки $(G_{pred} - \tilde{G}_{fact})^2$. Таким образом оптимизация функции потерь (1.11) в пределе эквивалентна минимизации среднеквадратической ошибки.

Однако при прогнозе редких явлений оценка общей оправдываемости оценка общей оправдываемости даёт преимущества категории отсутствия явления, то есть не является *беспристрастной*. В этом случае лучше использовать оценки по более сложным критериям Гильберта *ETS* (equitable threat score, критический индекс успешности) и Багрова – Хайдке *HSS* (heidke skill score) [131], [155], [17], [18], [244]:

$$\begin{aligned}
HSS &= \frac{a_{11}a_{22} - a_{12}a_{21}}{a_{31}a_{23} + a_{13}a_{32}}, \\
ETS &= \frac{a_{33}a_{11} - a_{13}a_{31}}{a_{33}(a_{33} - a_{22}) - a_{13}a_{31}},
\end{aligned} \tag{1.14}$$

где $a_{ik}, i, k = 1, 2, 3$ – элементы таблицы сопряженности (таблица 1.1). Будем использовать критерий Багрова – Хайдке HSS , поскольку он является беспристрастным, в отличие от ETS , который является беспристрастным лишь асимптотически, при стремлении объема архива к бесконечности [158].

Помимо оценок (1.14), далее так же приводятся оценки: предупрежденность явления $Pr = a_{11}/a_{13}$, общая оправдываемость $U = (a_{11} + a_{22})/a_{33}$, доля ложных тревог (false alarm rate) $FAR = a_{12}/a_{13}$.

В качестве метрики качества будем использовать средневзвешенное значение критериев HSS для нескольких градаций g_i :

$$HSS_{loss}(\bar{\theta}) = 2 \sum_{i=1}^{n_g} v_i \frac{a_{11i}a_{22i} - a_{12i}a_{21i}}{a_{31i}a_{23i} + a_{13i}a_{32i}} \rightarrow \max_{\bar{\theta}}, \tag{1.15}$$

где $a_{jki}, j, k = 1, 2, 3$ – таблица сопряженности (таблица 1.1) для градации g_i , v_i – безразмерные весовые коэффициенты, описывающие важность i -й градации. Прямая оптимизация качества получаемого прогноза по критерию HSS_{loss} (и по критерию HSS) методами машинного обучения невозможна: этот критерий не является непрерывным, а тем более кусочно-гладким.

Таблица 1.1 – Таблица сопряженности

	Прогнозировалось $\geq g$	Прогнозировалось $< g$	Всего
Наблюдалось $\geq g$	a_{11}	a_{21}	a_{31}
Наблюдалось $< g$	a_{12}	a_{22}	a_{32}
Всего	a_{13}	a_{23}	a_{33}

Аналогично тому, как выше был введен гладкий аналог оправдываемости (1.11), перейдем от кусочно-постоянного критерия HSS_{loss} к гладкому критерию $HSS_{loss,\delta}$. Будем оптимизировать критерий $HSS_{loss,\delta}$ вычисляемый по формуле (1.15), но вместо частотной матрицы a_{lk} будем использовать матрицу $a_{lk\delta}$ с элементами:

$$a_{lk\delta} = \sum_{j=1}^M \sigma_l \left[\delta(G_{pred,j} - g) \right] \sigma_k \left[\delta(\tilde{G}_{fact,j} - g) \right],$$

где используются сглаживающие логистические функции $\sigma_1(x) = \sigma(x)$, $\sigma_2(x) = 1 - \sigma(x)$, $\sigma_3(x) \equiv 1$. Очевидно, что если пренебречь случаями $G_{pred} = g$ и $\tilde{G}_{fact} = g$, то при $\delta_i \rightarrow +\infty$ элементы матрицы $a_{lk\delta}$ стремятся к соответствующим элементам матрицы a_{lk} .

При решении задачи прогноза скорости порывов ветра градиентными методами машинного обучения будем оптимизировать функционал

$$G_{loss}(\tilde{G}_{fact}, G_{pred}) = \sum_{i=1}^{n_g} v_i (1 - HSS_{loss, \delta_i}(g_i)), \quad (1.16)$$

где i – номер градации, а константы были выбраны эмпирически: $\delta_i^{-1} = 1$ м/с, $v_i = 1$.

В таблице 1.2 приведено сравнение предлагаемого функционала G_{loss} с некоторыми другими. При этом сравнении оптимизировался прогноз скорости порывов ветра как функция одного переменного $G_{pred} = G_{pred}(G_{COSMO-Ru6-ENA})$ от аргумента $G_{COSMO-Ru6-ENA}$ – прогноза скорости порывов ветра по модели COSMO-Ru в конфигурации COSMO-Ru6ENA. Жирным в таблице выделены наилучшие достигнутые результаты. Каждая строка соответствует оптимизации одного из функционалов. Видно, что среди рассмотренных функционалов наилучшие значения и критерия Багрова – Хайдке HSS , и предупрежденности порывов более 21 м/с достигнуты именно при оптимизации функционала G_{loss} .

Для задачи постпроцессинга прогнозов количества осадков можно предложить критерий качества, являющийся аналогом критерия G_{loss} качества порывов ветра (1.16). Однако, при оценках качества прогнозов количества осадков традиционно используются не только критерии качества, основанные на таблицах сопряженности, но и оценки средней абсолютной ошибки MAE и средней ошибки (смещения) ME . Последние особенно важны для применений метеорологических прогнозов в гидрологии [39], [21] и агрометеорологии [83], [78].

Таблица 1.2 – Результаты сравнения некоторых функций потерь в задаче прогноза порывов ветра

Оценивается \ Оптимизируется	G_{loss}	$U_{15\text{ м/с}}$	$U_{21\text{ м/с}}$	$Pr_{15\text{ м/с}}$	$Pr_{21\text{ м/с}}$	$HSS_{15\text{ м/с}}$	$HSS_{21\text{ м/с}}$
RMS	0.936	0.974	0.996	0.078	0.020	0.135	0.038
G_{loss}	0.747	0.965	0.994	0.377	0.268	0.348	0.272
MAE	0.918	0.974	0.996	0.097	0.039	0.161	0.074
$MAE + ME $	0.934	0.974	0.996	0.081	0.021	0.139	0.040
Исходный прогноз	0.776	0.959	0.995	0.437	0.216	0.344	0.267

У диссертанта не получилось выбрать единый критерий качества прогнозов осадков, при оптимизации по которому улучшаются критерии качества прогноза осадков HSS , MAE и ME : эти критерии конфликтуют между собой. Оптимизация HSS для уровней отсечения более 10 мм за 12 часов приводит к увеличению количества предсказаний соответствующего количества осадков, что увеличивает MAE и ME . Оптимизация же критерия MAE приводит к занижению сильных осадков и завышению количества фактов выпадения осадков.

На сайте комплексного прогноза [9] демонстрируются прогнозы 6ч сумм осадков $B_{6,pred}$, полученные диссертантом при оптимизации следующего критерия:

$$B_{loss}(B_{fact}, B_{pred}) = MAE + |ME| - \sum_{i=1}^9 v_i HSS_{loss, \delta_i}(g_i),$$

где $B_{pred} = B_{6,pred}(t, \tau) + B_{6,pred}(t, \tau - 6ч)$, $B_{fact} = B_{12,fact}(t + \tau)$ – спрогнозированные и измеренные 12 ч суммы осадков соответственно, а параметры $\delta_i, v_i, g_i, i = 1...9$ были выбраны следующим образом:

$$\begin{aligned} g_i &= 2^{i-4} \text{ мм/12 ч,} \\ v_i &= \left[1 + 2.5 \left(\frac{g_i}{1 \text{ мм/12 ч}} \right)^{1.3} \right] \text{ мм/12 ч,} \\ \delta_i &= 7 / g_i. \end{aligned}$$

Таким образом, максимальная учитываемая градация количества осадков была равна $g_9 = 32 \text{ мм/12 ч}$.

По мнению диссертанта, без решения проблемы «двойного штрафа», возникающей при промахе прогноза по пространству, невозможно составить хороший критерий качества. Для решения проблемы «двойного штрафа» необходимо иметь архив прогностических полей высокого разрешения. Подобный архив был недоступен. Поэтому далее вопрос прогноза осадков не рассматривается.

1.1.3. Теоретические основы нейронных сетей

Во многих прикладных задачах возникает необходимость аппроксимировать непрерывные функции многих переменных. Опишем конструкцию, называемую нейронной сетью, которая позволяет получить непрерывную аппроксимацию (в метрике C) любой непрерывной функции. Далее (параграф 1.3.2) будем встраивать нейронные сети в более сложные алгоритмы, находя, таким образом, оптимальные параметрические (статистические и функциональные) зависимости, используемые в этих алгоритмах.

Согласно теореме Колмогорова [53] любая непрерывная вещественнозначная функция на n -мерном кубе $F : [0,1]^n \rightarrow \mathbb{R}$ может быть представлена в виде суперпозиции функций одного переменного

$$F(x_1, \dots, x_n) = \sum_{q=1}^{2n+1} \chi_q \left[\sum_{p=1}^n w_{pq}(x_p) \right], \quad (1.17)$$

где «внешние» функции χ_q непрерывны, а «внутренние» w_{pq} непрерывны и монотонны.

Таким образом, задачу аппроксимации непрерывной функции n переменных можно свести к более простым задачам аппроксимации $(2n+1)n$ непрерывных функций одной переменной. Отметим, что нет аналога теоремы Колмогорова для гладких и аналитических функций: можно привести примеры гладких (аналитических) функций трех переменных, не представимых в виде суперпозиции гладких (аналитических) функций одной переменной [33], [89].

В простейшем случае функции w_{pq} линейны, а все внешние функции пропорциональны: $\chi_q(x) = c_q f(x)$, где c_q – константы, а f – какая-то фиксированная непрерывная кусочно-гладкая нелинейная функция. Тогда формула (1.17) совпадает с формулой для вычисления классического варианта нейронной сети – так называемого двухслойного персептрона (подробнее ниже, а также [85]):

$$F(\vec{X}_{input}) = W_2 f(W_1 \vec{X}_{input} + \vec{b}), \quad (1.18)$$

где функция f применяется к компонентам вектора поэлементно, $W_1 : \mathbb{R}^n \rightarrow \mathbb{R}^l$ и $W_2 : \mathbb{R}^l \rightarrow \mathbb{R}$ – линейные операторы, вектор $\vec{b} \in \mathbb{R}^l$. Коэффициенты вектора $\vec{b}$ и операторов W_1 и W_2 (всего $nl + 2l$ коэффициентов) подбираются по обучающему архиву в процессе оптимизации с целью минимизации функционала (1.1), вычисляющего среднюю погрешность прогноза. В формуле (1.17) использовалось $l = 2n + 1$, но в формуле (1.18) можно выбрать и любое другое $l \in \mathbb{N}$. Наибольшая вычислительная эффективность (наибольшая производительность, выраженная в количестве операций умножения с плавающей точкой) на современных компьютерах достигается при размерностях l и n кратных 32.

Элементарный пример обучаемой искусственной нейронной сети – линейный оператор – реализуется при тождественной функции активации f . Простейшая же нелинейная нейронная сеть – двухслойный персептрон (1.18) – была предложена Розенблаттом [215], [73]. Пример схемы двухслойного персептрона показан на рисунке 1.2.

Для удобства описания и программирования нейронные сети рассматривают состоящими из стандартных блоков, называемых слоями. Под глубиной K нейронной сети имеют ввиду максимальное количество слоёв, которое должен пройти сигнал внутри сети от входа до выхода. При наличии дополнительных знаний о задаче, можно рассматривать какие-то специальные слои. Но при отсутствии таких знаний логично выбирать наиболее вычислительно эффективные слои.

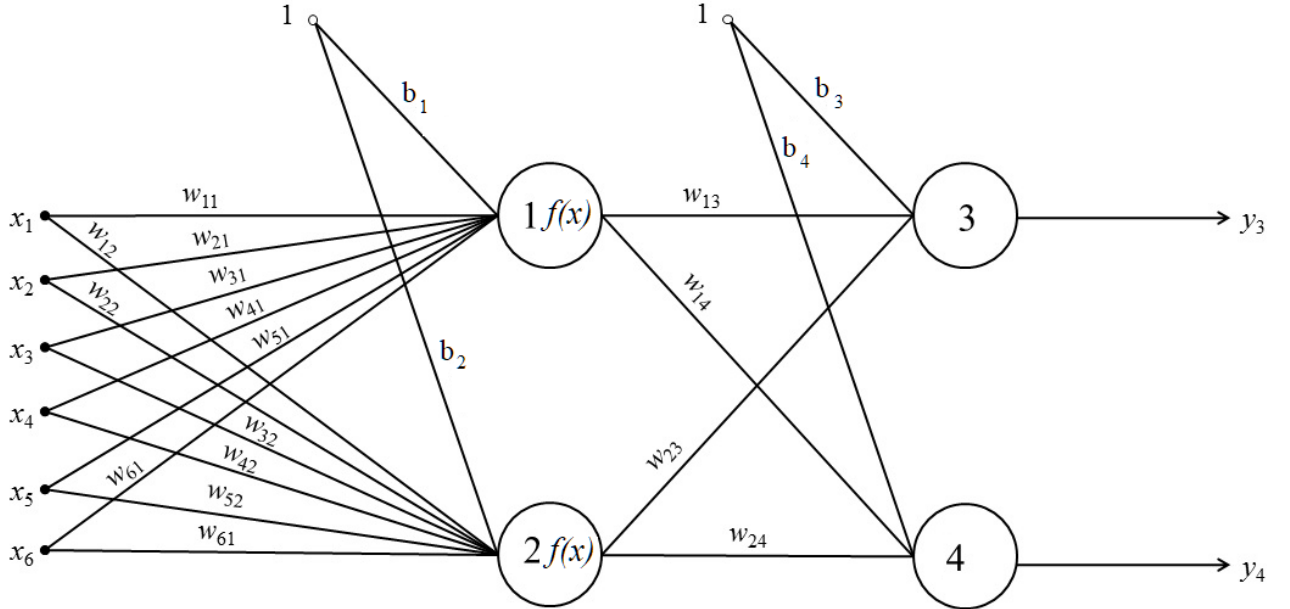


Рисунок 1.2 – Пример схемы двухслойного персептрона с 6 аргументами, 2 нейронами на скрытом слое и 2 выходами. Используются обозначения из формулы (1.18)

Высокой вычислительной эффективностью обладают так называемые полносвязные слои (fully-connected, далее FC), осуществляющие последовательное применение к входному вектору:

- а) линейного оператора $W: \mathbb{R}^n \rightarrow \mathbb{R}^l$;
- б) сдвига на вектор $\vec{b} \in \mathbb{R}^l$;
- в) заранее выбранной нелинейной вектор-функции f (называемую функцией активации).

Применение полносвязного слоя можно записать в виде формулы

$$\vec{x} \mapsto \vec{y} = F(\vec{x}, \{W, \vec{b}\}) = f(W\vec{x} + \vec{b}). \quad (1.19)$$

Параметрами слоя (1.19) являются коэффициенты линейного оператора W и вектора $\vec{b} \in \mathbb{R}^l$. Если не предполагается выполнения каких-либо дополнительных условий на линейный оператор W , то такой слой называют полносвязным с l нейронами и функцией активации f . Двухслойный персептрон (1.18) можно представить в виде двух FC слоёв (1.19):

$$\vec{X}_{input} \mapsto \vec{X}_1 = f(W_1 \vec{X}_{input} + \vec{b}) \mapsto \vec{X}_{pred} = W_2 \vec{X}_1.$$

Если слои нейронной сети применяются последовательно, то такая нейронная сеть называют сетью прямого распространения. Сеть прямого распространения, составленную из K

полносвязных слоёв, называют многослойным персептроном (multilayer perceptron, далее MLP), [51], [58]. В последнем слое (при $k = K$) MLP всегда выбираем функцию активации тождественной: $f_K(x) = x$. При $K = 2$ получаем двухслойный персептрон (1.18).

Рассмотрим функцию активации $f(x) = \text{ReLU}(x) = \max(0, x)$, тогда двухслойный персептрон (1.18) вычисляет некоторую непрерывную кусочно-линейную функцию. И наоборот: любую непрерывную кусочно-линейную функцию одного переменного с $l-1$ изломами можно представить в виде (1.18). Таким образом, любую непрерывную функцию одного переменного можно сколь угодно точно приблизить двухслойным персептроном (1.18), а увеличение точности этого приближения обеспечивается увеличением числа нейронов l .

Все функции χ_q и w_{pq} из разложения Колмогорова (1.17) можно приблизить двухслойными персептронами. Следовательно, любую непрерывную функцию нескольких переменных можно приблизить в метрике C четырехслойным персептроном с функциями активации ReLU, Id, ReLU, Id, где Id – тождественная функция. Поскольку произведение линейных операторов – линейный оператор, то такой персептрон эквивалентен трехслойному с функциями активации ReLU, ReLU, Id. Таким образом, любую непрерывную функцию нескольких переменных можно приблизить трехслойным персептроном с функциями активации ReLU, ReLU, Id.

Непрерывная монотонная функция $f(x)$ называется сигмоидальной, если выполнены предельные соотношения $\lim_{x \rightarrow -\infty} f(x) = 0$ и $\lim_{x \rightarrow +\infty} f(x) = 1$. Цыбенко [128] доказал, что если $f(x)$ – сигмоидальная функция, то двухслойный персептрон (1.18) с функцией активации $f(x)$ может приблизить любую непрерывную функцию на n -мерном кубе в метрике C .

При количестве слоёв $K > 2$ нейронную сеть называют глубокой. Появившиеся в последние несколько лет новые методы (например, [143], [123], [132] а так же далее 1.1.7-1.1.10) позволили вычислительно эффективно обучать глубокие нейронные сети (оптимизировать их коэффициенты). В данной работе глубокие нейронные сети используются в главе 3 для поиска оптимальных поправок к краткосрочным прогнозам приземных метеорологических параметров.

Используя тождество $4xy = (x+y)^2 - (x-y)^2$, легко показать, что любой многочлен n переменных степени не более 2^m может быть вычислен многослойным персептроном с $2m$ слоями и функциями активации $f_k(x) = x^2$. Действительно, за первые m слоёв можно

вычислить все мономы вида $x_1^{2^{k_1}} \dots x_n^{2^{k_n}}$, где $2^{k_1} + \dots + 2^{k_n} \leq 2^m$, а за следующие $m-1$ слоёв – все мономы степени не более 2^m . Последний слой по этим мономам вычислит искомый многочлен.

Применив разложение (1.17) согласно теореме Колмогорова [53] и теорему Вейерштрасса – Стоуна [48], приходим к выводу, что многослойными персептронами можно приблизить любую непрерывную функцию на n -мерном кубе. Для достижения наилучшей для данной задачи точности этих приближений, возможно, потребуется увеличивать количество слоёв нейронной сети K (то есть придется рассматривать более глубокие нейронные сети).

Используя обобщенную теорему Вейерштрасса – Стоуна [223] можно доказать [254], [40], что и многослойные персептроны с любой кусочно-гладкой непрерывной нелинейной функцией активации плотны в пространстве непрерывных функций на n -мерном кубе.

Суммируя вышесказанное, приходим к заключению, что трехслойного (двухслойного при сигмоидальной функции активации [128]) персептрона теоретически достаточно для приближения любой функции на n -мерном кубе в метрике C . Однако, в некоторых практических задачах, более многослойные нейронные сети по сравнению с такими персептронами могут иметь преимущество [108], [64]. Далее для K -слойного MLP, k -й слой которого имеет N_k нейронов, будем использовать обозначение $N_1 \dots N_{K-1}$.

1.1.4. Технология нейронных сетей

Искусственная нейронная сеть в общем виде представляет собой вычислительный алгоритм, представленный в виде (связного) ориентированного графа без циклов (directed acyclic graph, DAG) [108] и действующий над объектами определенного типа, называемыми тензорами (далее DAG-сеть). Под словом тензор здесь подразумевается (в отличие от классического тензорного анализа) специальный тип многомерных массивов. Вершины этого графа есть непрерывные кусочно-гладкие функции (называемые слоями или блоками) над наборами тензоров, а ребра задают потоки данных (тензоров). Этот ориентированный граф вычислений нейронной сети также называют архитектурой нейронной сети. Слои, из которых выходят ребра, (то есть они передают тензоры другим слоям) называют скрытыми. Слои, из которых не выходят ребра, возвращают набор векторов (тензоров), являющийся результатом (выходом, output) нейронной сети. Слои, в которые не входят ребра, принимают на вход аргументы (вход, input) нейронной сети.

Отметим, что в наиболее общем случае, количество и размерности используемых тензоров могут быть не заданы программистом заранее: один и тот же алгоритм может принимать на вход

переменное количество тензоров переменной размерности, причем меняться может не только размер тензоров, но и количество их индексов (так называемый *broadcasting*, [115]).

В классических нейронных сетях прямого распространения (*feed-forward*) [85] под слоем понимают применение линейного оператора, а затем нелинейной функции, формула (1.19). Если же мы говорим об алгоритме, то один алгоритм можно по-разному разбить на блоки (слои). Это приводит к неоднозначности в определении количества слоёв у данной DAG сети.

Дополнительную путаницу вносит тот факт, что само понятие слоя DAG сети допускает вложенность и часто определённые группы слоёв рассматривают как один, более сложный слой (блок слоёв). Некоторые пакеты, например PyTorch [205], позволяют рассматривать алгоритмы, содержащие ветвления (*if*) и циклы (*while*), что приводит к DAG сетям, граф вычислений (число слоёв) которых меняется (например, в зависимости от входных данных).

Часть тензоров, подаваемых на вход нейронной сети, будем полагать аргументами, а остальные – параметрами. При оптимизации (обучении) нейронной сети будем оптимизировать тензоры параметров. Предполагается, что затем, при использовании обученной нейронной сети (выводе, *inference*) тензоры аргументов будут подаваться на вход в оперативном режиме. С алгоритмической точки зрения тензоры параметров и аргументов нейронной сети являются объектами одного класса. Как правило тензор параметров используется в только одном слое нейронной сети, но иногда рассматривают нейронные сети, различные слои которых имеют общие параметры.

Чтобы обучить нейронную сеть, необходимо задать некоторый штрафной функционал, при вычислении которого используется набор выходов нейронной сети, например, при решении задачи регрессии это штрафной функционал L , определенный по формуле (1.1). Обучение нейронной сети минимизацией (1.1), как правило, проводят методом градиентного спуска (параграф 1.1.7), для реализации которого нужно оценивать градиент оптимизируемого функционала $\nabla_{\vec{\theta}} L$.

В технологии нейронных сетей градиент функционала L как правило вычисляют методом обратного распространения ошибки [34], [216]. Метод обратного распространения ошибки для каждого из слоёв F_k нейронной сети, последовательно, двигаясь по графу вычислений в противоположную к ориентации графа сторону, вычисляет градиент функционала L по всем аргументам $\vec{x}_{k-1}$ и параметрам $\vec{\theta}_k$ данного слоя. Если слой F_k осуществляет отображение вида:

$$\vec{x}_{k-1} \mapsto \vec{x}_k = F_k(\vec{x}_{k-1}, \vec{\theta}_k), \quad (1.20)$$

то, используя формулу дифференцирования сложной функции, легко вычислить:

$$\begin{aligned}\nabla_{\bar{x}_{k-1}} L &= \nabla_{\bar{x}_{k-1}} F_k \otimes \nabla_{\bar{x}_k} L, \\ \nabla_{\bar{\theta}_k} L &= \nabla_{\bar{\theta}_k} F_k \otimes \nabla_{\bar{x}_k} L.\end{aligned}\tag{1.21}$$

Упомянем специальные свободно распространяемые пакеты для работы с тензорами, такие как PyTorch [205], TensorFlow [228], содержащие библиотеки слоёв нейронных сетей. При использовании этих библиотек нет необходимости в программировании алгоритма для вычисления градиентов (1.21): они вычисляются автоматически (так называемая технология автоматического дифференцирования, autograd). При необходимости (например, для избегания арифметических операций с большими числами или экономии памяти) функцию вычисления градиентов можно запрограммировать самостоятельно. Рассмотренные в данной работе алгоритмы были запрограммированы диссертантом с использованием пакета PyTorch [205].

Хотя увеличение глубины нейронных сетей часто позволяет получать более точные приближения при меньшем суммарном количестве коэффициентов, но и обучать более глубокие сети часто оказывается сложнее: возникает так называемая «проблема малых градиентов». Проблема малых градиентов – общая для задач по минимизации функций многих переменных, когда градиенты по различным оптимизируемым параметрам отличаются друг от друга по порядку величин:

$$\exists i, j: \left| \frac{\partial L}{\partial \theta_j} \right| \ll \left| \frac{\partial L}{\partial \theta_i} \right|.\tag{1.22}$$

При выполнении соотношения (1.22) часто наблюдается низкая скорость сходимости методов оптимизации, таких как метод градиентного спуска [37], [63].

Согласно параграфу 1.1.3, чтобы уменьшить минимальное возможное для рассматриваемой нейронной сети значение функционала L можно либо увеличивать глубину нейронной сети, либо увеличивать количество нейронов на скрытых слоях сети. При увеличении количества нейронов в полносвязных слоях (1.19), количество параметров сети растет квадратично, что сильно увеличивает вычислительную сложность.

При выборе оптимальной для рассматриваемой задачи конфигурации нейронной сети желательно найти оптимальное количество слоёв (а также типы этих слоёв) и количество нейронов на каждом слое.

1.1.5. Вычисление градиентов многослойного персептрона

Рассмотрим K -слойный персептрон. Его k -й слой полносвязный (формула (1.19)) с N_k нейронами, то есть он может быть записан как отображение $\mathbb{R}^{N_{k-1}} \rightarrow \mathbb{R}^{N_k}$ вида (1.20):

$$\vec{x}_{k-1} \mapsto \vec{x}_k = F_k(\vec{x}_{k-1}, \vec{\theta}_k) = f_k(W_k \vec{x}_{k-1} + \vec{b}_k), \quad (1.23)$$

где вектор параметров слоя $\vec{\theta}_k = \{W_k, \vec{b}_k\}$ состоит из компонент оператора W_k и вектора $\vec{b}_k$.

Для оптимизации параметров $\vec{\theta}_k$ методом градиентного спуска необходимо уметь вычислять градиент $\nabla_{\vec{\theta}_k} L$ оптимизируемого функционала L , вычисляемого по формуле (1.1). В случае полносвязного слоя (1.23) необходимо вычислить градиенты $\nabla_{W_k} L$ и $\nabla_{\vec{b}_k} L$. Их легко найти из формул (1.1), (1.23), используя правило дифференцирования сложной функции (то есть, подставляя (1.23) в (1.21)):

$$\begin{aligned} \nabla_{W_k} L &= \frac{1}{M} \sum_{j=1}^M \left(f'_k(W_k \vec{x}_{k-1,j} + \vec{b}_k) \circ \nabla_{\vec{x}_{k,j}} e \right) \vec{x}_{k-1,j}^T, \\ \nabla_{\vec{b}_k} L &= \frac{1}{M} \sum_{j=1}^M f'_k(W_k \vec{x}_{k-1,j} + \vec{b}_k) \circ \nabla_{\vec{x}_{k,j}} e, \end{aligned} \quad (1.24)$$

где $\nabla_{W_k} L$ – матрица, составленная из частных производных штрафного функционала L по соответствующему элементу матрицы W_k , а $\circ$ – покомпонентное умножение векторов.

При вычислении градиентов $\nabla_{W_k} L$ и $\nabla_{\vec{b}_k} L$ по формулам (1.24) используется градиент $\nabla_{\vec{x}_k} e$. Для последнего слоя $k = K$ выполнено $\vec{x}_K = \vec{X}_{pred}$, а значит, градиент $\nabla_{\vec{x}_K} e$ вычисляется дифференцированием функции потерь $e(X_{fact}, X_{pred})$. Например, если используется функция потерь Хьюбера (1.4), то градиент $\nabla_{\vec{X}_{pred}} e$ может быть вычислен по формуле (1.8).

Для скрытых слоёв (то есть при $k < K$) вычислить $\nabla_{\vec{x}_k} e$ позволяет метод обратного распространения ошибки [34], формула (1.21). Например, для полносвязного слоя (1.23) метод обратного распространения ошибки примет вид:

$$\nabla_{\vec{x}_{k-1}} e = W_k^T \left(f'_k(W_k \vec{x}_{k-1} + \vec{b}_k) \circ \nabla_{\vec{x}_k} e \right). \quad (1.25)$$

Для вычисления каждого значения многослойного персептрона $F(\vec{x}, \vec{\theta})$ с n_θ параметрами по формуле (1.23) требуется: n_θ умножений, n_θ сложений, и $N_f = \sum_{k=1}^K N_k$ вычислений функции активации f . Для однократного вычисления градиентов по формулам (1.24), (1.25) потребуется $n_\theta + N_f$ умножений, $2n_\theta$ сложений и N_f вычислений производной функции активации f' . При

использовании функции активации ReLU временем, потраченным на вычисление f_k и f'_k можно пренебречь.

Таким образом, процесс оптимизации параметров нейронной сети относительно дешев: вычисление компьютером градиентов $\nabla_{\vec{\theta}} L$ лишь на 20-30% дороже вычисления нейронной сети F , а применение градиентного спуска при известных градиентах как правило заметно дешевле, поскольку производится сразу для целого набора реализаций входных тензоров $\vec{X}_{input}$, а не для каждой реализации отдельно (параграф 1.1.7).

Поскольку для быстрого вычисления градиента $\nabla_{W_k} L$ по первой из формул (1.24) требуется хранить значения векторов $\vec{x}_{k,j}$, то при обучении нейронной сети требуется существенно больше памяти (и времени на обращение в память), чем при её применении.

Оказывается, что при использовании монотонных функций активации f_k можно сократить количество используемой при обучении памяти компьютера. Так как монотонные функции обратимы, то производную f'_k можно выразить через значение функции f_k как некоторую функцию g_k :

$$f'_k(W_k \vec{x}_{k-1} + \vec{b}_k) = g_k(\vec{x}_k) = g_k(f_k(W_k \vec{x}_{k-1} + \vec{b}_k)). \quad (1.26)$$

Если существует функция g_k (а такая функция всегда существует при монотонных f_k), удовлетворяющая (1.26), то в процессе обучения отсутствует потребность в хранении помимо значений $\vec{x}_k$ так же и значений $W_k \vec{x}_{k-1} + \vec{b}_k$. Отметим, что согласно приведенной выше теореме Колмогорова (1.17), «внутренние» функции w_{pq} могут быть выбраны монотонными.

Примерами функций активации, для которые выражение для g_k принимает особенно простой вид служат: логистическая функция σ , гиперболический тангенс th и ReLU. Все они суть решения обыкновенных автономных дифференциальных уравнений первого порядка, то есть их производные можно выразить через их значение:

$$\begin{aligned} \sigma'(x) &= \sigma(x) - \sigma(x)^2, \\ th'(x) &= 1 - th(x)^2, \\ \text{ReLU}'(x) &= I_{x>0}(\text{ReLU}(x)). \end{aligned} \quad (1.27)$$

Обсудим некоторые практические особенности использования функций активации (1.27). Логистическая функция σ удовлетворяет условиям теоремы Цыбенко [128], параграф 1.1.3. Таким образом, при использовании функций активации σ (а также th), можно сократить (по сравнению с ReLU) количество слоёв нейронной сети. Однако, производные функций σ и th

очень быстро убывают с ростом аргумента, что на практике часто приводит к проблеме малых градиентов (1.22) и сильно замедляет сходимость градиентного спуска в пространстве параметров $\vec{\theta}$ нейронной сети. Таким образом, нельзя дать однозначных рекомендаций о использовании одной из функций активации (1.27) и для каждой практической задачи необходимо выбирать наилучшую (с точки зрения пользователя продукции модели МО) функцию активации.

1.1.6. Слой нейронной сети типа *embedding*

В некоторых задачах помимо вещественнозначных предикторов требуется учитывать типизации (например, по регионам, по типам подстилающей поверхности). Такой учёт можно осуществить, рассмотрев набор предикторов, принимающих значения 0 и 1 в зависимости от принадлежности данного объекта одному из классов.

Действительно, пусть i -й предиктор принимает значения из некоторого множества $A_i = \{a_{ik}, k = 1 \dots M_i\}$, мощностью M_i , то есть $x_i \in A_i$. Рассмотрим набор предикторов $\chi_{ki}, k = 1 \dots M_i$, k -й из которых принимает значение 1, когда $x_i = a_{ik}$, то есть $\vec{\chi}_i = I_{a_{ik}}(x_i)$. Embedding слой нейронной сети осуществляет линейное отображение $A_i \rightarrow \mathbb{R}^{E_i}$:

$$x_i \rightarrow B\vec{\chi}_i, \quad (1.28)$$

где B – некоторый линейный оператор, E_i – размерность embedding. Таким образом, embedding слой отображает предиктор x_i в k -й столбец матрицы B .

Поскольку embedding слой принимает на вход номер класса $x_i = 1 \dots M_i$, градиент ∇_{x_i} не имеет смысла, а значит embedding слой может быть только первым слоем нейронной сети. Теоретически embedding слой можно рассматривать как специальный вид полносвязного слоя (1.23) с E_i нейронами, но специальная программная реализация embedding слоя будет существенно быстрее, поскольку нет необходимости умножать матрицы согласно формуле (1.28), а достаточно лишь выбрать соответствующий столбец матрицы B .

В третьей части главы 3 embedding слой использован для учёта индивидуальных параметров синоптических станций, аналогично работе [209].

1.1.7. Обучение нейронных сетей методами градиентного спуска

При единовременном вычислении градиента штрафной функции $\nabla_{\vec{\theta}} L(\vec{\theta})$ по всему имеющемуся архиву данных часто а) вычислительно слишком дорого; б) невозможно сохранить в памяти все величины, используемые при вычислении градиентов (параграф 1.1.5). Поэтому, разобьем весь архив для обучения размера M на случайные малые части размера m (так называемые minibatch, [123]). Оптимизацию нейронной сети будем осуществлять пошагово, на каждом шаге оценивая градиент $\nabla_{\vec{\theta}} L(\vec{\theta})$ на случайном minibatch'е. Алгоритм оптимизации будет пробегать по всему архиву из M реализаций за M/m шагов. Один такой проход называют эпохой обучения.

Отметить так называемый подход активного обучения [230], когда отбор реализаций из большого архива размера M осуществляется неслучайным образом. Например, можно более часто выбирать реализации, на которых ошибка обучаемой модели больше, чем на других.

Метод стохастического градиентного спуска (stochastic gradient descend, SGD) на s -ом шаге использует оценку градиента штрафной функции $\nabla_{\vec{\theta}}^m L(\vec{\theta})$ по некоторому minibatch'у обучающей выборки и меняет параметры согласно итеративной формуле:

$$\vec{\theta}_{s+1} = \vec{\theta}_s - \eta_s \nabla_{\vec{\theta}}^m L(\vec{\theta}_s), \quad (1.29)$$

где $\eta_s > 0$ – положительная константа, называемая скоростью градиентного спуска на s -ом шаге. При оптимизации модели МО параметр η_s называют скоростью обучения. Чтобы обеспечить сходимость градиентного спуска требуется, чтобы, во-первых, градиент $\nabla_{\vec{\theta}} L(\vec{\theta})$ был ограничен, а во-вторых, чтобы последовательность скоростей обучения η_s удовлетворяла предельному соотношению $\lim_{s \rightarrow +\infty} \eta_s = 0$. Ограниченность градиента $\nabla_{\vec{\theta}} F(\vec{X}_{input}, \vec{\theta})$ зависит от используемого семейства моделей МО F и для многослойного персептрона ограниченность следует из формул (1.24) и (1.25).

Очевидно, что для поиска минимума штрафного функционала $L(\vec{\theta})$ может потребоваться (в зависимости от решаемой задачи) пройти сколь угодно длинный путь в пространстве параметров $\vec{\theta}$, а значит, сумма скоростей обучения должна расти неограниченно с увеличением числа шагов алгоритма градиентного спуска: $\lim_{S \rightarrow +\infty} \sum_{s=1}^S \eta_s = +\infty$. Указанным предельным соотношениям удовлетворяют, например, последовательности скоростей обучения вида:

$$\eta_s = \frac{\eta_0}{1 + C(t_E - 1)}, \quad (1.30)$$

где константа C отвечает за скорость уменьшения шага обучения, t_E – номер эпохи. На практике не всегда применяют строго убывающие последовательности шагов обучения, например, в [220] предлагается использовать циклически меняющийся шаг обучения.

При использовании алгоритма стохастического градиентного спуска (1.29) по сравнению с классическим методом (при $m=M$) градиентного спуска получается использовать для обучения гораздо большие архивы данных. Может показаться, что выбор размера minibatch'a m ограничен сверху объемом доступной памяти, но при практической реализации можно накапливать сумму градиентов $\nabla_{\vec{\theta}} L(\vec{\theta})$ за несколько загрузок данных в память. Размер minibatch может существенно влиять на результат оптимизации, причем различные исследователи говорят как о преимуществе больших minibatch [146], так и преимуществе малых minibatch [98]. Выбор малых m может существенно увеличить время, затраченное программой на так называемые накладные расходы: пересылку данных, инициализацию объектов и так далее. В работе [221] предлагается в процессе оптимизации не уменьшать шаг градиентного спуска, а увеличивать размер minibatch'a. При обучении рассмотренных в диссертации нейронных сетей по результатам численных экспериментов выбран размер minibatch'a, вычисляемый по формуле $m = M / 400$.

Техника так называемого косинусного отжига (cosine annealing, [160]), предлагает периодически сначала постепенно уменьшать шаг обучения, а затем резко увеличивать шаг обучения, сохранив перед этим увеличением получившиеся параметры $\vec{\theta}$. Используя косинусный отжиг, можно, запустив единожды процесс обучения, получить несколько локальных минимумов функционала $L(\vec{\theta})$, не повторяя начальные шаги оптимизации. Используя несколько локально оптимальных наборов коэффициентов, можно получить некоторый ансамбль прогнозов $\vec{X}_{pred,k}$ для проведения дальнейшего анализа распределения этих прогнозов. В [160] показано, что усреднение нескольких найденных таким образом нейронных сетей может улучшить качество прогноза.

Рекуррентное соотношение метода Нестерова (известного как Nesterov Accelerated Gradient, [62], [224]) может быть записано в виде:

$$\begin{aligned} m_0 &= 0, \\ \vec{m}_s &= \alpha_s \vec{m}_s - \eta_s \nabla_{\vec{\theta}} L(\vec{\theta}_s + \alpha_s \vec{m}_s), \\ \vec{\theta}_{s+1} &= \vec{\theta}_s + m_s, \end{aligned} \quad (1.31)$$

где $\alpha_s \approx 1 - 3/(s+5)$. Можно показать [62], что для гладких выпуклых функций L при определенном выборе скорости обучения η_s метод Нестерова (1.31) сходится квадратично:

$$L(\vec{\theta}_s) - L(\vec{\theta}_*) \leq \frac{4C \|\vec{\theta}_0 - \vec{\theta}_*\|^2}{(s+2)^2},$$

где $\vec{\theta}_*$ — точка (единственного) минимума функции L с ограниченным градиентом $|\nabla_{\vec{\theta}} L| \leq C$.

Метод Ньютона — Рафсона, примененный к поиску решения уравнения $\nabla_{\vec{\theta}} L = \vec{0}$, использует наряду с градиентом $\nabla_{\vec{\theta}} L$ так же и гессиан $H_{\vec{\theta}}(L)$:

$$\vec{\theta}_{s+1} = \vec{\theta}_s - \eta_s \nabla_{\vec{\theta}}^m L(\vec{\theta}_s) \left[H_{\vec{\theta}}(L(\vec{\theta}_s)) \right]^{-1}. \quad (1.32)$$

В предыдущем пункте было показано, что вычисление градиента $\nabla_{\vec{\theta}} L$ для многослойного персептрона можно осуществлять достаточно дешево, а значит для обучения нейронной сети оправдано применение методов градиентного спуска первого порядка. Вместе с тем, поскольку в матрице Гессе $H_{\vec{\theta}}(L)$ число различных элементов равно $n_{\theta}(n_{\theta}+1)/2$, вычисление матрицы Гессе, а значит и применение методов оптимизации второго порядка в $O(n_{\theta})$ раз дороже, чем вычисление градиента $\nabla_{\vec{\theta}} L(\vec{\theta})$. Чтобы избежать слишком дорогого прямого вычисления матрицы Гессе $H_{\vec{\theta}}(L)$, в (1.32) можно использовать различные оценки для $H_{\vec{\theta}}(L)$. Например, в [247] для полносвязного слоя (1.23) предполагается, что смешанные вторые производные по параметрам различных слоёв равны нулю, а вторые производные по параметрам одного слоя оцениваются через ковариационную матрицу признаков этого слоя:

$$\hat{H}_{W_k}(L) = \text{Cov}(\vec{x}_{k-1}) + \lambda E = \frac{1}{M} \vec{x}_k^T \vec{x}_k + \lambda E \quad (1.33)$$

где $\lambda = 1 + 10^{-4}$ — параметр регуляризации; E — единичная матрица размера N_k . Тогда рекуррентное соотношение метода (1.32) можно переписать в виде:

$$W_{k,s+1} = W_{k,s} - \eta_s \nabla_{W_k}^m L(\vec{\theta}_s) \left[\hat{H}_{W_k}(L) \right]^{-1}, \quad (1.34)$$

Обратная матрица $\hat{H}_{W_k}(L)$ участвует в рекуррентном соотношении (1.34), таким образом, сходимость метода (1.34) может быть достигнута только при хорошей обусловленности матрицы $\hat{H}_{W_k}(L)$. Значение параметра $\lambda > 1$ в (1.33) обеспечивает число обусловленности матрицы $\hat{H}_{W_k}(L)$ не хуже $O(\lambda/(\lambda-1))$.

По результатам проведенных диссертантом численных экспериментов, замена $\nabla_{W_k}^m L(\bar{\theta}_s)$ на $\nabla_{W_k}^m L(\bar{\theta}_s) [\hat{H}_{W_k}(L)]^{-1}$ ускоряет сходимость не только стохастического градиентного спуска (1.29), но и приведенных ниже адаптивных методов градиентного спуска.

Так называемые адаптивные методы градиентного спуска используют оценку $H_{\bar{\theta}}(L)$ на основе динамики изменения градиента $\nabla_{\bar{\theta}}^m L(\bar{\theta}_s)$ в зависимости от шага градиентного спуска s . Алгоритм Adagrad (adaptive gradient, [133]) использует следующие рекуррентные соотношения:

$$\begin{aligned} E_0 &= 0, \\ E_{i,s} &= E_{i,s-1} + \left(\partial_{\theta_i}^m L(\bar{\theta}_s) \right)^2, \\ \theta_{i,s+1} &= \theta_{i,s} - \eta_s \frac{\partial_{\theta_i}^m L(\bar{\theta}_s)}{\sqrt{E_{i,s} + \varepsilon}}, \end{aligned} \quad (1.35)$$

Поскольку $E_{i,s} \geq E_{i,s-1}$, всегда выполнено хотя бы одно из предельных соотношений $\lim_{s \rightarrow +\infty} E_{i,s} = +\infty$ или $\lim_{s \rightarrow +\infty} \partial_{\theta_i}^m L(\bar{\theta}_s) = 0$. Следовательно, алгоритм (1.35) сходится и при $\lim_{s \rightarrow +\infty} \eta_s = \eta_\infty > 0$. Поэтому, как правило, в алгоритме Adagrad полагают $\eta_s \equiv \text{const}$.

Алгоритм Adam (adaptive moments, [171], [132]) использует экспоненциальное сглаживание градиента:

$$\begin{aligned} m_0 &= E_0 = 0, \\ m_{i,s} &= \alpha m_{i,s-1} + (1 - \alpha) \partial_{\theta_i}^m L(\bar{\theta}_s), \\ E_{i,s} &= \beta E_{i,s-1} + (1 - \beta) \left(\partial_{\theta_i}^m L(\bar{\theta}_s) \right)^2, \\ \theta_{i,s+1} &= \theta_{i,s} - \eta_s \sqrt{\frac{1 - \beta^s}{E_{i,s} + \varepsilon}} \frac{m_{i,s}}{1 - \alpha^s}, \end{aligned} \quad (1.36)$$

где обычно используют «стандартные» параметры: $\alpha = 0.9, \beta = 0.999, \varepsilon = 10^{-8}$. Частным случаем (при $\alpha = 0$) алгоритма Adam (1.36) является (появившийся раньше Adam) алгоритм RMSProp [147], [229], [224].

Алгоритмы Adam и RMSProp могут на первых шагах делать слишком большие шаги. Чтобы стабилизировать алгоритмы Adam и RMSProp используют так называемый прогрев (warmup). При прогреве на первых шагах алгоритма градиентного спуска уменьшают скорость обучения η_s . Это позволяет более точно оценить величины импульса m_i и энергии E_i согласно (1.36) и стабилизировать процесс обучения.

Классический линейный прогрев в первые s_w шагов уменьшает скорость обучения в соответствии с эмпирической формулой

$$\tilde{\eta}_s = \eta_s \min(1, s / s_w). \quad (1.37)$$

В работе [186] предложен алгоритм RAdam (Rectified Adam), реализующий «оптимальный» прогрев. В диссертации всегда использовался линейный прогрев (1.37) длительностью $s_w = 50$ minibatch.

Алгоритм AdaBelief [252] является модификацией метода Adam и часто демонстрирует более быструю сходимость, чем метод Adam и остальные рассмотренные в данном параграфе методы. Метод AdaBelief использует другую оценку для $E_{i,s}$:

$$E_{i,s} = \beta E_{i,s-1} + (1 - \beta) \left(\partial_{\theta_i}^m L(\vec{\theta}_s) - \frac{m_{i,s}}{1 - \alpha^s} \right)^2. \quad (1.38)$$

Для адаптивных алгоритмов градиентного спуска Adagrad (1.35), Adam (1.36) и AdaBelief (1.38) сходимость достигается и при $\lim_{s \rightarrow +\infty} \eta_s > 0$. Однако, по результатам численных экспериментов, без уменьшения шага обучения получаемые оценки оказываются несколько хуже, чем при использовании убывающих η_s , например, согласно стратегии обучения (1.30).

1.1.8. Регуляризация

Если функционал $L(\vec{\theta})$ не является выпуклым вверх и имеет много локальных минимумов, а получавшиеся нейронные сети $F(\vec{x}, \vec{\theta})$ будут неустойчивыми, например, при малой вариации начальных значений (инициализации, параграф 1.1.10) параметров $\vec{\theta}_0$ и (или) используемого архива данных.

Для получения более устойчивых моделей $F(\vec{x}, \vec{\theta})$ к минимизируемому функционалу (1.1) добавляют дополнительное слагаемое – так называемый регуляризатор [79]. Примером регуляризованного функционала служит L_2 :

$$L_2(\vec{\theta}) = L(\vec{\theta}) + \frac{1}{2} l_2 \|\vec{\theta}\|_{L_2}^2 \rightarrow \min_{\vec{\theta}}, \quad (1.39)$$

где l_2 – малое положительное число – параметр регуляризации. Очевидно, что градиенты функционалов (1.1) и (1.39) связаны соотношением $\nabla_{\vec{\theta}} L_2 = \nabla_{\vec{\theta}} L + l_2 \vec{\theta}$.

Предложенный в [187] метод AdamW, оптимизирует функционал (1.39), но, в отличие от метода Adam (1.36), AdamW не применяет экспоненциальное сглаживание с параметрами α и β к градиенту регуляризатора $l_2\vec{\theta}$, а добавляет его непосредственно в выражение для $\vec{\theta}_{s+1}$. Таким образом, рекуррентные соотношения метода AdamW для оптимизации (1.39) совпадают с (1.36) с использованием $\nabla_{\theta_i} L(\vec{\theta}_t)$, но вместо последней из формул (1.36) используется рекуррентное соотношение:

$$\theta_{i,s+1} = \theta_{i,s} - \eta_s \sqrt{\frac{1-\beta^s}{E_{i,s} + \varepsilon}} \frac{m_{i,s}}{1-\alpha^s} - \eta_s l_2 \vec{\theta}. \quad (1.40)$$

Метод AdamW во многих задачах позволяет достичь меньших средних значений функции потерь по сравнению с методом Adam. Везде далее, за исключением параграфа П4.2.1 (в котором сравниваются различные алгоритмы градиентного спуска), использовался алгоритм AdamW, формула (1.40).

Помимо классического L_2 -регуляризатора (1.39) могут применяться и более сложные регуляризаторы. Например, в [152] предлагается использовать регуляризатор (так называемый градиентный штраф, gradient penalty), штрафующий за отличие нормы градиента нейронной сети от единицы:

$$R_G(\vec{\theta}) = \frac{1}{M} \sum_{j=1}^M \left(\left\| \nabla_{\vec{X}_{input,j}} F(\vec{X}_{input,j}, \vec{\theta}) \right\| - 1 \right)^2. \quad (1.41)$$

При решении задачи коррекции прогноза диссертантом проводились эксперименты по использованию следующего (близкого к (1.41)) регуляризатора, штрафующего за слишком большую поправку:

$$R_{\Delta}(\vec{\theta}) = \frac{1}{M} \sum_{j=1}^M \left(F(\vec{X}_{input,j}, \vec{\theta}) - \vec{X}_{0,j} \right)^2,$$

где $\vec{X}_{0,j}$ – исходный (корректируемый) прогноз. Регуляризатор R_{Δ} оказался неэффективным: при его использовании достигнутые значения функционала $L(\vec{\theta})$ увеличились по сравнению с использованием регуляризатора L_2 .

Альтернативой (дополнением к) регуляризации можно считать переход от безусловной минимизации функционала $L(\vec{\theta})$ к его условной минимизации при выполнении некоторых условий на параметры. Например, в работе [248] предложена методика центрирования градиентов (gradient centralization, далее GC), которая для полносвязного слоя (параграф 1.1.3, формула (1.19)) предлагает «центрировать» градиенты:

$$\nabla_{W_{sp}} L \rightarrow \nabla_{W_{sp}} L - \frac{1}{N_{k-1}} \sum_{p=1}^{N_{k-1}} \nabla_{W_{sp}} L, \quad (1.42)$$

где $s=1\dots N_k$. Центрирование градиентов (1.42) эквивалентно условной минимизации при условии выполнения N_k линейных ограничений вида

$$\sum_{p=1}^{N_{k-1}} W_{sp} = \text{const.}$$

В параграфе П4.2.1 рассмотрен вопрос эффективности применения ГС в задаче коррекции прогнозов приземных метеорологических параметров.

1.1.9. Нормировка предикторов

Нормировка предикторов (признаков) часто оказывается полезна при применении методов машинного обучения. Случай, когда признаки полносвязного слоя (1.19) близки по порядку величины предпочтителен, поскольку в этом случае в силу формул (1.24), (1.25) компоненты градиента $\nabla_{\bar{\theta}} L(\bar{\theta})$ так же будут близки по порядку величины, а значит и метод градиентного спуска так же будет сходиться быстрее.

Нормирующие предикторы (признаки) слоя $x \rightarrow \bar{x} = N(\bar{x}, \{\bar{\mu}, \bar{\sigma}\})$ осуществляют отображение $\mathbb{R}^n \rightarrow \mathbb{R}^n$ вида:

$$x_i \rightarrow \bar{x}_i = \frac{x_i - \mu_i}{\sqrt{\sigma_i + \varepsilon}}, \quad (1.43)$$

где μ_i и σ_i – параметры слоя, $i=1, \dots, n$, а малый параметр ε обычно полагается равным 10^{-5} . В качестве параметров μ_i и σ_i на s -м шаге обучения нейронной сети использует оценки среднего и дисперсии признака x_i , полученные экспоненциальным сглаживанием:

$$\begin{aligned} m_{i,0} &= 0, E_{i,0} = 1, \\ m_{i,s} &= (1 - \beta) m_{i,s-1} + \frac{\beta}{m} \sum_{j=1}^m x_{ij}, \\ \mu_{i,s} &= \frac{m_{i,s}}{1 - \beta^s}, \\ E_{i,s} &= (1 - \beta) E_{i,s-1} + \frac{\beta}{m} \sum_{j=1}^m x_{ij}^2, \\ \sigma_{i,s} &= \frac{E_{i,s}}{1 - \beta^s} - \mu_{i,s}^2, \end{aligned} \quad (1.44)$$

где β – параметр экспоненциального сглаживания, m – размер minibatch. Как правило, параметр β выбирают близким к обратному количеству шагов градиентного спуска за одну эпоху обучения. В данной работе использовался параметр $\beta = 0.01$. Для подстройки параметров $\bar{\mu}$ и $\bar{\sigma}$, согласно (1.44), не используются градиенты ошибок. Таким образом в нормирующем слое (1.43) есть параметры, но они не являются обучаемыми. Во время обучения нейронной сети признаки, являющиеся выходами скрытых слоёв с оптимизируемыми параметрами, меняют своё распределение. Меняющиеся одновременно с этим параметры μ_i и σ_i нормирующего слоя (1.43) компенсируют это изменение.

Вариант нормирующего слоя, называемый BatchNorm, вычисляет линейную функцию от нормированного предиктора (признака) $\bar{x}$, а именно:

$$x_i \rightarrow \gamma_i \bar{x}_i + \beta_i = x_i \frac{\gamma_i}{\sqrt{\sigma_i + \varepsilon}} + \left[\beta_i - \frac{\mu_i}{\sqrt{\sigma_i + \varepsilon}} \right], \quad (1.45)$$

где γ_i, β_i – обучаемые параметры (всего $2n$ параметров), на первом шаге обучения, полагаемые равными $\gamma_{i,0} = 1$, $\beta_{i,0} = 0$.

В [163] было показано, что хотя добавление BatchNorm слоёв (1.45) никак не влияет на рассматриваемый класс функций $\nabla_{\bar{\theta}} F(\bar{X}_{input}, \bar{\theta})$ и BatchNorm слои не помогают достичь меньших значений штрафного функционала $L(\bar{\theta})$, однако нейронные сети с BatchNorm слоями обучаются быстрее.

Помимо BatchNorm слоёв существуют и другие варианты нормирующих слоёв. Слои типа LayerNorm [104] и IterNorm [161] пытаются одновременно уменьшить кросс-корреляции признаков x_i . Классический метод перехода к главным ортогональным компонентам тут не применим: в процессе обучения нейронной сети признаки могут становиться линейно зависимыми, а неограниченность нормы (и нормы градиента) оператора перехода к ортогональным компонентам может привести к потере сходимости метода градиентного спуска.

Во всех рассмотренных в данной работе нейронных сетях первый слой был нормирующим с параметрами, обновляемыми согласно алгоритму (1.44).

1.1.10. Инициализация параметров нейронной сети

Как и в любой другой задаче оптимизации, при обучении нейронной сети весьма важен выбор начального приближения $\bar{\theta}_0$, то есть инициализация параметров нейронной сети. Как

правило, параметры $\vec{\theta}$ инициализируются из случайного распределения с нулевым средним. При этом оказывается, что на скорость обучения нейронной сети сильно влияет дисперсия распределения, из которого были выбраны начальные значения параметров сети. В [143] для полносвязного слоя (1.19) с n нейронами и l аргументами предлагается использовать так называемую инициализацию Ксавье, когда параметры выбираются из равномерного распределения на отрезке:

$$W_{ij} \in \left[-g_0 \sqrt{\frac{6}{n+l}}, g_0 \sqrt{\frac{6}{n+l}} \right], \quad (1.46)$$

В работе [143] параметр g_0 был равен 1, и было показано, что в этом случае вариации векторов $Var(\vec{x})$ и $Var(W\vec{x})$ близки по величине. В диссертации использовалась инициализация (1.46) для всех слоёв, вопрос выбора оптимального значения g_0 рассмотрен в параграфе 3.2.2.

В [154] предлагается несколько улучшенная по сравнению с (1.46) инициализация в случае, если предшествующий слой имеет функцию активации

$$\text{LeakyReLU}(x) = \max(0, x) + a \min(0, x),$$

где $a \geq 0$ – константа. Тогда параметр g_0 в формуле (1.46) предлагается выбрать равным:

$$g_0 = \sqrt{1 + (1 - a)^2} \quad (1.47)$$

Согласно работам [143] и [154] если нейронная сеть инициализирована согласно формулам (1.46) и (1.47), то скрытые признаки (выходы скрытых слоёв) будут иметь близкий порядок величин даже без использования промежуточных нормирующих слоёв (1.43), а значит, нейронная сеть будет обучаться быстрее, чем при использовании инициализации значениями, выбранными из вероятностного распределения, не зависящего от n и l .

1.1.11. Оптимизация гиперпараметров

Хотя в формуле для вычисления прогноза нейронной сети $\vec{X}_{pred} = F(\vec{X}_{input}, \vec{\theta})$ напрямую участвуют только входные аргументы $\vec{X}_{input}$ и параметры $\vec{\theta}$, но в алгоритме машинного обучения модели МО F , который оптимизирует параметры $\vec{\theta}$, могут участвовать и другие величины $\vec{\mu}$, называемые гиперпараметрами, которые могут повлиять на найденный оптимальный набор параметров $\hat{\theta}_{\vec{\mu}}$, а значит и оценки качества получившегося прогноза.

Для K -слойного персептрона гиперпараметрами являются в том числе:

1. количества нейронов N_k на скрытых слоях сети (при $k = 1, \dots, K - 1$);
2. функции активации на скрытых слоях сети f_k ;
3. параметры регуляризации $R(\vec{\mu}, \vec{\theta}_{\vec{\mu}})$ (параграф 1.1.8), например, параметр l_2 в (1.39);
4. параметры, используемые в алгоритме градиентного спуска (параграф 1.1.7), в том числе стратегия выбора скорости обучения η_s на s -м шаге и критерий останова метода градиентного спуска;
5. используемая инициализация (параграф 1.1.10).

Гиперпараметры так же могут описывать используемую архитектуру нейронной сети и дополнительные параметры слоёв, например, параметр β в BatchNorm слое, формула (1.44).

Чтобы провести оптимизацию гиперпараметров разделяют имеющийся архив данных на два, как можно более независимых и репрезентативных, подархива: архив для обучения A_{train} и архив для валидации A_{valid} . При прогнозе непрерывных по пространству-времени величин, независимость можно гарантировать, только если подархивы разделены в пространстве-времени.

В метеорологических задачах разделение по пространству невозможно, так как различные точки Земли имеют различный климат (микроклимат) и полученные на архиве для обучения оценки для одного региона могут быть неадекватны для другого региона в архиве для валидации. Валидация по отдельным метеостанциям в рассматриваемом регионе невозможна, поскольку погрешности и поправки в точках близких метеостанций не независимы (соответствующие пространственные корреляционные функции рассматриваются в главе 4).

Таким образом, разделение архива возможно только по времени. При разделении по времени желательно учесть сезонность, то есть использовать архивы длительностью, кратной году.

Архив для обучения будем использовать для поиска параметров $\hat{\theta}_{\vec{\mu}}$, оптимальных в смысле штрафного функционала (1.1) для данного набора гиперпараметров $\vec{\mu}$:

$$\arg \min_{\vec{\theta}} \frac{1}{|A_{train}|} \sum_{j \in A_{train}} e\left(X_{fact,j}, F\left(\vec{X}_{input,j}, \vec{\theta}_{\vec{\mu}}\right)\right) + R\left(\vec{\mu}, \vec{\theta}_{\vec{\mu}}\right) = \hat{\theta}_{\vec{\mu}} \quad (1.48)$$

При оптимизации параметров $\hat{\theta}_{\vec{\mu}}$ в силу (1.48) используются методы градиентного спуска (параграф 1.1.8).

Архив для валидации A_{valid} будем использовать для оптимизации гиперпараметров:

$$\arg \min_{\vec{\mu}} \frac{1}{|A_{valid}|} \sum_{j \in A_{valid}} e\left(X_{fact,j}, F\left(\vec{X}_{input,j}, \hat{\theta}_{\vec{\mu}}\right)\right) = \hat{\mu} \quad (1.49)$$

При оптимизации гиперпараметров согласно (1.49) использовался метод поиска по сетке в пространстве гиперпараметров. Существуют и более продвинутые методы оптимизации гиперпараметров, например, метод *neural architecture search* для поиска архитектуры нейронной сети [107], [130], [203].

Аналогично переподгонке параметров под данные для обучения, возможна переподгонка гиперпараметров под данные для валидации. Для осуществления полностью независимой проверки получившейся модели машинного обучения $F(\vec{X}_{input,j}, \hat{\theta}_{\hat{\mu}})$ нужен независимый и репрезентативный тестовый подархив A_{test} . Перед проверкой модели машинного обучения на тестовом подархиве проводилась дополнительная оптимизация параметров на расширенном подархиве, являющемся объединением обучающего и валидационного подархива при найденных согласно (1.49) оптимальных гиперпараметрах $\hat{\mu}$:

$$\arg \min_{\vec{\theta}} \frac{1}{|A_{train} \cup A_{valid}|} \sum_{j \in A_{train} \cup A_{valid}} e(X_{fact,j}, F(\vec{X}_{input,j}, \vec{\theta}_{\hat{\mu}})) + R(\hat{\mu}, \vec{\theta}_{\hat{\mu}}) = \hat{\theta}_{\hat{\mu}} \quad (1.50)$$

Оптимизацию (1.50) проводят не всегда. Например, если подархив A_{valid} использован для так называемой ранней остановки обучения (обучение модели прекращается, если функционал качества на A_{valid} начал расти), то оптимизация (1.50) может приводить к переобучению. В данной работе периоды архивов A_{train} и A_{valid} были равны году, поэтому применялась оптимизация (1.50), а ранняя остановка не использовалась.

И наконец, применялась итоговая проверка на тестовом архиве:

$$L_{test} = \sum_{j \in A_{test}} e(X_{fact,j}, F(\vec{X}_{input,j}, \hat{\theta}_{\hat{\mu}})). \quad (1.51)$$

В данной работе проверка (1.51) осуществлялась в оперативном режиме счета, таким образом гарантировалось, что никакие данные наблюдений из будущего не использовались при вычислении поправок.

1.2. Используемые наборы данных

1.2.1. Данные наблюдений на синоптических станциях

Использовался архив наблюдений на синоптических станциях. Эти наблюдения, как правило, проводятся 4 или 8 раз в сутки в сроки, кратные, соответственно, 6 или 3 ч, начиная с 00:00 всемирного согласованного времени (ВСВ). Синоптическая станция в каждый срок наблюдения передаёт данные о температуре воздуха, точке росы, скорости и направлении ветра

и другие параметры. В Гидрометцентре России данные измерений температуры воздуха и точки росы проходят контроль, основанный на скользящем среднем, а данные о скорости и направлении ветра – более грубый горизонтальный контроль по первому приближению [14]. Наблюдения на автоматических станциях не использовались при обучении, поскольку они требуют дополнительного контроля. Однако, данные с автоматических станций использовались при тестировании прогнозов, рассчитанных в оперативном режиме счета.

1.2.2. Данные оперативных прогнозов модели COSMO-Ru

При интерполяции прогностических полей в точки наблюдений использовались архивы приземных прогностических полей модели COSMO-Ru различных конфигураций [122], [70], [69], [71], [72] в формате GRIB и специальная программа, предоставленная Г.С.Ривиным [20].

В таблице 1.3 приведен список используемых конфигураций модели COSMO-Ru и их основные характеристики. Рассмотрены конфигурации модели COSMO-Ru: COSMO-Ru13-ENA [71], COSMO-Ru6ENA [70], COSMO-RuBy, COSMO-Ru1Msk [69], [72].

На рисунке 1.3 приведены расчетные области этих конфигураций. Далее для будем обозначать архив прогнозов по конфигурации COSMO-A за период N как A-N, например архив прогнозов по конфигурации COSMO-Ru13ENA за период с мая 2019 г. по июнь 2020 г. как Ru13ENA-05/2019-06/2020. За 2021 г. здесь использовались архивы прогнозов до 31 марта 2021 г.

Таблица 1.3 – Используемые конфигурации модели COSMO-Ru

Название конфигурации	COSMO-Ru13ENA	COSMO-Ru6ENA	COSMO-RuBy	COSMO-Ru1Msk
Версия модели COSMO	5.03-5.09	5.03-5.09	5.03-5.09	5.00-5.05 urban
Шаг сетки	13.2 км	6.6 км	2.2 км	1.0 км
Размер сетки	1000×500	2000×1000	1400×1200	180×180
Используемые начальные сроки ВСВ, ч	00, 12	00, 06, 12, 18	00, 06, 12, 18	00, 06, 12, 18
Используемая максимальная заблаговременность, ч	99	120ч (от 00, 12 ВСВ) 78ч (от 6, 18 ВСВ)	48	48
Используется синоптических станций	3771	2796	778	39

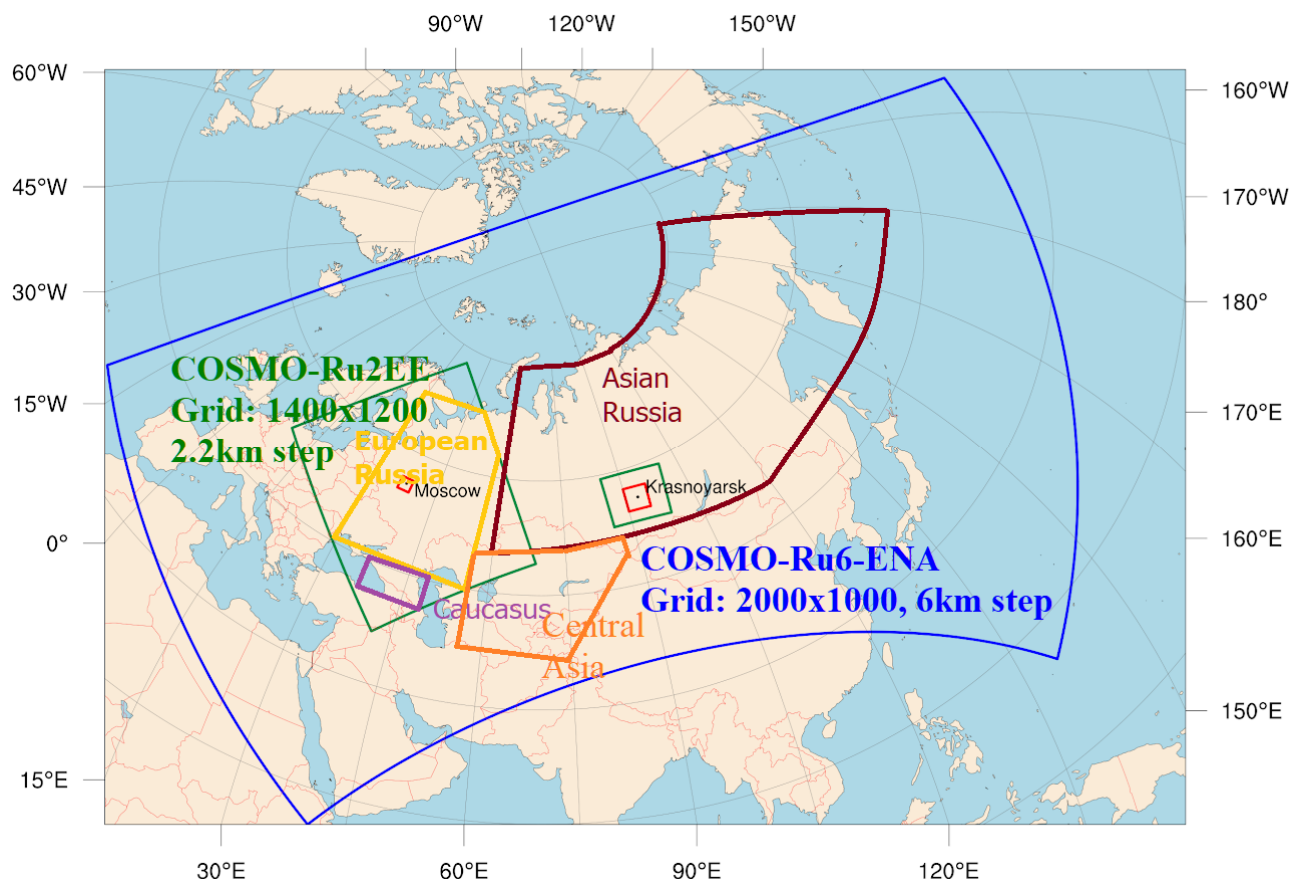


Рисунок 1.3 – Расчетные области используемых конфигураций (COSMO-Ru6ENA, COSMO-RuBy) модели COSMO-Ru и области (ЕЧР, АЧР, Кавказ, Центральная Азия, таблица 1.5), по которым рассчитывались оценки качества рассмотренных прогнозов

Перед проведением тестирования была проведена повторная оптимизации параметров на объединении архивов, использованных для обучения и валидации.

Для обучения, тестирования и валидации различных нейронных сетей использовались различные архивы прогнозов (таблица 1.4). Длина каждого из используемых архивов была кратна году. Начиная с 1 октября 2020 г. каждые 2 недели производилось дополнительное обучение на скользящем архиве, содержащем данные прогнозов модели COSMO-Ru по конфигурациям COSMO-Ru6ENA и COSMO-RuBy за последний год (объем архива для обучения составлял около 6.8×10^7 пар прогноз-факт), а начиная с 1 февраля 2021 г. – на архиве, содержащем архив прогнозов за последние 2 года (объем архива 1.3×10^8).

Архивы, используемые для обучения, содержали только прогнозы от 00:00 и 12:00 ВСВ. Тестирование постпроцессинга прогнозов модели COSMO-Ru производилось на прогнозах от 00:00, 06:00, 12:00, 18:00 ВСВ.

Тестирование предлагаемого постпроцессинга проводилось в автоматизированном оперативном режиме отдельно на территориях, указанных в таблице 1.5. Все эти территории лежат внутри области, в которой проходило обучение. Указанные в таблице 1.5 территории

показаны графически на рисунке 1.3. Постпроцессинг применялся в автоматизированном режиме счета для коррекции прогнозов для модели COSMO-Ru в конфигурациях COSMO-Ru6ENA [70], COSMO-RuBy, COSMO-Ru1Msk [69], [72].

Таблица 1.4 – Используемые для обучения, валидации и тестирования различных нейронных сетей архивы прогнозов модели COSMO-Ru

Задача	Обучение	Валидация	Тестирование
Квазилинейная систематическая коррекция (КСК, глава 2)	Ru13ENA-2017-2018	Ru6ENA-2019	Ru6ENA-2020
Интерполяция поправок (глава 4)	Ru6ENA-2019 RuBy-2019	Ru6ENA-2020 RuBy-2020	Ru6ENA-2021 RuBy-2021
Коррекция (КНС, глава 3)	Ru13ENA-2017	Ru13ENA-2018	Ru6ENA-01/2019-09/2020 RuBy-01/2019-09/2020
КСК + КНС	Скользящий последний год Ru6ENA и RuBy		Ru6ENA-10/2020-01/2021 RuBy-10/2020-01/2021
КСК + КНС	Скользящие 2 последних года Ru6ENA и RuBy		Ru6ENA-02/2021-03/2021 RuBy-02/2021-03/2021

Таблица 1.5 – Территории (рисунок 1.3), используемые при оценках качества прогнозов, и количество станций на них

Название	Количество станций	Северная широта	Восточная долгота
Европейская часть России (ЕЧР)	912	43-71	27-62
Азиатская часть России (АЧР)	774	50-75	62-180
Центральная Азия	326	36-56	53-88
Московского региона	39	54,8-56,4	36,1-38,9
Кавказ	60	39-44	42-49

1.2.3. Данные оперативных прогнозов ведущих мировых центров

В системе комплексного прогноза погоды используются прогностические поля, выпускаемые ведущими мировыми центрами. Прогностические поля поступают через сеть международного обмена метеорологической информацией и (или) сеть интернет и имеют шаг сетки в несколько раз больше, чем расчетный шаг соответствующих численных моделей прогноза погоды (ЧМПП). Доступный ассортимент приземных прогностических полей, их заблаговременности и горизонтальный шаг сетки указаны в таблица 9.2. Накопленный архив пар прогноз-факт включал только прогнозы с заблаговременностью до 144 часов. Рассматривались

начальные сроки 00:00 и 12:00 ВСВ. Всего в архиве за каждый год присутствовало около 9.2×10^7 фактических значений, которым соответствовал прогноз хотя бы по одной ЧМПП.

Методы интерполяции, применяемые в иностранных центрах прогноза, при интерполяции с расчетной сетки ЧМПП на сетку, в которой распространяются прогностические данные в сети международного обмена метеорологической информацией, автору неизвестны. При интерполяции же поступающих данных в точки синоптических станций использовался метод билинейной интерполяции.

Помимо прогностических полей мировых центров прогноза погоды в системе комплексного прогноза погоды также используются прогностические поля модели COSMO-Ru конфигурации COSMO-Ru6ENA, а также модели PLAV [82], [81]. Для интерполяции прогнозов COSMO-Ru в точки синоптических станций используется специальная программа, предоставленная Г.С.Ривиным [20], а для прогнозов модели PLAV – билинейная интерполяция.

Прогнозы всех перечисленных в таблице 1.6 ЧМПП использовались при расчете комплексного прогноза погоды (вторая часть главы 2 и глава 3) с использованием нейронных сетей. Комплексный прогноз рассчитывался для начальных сроков 00:00 и 12:00 ВСВ. Чтение прогностических полей используемых ЧМПП из баз данных Гидрометцентра России производилось через 05 часов 00 минут после соответствующего начального срока. При недоступности этого прогноза использовался прогноз с начальным сроком на 12 часов раньше и с заблаговременностью на 12 часов большей. Если же и он был недоступен, то данная ЧМПП не использовалась при расчете комплексного прогноза. Отметим, что для модели ECMWF всегда использовался прогноз на 12 часов большей заблаговременностью, поскольку прогностические поля этой модели поступают в базы данных примерно через 3 часа после расчета комплексного прогноза в оперативном режиме.

Таблица 1.6 – Доступные данные прогнозов ведущих мировых центров

Численная модель	CosmoRu6-ENA	PLAV	ICON-De	GFS	IFS	UKMO		JMA
Организация	Гидрометцентр России		DWD (Германия)	NCEP (США)	ECMWF	MetOffice (Великобритания)		JMA (Япония)
Шаг сетки	6.6 км	75км	0,25°	0,5°	0,5°	1°	2,5°	1,25°
Максимальная заблаговременность, шаг	120ч 3ч	120ч 3ч	168ч 6ч	180ч 3ч	168ч 6ч	48ч 6ч	120ч 12ч	72ч 6ч
Используемые поля	T, Td, U, V, P, G	T, Td, U, V, P	T, U, V, P, G	T, Td, U, V, P, G	T, U, V, P, G	T, Td, U, V, P		T, Td, U, V, P
Пар прогноз-факт в год	6.3×10^7	5.8×10^7	2.2×10^7	8.8×10^7	3.7×10^7	1.5×10^7		2.2×10^7

Для обучения, тестирования и валидации комплексного прогноза в различные периоды использовались различные архивы прогнозов, таблица 1.7. Тестирование качества комплексного прогноза производилось в оперативном режиме счета. Каждый из используемых архивов был кратен году. Начиная с 1 октября 2020 г. каждые 2 недели производилось дополнительное обучение на скользящем архиве, содержащем прогностические данные за последний год (объем архива для обучения составлял около 9.2×10^7 пар прогноз-факт), а начиная с 1 февраля 2021 г. – прогностические данные за последние 2 года (объем архива около 1.8×10^8).

Таблица 1.7 – Используемые периоды для обучения, валидации и тестирования комплексного прогноза погоды

Обучение	Валидация	Тестирование
04/2017-03/2018	04/2018-03/2019	04/2019-09/2020
Скользящий последний год		10/2020-01/2021
Скользящие 2 последних года		02/2021-03/2021

В таблице 1.8 приведена информация о количестве пропусков в данных прогнозов с заблаговременностью 48 часов. Видно, что случаи отсутствия (запаздывания) прогностических данных в базах данных Гидрометцентра России весьма часты и не стоит рассчитывать на то, что в момент составления комплексного прогноза будет доступна интересующая нас история прогнозов по всем рассматриваемым ЧМПП: так история за 15 суток доступна примерно в половине случаев.

Таблица 1.8 – Доля дней, таких что в k предшествующих дней через 05 часов 00 минут после начального срока были доступны все k полей прогнозов температуры воздуха на высоте 2 м с заблаговременностью 48 часов

Дней k	1	5	15	25	35
Модель					
CosmoRu6ENA	97.7%	90.4%	75.6%	62.7%	52.0%
ICON-DE	99.2%	97.0%	92.0%	88.4%	85.4%
ECMWF	98.5%	95.2%	88.3%	81.3%	74.0%
JMA	97.5%	97.5%	84.4%	77.3%	70.0%
GFS	99.5%	97.5%	83.0%	88.8%	84.4%
PLAV	86.0%	83.8%	78.2%	72.3%	66.0%
UKMO	99.7%	98.7%	96.9%	95.5%	94.0%
Все 7 моделей	81.3%	69.6%	50.6%	36.4%	24.9%

Поэтому любой практически значимый постпроцессинг, использующий историю прогнозов, должен приспосабливаться к наличию пропусков в данных. Более того, при

оптимизации параметров применяемой коррекции нельзя делать предположение о доступности всей интересующей нас информации.

1.3. Предпосылки и перспективы применения нейронных сетей в геофизике

1.3.1. Блочный подход к построению моделей машинного обучения

Опишем сначала преимущества и научно-технологическую основу нейросетевых методов, а затем перейдём к описанию конкретных результатов.

С теоретической точки зрения привлекательность использования именно нейронных сетей заключается в следующем. Нейронные сети гарантируют гладкость (кусочно-гладкость) решающей функции. Любую непрерывную функцию на n -мерном кубе можно приблизить нейронной сетью [128], [254], параграф 1.1.3. Нейронные сети хорошо подходят для обработки большого объёма данных и не предполагают, что значения входных параметров распределены согласно какому-то вероятностному распределению.

С технологической точки зрения привлекательность нейронных сетей (DAG-сетей) заключается в следующем. Современные прикладные тензорные библиотеки программных модулей, такие как [205] и [228], содержат большое количество блоков (параграф 1.1.4), которые позволяют легко конструировать, обучать и применять DAG-сети. Все эти блоки имеют режим вычисления градиентов. Практически любой вещественнозначный алгоритм A , вычисляющий некоторую функцию с абсолютно ограниченным градиентом, можно запрограммировать с использованием лишь блоков из этих библиотек. Такая программа позволит использовать этот алгоритм A при вычислении модели МО (параграф 1.1.4) и быстро (за примерно то же время, что требуется для применения алгоритма A) вычислять градиенты модели МО.

Таким образом, теоретическая проблема оптимизации параметров сложных алгоритмов сводится к задаче минимизации функций многих переменных с известным градиентом, которая уже может быть решена методами градиентного спуска первого порядка, а не методами нулевого порядка. Это важно потому, что применение градиентного спуска позволяет оптимизировать модели МО, количество параметров которых на 2 - 4 порядка больше, чем допустимо при применении методов оптимизации нулевого порядка. Однако практическая применимость такого представления может быть существенно ограничена количеством доступных данных и памяти (потребуется хранить результаты почти всех промежуточных вычислений), производительностью используемой вычислительной системы или невозможностью достичь достаточного уровня параллелизма для рассматриваемого алгоритма A .

Благодаря возможности использования внутри оптимизируемого алгоритма A нейросетевых модулей, становится возможной оптимизация частично интерпретируемых алгоритмов (так называемых «серых ящиков») методами градиентного спуска первого порядка. «Серым ящиком» называют алгоритм, «внешние» блоки которого известны, но некоторые из «внутренних» блоков представляют собой «черный ящик».

Чтобы найти наилучшую (согласно заранее заданному критерию качества) для решения рассматриваемой задачи модель МО $\vec{F}(\vec{X}_{input}, \vec{\theta})$ может потребоваться перебрать несколько конфигураций модели МО F . Можно конструировать каждую модель МО отдельно, программировать для каждой модели МО отдельный алгоритм для вычисления градиента штрафного функционала $\nabla_{\vec{\theta}} L$ и оптимизировать параметры $\vec{\theta}$.

Более практичным является блочный подход (параграф 1.1.4, [200], [108], [199], [64]), использующий метод обратного распространения ошибки (параграф 1.1.5), когда модель F строится из заранее запрограммированных блоков $\vec{F}_k(\vec{X}_k, \vec{\theta}_k)$. Чтобы включить блок $\vec{F}_k(\vec{X}_k, \vec{\theta}_k)$ в блочную модель градиентного МО необходимо, чтобы блок $\vec{F}_k(\vec{X}_k, \vec{\theta}_k)$, помимо прямого вычисления, поддерживал вычисление градиентов, а именно, вычисление произведений матриц Якоби $\nabla_{\vec{\theta}_k} \vec{F}_k$ и $\nabla_{\vec{X}_k} \vec{F}_k$ на некоторый вектор $\vec{z}_k$ той же размерности, что и выход блока $\vec{F}_k$.

В программном комплексе все блоки модели F являются объектами классов-наследников определенного абстрактного класса, например, абстрактного класса `torch.nn.Module` в пакете PyTorch [205]. Абстрактным называют базовый класс, экземпляры которого не создаются, а создаются экземпляры классов-наследников. Такая программистская конструкция даёт возможность при конструировании других моделей МО использовать эти блоки повторно (быть может, с оптимизируемыми заново параметрами) без написания дополнительного программного кода: для вычисления градиентов функционала потерь используется метод обратного распространения ошибки [34], параграф 1.1.5.

В современных свободно-распространяемых пакетах, таких как TensorFlow [228] и PyTorch [205] содержатся блоки, реализующие различные общепринятые слои нейронных сетей, предназначенных для преобразования данных специальных типов: векторов (полносвязные слои, full-connected, FC [73]); многомерных временных рядов (рекуррентные слои, recurrent neural network, RNN [157], [147]); наборов полей (сверточные слои, convolutional neural network, CNN [181]); плохо упорядоченных данных (слои transformer, attention [237]); классификаций (слои Embeddings).

Блоки модели МО (объекты различных классов-наследников класса `torch.nn.Module`) можно комбинировать внутри одной модели МО F без существенных ограничений, а затем оптимизировать параметры всех блоков методом обратного распространения ошибки [34]. Кроме того, разработанные ранее алгоритмы могут быть перепрограммированы в виде блоков и таким образом встроены в модель МО. Это даёт возможность эффективно оптимизировать параметры этих блоков. Примером такого встраивания является блок NeuralODE [121] – результат встраивания метода Рунге-Кутты для решения обыкновенных дифференциальных уравнений (ОДУ) в DAG-сеть. Таким образом, при применении одного из блоков $\vec{F}_k$ модели F приближенно решается ОДУ с нейронной сетью в правой части.

Таким образом, DAG-сети являются наиболее технологичным подходом к обработке практически любых данных, в том числе разнородных, позволяют относительно легко учитывать особенности обрабатываемых данных, оптимизировать (обучать) параметры разработанных ранее методов и комплексов программ на новом уровне абстракции.

Недостаток нейронных сетей: они часто являются черным ящиком (black box): их ответы сложно интерпретировать, то есть понять причины полученного ответа. Это может заметно осложнить дальнейшую оптимизацию качества и вычислительной эффективности модели МО. В задаче коррекции прогнозов ЧМПП интерпретируемость полученной модели машинного обучения весьма важна, поскольку она может помочь в последующем улучшении корректируемой ЧМПП.

1.3.2. Синергетический эффект нейронных сетей

Ещё одно преимущество нейронных сетей (DAG-сетей) заключается в том, что методы, основанные на нейронных сетях для смежных задач, легко комбинировать и обучать совместно, что должно приводить к **синергетическому эффекту** по сравнению с отдельным обучением.

В настоящее время различные компоненты системы прогноза погоды оптимизируются независимо друг от друга, и существующая система может быть неоптимальна в целом. Это проявляется, например, в так называемом эффекте spin-up: при малых заблаговременностях (до 6-12 часов) погрешность ЧМПП падает (а не растёт) с ростом заблаговременности. Эффект spin-up объясняется тем, что начальные метеорологические поля недостаточно согласованы между собой. В перспективе блочный подход к построению моделей МО позволит построить оптимальную в целом систему прогноза погоды. Для этого можно объединить в единую систему блоки пре- и постпроцессинга и физическую модель, поддерживающую режим вычисления градиентов (матриц Якоби).

Объединенная система позволит оптимизировать качество выходной продукции напрямую, оптимизируя все параметры физической модели и её пре- и постпроцессинга одновременно, не используя дополнительные предположения о вероятностно-статистических распределениях полей физических параметров. Создание объединенной системы прогноза проще начинать с конца технологической линии, то есть с создания нейросетевого постпроцессинга: для постпроцессинга легко задать критерии качества (параграфы 1.1.1, 1.1.2). Примером применения такого подхода является задача оптимизации системы усвоения данных для данной математической модели, рассмотренная в [111], [142].

1.3.3. Полностью нейросетевые модели прогноза

В настоящее время существуют прототипы **полностью нейросетевых глобальных моделей прогноза погоды**, например, [100], [210], [241]. Эти модели работают с метеорологическими полями на сетке с горизонтальным разрешением 200 - 600 км и являются черными ящиками. Качество их прогнозов соизмеримо с качеством прогнозов по классическим физически обоснованным глобальным ЧМПП с горизонтальным разрешением около 200 км.

Важнейшим преимуществом полностью нейросетевых моделей прогноза является скорость вычислений: согласно работе [241] нейросетевая модель при сопоставимом качестве прогноза вычисляет его в 7200 раз быстрее классической численной модели (приблизительно решающей дифференциальные уравнения). Отметим, что существующие нейросетевые модели прогноза рассматривают существенно меньше метеорологических полей и уровней по вертикальной координате, например, модель из [210] предсказывает всего 4 двумерных метеорологических поля: приземную температуру, осадки, высоту уровня 500 гПа H500 и температуру на уровне 850 гПа T850.

Уменьшение шага сетки в нейросетевых моделях потребует увеличения количества параметров модели, а значит и увеличения длины используемого архива данных. Основным препятствием для дальнейшего развития нейросетевых моделей прогноза является невозможность существенного увеличения архива данных, используемого для обучения: в рассмотренных выше работах, использовались архивы за около 40 и более лет (например, набор данных weatherbench [201]) и их пополнение более старыми данными невозможно (за год к архиву прибавляется лишь этот один год). Ограниченность используемых архивов может привести к некорректному прогнозированию очень редких в данной местности явлений.

Вероятным решением этой проблемы может быть подход искусственного размножения имеющихся данных, так называемые аугментации (augmentation) [234], [127], на настоящее время

требующий дополнительной проработки для применения к метеорологическим данным. Другим решением проблемы недостатка данных является рассмотрение не моделей, являющихся черными ящиками, а нейросетевых моделей, учитывающих физические ограничения.

В последнее время различными авторами предложены нейросетевые физические модели, рассматривающие дифференциальные уравнения, параметризованные с помощью нейронных сетей. Обзор применения нейросетевых моделей в геофизике сделан в [170].

В пионерской работе в этом направлении для интегрирования используется метод Эйлера [74]. Метод оптимизации параметров моделей, решающий обыкновенные дифференциальные уравнения методом Рунге-Кутты, – так называемые *нейронные обыкновенные дифференциальные уравнения* (нейронные ОДУ) предложен в [121], а для гамильтоновых систем в [149]. Нейронные уравнения в частных производных рассматривались в работах [207], [151], [182]. Метод нейронных ОДУ [121] использует явные методы Рунге-Кутты, которые для некоторых краевых задач могут не обеспечить требуемой точности решения системы ОДУ. Однако, для краевых задач возможно применение подхода, предложенного для уравнений в частных производных из [207]. В подходе [207] решение нескольких уравнений (Шрёдингера, Бургерса, Аллена – Кана, Кортевега – де Фриза, двумерной системы Навье – Стокса) ищется в виде нейронной сети, удовлетворяющей системе уравнений неявного метода Рунге-Кутты до 1000-го порядка точности.

Другой способ использования методов МО задачах моделирования погоды – замена некоторых вычислительно дорогих блоков существующих ЧМПП на более дешевые блоки, вычисляемые методами МО [177]. Эти подходы применялись к блокам ЧМПП: блоку радиации [175], [176] и физики облаков [180], что позволило ускорить вычисления этих блоков в 10-150 раз.

1.3.4. Комбинированные и квазилинейные методы

Традиционные линейные статистические методы позволяют (используя рассчитанную заранее ковариационную матрицу признаков) применять постпроцессинг в том числе и в тех случаях, когда часть информация о значениях некоторых из предикторов $\vec{x}$ недоступна. Общие нелинейные методы в таком случае, как правило, не применимы: для расчета оптимальных параметров нелинейного метода при условии использования лишь некоторых из предикторов $\vec{x}$ оптимизацию параметров $\vec{\theta}$ проводят заново. В некоторых случаях невозможно предусмотреть заранее все возможные варианты наличия/отсутствия входной информации, потому что количество таких вариантов может быть слишком большим. Например, если наш алгоритм

использует все наблюдения за последние 7 суток с шагом 3 часа в рассматриваемой точке, то число возможных случаев наличия/отсутствия входной информации даже одного метеозлемента равно $2^{56} > 7 \times 10^{16}$. Рассмотреть заранее такое большое количество случаев не представляется возможным. Разработка комбинированных методов, использующих преимущества линейных и нелинейных методов, является актуальной задачей.

Простейший способ объединить преимущества линейных и нелинейных методов заключается в рассмотрении квазилинейных методов. Комбинируя квазилинейные методы и подход блочного машинного обучения приходим к рассмотрению квазилинейных методов, коэффициенты которых вычисляются методами машинного обучения. Отметим, что квазилинейные методы являются частично интерпретируемыми («серый ящик»), а то время как полностью нелинейные методы зачастую представляют собой черный ящик. Это является важным преимуществом квазилинейных методов над полностью нелинейными методами.

При оперативной интерпретации прогнозов нескольких ЧМПП, используемые модели МО должны быть применимы в том числе в случаях, когда информация о части запланированных наблюдений недоступна или была отброшена при контроле качества. Из-за возможных технических накладок (при передачи данных) при использовании нескольких ЧМПП проблема доступности прогностической информации в полном объеме возникает существенно чаще (таблица 1.8), чем при использовании прогнозов одной ЧМПП: часть прогностических полей может ещё не поступить в используемые базы данных к моменту составления комплексного прогноза.

1.3.5. Задача интерполяции и машинное обучение

При решении задачи постпроцессинга прогноза метеорологических параметров желательно использовать данные измерений, которые известны только в точках метеорологических станций. Чтобы применить метод коррекции в точках, отличных от точек метеорологических станций, можно либо при применении метода коррекции не использовать данные последних измерений (что естественно ухудшит качество откорректированного прогностического поля); либо после вычисления поправок в точках синоптических станций проинтерполировать полученные поправки по горизонтальным (а возможно, и по вертикальной) координатам. Задача интерполяции поправок традиционно решается в препроцессинге ЧМПП.

При интерполяции полей метеорологических параметров, измеренных на метеорологических станциях, по горизонтальным координатам с помощью традиционных линейных методов (таких как оптимальная интерполяция [36], вариационное усвоение данных

[106], фильтр Калмана [159], [135]), как правило, делают весьма существенные предположения. Ключевые из этих предположений, существенно влияющие на качество интерполяции: однородность и изотропность интерполируемого поля по горизонтальным координатам; распределение значений интерполируемого поля согласно гауссовскому вероятностному распределению; линейность оператора интерполяции.

Отказ от этих предположений или замена их на более слабые является актуальной задачей, поскольку может существенно улучшить качество интерполяции полей метеорологических параметров.

Для интерполяции поправок из точек синоптических станций используется статистическая структура этих поправок. Задача изучения статистической структуры этих поправок может оказаться проще, чем решаемая в блоке препроцессинга ЧМПП задача изучения статистической структуры погрешностей прогнозов ЧМПП минимальной заблаговременности. Во-первых, поскольку поправки приближают среднюю погрешность прогноза, то поле поправок заметно менее зашумленное, чем поле погрешностей. Во-вторых, количество доступных данных может быть заметно больше, чем в случае изучения статистической структуры погрешностей прогнозов ЧМПП минимальной заблаговременности: поправки интерполируются и при больших заблаговременностях прогноза. При этом разработанные для интерполяции поправок подходы могут быть применены также и в пре-процессинге ЧМПП.

1.4. Технические подробности реализации предлагаемого постпроцессинга

Запуск разработанного постпроцессинга в автоматизированном оперативном режиме осуществлялся в сотрудничестве с несколькими сотрудниками ФГБУ «Гидрометцентр России». Все предлагаемые модели МО, используемые в постпроцессинге, были реализованы на языке программирования Python [204] с использованием пакетов из дистрибутива Anaconda3-2019.10 [96], а также пакета PyTorch [205] версии 1.4.1. Проверялась работоспособность разработанных python-скриптов на следующих вычислительных системах:

- на суперкомпьютерах Cray XC40-LC и T-Platforms V6000 под управлением ОС SUSE Linux 15.1;
- на личном и рабочем персональных компьютерах под управлением ОС Windows 10;
- на персональном компьютере под управлением ОС SUSE Linux 15.2, удаленный доступ к которому был обеспечен Ю.А.Степановым;
- в облачной среде Google Colaboratory [145].

Автор благодарит Ю.А.Степанова за помощь в выборе наиболее портируемых версий python-пакетов. Автор благодарит директора ФГБУ «ГВЦ Росгидромета» С.В.Лубова за помощь с установкой дистрибутива Anaconda3-2019.10 и других используемых python-пакетов на суперкомпьютерные системы Cray XC40-LC и T-Platforms V6000 (модуль common/general/python/Anaconda3-2019.10), а также заместителя директора ФГБУ «Гидрометцентр России» Д.Б.Киктева и начальника лаборатории Г.С.Ривина за ходатайство о необходимости провести эту установку.

1.4.1. Подробности реализации

Для обращения к базам данных «Прогноз» Гидрометцентра России (далее БД) автором был создан специальный python-модуль `hmcdbase`, в котором реализован клиентский доступ как к локальным БД, так и к удаленным БД. Автор благодарит А.Ю.Недачину за реализацию используемого интерфейса удаленного доступа на стороне сервера БД, а также Ю.А.Степанова и И.И.Кулакову за предоставление подключаемой `.so` (shared object) версии библиотеки `bankload`, используемой для локального доступа к БД из python-модуля `hmcdbase`.

При работе на вычислительных узлах суперкомпьютеров Cray XC40-LC и T-Platforms V6000 доступ к удаленным БД невозможен, поскольку вычислительные узлы не имеют внешнего сетевого адреса. Перед запуском задания на вычислительных узлах необходимо скопировать требуемую информацию с удаленных машин в локальные базы данных, размещенные в дисковой системе Lustre3.2. Для этого на front-end запускается python-скрипт, использующий модуль `hmcdbase` в режиме удаленного доступа. После завершения выполнения задания на вычислительных узлах другим скриптом выполняется копирование в обратном направлении.

Для чтения прогностических полей модели COSMO-Ru в формате GRIB версии 1 использовался пакет `pygrib` (PUre PYthon GRIB), [164]. Поля в формате GRIB версии 2 преобразовывались в формат GRIB версии 1 специальной утилитой `fieldextra`, используемой консорциумом COSMO [122], [20].

Для хранения различных метеорологических (прогностических и фактических) и прочих (таких как географическая широта и долгота) параметров был реализован специальный класс `DataDict` – наследник стандартного класса `UserDict`. Реализация класса `DataDict` обусловлена тем, что для различных алгоритмов удобно переупорядочивать многомерные массивы значений параметров различным образом, что может приводить к ошибкам при недостаточно аккуратном последовательном применении нескольких алгоритмов.

Для подготовки minibatch (параграф 1.1.5) был реализован специальный класс DatasetCor – наследник класса torch.utils.data.Dataset. Реализация этого класса обусловлена тем, что стандартная реализация для генерации каждой реализации предикторов и целевых значений запускает свой subprocess, что приводит к большим накладным расходам. Класс DatasetCor однократно запускает фиксированное число нитей (потоков, thread) для генерации одного minibatch, а также запускает генерацию следующего minibatch в фоновом режиме сразу после возврата предыдущего minibatch. Это позволяет в некоторых случаях (особенно для моделей МО, рассматриваемых в главе 3) увеличить производительность генерации minibatch.

Все предложенные модели МО были реализованы в виде классов-наследников класса torch.nn.Module. Такая реализация позволит в дальнейшем использовать эти модели МО в качестве блоков при построении более сложных технологий обработки данных.

Для конфигурирования моделей МО был реализован специальный модуль configure, в котором задаётся конфигурация используемой модели МО (словари его гиперпараметров). Процедура обучения реализована таким образом, чтобы можно было обучать несколько однотипных (рассматриваемых в одной главе и использующих совпадающий набор предикторов) моделей различных конфигураций одновременно, без повторной генерации minibatch. Это позволяет производить более корректную валидацию качества прогнозов, вычисляемых данной конфигурацией модели МО, поскольку гарантируется, что все псевдослучайные minibatch совпадали для различных конфигураций.

Для удобства для каждой рассматриваемой конфигурации модели МО создается объект специального класса Trainer, в котором хранятся в том числе: информация о конфигурации; обучаемая модель МО, то есть объект класса-наследника класса torch.nn.Module; информация о процессе обучения данной модели МО; результаты применения модели МО (предсказанные значения) – при применении модели МО (инференс, inference).

Для запуска в оперативном режиме используется автоматизированная система сбора и обработки информации ФГБУ «Гидрометцентр России» (далее АССОИ) и планировщик crontab. Для запуска заданий на вычислительных узлах – команды qsub и arun.

В таблице 1.9 приведено расписание запуска компонент технологической линии выпуска комплексного прогноза погоды, в которой используются предлагаемый постпроцессинг в оперативном режиме. В первом столбце указано время запуска и завершения работы для начального срока 00:00 ВСВ.

Таблица 1.9 – Расписание запуска предлагаемого постпроцессинга в оперативном режиме

Время запуска – время завершения ВСВ / раз в сутки	Задача	Используемые БД
04:54 – 04:58 / 4	Python-скрипт записи прогнозов моделей COSMO-Ru конфигураций COSMO-Ru6ENA, COSMO-RuBy и COSMO-Ru1Msk, а также их откорректированных версий в БД COBB	COBB, SHOT
05:00 – 05:04 / 2	Задачи АССОИ fores (интерполяция прогностических полей) и zapnabl (запись соответствующих данных наблюдений) – автор А.Н.Багров	MOBB, SHOT, EGRG, RGRR, NCVP, EDZN, RJTD, SLMA, LAG2, LMW7, MAKТ, RGPR, ECMG, CAND, MABM
05:05 – 05:12 / 2	Python-скрипт расчета комплексного прогноза (T-Platforms V6000)	MOBB, COBB, SHOT
05:05 – 05:15 / 2	Задачи АССОИ ansdop2 – резервная fortran-версия комплексного прогноза	MOBB, COBB
05:20 – 05:35 / 2	Рисование метеограмм и синоптических карт для сайтов методического кабинета [9] и проекта по прогнозированию суровых погодных условий в Центральной Азии (авторы Н.А. Светлова Е.Н.Шакотько)	MOBB
–	Расчет откорректированных прогностических полей с первым приближением по модели COSMO-Ru в конфигурациях COSMO-Ru6ENA (скрипт передан Д.В.Блинову для визуализации откорректированных приземных карт и использования в системе усвоения приземных данных)	SHOT, COBB, MOBB

1.4.2. Ускорение на GPU

Диссертантом были проведены эксперименты по тестированию производительности вычислений на модуле с двумя 18-ядерными CPU E5-2697v4 (суперкомпьютер Cray XC40-LC) и на модуле V220F с двумя 20-ядерными CPU Xeon Gold 6148 и двумя GPU-ускорителями NVidia Tesla V100 PCI-E (суперкомпьютер T-Platforms V6000). Результаты сравнения производительности приведены в таблице 1.10. Во всех случаях, за исключением строки «Применение КНОИ (1 поле 2000x1000)» указано реальное время, затрачиваемое на вычисление всех рассматриваемых приземных полей (T , T_d , U , G , P_0) без учёта загрузки данных. При обучении всех моделей МО, используемых при постпроцессинге, использовались двухлетние архивы по 2800 станциям, объемом порядка $1.3-1.8 \times 10^8$ пар прогноз-факт. Для распараллеливания алгоритма на два GPU применялась функция `torch.nn.parallel.data_parallel`.

Из таблицы 1.10 видно, что использование GPU позволяет существенно ускорить обучение и применение предлагаемых моделей МО, но величина ускорения сильно зависит от применяемой модели МО: для КНОИ (глава 4) ускорение составляет до 115 раз, а для КНС (глава 3) – от 1.6 раза. Такая большая разница обусловлена тем, что КНС требует относительно мало арифметических операций, и CPU не успевает подготавливать данные (наборы предикторов) для модели КНС.

Таблица 1.10 – Сравнение производительности различных моделей МО на одном CPU-модуле (суперкомпьютер Cray XC40-LC) и на одном GPU-модуле V220F (суперкомпьютер T-Platforms V6000). В скобках указано ускорение GPU относительно CPU

Задача	CPU	1 GPU V100	2 GPU V100
Обучение КНОИ (1 эпоха из 5)	12 ч	40 мин (x18)	25 мин (x28)
Применение КНОИ (1 поле 2000x1000)	190 с	3.9 с (x49)	2.1 с (x90)
Применение КНОИ (1 прогноз до 120ч с шагом 3ч сетка 2000x1000)	8 ч	8 мин (x60)	250 с (x115)
Обучение КСК (1 эпоха из 5)	4 ч 25мин	22 мин (x12)	12.5 мин (x21)
Применение КСК (1 прогноз до 120ч с шагом 3ч по 2800 станциям)	85 с	20 с (x4.2)	12 с (x7)
Обучение ККП (1 эпоха из 5)	32 ч	3ч 45 мин (x8.5)	2 ч (x15)
Применение ККП (1 прогноз до 144ч с шагом 3ч по 2800 станциям)	280 с	92 с (x3)	63 с (x4.5)
Обучение КНС (1 эпоха из 20)	490 с	310 с (x1.6)	330 с (x1.5)
Применение КНС (1 прогноз до 144ч с шагом 3ч по 2800 станциям)	8 с	1.5 с (x5)	0.8 с (x10)

Выводы из главы 1

В первой части главы приведено теоретическое обоснование и описаны особенности практической реализации технологии нейронных сетей. Рассмотрены процессы конструирования, обучения параметров и оптимизации гиперпараметров нейронной сети. Приведена серия наиболее практически значимых результатов отечественных и иностранных авторов.

Во второй части главы описаны используемые в данной работе прогностические и фактические данные. Для проведенных экспериментов использовались прогностические данные отечественных ЧМПП (COSMO-Ru, PLAV), а также нескольких иностранных ЧМПП (ICON-De, GFS, IFS, UKMO, JMA), поступающие в Гидрометцентр России через сеть международного обмена гидрометеорологической информацией.

В третьей части главы обсуждаются предпосылки и перспективы использования предлагаемых в диссертации моделей МО. Обосновывается использование именно квазилинейных методов и пути дальнейшего развития применений нейронных сетей в геофизике.

В четвертой части приведены технические подробности: подробности программной реализации; количество затрачиваемых вычислительных ресурсов и расписание работы в предлагаемых моделях МО в оперативном режиме. При тестировании производительности при использовании GPU ускорение по сравнению с CPU составило от 1.6 (для моделей МО из главы 3) до 115 раз (для моделей МО из главы 4). Применение моделей из глав 2 и 3 возможно в оперативном режиме и при использовании CPU и при использовании GPU, а при интерполяции из главы 4 – только на GPU.

ГЛАВА 2. Систематическая коррекция и комплексификация прогнозов

Методы систематической коррекции (первая часть главы) и комплексификации (вторая часть главы) прогнозов погоды используют идею оценки статистических характеристик погрешностей прогноза в точке рассматриваемой метеорологической станции по последним известным погрешностям, рассматриваемой (рассматриваемых) численных моделей прогноза погоды (ЧМПП).

Техника систематической коррекции прогнозов погоды основана на оценке систематического смещения модели [167], [239]. В отечественной практике она широко применяется к среднесрочным прогнозам погоды в ФГБУ «СибНИГМИ» [50], [49], [66].

Во второй части главы описывается развитие системы комплексного прогноза в Гидрометцентре России [15], [10], [11], [12], [13]. Идея статистической комплексификации прогнозов различных ЧМПП основана на оценках дисперсии и систематического смещения рассматриваемых ЧМПП. В отечественной практике комплексификация находила применение и в краткосрочном и в среднесрочном прогнозе [15], [10] и в долгосрочных [86], [32] прогнозах погоды. Северо-Евразийский климатический центр (СЕАКЦ) [76] применяет комплексификацию к долгосрочным прогнозам погоды [80].

Применение систематической коррекции и комплексификации после резкой смены погоды может ухудшить прогноз, поэтому в данной работе при вычислении систематической коррекции будет учитываться то, насколько корректируемый прогноз похож на прогнозы с более ранним начальным сроком, погрешности которых известны. В приведенных выше работах это не учитывалось при систематической коррекции и комплексификации.

Отметим работу [178], в которой предложена комплексификация прогнозов осадков с помощью нейронной сети.

2.1. Систематическая коррекция

Для нахождения систематических поправок в точке рассматриваемой синоптической станции этим методом используем период обучения, включающий последние t_l суток. Обозначим R – оператор вычисления скользящего среднего от произвольной функции $X(t, \tau)$ дискретного переменного с некоторыми положительными весами $w_X(\delta t, t, \tau) > 0$ за последние t_l суток:

$$R\{X(t, \tau), w_X(\delta t, t, \tau)\} = \frac{\sum_{\delta t=\lfloor \tau \rfloor}^{\lfloor \tau \rfloor + t_l} w_X(\delta t, t, \tau) X(t - \delta t, \tau)}{\sum_{\delta t=\lfloor \tau \rfloor}^{\lfloor \tau \rfloor + t_l} w_X(\delta t, t, \tau)}, \quad (2.1)$$

где $\lfloor \tau \rfloor$ – округление заблаговременности вверх до целого количества суток. Заметим, что в (2.1) весовая функция $w_X(\delta t, t, \tau)$ определена с точностью до положительной мультипликативной константы $C_X(t, \tau)$.

С помощью оператора R независимо для каждого из метеопараметров $X = TE, T, Td, P_0, U, V, |\vec{U}|, G$ для каждой станции и для каждого t и τ будем оценивать смещение прогнозов $X_0(t, \tau)$ относительно измерений $X_{fact}(t + \tau)$, то есть вычислять $R\{BIAS_X(t, \tau)\}$, где $BIAS_X$ – погрешность прогноза

$$BIAS_X(t, \tau) = X_{fact}(t + \tau) - X_0(t, \tau).$$

Поскольку информация о результатах численного моделирования экстремальных температур TE была недоступна (таблица 1.6), то при прогнозировании экстремальной температуры в качестве корректируемого прогноза использовалась «срочная» температура $TE_0(t, \tau) = T_0(t, \tau)$. При коррекции четырех параметров, характеризующих скорость ветра, весовые функции совпадали: $w_U = w_V = w_{|\vec{U}|} = w_G$.

Поле $X_{C0} = X_0(t, \tau) + R\{BIAS_X(t, \tau)\}$ является оценкой для поля $X_{fact}(t, \tau)$, однако некоторые из полей X_{C0} , могут быть не согласованы между собой, то есть описывать физически невозможное состояние системы, например оценка точки росы может быть выше оценки температуры $Td_{C0} > T_{C0}$. Поэтому, следуя параграфу 1.1.2, применим корректировки по формулам (1.3) и будем называть систематически поправленными прогнозы приземных метеорологических параметров, вычисленные по формулам:

$$X_C(t, \tau) = adapt(X_0(t, \tau) + R\{BIAS_X(t, \tau)\}), \quad (2.2)$$

где $adapt$ – оператор согласования (1.3).

Поскольку в коде КН-01 SYNOP [84] информация со станций о наблюдавшихся порывах ветра передается не всегда, то при коррекции прогнозов порывов ветра в случае отсутствия достаточного количества информации (когда число слагаемых в сумме (2.1) менее 70% от t_l) о

наблюдениях за порывами ветра на последние t_l дней, для расчета поправок к скорости порывов ветра использовалась формула:

$$G_{C0} = G_0(t, \tau) + \gamma \cdot R \left\{ \text{BIAS}_{|\bar{U}|}(t, \tau) \right\},$$

где $\gamma > 1$ – дополнительный оптимизируемый в силу минимизации функции потерь G_{loss} (параграф 1.1.2) параметр. По результатам проведенной оптимизации параметр γ оказался равен 2.57.

Обозначим ΔRMS – величина, на которую уменьшается среднеквадратичная погрешность прогноза при коррекции (2.1), то есть:

$$\Delta RMS = \sqrt{\frac{1}{N} \sum_{i=1}^N (X_0 - X_{fact})^2} - \sqrt{\frac{1}{N} \sum_{i=1}^N (X_C - X_{fact})^2}.$$

2.1.1. Оптимизация параметров классической систематической коррекции

Величина ΔRMS , в зависимости от длины периода обучения t_l при постоянной весовой функции $w \equiv 1$ показана на рисунке 2.1. Видно, что при «грубом» выборе весовой функции $w \equiv 1$ наименьшую среднеквадратическую ошибку RMS имеют прогнозы, прошедшие систематическую коррекцию с периодом обучения t_l от 10 до 25 суток в зависимости от корректируемого метеорологического параметра.

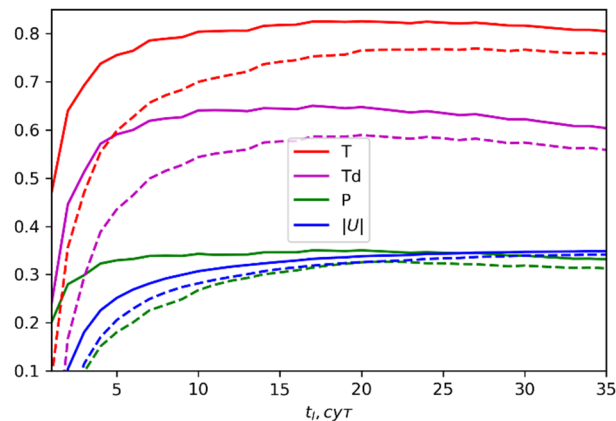


Рисунок 2.1 – Изменение среднеквадратичной погрешности прогнозов различных метеорологических параметров из архива Ru13-2018 после применения систематической коррекции (2.1) в зависимости от длины периода обучения при постоянной весовой функции

$w \equiv 1$ в зависимости от периода обучения t_l :

сплошные линии – заблаговременность 24 ч, пунктир – 48 ч

Далее в данной работе длина периода обучения была равна $t_l = 35$ суток, а формулы (2.1) применялась только если в (2.1) были известны как минимум 25 слагаемых, то есть за последние t_l суток отсутствовало не более 10 пар прогноз-факт.

На качество поправленного прогноза $X_C(t, \tau)$ помимо длины периода обучения t_l , влияет используемая весовая функция $w_X(\delta t, t, \tau)$. Часто используют так называемое *экспоненциальное сглаживание*:

$$w_X(\delta t, t, \tau) = \exp(-\lambda \delta t), \quad (2.3)$$

где λ – коэффициент «забывания истории», который будем подбирать экспериментально. В случае экспоненциального сглаживания использование увеличение длины периода обучения при t_l сверх $4/\lambda$ практически не влияет на поправленный прогноз, так как соответствующие веса слишком малы ($e^{-4} \approx 0.018$). Как правило, оптимальный коэффициент λ близок к обратной величине характерного временного периода для рассматриваемого процесса, в данном случае следует ожидать, что она окажется близкой к обратной величине так называемого *естественного синоптического периода* (типичной продолжительностью около 6 суток).

На рисунке 2.2 приведена экспериментально найденная зависимость уменьшения среднеквадратичной погрешности прогноза в зависимости от параметра λ . Проведенные эксперименты показывают, что оптимальным значением (среднеквадратичная погрешности прогноза уменьшилась сильнее всего) является $\lambda_0 \approx 0.13 \text{сут}^{-1}$. В дальнейшем прогнозы, поправленные с помощью экспоненциального сглаживания (2.3) с $\lambda = \lambda_0$, будем называть экспоненциально скорректированными (ЭК).

Рассмотрим некоторые альтернативные способы вычисления весовой функции $w_X(\delta t, t, \tau)$. Разумно предположить, что, например, так как после прохождения холодного атмосферного фронта синоптическая станция уже будет находиться в холодной воздушной массе, то оценки смещения, рассчитанные по данным, измеренным в теплой воздушной массе, будут несостоятельными.

В работах диссертанта [3] и [26] оценивались корреляционные функции в зависимости от того находятся ли две рассматриваемые точки по одну сторону от атмосферного фронта или по разные. Было показано, что если точки разделены атмосферным фронтом, то значения корреляционных функции при том же расстоянии между точками заметно меньше (рисунок 2.3). Поскольку это различие весьма существенно, то можно предположить, что и ковариационные функции погрешностей ЧМПП, а значит и оптимальная весовая функция $w_X(\delta t, t, \tau)$, должны

быть существенно меньше в случае, когда между моментами времени $t - \delta t + \tau$ и $t + \tau$ через рассматриваемую точку проходил атмосферный фронт.

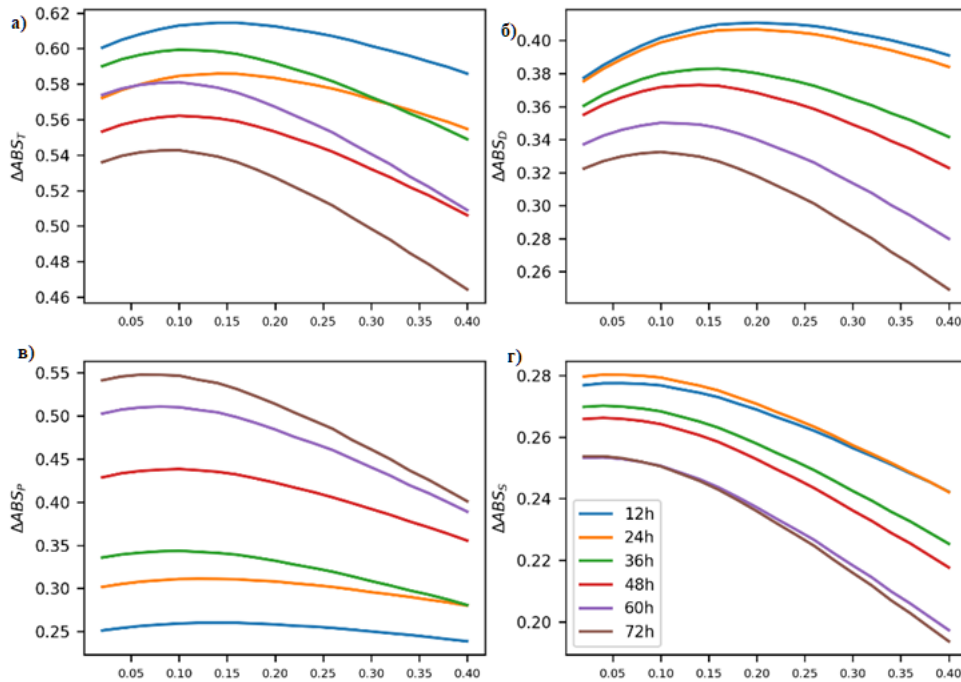


Рисунок 2.2 – Изменение средней абсолютной погрешности прогнозов различных приземных метеорологических параметров различных заблаговременностей из архива Ru13-2018 после применения экспоненциальной систематической коррекции (2.3) в зависимости параметра λ (сутки⁻¹): а) прогноз температуры воздуха; б) прогноз точки росы; в) прогноз давления, приведенного к уровню моря; г) прогноз модуля скорости ветра

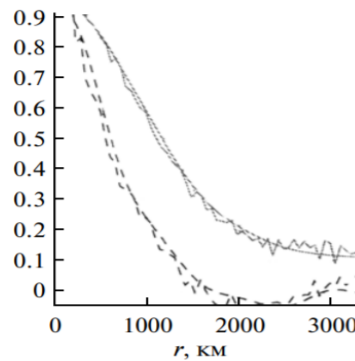


Рисунок 2.3 – Автокорреляционные функции температуры воздуха: для точек, разделенных атмосферным фронтом (пунктир), для точек, не разделенных атмосферным фронтом (сплошные линии) в зависимости от расстояния

Предположим, что оптимальные величины весов $w_X(\delta t, t, \tau) > 0$, используемые при вычислении поправленного прогноза $X_C(t, \tau)$, зависят от разности между текущим прогнозом от срока t и прогнозом от срока $t + \delta t$, вес погрешности которого вычисляется:

$$w_X(\delta t, t, \tau) = w_X(\delta t, |\Delta_{\delta t} X_0|),$$

где оператор $\Delta_{\delta t}$ вычисляет изменение за δt дней:

$$\Delta_{\delta t} X_0(t, \tau) = X_0(t, \tau) - X_0(t - \delta t, \tau).$$

Если $w_X(\delta t, |\Delta_{\delta t} X_0|)$ – убывающая (с ростом обеих переменных) функция, то формулу (2.1) можно рассматривать как частный случай взвешенного метода k ближайших соседей (k nearest neighbors, [227]), где $k = t_l$.

Рассмотрим следующее параметрическое семейство убывающих весовых функций:

$$w_X(\delta t, |\Delta_{\delta t} X_0|) = \frac{\exp(-\lambda \delta t)}{1 + \beta |\Delta_{\delta t} X_0|^2}, \quad (2.4)$$

где λ и β – некоторые положительные параметры, которые будем оптимизировать.

Найденные оптимальные параметры для различных метеорологических параметров приведены в таблице 2.1. Видно, что оптимальные параметры весовой функции вида (2.4) для систематической коррекции температуры и точки росы близки, а оптимальная весовая функция для систематической коррекции скорости ветра убывает с увеличением δt на порядок медленнее. На рисунке 2.4 приведены изолинии оптимальной весовой функции w_T вида (2.4) для систематической коррекции прогнозов температуры. Видно, что на оптимальную весовую функцию существенно влияет изменение прогнозируемой величины $\Delta_{\delta t} X_0$, а не только сдвиг по времени δt . В дальнейшем систематическую коррекцию с весовыми функциями (2.4) будем называть оптимизированной систематической коррекцией (далее ОСК). Результаты сравнения систематической коррекции в реализациях ЭК и ОСК представлены на рисунке 2.9.

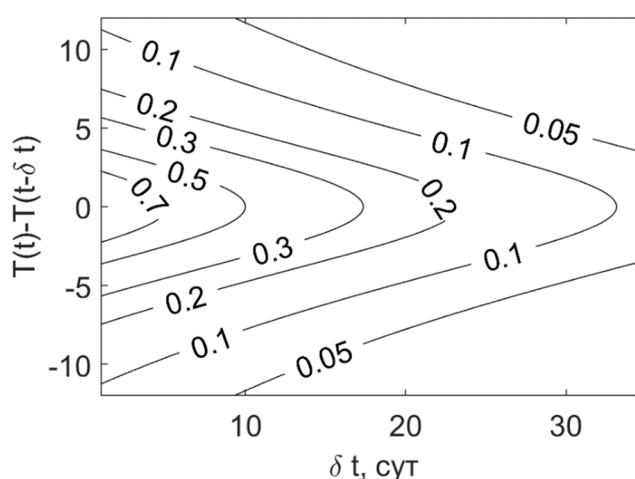


Рисунок 2.4 – Изолинии весовой функции w_T вида (2.4), оптимальной для систематической коррекции прогнозов температуры

Таблица 2.1 – Оптимальные параметры весовой функции w_X вида (2.4), оптимальной для систематической коррекции прогнозов различных метеорологических параметров

Параметр X	λ , сут ⁻¹	β
T	0.0694	0.066 К ⁻²
Td	0.0692	0.076 К ⁻²
P_0	0.0360	0.052 гПа ⁻²
$ \vec{U} $	0.0052	0.252 с ² /м ²

2.1.2. Систематическая коррекция как специальный слой нейронной сети

Перейдем к более общему случаю систематической коррекции. Рассмотрим положительную весовую функцию w_X , зависящую также от заблаговременности и конкретных прогностических значений $\bar{\Psi}_0(t, \tau)$, в том числе и метеорологических параметров, отличных от корректируемого $X_0(t, \tau)$:

$$\begin{aligned} w_X(\delta t, t, \tau) &= \exp(\omega_X(\vec{Y})) = \\ &= \exp\{\omega_X(\delta t, \tau, \varphi(t + \tau), \bar{\Psi}_0(t - \delta t, \tau), \bar{\Psi}_0(t, \tau), \bar{\Psi}_{fact}(t - \delta t, \tau))\}. \end{aligned} \quad (2.5)$$

Далее функцию $\omega_X = \ln w_X$ будем называть логарифмом весовой функции (ЛВФ). Разумно предположить, чем более близки прогнозы $\bar{\Psi}_0(t - \delta t, \tau)$ и $\bar{\Psi}_0(t, \tau)$, тем более близки погрешности текущего $X_0(t, \tau)$ и предшествующего прогноза $X_0(t - \delta t, \tau)$. Тогда больше и оптимальная ЛВФ ω_X . Однако, без дополнительных предположений определить конкретный вид ЛВФ ω_X не представляется возможным.

Заметим, что оптимальная ЛВФ ω_X определена с точностью до аддитивной константы $C_X(t, \tau)$, а значит, если последний слой ($k = K$) нейронной сети ω_X – полносвязный (параграф 1.1.3, формула (1.23)), то можно без ограничения общности положить смещения в последнем слое равными нулю $b_K \equiv 0$.

Для оценки весовой функции w_X можно было бы учитывать взаимные корреляции погрешностями с различными δt . Для этого нам нужно оценивать корреляционную функцию (КФ) ошибок прогноза $BIAS_X(t - \delta t, \tau)$ в зависимости от нескольких предикторов, объединенных в вектор $\vec{Y}$, как, например, в (2.5):

$$\text{cov}(BIAS_X(t, \tau), BIAS_X(t - \delta t, \tau)) \approx C(\vec{Y}).$$

В список предикторов $\vec{Y}$ может входить любые известные в момент составления прогноза величины (используемые предикторы описаны в параграфе 2.1.3). При этом нужно обеспечить положительную определенность КФ $C(\vec{Y})$ [38], [59]. Сложность задачи оценки положительно определенной КФ $C(\vec{Y})$ быстро растет с ростом количества аргументов (размерности вектора $\vec{Y}$). Используя оценку этой КФ, можно затем вычислить оптимальные веса w_X например, методом оптимальной интерполяции, [35], глава 4. Однако, в данном случае погрешности $BIAS_X$ экстраполируются по времени, а для применения оптимальной интерполяции желательно, чтобы влияющие точки окружали со всех сторон точку, в которой оценивается прогноз [36], [41]. В случае экстраполяции метод оптимальной интерполяции часто приводит к отрицательным значениям некоторых весов w_X , что увеличивает погрешности поправленного прогноза X_C . Таким образом, учёт взаимных корреляций погрешностей с различными сдвигами начального срока δt является нетривиальной задачей.

Прямое вычисление ЛВФ $\omega_X(\vec{Y})$ как нейронной сети, принимающей на вход вектор $\vec{Y}$ не требует обеспечения положительной определенности. Кроме того, сложность расчета ЛВФ ω_X помощью нейронных сетей растет с ростом размерности вектора $\vec{Y}$ лишь линейно.

Будем аппроксимировать ЛВФ ω_X некоторой нейронной сетью. Тогда для вычисления одного прогноза (2.1) с использованием нейронной сети ω_X необходимо t_l раз рассчитать веса $w_X(\delta t, t, \tau)$, а затем для расчета поправленного прогноза $X_C(t, \tau)$ применить формулы (2.1), (2.2). Далее метод систематической коррекции, оценивающий ЛВФ некоторой нейронной сетью, будем называть квазилинейной систематической коррекцией (КСК).

Если ЛВФ $\omega_X(\delta t, t, \tau)$ зависит от погрешности $BIAS_X(t - \delta t, \tau)$, но поправка (2.1) будет нелинейно зависеть от погрешностей $BIAS_X(t - \delta t, \tau)$.

Чтобы обучить нейронную сеть ω_X достаточно (зная градиент $\nabla_{X_C} e$ функции потерь e), уметь вычислять (формула (1.22), параграфы 1.1.5 и 1.1.7) градиент ошибки $\nabla_{\omega_X} e$. Градиент функции потерь $\nabla_{\omega_p} e$ можно вычислить исходя из правила дифференцирования сложной функции и формул (2.1), (2.2), (2.5):

$$\nabla_{\omega_P(\delta t, t, \tau)} e(P_{fact}(t+\tau), P_C(t, \tau)) = \frac{w_P(\delta t, t, \tau)}{\sum_{\delta s=\lfloor \tau \rfloor}^{t_l+\lfloor \tau \rfloor} w_P(\delta s, t, \tau)} \nabla_{P_C(t, \tau)} e(P_{fact}(t+\tau), P_C(t, \tau)) \cdot [BIAS_P(t-\delta t, \tau) - R\{BIAS_P(t, \tau)\}]. \quad (2.6)$$

Заметим, что вычисление градиента по формуле (2.6) весьма экономично: требуется всего 3 умножения и 1 вычитание (знаменатель дроби уже вычислялся ранее в (2.1)).

Из-за введения дополнительных поправок в систематическую коррекцию прогнозов $T, Td, \vec{U}$ (формулы (1.3) и (2.2)) формулы для вычисления градиентов $\nabla_{\omega_T} e$ и $\nabla_{\omega_U} e$ отличаются от (2.6) и вычисляются по формулам:

$$\begin{aligned} \nabla_{\omega_U(\delta t, t, \tau)} e(\vec{U}_{fact}(t+\tau), \vec{U}_C(t, \tau)) &= \frac{w_U(\delta t, t, \tau)}{\sum_{\delta s=\lfloor \tau \rfloor}^{t_l+\lfloor \tau \rfloor} w_U(\delta s, t, \tau)} \frac{\max\left(0, |\vec{U}|_0(t, \tau) + R\{BIAS_{|\vec{U}|}(t, \tau)\}\right)}{S_C(t, \tau)^3} \\ &\left\{ \nabla_{U_C(t, \tau)} e(\vec{U}_{fact}(t+\tau), \vec{U}_C(t, \tau)) \cdot V_C(t, \tau) \cdot [BIAS_U(t-\delta t, \tau) - R\{BIAS_U(t, \tau)\}] - \right. \\ &\left. - \nabla_{V_C(t, \tau)} e(\vec{U}_{fact}(t+\tau), \vec{U}_C(t, \tau)) \cdot U_C(t, \tau) \cdot [BIAS_V(t-\delta t, \tau) - R\{BIAS_V(t, \tau)\}] \right\}, \quad (2.7) \\ \nabla_{\omega_T(\delta t, t, \tau)} e(T_{fact}(t+\tau), T_C(t, \tau), Td_{fact}(t+\tau), Td_C(t, \tau)) &= \frac{w_T(\delta t, t, \tau)}{\sum_{\delta s=\lfloor \tau \rfloor}^{t_l+\lfloor \tau \rfloor} w_T(\delta s, t, \tau)} \left\{ \left(\nabla_{T_C(t, \tau)} e + I_{Td>T} \nabla_{Td_C(t, \tau)} e \right) \cdot \right. \\ &\cdot [BIAS_T(t-\delta t, \tau) - R\{BIAS_T(t, \tau)\}] + (1 - I_{Td>T}) \nabla_{Td_C(t, \tau)} e \cdot [BIAS_{Td}(t-\delta t, \tau) - R\{BIAS_{Td}(t, \tau)\}] \left. \right\}, \quad (2.8) \end{aligned}$$

где $I_{Td>T}$ – индикатор множества $Td_0(t, \tau) + R\{BIAS_{Td}(t, \tau)\} > T_C(t, \tau)$.

Формулы (2.6), (2.7), (2.8) позволяют обучать ЛВФ $\omega_P, \omega_U, \omega_T$, а также рассматривать систематическую коррекцию (2.1) с весовыми функциями (2.5) как специальный слой как DAG-сети (параграф 1.1.4), который в дальнейшем будем называть слоем систематической коррекции.

2.1.3. Аргументы весовой функции

Ввиду формул расчета градиента (1.24) и (1.25) для обучения нейронной сети ω_X , которая оценивает логарифм весовой функции (2.5), достаточно уметь вычислять градиент ∇_{ω_X} согласно формулам (2.6), (2.7), (2.8) и использовать метод градиентного спуска.

В качестве аргументов ω_T диссертантом были рассмотрены всевозможные подмножества (7 вариантов) следующего множества аргументов:

1. δt – сдвиг по времени;
2. $\Delta_{\delta t} T_0(t, \tau)$ – изменение температуры за δt дней;
3. $BIAS_T(t - \delta t, \tau) = T_{fact}(t - \delta t, \tau) - T_0(t - \delta t, \tau)$ – погрешность прогноза, данного ровно δt дней назад.

Результаты тестирования КСК на архиве RuBENA-2019 (таблица 1.3) с этими 7 наборами аргументов представлены на рисунке 2.5, где за 0 принята погрешность в случае, если ЛВФ ω_X зависит только от δt . Видно, что наиболее существенным из рассмотренных параметров является сдвиг по времени δt . Затем, добавление к δt аргумента $\Delta_{\delta t} T_0(t, \tau)$ заметно улучшает качество прогноза, а добавление аргумента $BIAS_T$, делающего нашу систематическую коррекцию квазилинейной, эффективно только если при расчете ЛВФ ω_X учитывается $\Delta_{\delta t} T_0(t, \tau)$.

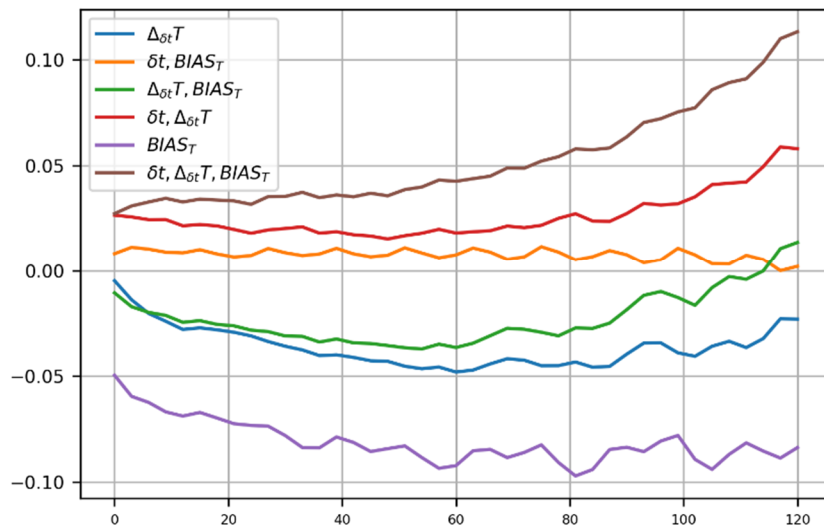


Рисунок 2.5 – Разность (в зависимости от заблаговременности прогноза) между средней абсолютной погрешностью прогноза температуры, рассчитанного КСК при ω_X , зависящем только от δt , и при ω_X , зависящем от других наборов параметров. Использовался тестовый набор данных RuBENA-2019

При проведении дальнейших численных экспериментов были рассмотрены перечисленные ниже наборы входных аргументов нейронной сети ω_X .

При коррекции прогнозов давления, приведенного к уровню моря ($X = P_0$), набор аргументов включал следующие 3 величины:

1. δt – сдвиг по времени;
2. $\Delta_{\delta t} P_0(t, \tau)$ – изменение давления за δt дней;
3. $BIAS_P(t - \delta t, \tau) = P_{fact}(t - \delta t, \tau) - P_0(t - \delta t, \tau)$ – погрешность прогноза, данного δt дней назад.

При коррекции прогнозов температуры воздуха и точки росы ($X = T, Td$) использовались разности $\Delta_{\delta t}$ и $BIAS$ для обоих метеорологических параметров, таким образом, набор аргументов включал 5 величин:

1. δt – сдвиг по времени;
2. $\Delta_{\delta t} T(t, \tau), \Delta_{\delta t} Td(t, \tau)$ – изменение температуры воздуха и точки росы за δt дней;
3. $BIAS_T(t - \delta t, \tau), BIAS_{Td}(t - \delta t, \tau)$ – погрешности прогноза, данного δt дней назад.

При коррекции скорости ветра ($X = \vec{U}$) набор аргументов включал следующие 5 величин:

1. δt – сдвиг по времени;
2. $\Delta_{\delta t} S(t, \tau) = \sqrt{U(t, \tau)^2 + V(t, \tau)^2} - \sqrt{U(t - \delta t, \tau)^2 + V(t - \delta t, \tau)^2}$ – изменение модуля скорости ветра за δt дней;
3. $|\Delta_{\delta t} \vec{U}(t, \tau)| = \sqrt{\Delta_{\delta t} U(t, \tau)^2 + \Delta_{\delta t} V(t, \tau)^2}$ – модуль изменения скорости ветра за δt дней;
4. $BIAS_S(t - \delta t, \tau) = \sqrt{U_{fact}(t - \delta t, \tau)^2 + V_{fact}(t - \delta t, \tau)^2} - \sqrt{U(t - \delta t, \tau)^2 + V(t - \delta t, \tau)^2}$ – погрешность прогноза модуля скорости ветра, данного δt дней назад;
5. $|BIAS_{\vec{U}}(t - \delta t, \tau)| = \sqrt{BIAS_U(t - \delta t, \tau)^2 + BIAS_V(t - \delta t, \tau)^2}$ – модуль погрешности прогноза вектора скорости ветра, данного δt дней назад.

Далее систематическую коррекцию, использующую нейронную сеть, принимающую на вход указанные наборы из 5 предикторов (3 предикторов для коррекции давления, приведенного к уровню моря) будем называть квазилинейной систематической коррекцией (КСК).

Заметим, что, так как вычисляемый вес $w_X = \exp(\omega_X)$ в (2.1) умножается на слагаемое $BIAS_X(t - \delta t, \tau)$, то величина $BIAS_X(t - \delta t, \tau)$ известна нам тогда и только тогда, когда применяется нейронная сеть ω_X . Поправки (2.1), рассчитываемые КСК, нелинейно зависят от погрешностей корректируемого прогноза $BIAS_X(t - \delta t, \tau)$.

Согласно (2.1), при вычислении одного прогностического значения используются t_l значений весов $w(\delta t, t, \tau)$, то есть t_l раз применяется нейронная сеть ω_X с n_θ параметрами.

Таким образом, вычислительная сложность для различных вариантов систематической коррекции составляет: $O(t_l n_\theta)$ для КСК; $O(t_l)$ для ОСК; $O(1)$ для ЭК. По этой причине оптимизация гиперпараметров заключалась лишь в проведении экспериментов с двух и трехслойными MLP с функциями активации ReLU и tanh с 24, 32, 32-16, 24-12, 16-8 нейронами на скрытых слоях. В результате этих численных экспериментов была выбран трехслойный персептрон с 24 и 12 нейронами на скрытых слоях и функциями активации ReLU, поскольку MLP с 32-16 при большей вычислительной стоимости давали близкие оценки качества, а с 16-8, 24 и 32 нейронами – несколько худшие оценки при незначительном уменьшении времени, затраченного на вычисления. Нейронная сеть ω_X имела $n_\theta=408$ параметров при коррекции P_0 и $n_\theta=456$ параметров при коррекции $X = T, Td, \vec{U}$.

Результаты сравнительного тестирования КСК на архиве Ru6ENA-2019 приведены на рисунке 2.9. Обучение происходило на объединении архивов Ru13ENA-2017 и Ru13ENA-2018.

Так как КСК использует всего 3 или 5 аргументов, а вычисление всех весов с помощью нейронной сети требует $O(t_l n_\theta)$ умножений, то при применении КСК для вычисления весов ω_X эффективно использовать заранее рассчитанные многомерные таблицы значений ω_X . Альтернативный вариант ускорения расчётов: приблизить полученные веса более простой параметрической функцией 3 или 5 аргументов.

При оптимизации параметров (обучении) нейронной сети ω_X в процессе обучения помимо уменьшения скорости обучения η_s (параграф 1.1.7) увеличивалась длина периода оценки t_l . А именно в первую эпоху обучения величина t_l увеличивалась с 5 до 35 дней. Это позволило не только существенно (в ~ 1.75 раза) увеличить скорость обучения, но и достичь несколько меньших значений функционала $L(\vec{\theta})$ по сравнению с алгоритмом, когда t_l постоянен. Набор для обучения minibatch размера m формировался из временных рядов $\bar{\Psi}(t, \tau_j, \vec{x}_j)$, где $(\tau_j, \vec{x}_j), j=1\dots m$ – псевдослучайные пары заблаговременности и номера станции.

2.1.4. Результаты оптимизации весовой функции систематической коррекции

Изолинии на некоторых сечениях ЛВФ ω_T , используемой КСК для коррекции температуры воздуха и точки росы, приведены на рисунке 2.6. Изолинии ЛВФ ω_P для коррекции давления на уровне моря – на рисунке 2.7, а ЛВФ ω_U для коррекции скорости ветра – на рисунке

2.8. С целью упрощения числовых обозначений в подписях приводятся не изолинии весовой функции w_X , а совпадающие с ними изолинии её логарифма (ЛВФ) ω_X .

Поскольку оптимальная ЛВФ ω_X определена с точностью до аддитивной константы, то на рисунках за 0 взято значение найденной ω_X в точке, в которой $\delta t = 1$ сут, а все остальные аргументы равны 0. Различие в значениях ω_X равное 3 соответствует в $e^3 \approx 20$ раз отличающемуся весу в формуле (2.1).

Видно, что все найденные ЛВФ существенно зависят от погрешности прогноза $BIAS_X(t - \delta t, \tau)$, то есть рассматриваемые коррекции не являются линейными. Видно, что ω_X принимает большие значения, когда $BIAS_X$ и $\Delta_{\delta t} X$ близки к 0. Таким образом, КСК уменьшает вес при больших погрешностях $BIAS_X$ и при сильных изменениях корректируемого метеорологического параметра. Заметим, что согласно рисункам 2.6а, 2.7а, 2.8а функция ω_X принимает большие (по сравнению с другими прямыми, проходящими через 0) значения на прямой $x = y$ на соответствующем рисунке:

$$X_{fact}(t - \delta t + \tau) - X(t - \delta t, \tau) = BIAS_X(t - \delta t, \tau) = \Delta_{\delta t} X(t, \tau) = X(t, \tau) - X(t - \delta t, \tau),$$

то есть, когда корректируемое значение равно измеренному ранее: $X_{fact}(t - \delta t + \tau) = X(t, \tau)$.

Из рисунка 2.6б видно, что если в последние дни наблюдалась большая межсуточная изменчивость $\Delta_{\delta t} X$, то КСК (в отличие от ЭК) использует более старую историю.

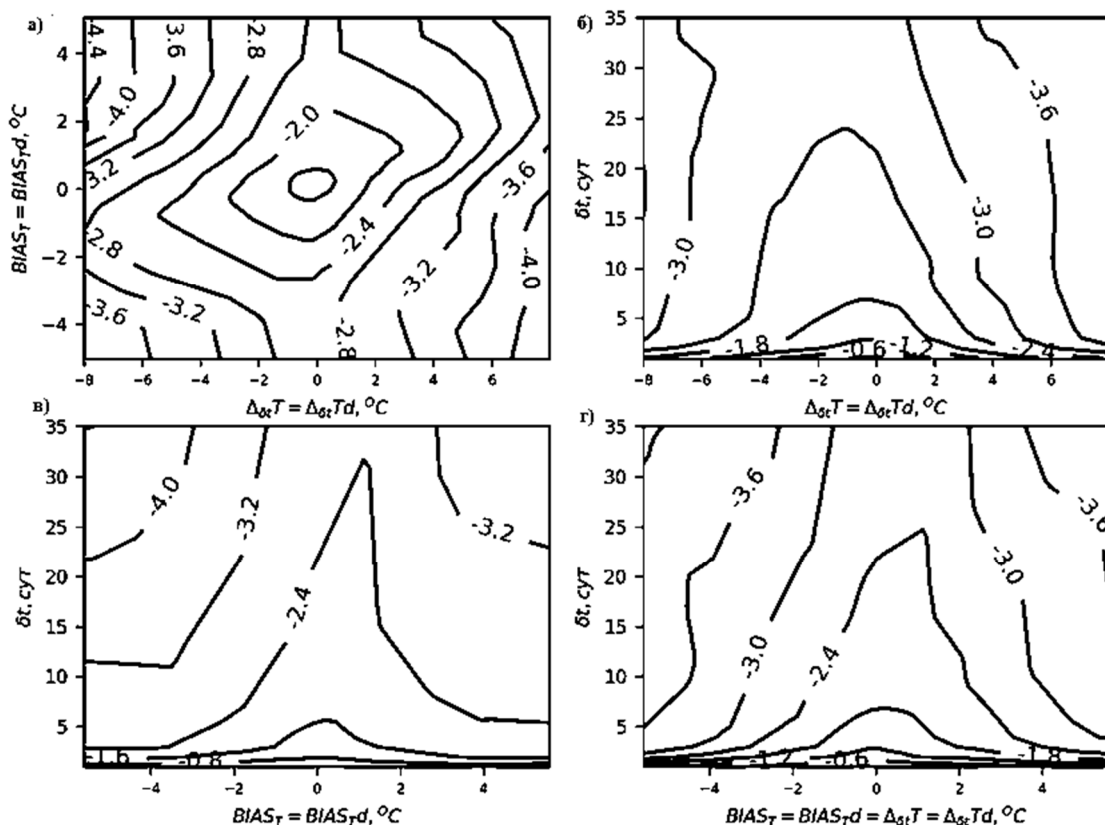


Рисунок 2.6 – Изолинии оптимальной ЛВФ ω_T на различных двумерных сечениях:

- а) $\delta t = 5 \text{ cym}$, $\Delta_{\delta t} T(t, \tau) = \Delta_{\delta t} T_d(t, \tau)$, $BIAS_T(t - \delta t, \tau) = BIAS_{Td}(t - \delta t, \tau)$;
 б) $\Delta_{\delta t} T(t, \tau) = \Delta_{\delta t} T_d(t, \tau)$, $BIAS_T(t - \delta t, \tau) = BIAS_{Td}(t - \delta t, \tau) = 0$;
 в) $BIAS_T(t - \delta t, \tau) = BIAS_{Td}(t - \delta t, \tau)$, $\Delta_{\delta t} T(t, \tau) = \Delta_{\delta t} T_d(t, \tau) = 0$;
 г) $BIAS_T(t - \delta t, \tau) = BIAS_{Td}(t - \delta t, \tau) = \Delta_{\delta t} T(t, \tau) = \Delta_{\delta t} T_d(t, \tau)$

ЛВФ ω_P для коррекции P_0 с ростом δt убывает быстрее, рисунок 2.7, чем ω_T и ω_U (рисунки 2.6, 2.8), то есть эффект памяти в поле давления проявляется существенно меньше. Это результат несколько неожиданный: поле давления гораздо более гладкое.

Из рисунка 2.8 видно, что при коррекции скорости ветра аргумент δt оказывает существенно меньшее влияние на ЛВФ ω_U , чем при коррекции других метеорологических параметров, а большее значение имеет изменение скорости ветра за δt дней. Можно сделать предположение, что основное влияние на структуру погрешностей в точке прогноза оказывает обтекание препятствий ветром определенной скорости и направления, а не эффект памяти.

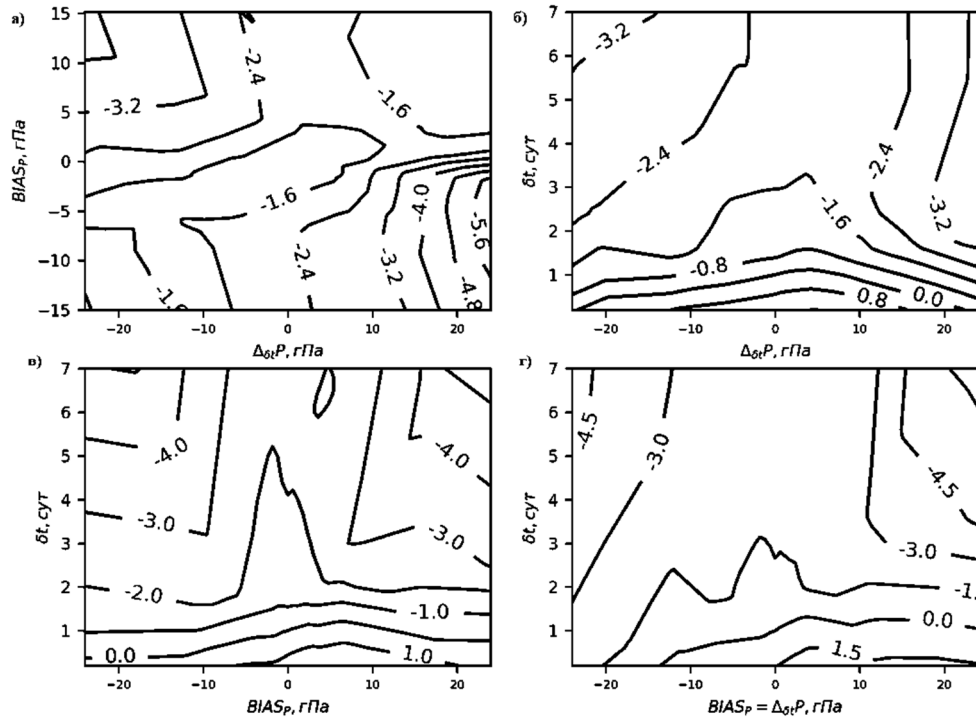


Рисунок 2.7 – Изолинии оптимальной ЛВФ ω_P на различных двумерных сечениях:

а) $\delta t = 5 \text{ сут}$; б) $BIAS_P(t - \delta t, \tau) = 0$; в) $\Delta_{\delta t} P(t, \tau) = 0$; г) $BIAS_P(t - \delta t, \tau) = \Delta_{\delta t} P(t, \tau)$

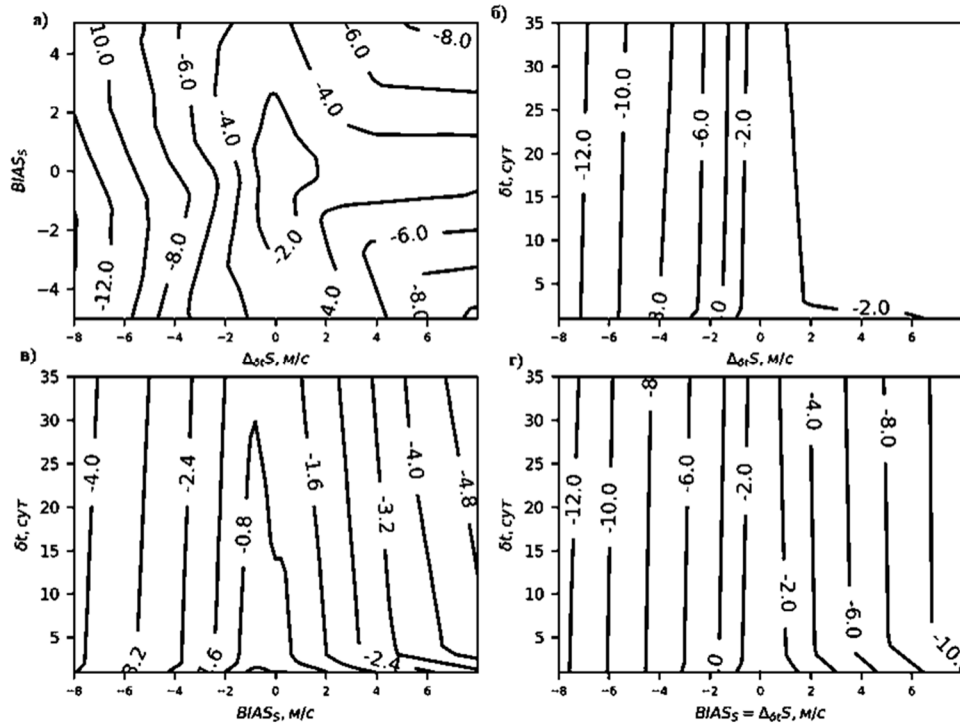


Рисунок 2.8. Изолинии оптимальной ЛВФ ω_U на различных двумерных сечениях:

а) $\delta t = 5 \text{ сут}$, $2|\Delta_{\delta t} S(t, \tau)| = |\Delta_{\delta t} \vec{U}(t, \tau)|$, $2|BIAS_S(t - \delta t, \tau)| = |BIAS_{\vec{U}}(t - \delta t, \tau)|$;

б) $2|\Delta_{\delta t} S(t, \tau)| = |\Delta_{\delta t} \vec{U}(t, \tau)|$, $BIAS_S(t - \delta t, \tau) = |BIAS_{\vec{U}}(t - \delta t, \tau)| = 0$;

в) $\Delta_{\delta t} S(t, \tau) = |\Delta_{\delta t} \vec{U}(t, \tau)| = 0$, $2|BIAS_S(t - \delta t, \tau)| = |BIAS_{\vec{U}}(t - \delta t, \tau)|$;

г) $\Delta_{\delta t} S(t, \tau) = BIAS_S(t - \delta t, \tau)$, $2|\Delta_{\delta t} S(t, \tau)| = |\Delta_{\delta t} \vec{U}(t, \tau)|$, $2|BIAS_S(t - \delta t, \tau)| = |BIAS_{\vec{U}}(t - \delta t, \tau)|$

2.1.5. Сравнение результатов систематической коррекции с различными весовыми функциями

На рисунке 2.9 показаны результаты тестирования качества систематически поправленных прогнозов с различными весовыми функциями: с постоянной $w=1$, а также с весовыми функциями (2.3), (2.4), (2.5), соответствующим ЭК, ОСК, КСК соответственно. Приведена средняя абсолютная ошибка откорректированных прогнозов на архиве Ru6ENA-08/2019-07/2020 отдельно каждого месяца в зависимости от заблаговременности. Оптимизация параметров происходила на объединении архивов Ru13ENA-2017 и Ru13ENA-2018. Видно, что выигрыш от применения систематической коррекции с $w=1$, ЭК и ОСК весьма близки, а выигрыш от применения КСК до 2 раз больше. Применение нейронной сети улучшило по сравнению с ЭК качество прогнозов на величину, соответствующую уменьшению заблаговременности на 1 сутки. Причем эффект от КСК в среднем не убывает с ростом заблаговременности, в отличие от других рассмотренных реализаций систематической коррекции.

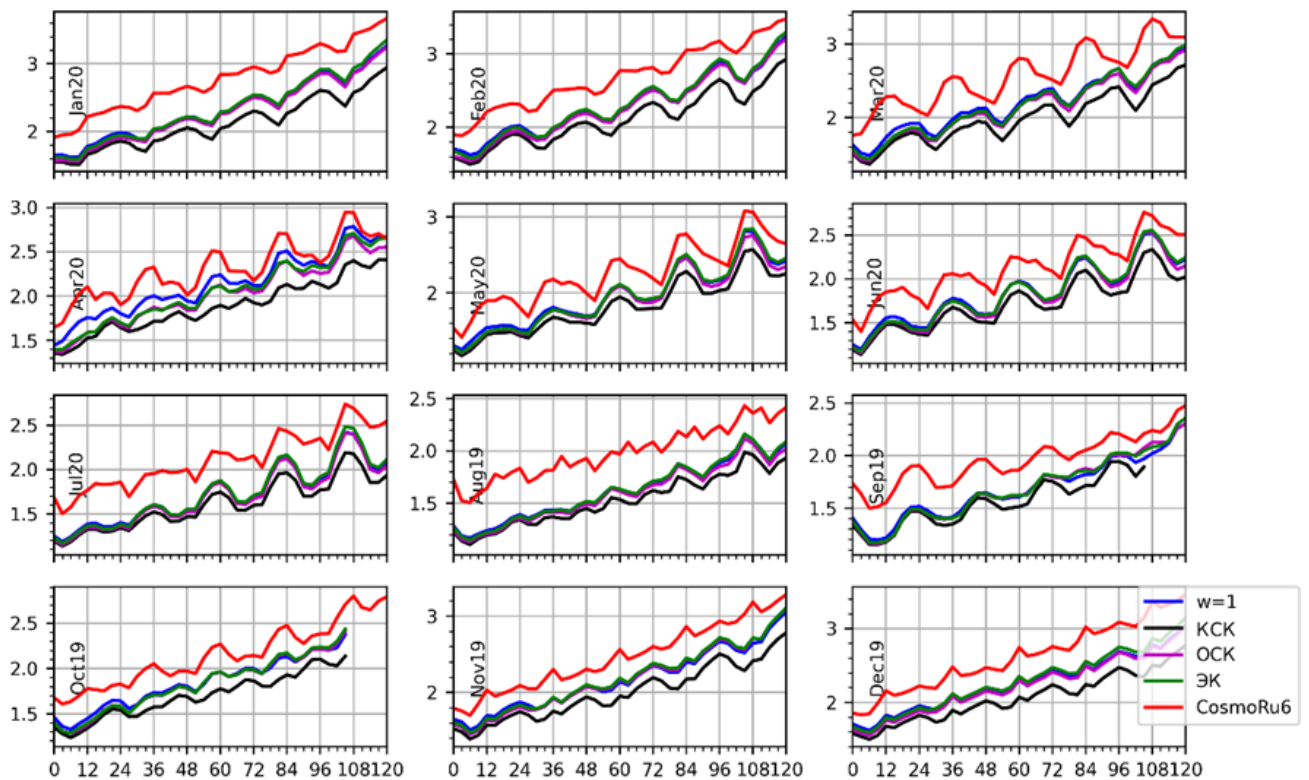


Рисунок 2.9 – Средняя абсолютная ошибка (в зависимости от заблаговременности прогноза) различным образом поправленных прогнозов из архива Ru6ENA-08/2019-07/2020 от начального срока 00:00 ВСВ отдельно для каждого месяца: с постоянной весовой функцией $w=1$ и с весовыми функциями при различной коррекции: ЭК, ОСК, КСК

2.2. Комплексный (мультимодельный) прогноз

Предположим, что известны прогнозы $X_i(t, \tau), i = 1 \dots n_{\text{mod}}$ нескольких различных ЧМПП. Обозначим $I(t, \tau) \subseteq \{1, \dots, n_{\text{mod}}\}$ – множество тех ЧМПП, прогнозы которых с заблаговременностью τ известны в момент составления прогноза от начального срока t . Список рассматриваемых ЧМПП приведен в таблице 1.6.

Поставим задачу поиска оптимальной линейной комбинации этих прогнозов в смысле минимальной среднеквадратичной погрешности:

$$\begin{aligned} X_{\text{comp0}}(t, \tau) &= b + \sum_{i \in I(t, \tau)} a_i X_i(t, \tau), \\ \sum_{j=1}^M (X_{\text{comp0}} - X_{\text{fact}})^2 &\rightarrow \min_{b, a_i} \end{aligned} \quad (2.9)$$

Предположим, что погрешности $X_{\text{fact}}(t + \tau) - X_i(t, \tau)$ различных ЧМПП распределены согласно нормальному распределению: $N(\mu_i, \sigma_i^2)$ и погрешности различных ЧМПП независимы, то есть их ковариационная матрица имеет вид:

$$E[(X_{\text{fact}} - X_i)(X_{\text{fact}} - X_j)] = \delta_{ij} \sigma_i^2.$$

Тогда решением оптимизационной задачи (2.9) следующее:

$$X_{\text{comp0}}(t, \tau, I(t, \tau)) = \frac{\sum_{i \in I(t, \tau)} \frac{X_i(t, \tau) + \mu_i(t, \tau)}{\sigma_i^2(t, \tau)}}{\sum_{i \in I(t, \tau)} \sigma_i^{-2}(t, \tau)}. \quad (2.10)$$

Параметры μ_i, σ_i будем оценивать аналогично параграфу 2.1.1, используя оператор R , формула (2.1), но для увеличения устойчивости формулы (2.10) завязим оценку используемую дисперсий на некоторую положительную константу α_τ^2 :

$$\begin{aligned} \mu_i(t, \tau) &= R\{X_{\text{fact}}(t + \tau) - X_i(t, \tau)\} \\ \sigma_i^2(t, \tau) &= \alpha_\tau^2 + R\left\{\left(X_{\text{fact}}(t + \tau) - X_i(t, \tau)\right)^2\right\} - \left(\mu_i(t, \tau)\right)^2. \end{aligned} \quad (2.11)$$

Параметр α_τ (имеющий ту же размерность, что поле X), играет роль регуляризатора в (2.10) и зависит от заблаговременности τ .

Параметр α_τ будем оптимизировать. При оптимизации начальными значениями α_τ полагались для температуры и точки росы $\alpha_\tau = 1^\circ\text{C}$, для скорости ветра $\alpha_\tau = 1 \text{ м/с}$, для давления $\alpha_\tau = 0.2 \text{ гПа}$. Оптимальные параметры α_τ показаны ниже на рисунке 2.11.

Были проведены численные эксперименты, когда параметр α_τ индивидуальны для различных ЧМПП, но получившиеся в этом случае оценки качества прогнозов X_{comp0} на независимой выборке оказались не лучше, чем в случае, когда для всех ЧМПП использовались одни и те же α_τ .

Вероятная причина этого результата в том, что при выборе различных α_τ для различных ЧМПП при комплексификации (в среднем) самая неточная ЧМПП практически не учитывается, а тем не менее существуют частные случаи, в которых в среднем более точные ЧМПП ошибаются сильнее в среднем менее точной ЧМПП. При использовании одних и тех же α_τ для всех ЧМПП веса a_i в (2.9) зависят только от ошибок i -й ЧМПП за последние t_l суток и менее точная в среднем ЧМПП может в некоторых случаях получить большой вес в комплексном прогнозе, вычисляемом по формулам (2.9), (2.11).

2.2.1. Комплексификация при недоступности части прогностических данных

Для станций, ведущих наблюдения только 4 раза в сутки, невозможно оценить значения смещения μ_i и стандартного отклонения σ_i согласно формуле (2.11) в дополнительные сроки ($t + \tau = 3, 9, 15, 21 \text{ ч ВСВ}$). Для этих станций значения μ_i и σ_i в дополнительные сроки будем вычислять линейной интерполяцией по времени:

$$\begin{aligned}\mu_i(t, \tau) &= 0.5 \left(\mu_i(t, \tau - 3\tau) + \mu_i(t, \tau + 3\tau) \right); \\ \sigma_i^2(t, \tau) &= 0.5 \left(\sigma_i^2(t, \tau - 3\tau) + \sigma_i^2(t, \tau + 3\tau) \right).\end{aligned}\tag{2.12}$$

В выражении (2.1) для оператора R участвует весовая функция $w_X(\delta t, t, \tau)$, которую будем оптимизировать аналогично первой части настоящей главы. При оценке весовой функции $w_X(\delta t, t, \tau)$ по формулам ОСК (2.4) и КСК (2.5) результат зависит от прогностических метеорологических параметров, а значит при вычислении комплексного прогноза возникает неоднозначность в определении весовой функции $w_X(\delta t, t, \tau)$: можно использовать прогнозы нескольких комплексифицируемых ЧМПП. В данной работе использовалась весовая функция, соответствующая средневзвешенному значению логарифма весовых функций ω_{X_i} , вычисленных

с использованием прогнозов различных ЧМПП. Итоговая весовая функция рассчитывается по формуле:

$$w_{X_{comp}}(\delta t, t, \tau) = \exp \left[\left(\sum_{i \in I(t, \tau)} \rho_{i\tau} \right)^{-1} \sum_{i \in I(t, \tau)} \rho_{i\tau} \omega_{X_i}(\delta t, t, \tau) \right], \quad (2.13)$$

где $\rho_{i\tau} > 0$ – веса различных ЧМПП в определении весовой функции. Оптимальные $\rho_{i\tau}$ для каждого из рассматриваемых метеорологических параметров, определены с точностью до общей (положительной) мультипликативной константы, показаны далее на рисунке 2.12.

Заметим, что формулы (2.10) и (2.13) остаются применимыми в ситуации, когда значения прогнозов некоторых из ЧМПП недоступны (например, по техническим причинам): в этом случае соответствующие слагаемые в обеих суммах опускаются.

Поскольку не для всех ЧМПП были доступны прогностические поля точки росы Td (таблица 1.6), то при расчетах $\omega_{T_{DWD}}$ и $\omega_{T_{ECMWF}}$ вместо Td_{DWD} и Td_{ECMWF} использовались прогностические значения Td_{NCEP} .

Поскольку на практике прогнозы не всех ЧМПП нам известны со всеми рассматриваемыми заблаговременностями (таблица 9.2), то при вычислении комплексного прогноза по формуле (2.10) число участвующих ЧМПП сильно меняется от заблаговременности к заблаговременности. Это может приводить к необоснованному колебанию прогностических значений $X_{comp0}(t, \tau)$ в зависимости от заблаговременности. Синтетический (сгенерированные программно) пример таких колебаний для случая двух ЧМПП приведен на рисунке 2.10.

Учитывать эту неравномерность будем, вводя дополнительную поправку:

$$\begin{aligned} X_{comp}(t, \tau) = & X_{comp0}(t, \tau, I(t, \tau)) + \\ & + \frac{1}{2} \{ X_{comp0}(t, \tau - \Delta\tau, I(t, \tau - \Delta\tau)) - X_{comp0}(t, \tau - \Delta\tau, I(t, \tau) \cap I(t, \tau - \Delta\tau)) + \\ & + X_{comp0}(t, \tau + \Delta\tau, I(t, \tau + \Delta\tau)) - X_{comp0}(t, \tau + \Delta\tau, I(t, \tau) \cap I(t, \tau + \Delta\tau)) \}, \end{aligned} \quad (2.14)$$

где $\Delta\tau$ – шаг по заблаговременности. Поправки (2.14) применялись в соответствии с наличием прогнозов определенных заблаговременностей (таблица 1.6) дважды: в первый раз с $\Delta\tau = 6ч$ при $\tau = 54, 66, \dots, 138ч$, а второй раз с $\Delta\tau = 3ч$ при $\tau = 3, 9, \dots, 141ч$.

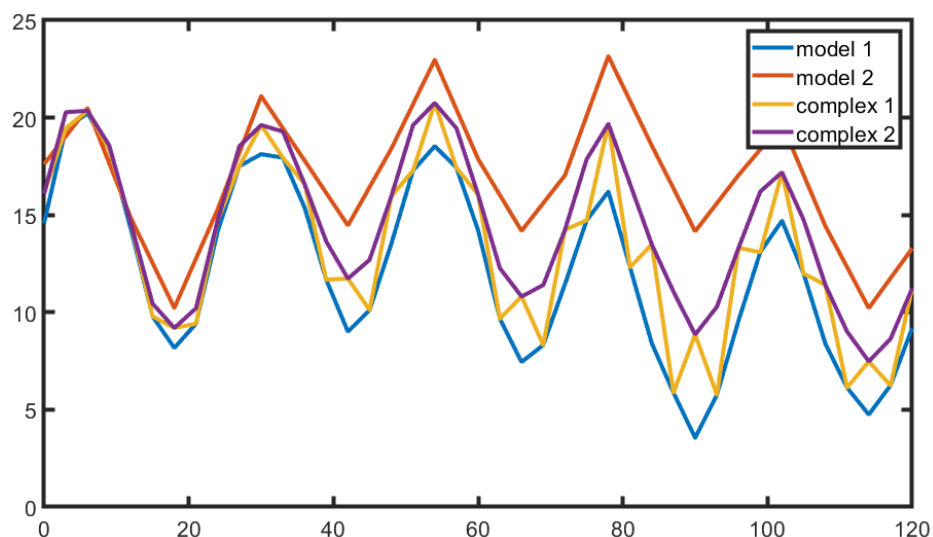


Рисунок 2.10. Синтетический пример избавления от вычислительных шумов, вызванных недоступностью прогнозов численной модели 2 в дополнительные сроки:
 complex 1 – результат комплексификации моделей 1 и 2 согласно (2.10);
 complex 2 – после применения поправок (2.14)

Как и при систематической коррекции непосредственно перед вычислением функции потерь (параграфы 1.1.1 и 1.1.2) к комплексному прогнозу X_{comp} применялся оператор согласования прогнозов различных метеорологических параметров $adapt$, формула (1.3). Полученный таким образом прогноз будем называть квазилинейным комплексным прогнозом (ККП).

2.2.2. Результаты оптимизации параметров комплексного прогноза

Оптимальные параметры регуляризации α_τ , формула (2.11), показаны на рисунке 2.11, а оптимальные веса ρ_{it} , формула (2.13) – на рисунке 2.12. Видно, что найденные параметры как функции заблаговременности τ имеют вычислительные шумы, вызванные отсутствием прогнозов некоторых ЧМПП некоторых заблаговременностей, таблица 1.6.

Использование параметра регуляризации α_τ (рисунок 2.11) в том числе уравнивает влияние различных ЧМПП на комплексный прогноз. Видно, что за исключением случая комплексификации скорости ветра U_{10m} оптимальное значение параметра регуляризации α_τ в среднем растет с ростом заблаговременности.

Отметим, что качество прогнозов используемой ЧМПП (глава 5) никак не коррелирует с соответствующими весами ρ_{it} . Например, согласно рисункам 2.12а и 2.12б модель PLAV,

качество прогнозов температуры воздуха которой как правило хуже всех остальных рассматриваемых ЧМПП, входит в ЛВФ $w_{\chi_{comp}}$, формула (2.13) с весьма существенным весом.

Отметим также, что согласно рисунку 2.12, прогнозы модели UKMO с заблаговременностью 84, 96, 108 и 120ч, которые имеют наихудшее горизонтальное разрешение (таблица 1.6), имеют наибольший вес в ЛВФ для всех прогнозов, за исключением прогноза скорости ветра. Можно предположить, что для определения синоптической ситуации (типизации физических процессов, влияющих на ошибки прогнозов) лучше использовать прогностические поля с грубым (200 км и более) горизонтальным разрешением.

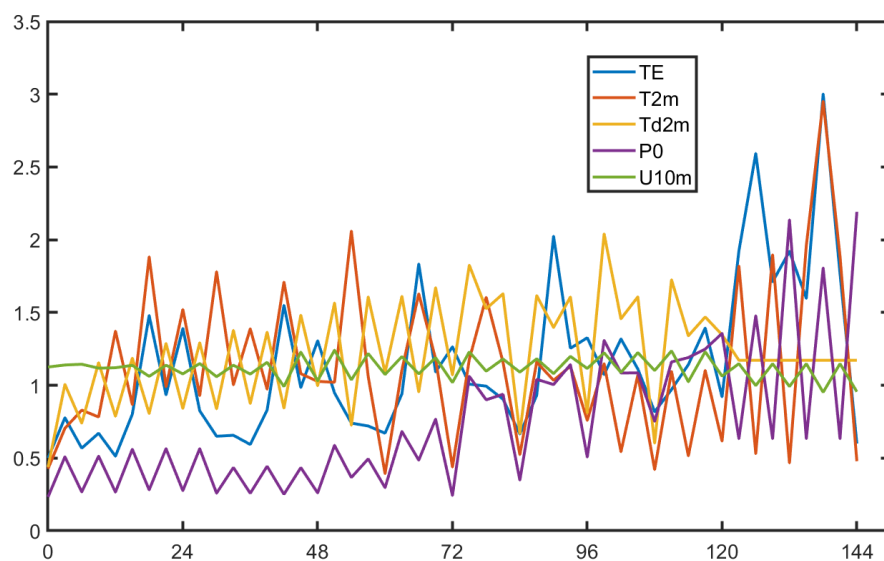
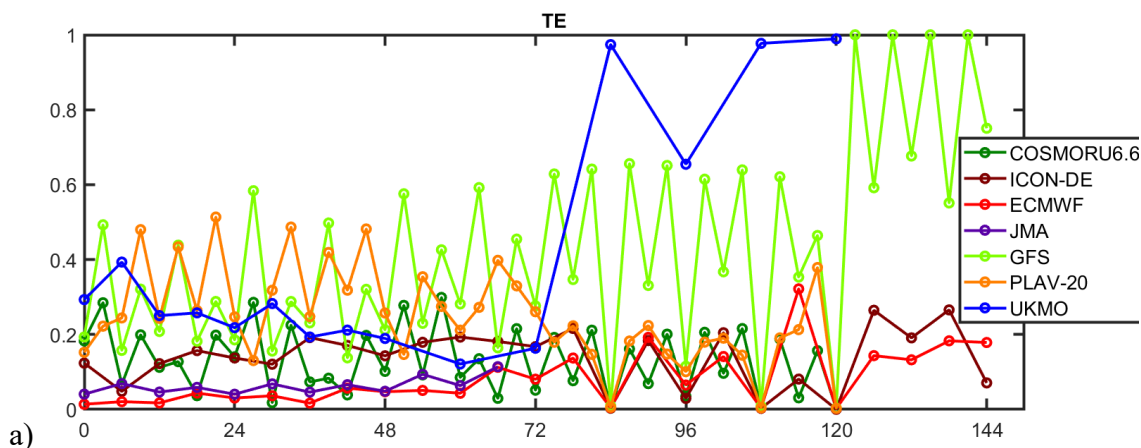
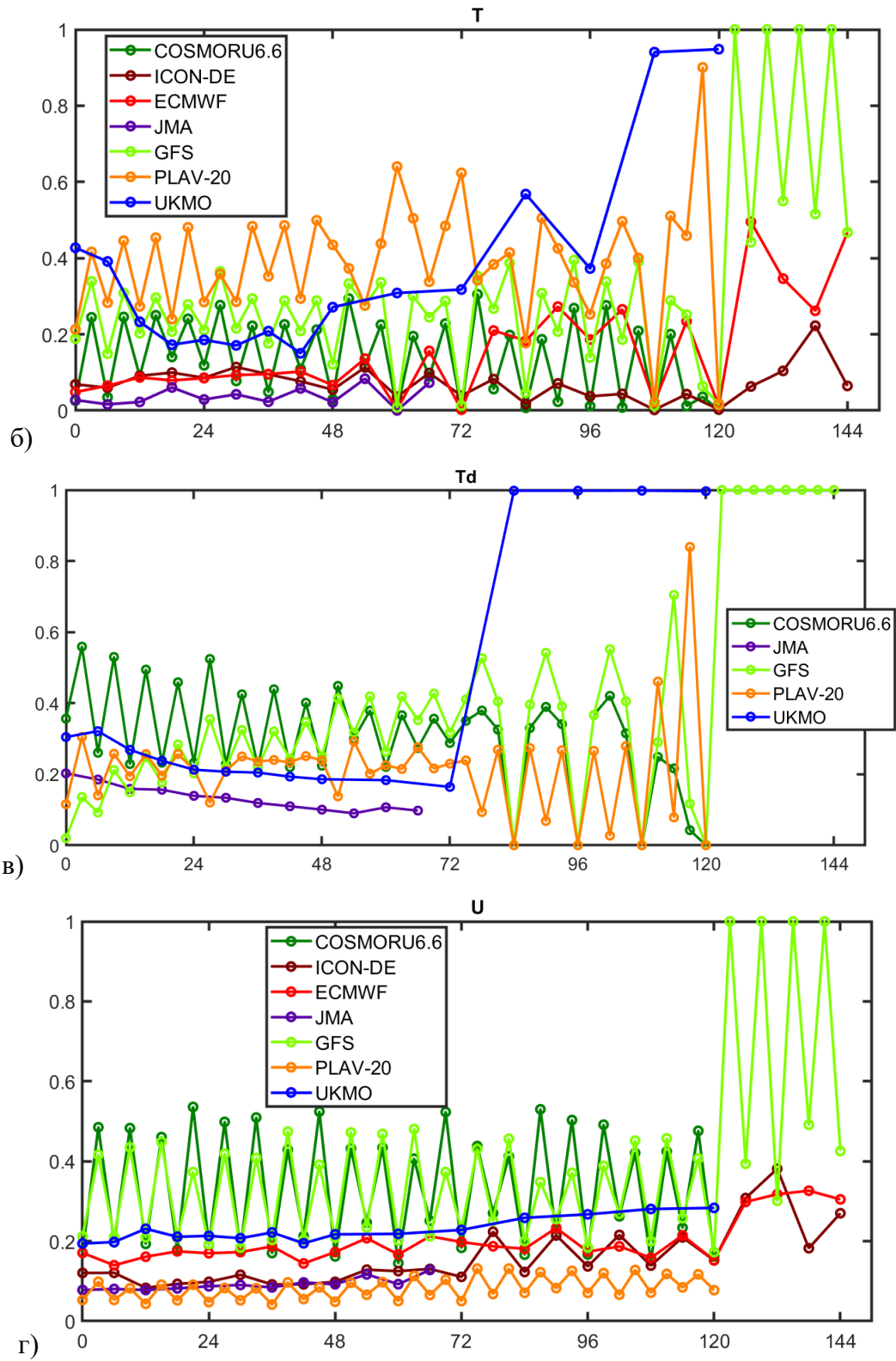


Рисунок 2.11 – Найденные оптимальные параметры регуляризации α_τ ККП, используемые в формуле (2.11) для различных приземных метеорологических параметров в зависимости от заблаговременности τ . Единицы измерения: $^{\circ}\text{C}$, гПа, м/с





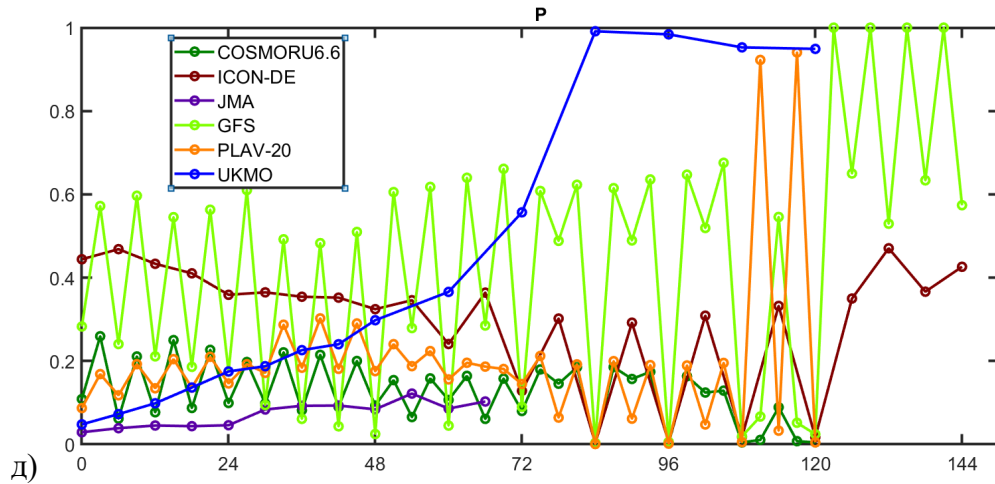


Рисунок 2.12 – Найденные оптимальные безразмерные параметры ККП: веса ρ_{it} при ЛВФ ω_{X_i} , вычисленных по различным ЧМПП, формула (2.13), в зависимости от заблаговременности τ . Для различных приземных метеорологических параметров: а) экстремальная температуры воздуха TE ; б) температура воздуха на высоте 2 м T_{2m} ; в) точка росы Td_{2m} ; г) скорость ветра $\bar{U}_{10m}$; д) давления, приведенного к уровню моря P_0

Получившиеся при оптимизации параметры α_τ и ρ_{it} ККП были приведены на рисунках 2.11 и 2.12 соответственно. Поскольку в соответствии с таблицей 1.6 были доступны прогнозы не для всех ЧМПП всех заблаговременностей, поэтому веса ρ_{it} определены не для всех i и τ . На рисунке 2.12 веса ρ_{it} нормированы таким образом, чтобы для всех заблаговременностей τ было выполнено равенство $\sum_i \rho_{it} = 1$.

Из рисунка 2.12 видно, что из рассмотренных ЧМПП наибольший вклад (наибольшее ρ_{it}) в оценку ЛВФ (2.13) дают:

- при комплексификации прогнозов температуры T_{2m} модели GFS, PLAV, UKMO;
- при комплексификации прогнозов точки росы Td_{2m} модели GFS, PLAV, COSMO-Ru6ENA;
- при комплексификации прогнозов давления на уровне моря P_0 модели GFS, ICON-DE, UKMO;
- при комплексификации прогнозов скорости ветра U_{10m} модели GFS и COSMO-Ru6ENA.

Можно сделать вывод, что указанные модели лучше типизируют синоптические ситуации, важные для прогноза указанных приземных метеорологических параметров.

2.2.3. Комплексный прогноз как специальный слой нейронной сети

Описанный в предыдущем параграфе 2.2.2 комплексный прогноз можно рассматривать как DAG-сеть (параграф 1.1.4). Это позволяет оптимизировать ЛВФ $\omega_{X_i}(\delta t, t, \tau)$ с целью оптимизации качества прогноза X_{comp} .

Предлагаемый алгоритм вычисления ККП состоит из следующих шагов:

1. вычисление оценок логарифмической весовой функции (ЛВФ) нейронной сетью для каждой из участвующих ЧМПП;
2. вычисление взвешенной оценки весовой функции $w_{X_{comp}}$ по формуле (2.13);
3. вычисление оценок параметров μ_i, σ_i согласно формулам (2.11) и (2.1) с весовой функцией $w_{X_{comp}}$, найденной на шаге 2;
4. учёт поправок (2.12) для станций, производящих измерения только 4 раза в сутки;
5. вычисление комплексификации (2.10);
6. учёт поправок (2.14), использующих прогнозы различных заблаговременностей на данной станции;
7. применение оператора согласования *adapt* по формулам (1.3).

Прогноз ККП зависит от следующих параметров $\vec{\theta}$, которые будем оптимизировать:

- а) параметры нейронной сети (сетей) ЛВФ $\omega_{X_i}(\delta t, t, \tau)$, которые могут быть как индивидуальны для различных используемых ЧМПП, так и общими для всех или некоторых из используемых ЧМПП;
- б) параметры весов различных ЧМПП ρ_{it} в ЛВФ, используемые в формуле (2.13);
- в) параметры регуляризации α_τ , используемые в формуле (2.11).

В данной работе ЛВФ $\omega_{X_i}(\delta t, t, \tau)$ вычислялись одной нейронной сетью, но принимающей на вход, помимо аргументов ЛВФ, описанных в параграфе 2.1.3, также дополнительные n_{mod} аргументов – индикаторных функций номера ЧМПП i . Например, при коррекции температуры T и точки росы Td набор аргументов $\omega_{X_i}(\delta t, t, \tau)$ включал 12 величин:

1. δt – сдвиг по времени;
2. $\Delta_{\delta t} T(t, \tau), \Delta_{\delta t} Td(t, \tau)$ – изменение температуры воздуха и точки росы за δt дней согласно i -й ЧМПП;
3. $BIAS_T(t - \delta t, \tau), BIAS_{Td}(t - \delta t, \tau)$ – погрешности прогноза i -й ЧМПП, данного δt дней назад;

4. $\delta_j^i, j=1, \dots, n_{\text{mod}}$ – индикатор (символ Корнекера) используемой ЧМПП ($n_{\text{mod}}=7$ аргументов).

Заметим, что если при вычислении ЛВФ $\omega_{X_i}(\delta t, t, \tau)$ не использовать δ_j^i , то множество обучаемых параметров ККП будет отличаться от множества обучаемых параметров КСК лишь наличием в ККП параметров ρ_{it} и α_τ . Это позволяет протестировать ККП, не проводя дополнительные оптимизации, положив, например, $\rho_{it} \equiv 1$, $\alpha_\tau \equiv 0.1$. Таким образом, можно применять ККП, не накапливая архивы прогнозов всех используемых в ККП численных моделей (отчасти пожертвовав качеством комплексного прогноза).

Рассмотрим ККП как DAG-сеть (параграф 1.1.4). Для итогового прогноза $\vec{X}_{\text{comp}}$, формула (2.14), штрафной функционал $L(\vec{\theta})$, формула (1.1), запишется как:

$$L(\vec{\theta}) = \frac{1}{M} \sum_{j=1}^M e\left(\vec{X}_{\text{fact},j}, \vec{X}_{\text{comp},j}(\vec{\theta})\right) \rightarrow \min_{\vec{\theta}}.$$

Все параметры ККП $\vec{\theta}$ оптимизируем методом градиентного спуска, вычисляя градиенты методом обратного распространения ошибки. Общее количество оптимизируемых параметров ККП $\vec{\theta}$ при комплексификации прогнозов давления, приведенного к уровню моря, P_0 было равно 968, при комплексификации прогнозов температуры и точки росы – 1995, при комплексификации прогнозов скорости ветра и порывов ветра – 1108.

Хорошее первое приближение оказалось очень важно при оптимизации параметров ККП. В первую эпоху обучения при этой оптимизации использовался ККП, вычисленный по упрощенному алгоритму:

1. Наборы minibatch размера m формировались аналогично параграфу 1.1.7, а именно из одномерных временных рядов $\vec{\Psi}_i(t, \tau_j, \vec{x}_j)$, где $(\tau_j, \vec{x}_j), j=1 \dots m$ – псевдослучайные пары, состоящие из заблаговременности прогноза τ_j и номера станции $\vec{x}_j$;
2. Оценки дисперсий σ_i^2 полагались равными для всех ЧМПП: $\sigma_i^2 \equiv 1$;
3. Параметры ρ_i были фиксированы $\rho_i \equiv 1$ и не оптимизировались;
4. Длина t_l периода для оценки смещения μ_i и стандартного отклонения σ_i увеличивалась (на разных шагах градиентного спуска) с 5 до 35 дней;
5. Поправки (2.12) и (2.14) для промежуточных заблаговременностей не применялись.

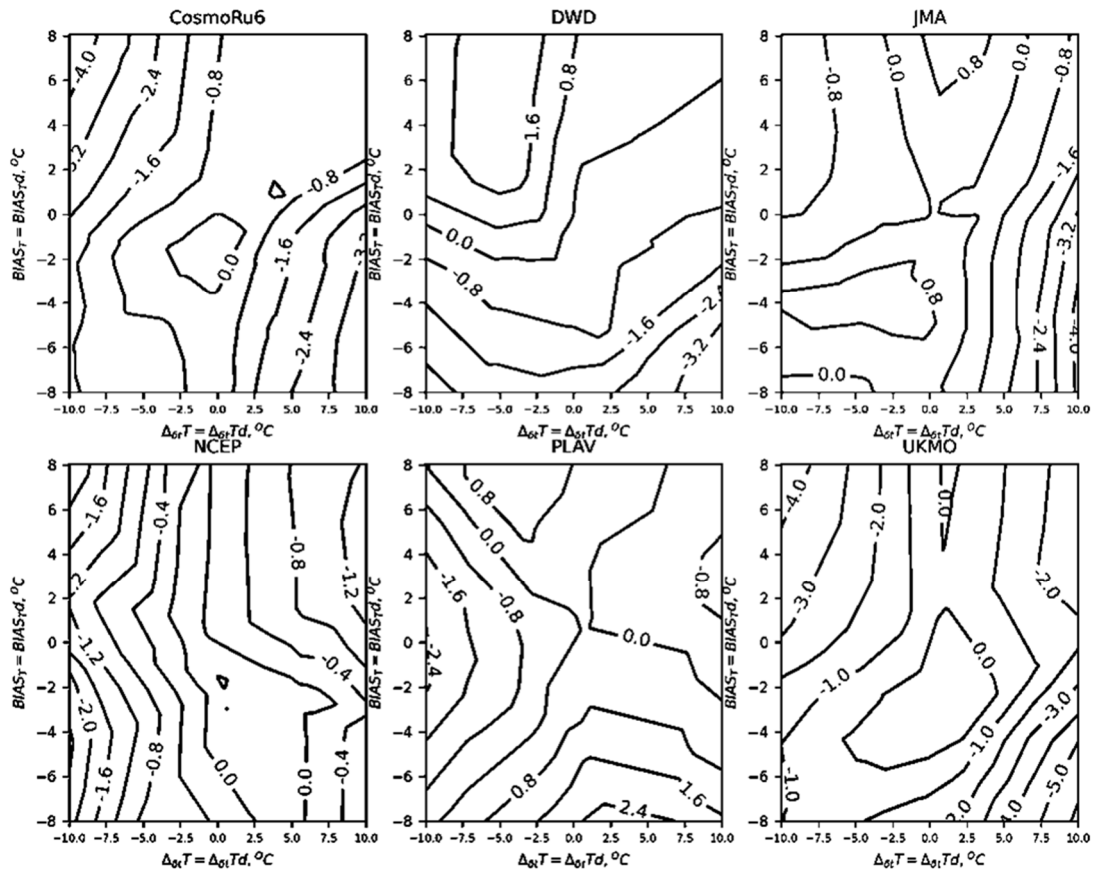
Без упрощений 2 и 3 комплексный прогноз часто в процессе оптимизации устремлялся в область параметров, в которой на комплексный прогноз влияет лишь одна - две лучшие ЧМПП.

Выход же из этой области параметров может потребовать очень большого количества итераций метода оптимизации.

В вторую и последующие эпохи обучения оптимизировался ККП, вычисленный по описанному выше алгоритму:

1. Наборы minibatch размера m формировались из двумерных временных рядов $\bar{\Psi}_i(t, \tau, \bar{x}_j)$, где $\bar{x}_j, j = 1 \dots m$ – псевдослучайная станция;
2. Применялся описанный выше алгоритм вычисления ККП.

На рисунке 2.13 представлены изолинии ЛВФ температуры ω_{T_i} пяти аргументов на сечениях $\delta t = 5 \text{ сут}$, $\Delta_{\delta t} T_i(t, \tau) = \Delta_{\delta t} T d_i(t, \tau)$, $BIAS_{T_i}(t - \delta t, \tau) = BIAS_{T d_i}(t - \delta t, \tau)$. Так как оптимальные ω_{T_i} определены с точностью до аддитивной константы, то за 0 функции ω_{T_i} взято значение на данном сечении в точке $(0, 0)$.



Из рисунка 2.13 видно, что для всех рассмотренных численных моделей ЛВФ ω_{T_i} на данном сечении имеет либо один максимум (2.13а COSMO-Ru6ENA, 2.13б DWD, 2.13г NCEP, 2.13е UKMO), либо два максимума (2.13в JMA, 2.13д PLAV). Максимум ЛВФ $\omega_{T_{DWD}}$ сильно отличается от 0. Вероятной причиной служит то, что так как прогнозы Td_{DWD} были недоступны (таблица 1.6), то при оптимизации функции $\omega_{T_{DWD}}$ использовались прогнозы Td_{NCEP} .

На рисунке 2.14 представлены изолинии оптимальной ЛВФ скорости ветра ω_{U_i} на сечениях $\delta t = 5 \text{ сут}$, $2|\Delta_{\delta t} S(t, \tau)| = |\Delta_{\delta t} \vec{U}(t, \tau)|$, $2|BIAS_S(t - \delta t, \tau)| = |BIAS_{\vec{U}}(t - \delta t, \tau)|$ для различных ЧМПП. Видно, что оптимальные функции ω_{U_i} для различных ЧМПП отличаются друг от друга существенно меньше, чем оптимальные функции ω_{T_i} .

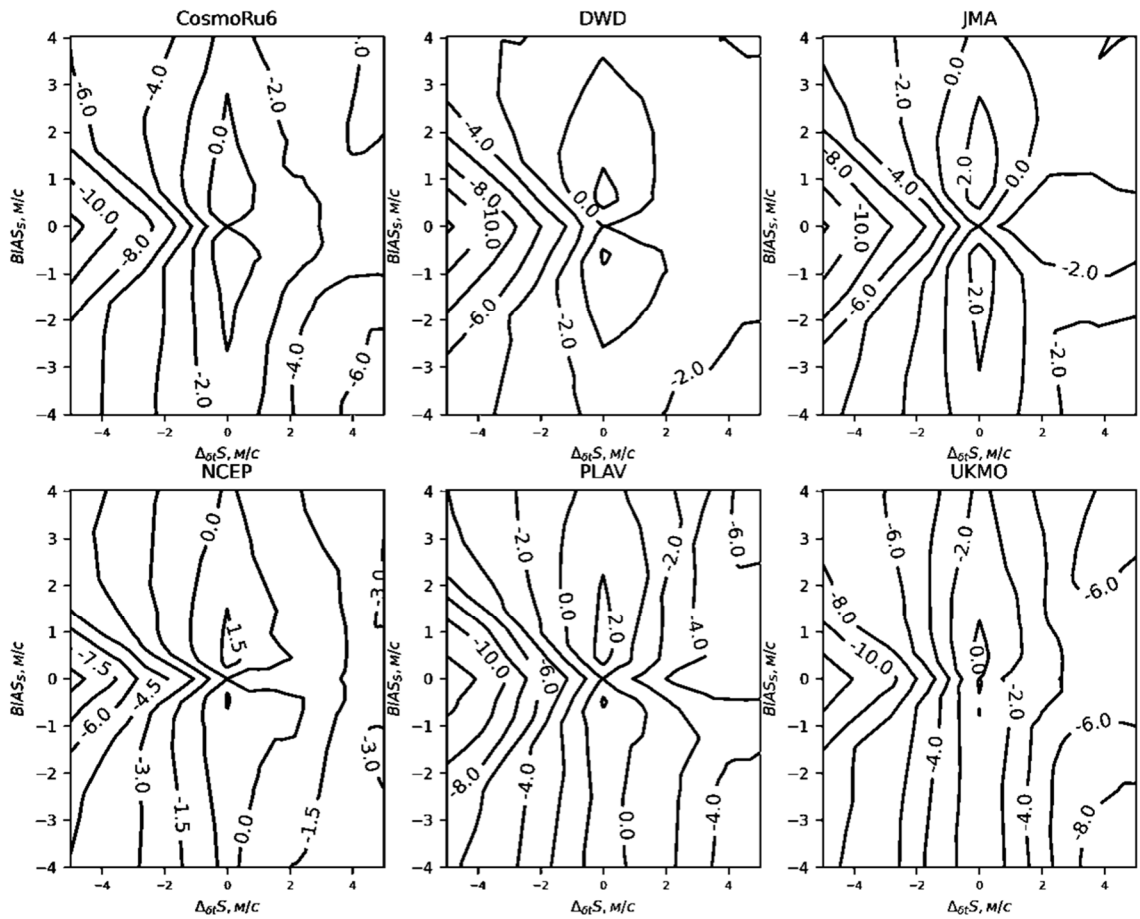


Рисунок 2.14 – Изолинии ЛВФ $\omega_{U_i}(\delta t, t, \tau)$ на сечениях $\delta t = 5 \text{ сут}$, $2|\Delta_{\delta t} S(t, \tau)| = |\Delta_{\delta t} \vec{U}(t, \tau)|$, $2|BIAS_S(t - \delta t, \tau)| = |BIAS_{\vec{U}}(t - \delta t, \tau)|$ для различных ЧМПП: а) COSMO-Ru6ENA; б) DWD; в) JMA; г) NCEP; д) PLAV; е) UKMO

На рисунке 2.15 приведен пример динамики температуры и логарифмических весов $\omega_{T_{COSMO-Ru6ENA}}(\delta t, t, \tau)$ и $\omega_{T_{UKMO}}(\delta t, t, \tau)$, используемых ККП для моделей COSMO-Ru6ENA и UKMO соответственно при заблаговременности $\tau = 60ч$ на станции 29637 Новосибирск за период с июня 2019 г. по июнь 2020 г. Видно, что найденная оптимальная весовая функция, как правило, не является монотонно убывающей по переменной δt и в среднем ЛВФ $\omega_{T_{COSMO-Ru6ENA}}$ убывает медленнее, чем ЛВФ $\omega_{T_{UKMO}}$. В зимний период с ростом δt ЛВФ ω_{T_i} в среднем убывают медленнее, чем в летний. Отметим отдельно, что периоды с сильными аномалиями температуры в Новосибирске, например 16-20 сентября 2019 г. и 20-24 апреля 2020 г., имеют здесь меньший вес при вычислении поправок, то есть погрешности в период больших аномалий дают минимальный вклад в систематическую коррекцию для последующих дней (особенно это заметно для модели UKMO, рисунок 2.15б).

На рисунке 2.16 приведена аналогичная динамика скорости ветра и ЛВФ $\omega_{U_{COSMO-Ru6ENA}}(\delta t, t, \tau)$ и $\omega_{U_{UKMO}}(\delta t, t, \tau)$, используемых ККП скорости ветра на станции 37006 Новороссийск за период с июня 2019 г. по июнь 2020 г. Видно, что логарифмические веса, используемых ККП с увеличением δt убывают медленно. Это согласуется с рисунками 2.14а и 2.14е, на которых видно, что $\omega_{U_{COSMO-Ru6ENA}}(\delta t, t, \tau)$ и $\omega_{U_{UKMO}}(\delta t, t, \tau)$ гораздо существеннее зависят от изменения скорости ветра $\Delta_{\delta t} S(t, \tau)$, чем от сдвига по времени δt .

На рисунках 2.15 и 2.16 белые полосы соответствуют отсутствию в используемом архиве каких-либо прогностических или фактических значений, используемых для расчета соответствующих величин. Этих значений не так много (1.1%), но, если рассматривать только случаи, когда есть весь набор прогнозов и измерений за предшествующие $t_l = 35$ дней, то количество рассматриваемых случаев существенно сократится (таблица 1.8). Поэтому, на практике применимы исключительно алгоритмы, допускающие пропуски в данных.

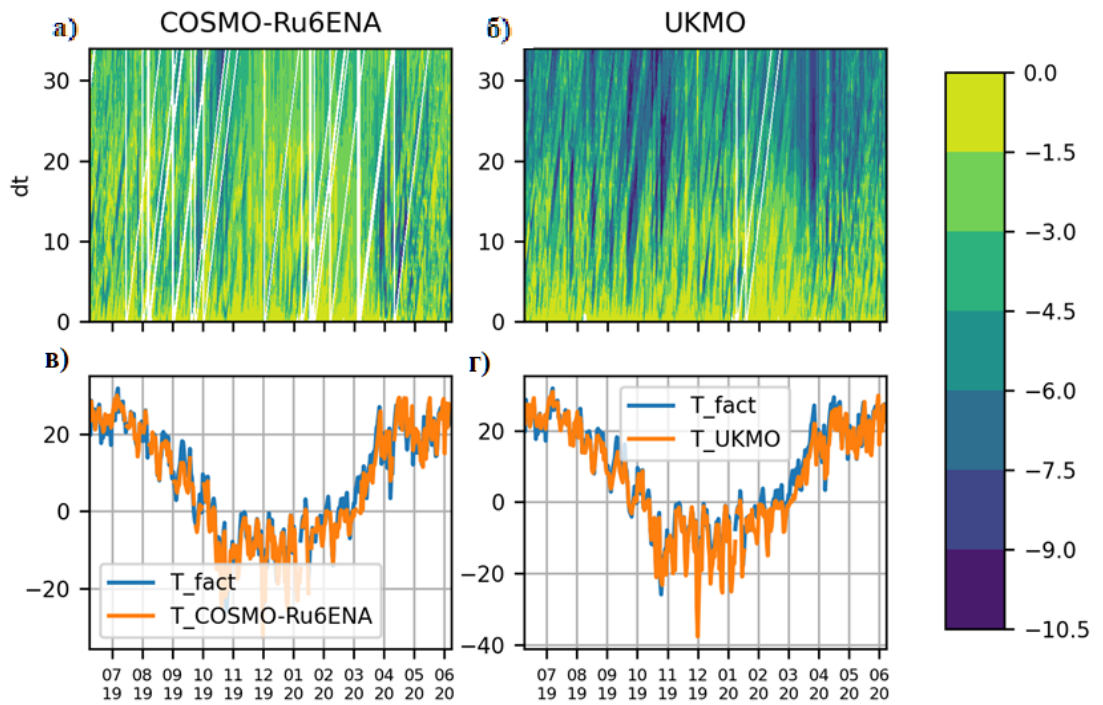


Рисунок 2.15 – Динамика: а), б) ЛВФ $\omega_{T_i}(\delta t, t, \tau)$, вычисленные ККП на станции 29637 Новосибирск для заблаговременности $\tau = 60ч$ с начальным сроком 00:00 ВСВ и в), г) соответствующих фактических и прогностических температур для моделей COSMO-Ru6ENA (а, в) и UKMO (б, г). На оси абсцисс отмечены первые числа месяцев 2019 - 2020 гг

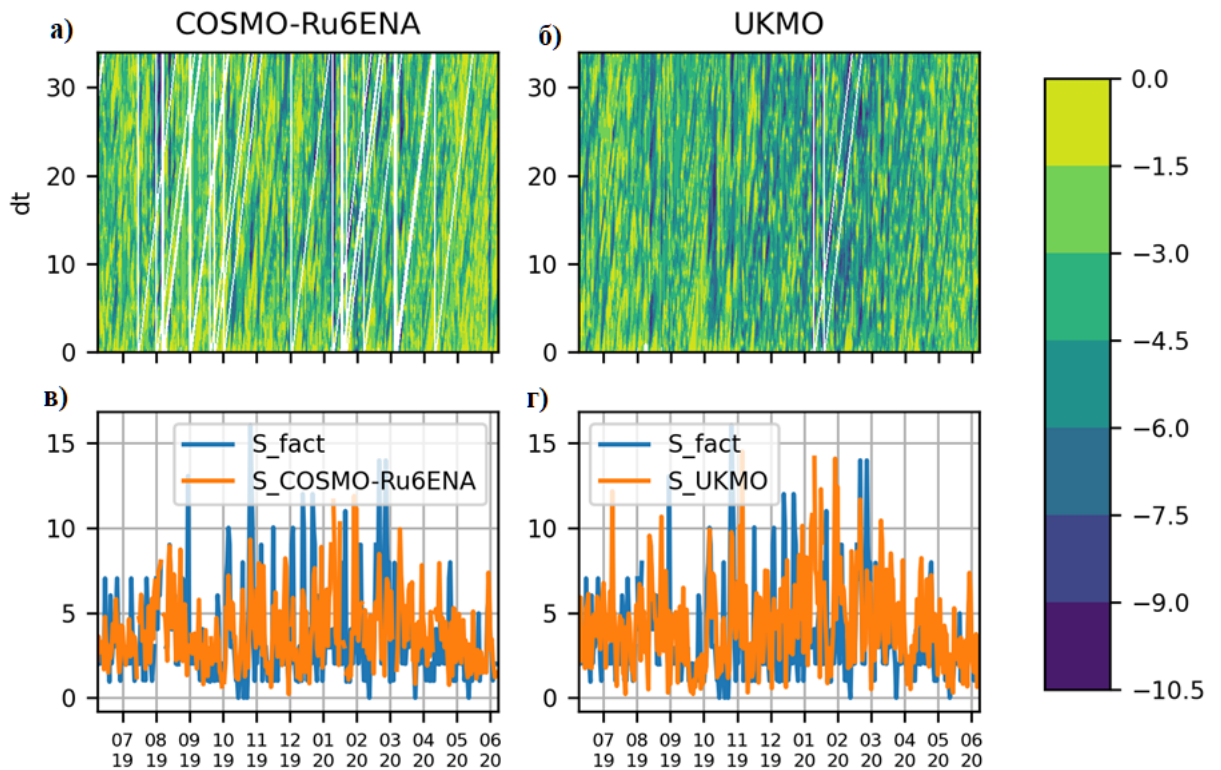
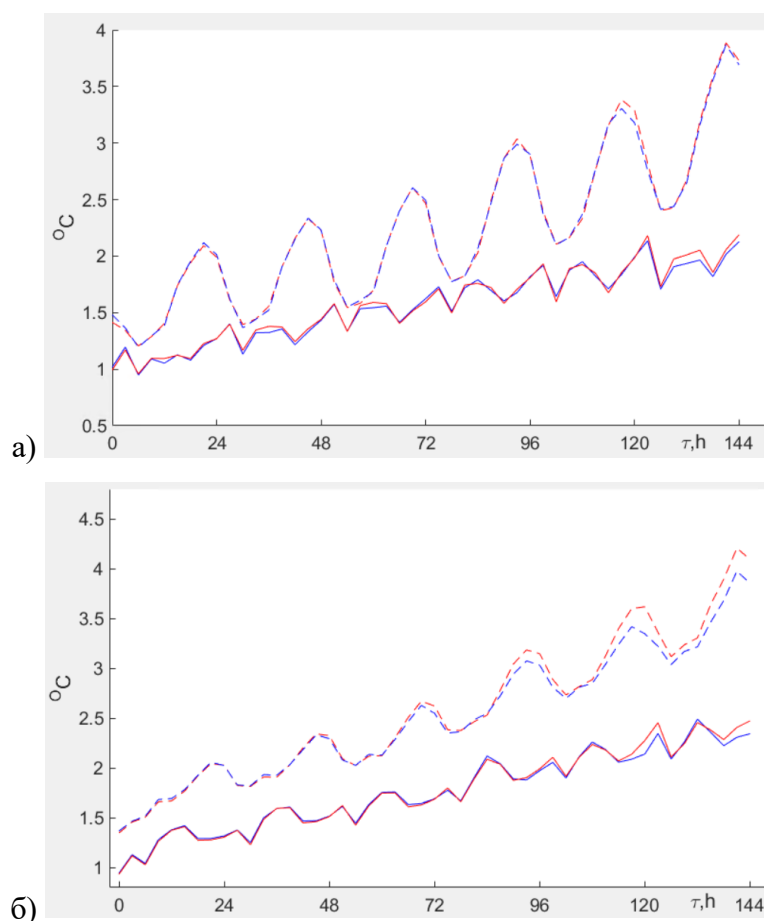


Рисунок 2.16 – Динамика: а), б) ЛВФ $\omega_{U_i}(\delta t, t, \tau)$, вычисленные ККП на станции 37006 Новороссийск для заблаговременности $\tau = 60ч$ с начальным сроком 00:00 ВСВ и в), г) соответствующих фактических и прогностических скоростей ветра для моделей COSMO-Ru6ENA (а, в) и UKMO (б, г)

2.2.4. Сравнение линейной и квазилинейной версий комплексного прогноза

Проведем сравнение средней абсолютной погрешности прогнозов, вычисленных по двум версиям комплексного прогноза: 1) квазилинейного ККП и 2) «линейного» комплексного прогноза, то есть с весовой функцией, вычисляемой нейронной сетью, но не зависящей от известных погрешностей $BIAS_X$. Результаты приведены на рисунке 2.17 отдельно для ЕЧР (сплошные линии) и для АЧР (пунктиры). Видно, что квазилинейная версия имеет преимущество над линейной на обеих рассмотренных территориях. Преимущество квазилинейной комплексификации над линейной существенно при заблаговременностях τ более 84ч.

Выигрыш от перехода к квазилинейности на рисунке 2.17 существенно меньше, чем на рисунке 2.9, поскольку тут и в линейной версии веса w_X вычислялись нейронной сетью, а не (как на рисунке 2.9) в виде функции из двухпараметрического семейства (2.4).



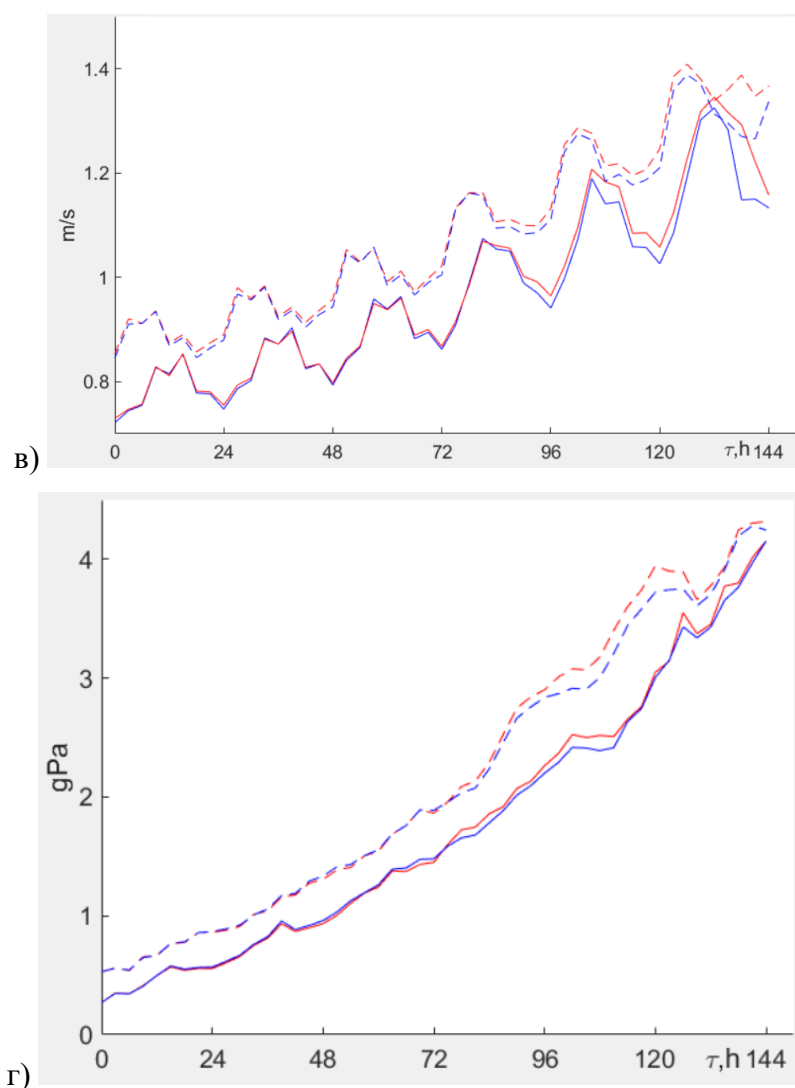


Рисунок 2.17 – Сравнение средней абсолютной погрешности различных версий ККП различных метеорологических параметров за март 2021 г. от начального срока 00:00 ВСВ в зависимости от заблаговременности: а) для температуры воздуха T , °С; б) для точки росы Td , °С; в) для модуля скорости ветра $|\vec{U}|$, м/с; г) давления, приведенного к уровню моря P , гПа.

Синие линии – ККП, красные линии – «линейная» версия комплексного прогноза.

Сплошные линии – по территории ЕЧР, пунктиры – по территории АЧР

2.2.5. Оценка вклада различных численных моделей

Увеличение числа используемых в комплексном прогнозе ЧМПП удорожает вычисления и делает систему более уязвимой к отсутствию части данных, но не всегда приводит к улучшению качества финального результата. Желательно проверить, насколько каждая из используемых в комплексном прогнозе ЧМПП улучшают его качество. Для этого будем по очереди исключать из комплексного прогноза одну из ЧМПП и оценивать качество получающихся прогнозов.

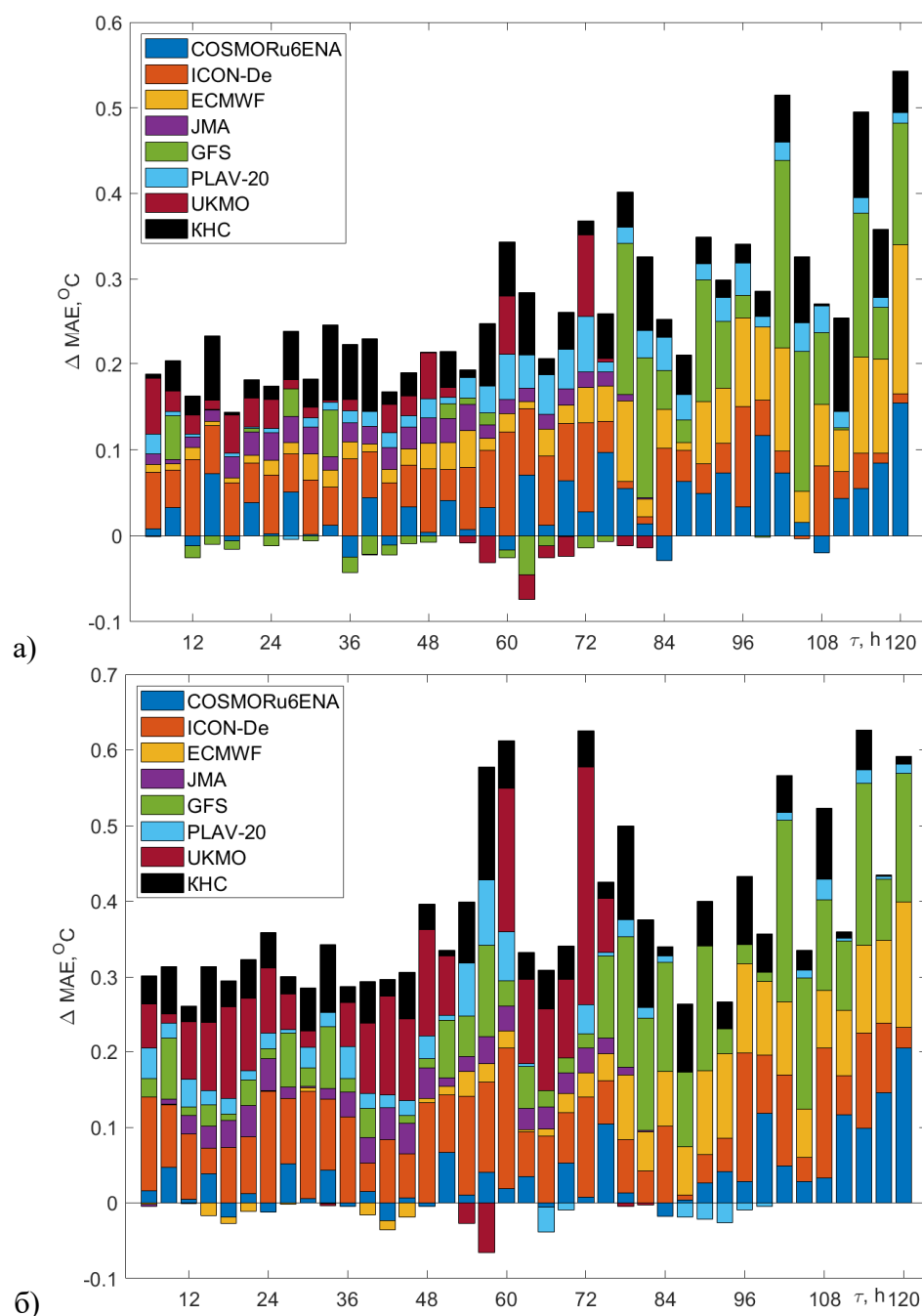


Рисунок 2.18 – Величина, на которую увеличивается средняя абсолютная ошибка прогноза (высота прямоугольников соответствующего цвета) за период с января по март 2021 г. температуры воздуха $T, ^\circ C$ от начального срока 00:00 ВСВ при исключении из комплексного прогноза одной из ЧМПП в зависимости от заблаговременности:
а) в ЕЧР; б) в АЧР

Результаты такой проверки на архиве за период с января по март 2021 г. показаны на рисунке 2.18, где высота цветных прямоугольников равна величине ΔMAE , на которую увеличивается средняя абсолютная ошибка прогноза температуры воздуха при исключении одной из численные модели. Видно, что все участвующие ЧМПП в среднем улучшали качество комплексного прогноза, но не для всех заблаговременностей. Наибольший положительный вклад вносят модели GFS, ICON-De и ECMWF. Для сравнения показан эффект от использования

дополнительной коррекции, основанной на нейронных сетях (КНС, черные прямоугольники), и описанной далее в главе 3. Видно, что эффект от использования КНС соизмерим с добавлением прогностических данных по одной из лучших ЧМПП.

Выводы из главы 2

В главе 2 предложены следующие модели МО: квазилинейная систематическая коррекция (КСК) и квазилинейная комплексификация прогнозов (ККП). При систематической коррекции будущее смещение ЧМПП в точке рассматриваемой синоптической станции оценивается по её известным погрешностям на той же станции. При комплексификации помимо смещения оценивается и дисперсия прогнозов ЧМПП.

КСК и ККП для оценки смещения и дисперсии к известным погрешностям ЧМПП применяется квазилинейный оператор с положительными коэффициентами. Логарифм коэффициентов этого оператора вычисляется нейронной сетью. Если нейронная сеть принимает на вход известные погрешности прогнозов, то модель постпроцессинга является квазилинейной по известным погрешностям ЧМПП.

В параграфе 2.1.4 проанализированы изолинии нейронной сети на некоторых сечениях в пространстве её аргументов. Продемонстрировано (рисунок 2.8), что при коррекции прогнозов скорости ветра КСК выбирает случаи как можно более близкой скорости и направления ветра, не учитывая сдвиг по времени, что говорит о том, что решающее влияние на поле ошибок прогноза ветра на высоте 10 м оказывает обтекание препятствий на поверхности Земли. При коррекции полей температуры, точки росы и давления сдвиг по времени имеет значение (рисунки 2.6, 2.7), то есть можно говорить об сохранении знака ошибки в течении некоторого времени (инерции ошибки, эффекте памяти). Аналогичный анализ весов, используемых в ККП (рисунок 2.14), показал, что структура погрешностей отличается для различных ЧМПП.

Предложенные модели МО применимы в случае недоступности части прогностической информации и (или) данных наблюдений на рассматриваемой синоптической станции. КСК и ККП были применены к прогнозам приземных метеорологических параметров (температуры воздуха, точки росы, давления, приведенного к уровню моря, скорости и направления ветра, порывов ветра) по различным ЧМПП. В параграфах 2.1.5 и 2.2.4 показано, что квазилинейные версии КСК и ККП имеют преимущество над их линейными версиями, при этом использование нейронной сети для вычисления коэффициентов систематической коррекции улучшает (по сравнению с простой экспоненциальной коррекцией) качество прогнозов на величину, соответствующую уменьшению заблаговременности на 1 сутки.

ГЛАВА 3. Нейросетевая коррекция

В главе 2 корректировались прогнозы с учётом последних измерений на рассматриваемой синоптической станции, и были получены статистически несмещенные прогнозы. В этой главе рассмотрим коррекцию, основанную на нейронных сетях, (КНС) и попытаемся с её помощью улучшить качество прогнозов приземных метеорологических параметров, полученных методами систематической коррекции и комплексификации (глава 2). Предлагаемая модель МО является непосредственным развитием методики, предложенной автором в работах [116], [28], и применяется не только к откорректированным (комплексифицированным), но и к оригинальным прогнозам по модели COSMO-Ru.

3.1. Используемые предикторы

Напомним наши обозначения. $\vec{\Psi}_0(t, \tau, \vec{x})$ – прогноз k приземных метеорологических параметров в точке $\vec{x}$ области $\Omega \subset \mathbb{S}^2$ от срока t на момент времени $t + \tau$. В данной главе будем искать оптимальные поправки к прогнозу $\vec{\Psi}_0$. В качестве корректируемого прогноза $\vec{\Psi}_0$ рассматривались:

- прогнозы по модели COSMO-Ru в конфигурациях COSMO-Ru6ENA, COSMO-RuBy;
- прогнозы по модели COSMO-Ru в конфигурациях COSMO-Ru6ENA, COSMO-RuBy, подвергнутые процедуре КСК (раздел 1 главы 2);
- комплексифицированные по алгоритму ККП прогнозы (раздел 2 главы 2).

Обозначим вспомогательные (то есть не являющиеся прогностическими переменными) предикторы, такие как день года, местное время суток и синус широты, вектор-функцией $\vec{\varphi}(t, \tau, \vec{x})$, число её компонент обозначим d .

Оптимальные аддитивные поправки к прогнозам будем вычислять в различных точках $\vec{x} \in \Omega$ независимо, используя серию прогнозов в точке $\vec{x}$. Поправленный прогноз $\vec{\Psi}_N$ представим в виде:

$$\vec{\Psi}_N(t, \tau, \vec{x}) = \vec{\Psi}_0(t, \tau, \vec{x}) + \vec{F}(\vec{\varphi}(t, \tau, \vec{x}), \tau, \vec{\Psi}_0(t - \Delta t_1, \tau + \Delta \tau_1, \vec{x}), \dots, \vec{\Psi}_0(t - \Delta t_m, \tau + \Delta \tau_m, \vec{x})), \quad (3.1)$$

где сдвиги по времени $\Delta \tau_j \in \mathbb{R}$ и $\Delta t_j \geq 0, j = 1, \dots, m$ будут выбраны заранее, $\vec{F}$ – искомая вектор-функция $n = d + 1 + mk$ аргументов, $k = 7$ количество рассматриваемых метеорологических параметров.

При практическом использовании функция $\vec{F}$ должна осуществлять интерполяцию, а не экстраполяцию в пространстве-времени, поэтому, согласно формуле (3.1), аргумент $\vec{F}$ прогностического срока $t + \tau$ используется не напрямую.

При выборе сдвигов, связанных соотношением $\Delta\tau_j = \Delta t_j$, то есть при использовании более раннего прогноза (прогнозов) на тот же срок $t + \tau$ от более раннего начального срока $t - \Delta t_j$ с большей заблаговременностью $\tau + \Delta t_j$, получим так называемое запаздывающее прогнозирование (lagged forecasting) [167]. Если максимальная заблаговременность исходного прогноза Ψ_0 равна $\tau_{\max}$ (глава 2, таблицы 1.3 и 1.6), то коррекция (3.1) применима только при $\tau \in \left[-\min_j \Delta\tau_j; \tau_{\max} - \max_j \Delta\tau_j \right]$. Поэтому выбирать большие по модулю сдвиги по заблаговременности $\Delta\tau_j \neq 0$ может быть не целесообразно.

Погрешности, которые можно исправить с помощью поправок вида (3.1), могут возникать в ЧМПП, прежде всего, из-за систематически смещенных начальных данных или из-за неточной параметризации физических процессов, приводящей к систематически смещенной скорости развития и (или) интенсивности этих процессов.

Для того чтобы иметь возможность исправить систематические смещения по пространству, можно в (3.1) добавить возможность сдвига по пространству, но это многократно увеличит количество аргументов функции F и требования к вычислительным ресурсам и к количеству используемой памяти. В настоящей работе сдвиги по пространству не рассматривались.

Откорректированный прогноз Ψ_N , вычисленный нашей моделью МО, не должен резко меняться при небольшом изменении корректируемых прогнозов Ψ_0 , а значит функция F в (3.1) должна быть непрерывна. Ниже аппроксимацию непрерывной функции F будем осуществлять глубокой нейронной сетью.

Для описания наборов сдвигов $\Delta\tau_j, \Delta t_j$ используем два параметра: параметр $N_{history}$ описывает количество суток, предшествующих данным, прогностические параметры за которые используются; параметр N_{old} описывает количество учитываемых более ранних прогнозов на тот же срок $t + \tau$. Итак, нейронная сеть коррекции принимает на вход следующие $42 + 21N_{old} + 18N_{history}$ аргументов:

1. τ – заблаговременность (1 аргумент);
2. $\sin \varphi$ – синус широты (1 аргумент);

3. $\sin\text{Hour} = \sin\left(\lambda + 2\pi \frac{t}{24\text{ч}}\right)$ – синус локального времени суток (1 аргумент);
4. δx – горизонтальный шаг сетки корректируемой конфигурации модели COSMO-Ru в градусах (1 аргумент);
5. $\alpha_{\text{sun}}(t + \tau)$ – высота Солнца, то есть синус угла (со знаком) между направлением на Солнце и горизонтом (1 аргумент);
6. $\alpha_{\text{sun}}^{\max}(t), \alpha_{\text{sun}}^{\min}(t)$ – высота Солнца в зените и надире за текущие сутки – зависят от дня года и широты (2 аргумента);
7. $Z(t, \tau)$, где $Z = P_0, T_0, D_0, G_0, S_0, U_0, V_0$ – прогностические метеорологические величины (7 аргументов);
8. $Z(t, \tau) - Z(t, \tau + \Delta\tau)$, где $Z = P_0, T_0, D_0, G_0, S_0, U_0, V_0$, при $\Delta\tau = -6, -3, 3, 6\text{ч}$ – изменение прогностических метеорологических параметров за $\Delta\tau$ (28 аргументов);
9. $Z(t, \tau) - Z(t - \delta t, \tau + \Delta\tau)$, где $Z = P_0, T_0, D_0, S_0, U_0, V_0$, при $\Delta\tau = -6, 0, 6\text{ч}$ – изменение прогностических метеорологических параметров за $\delta t = 1, \dots, N_{\text{history}}$ дней ($18N_{\text{history}}$ аргументов);
10. $Z(t, \tau + \Delta\tau) - Z(t - \delta t, \tau + \delta t + \Delta\tau)$, где $Z = P_0, T_0, D_0, G_0, S_0, U_0, V_0$, при $\Delta\tau = -6, 0, 6\text{ч}$ – изменение прогностических метеорологических параметров относительно прогнозов, данных $\delta t = 12\text{ч}, \dots, 12\text{ч} \cdot N_{\text{old}}$ часов назад ($21N_{\text{old}}$ аргументов).

При коррекции комплексифицированных прогнозов ККП предиктор δx (горизонтальный шаг сетки) не использовался.

Максимальная заблаговременность прогноза, для которой известны все перечисленные величины, на $12N_{\text{old}} + 6$ часов меньше максимальной заблаговременности исходных прогнозов.

Рассматривались следующие пары $(N_{\text{old}}, N_{\text{history}})$:

$$(N_{\text{old}}, N_{\text{history}}) = (3, 1), (2, 1), (1, 2), (0, 2), (0, 0). \quad (3.2)$$

В итоговых оценках (глава 5) оценивался прогноз, составленный из прогнозов с различными парами $(N_{\text{old}}, N_{\text{history}})$ с приоритетом выбора поправок, полученных для пары $(N_{\text{old}}, N_{\text{history}})$, стоящей левее в ряду (3.2). Если пропуски в прогностических данных за последние дни отсутствовали, то:

1. при $\tau \leq \tau_{\text{max}} - 42\text{ч}$ использовались $N_{\text{old}} = 3, N_{\text{history}} = 1$ (123 аргумента);

2. при $\tau_{\max} - 30ч \leq \tau < \tau_{\max} - 42ч$ использовались $N_{old} = 2, N_{history} = 1$ (102 аргумента);
3. при $\tau_{\max} - 18ч \leq \tau < \tau_{\max} - 30ч$ использовались $N_{old} = 1, N_{history} = 2$ (99 аргументов);
4. при $\tau_{\max} - 6ч \leq \tau < \tau_{\max} - 18ч$ использовались $N_{old} = 0, N_{history} = 2$ (78 аргументов);
5. при $\tau < 6ч$ или $\tau_{\max} - 6ч < \tau$ использовались $N_{old} = 0, N_{history} = 0$ (42 аргумента).

При оптимизации гиперпараметров (параграф 1.1.11) нейронной сети коррекции использовались $N_{history} = 2, N_{old} = 0$, то есть набор из 78 входных аргументов. При оптимизации гиперпараметров для обучения использовался архивы Ru13-2017 и Ru6ENA-2019, а для валидации архивы Ru13-2018 и Ru6ENA-2020 (таблица 1.4).

3.2. Оптимизация нейронной сети

В этой части главы обучение проводилось по архиву Ru13ENA-2017, а валидации – на архиве Ru13ENA-2018 (таблица 1.4). Объём каждого из архивов составлял около $6,8 \times 10^7$ реализаций.

3.2.1. Оптимизация стратегии обучения нейронной сети

Далее сравнения точности прогнозов производятся в терминах изменения Δe функции потерь Хьюбера (параграф 1.1.1, формула (1.4)), а также изменения среднеквадратической ошибки ΔRMS при применении поправок (3.1), то есть:

$$\begin{aligned} \Delta e &= \frac{1}{N} \sum_{j=1}^N e(X_{0,j}, X_{fact,j}) - e(X_{C,j}, X_{fact,j}), \\ \Delta RMS &= \sqrt{\frac{1}{N} \sum_{j=1}^N |X_{0,j} - X_{fact,j}|} - \sqrt{\frac{1}{N} \sum_{j=1}^N |X_{C,j} - X_{fact,j}|}. \end{aligned} \quad (3.3)$$

Рекуррентные соотношения методов обучения нейронных сетей (параграф 1.1.7) имеют параметр, называемый скоростью обучения η_s и пропорциональны скорости изменения параметров за s -й шаг градиентного спуска. Алгоритм изменения скорости обучения называют **стратегией обучения**.

Диссертантом были проведены численные эксперименты по обучению трехслойного персептрона с 48-24 нейронами на скрытых слоях и с функциями активации ReLU для коррекции температуры методом AdamW (формула (1.40)) со стратегиями обучения из двухпараметрического семейства (1.30) при различных параметрах η_0 и C .

Результаты численных экспериментов по коррекции прогнозов температуры с minibatch размера $m=M/400=1.39 \times 10^5$ (то есть одна эпоха состояла из 400 шагов) и с одной и той же псевдослучайной инициализацией параметров нейронной сети приведены в таблице 3.1. Такой выбор размера minibatch m обусловлен тем, что при меньших значениях m увеличиваются так называемые накладные расходы, а при увеличении – существенно возрастает объем используемой при оптимизации памяти. Численные эксперименты показали, что стратегия обучения, достигшая лучшего результата на обучающей выборке, не всегда даёт наилучший результат на валидационном архиве. В таблице 3.1 приведены результаты только для 50-й эпохи обучения, поскольку на этой эпохе обучения выбранные диссертантом стратегии достигли своего оптимума. Жирным выделены результаты, отличающиеся по ΔRMS (формула (3.3)) от наилучших рассмотренных не более чем на 0.02°C . Результаты по некоторым из рассмотренных стратегий в другие эпохи обучения приведены на рисунке 3.1.

Таблица 3.1 – Изменение среднеквадратической ошибки ΔRMS (формула (3.3)) прогноза приземной температуры на обучающей Ru13-2017 и на валидационной Ru13-2018 выборках при различных значениях параметров η_0 и C стратегии изменения скорости обучения (формула (1.30)) на 50-й эпохе обучения методом AdamW

№	η_0	C	ΔRMS_{train}	ΔRMS_{valid}
1	0.0025	1.5	0.427	0.332
2	0.005	1.5	0.440	0.333
3	0.01	1.5	0.453	0.322
4	0.02	1.5	0.457	0.320
5	0.0025	3	0.409	0.325
6	0.005	3	0.428	0.332
7	0.01	3	0.444	0.326
8	0.02	3	0.448	0.326
9	0.0025	6	0.388	0.314
10	0.005	6	0.410	0.327
11	0.01	6	0.432	0.325
12	0.02	6	0.437	0.327

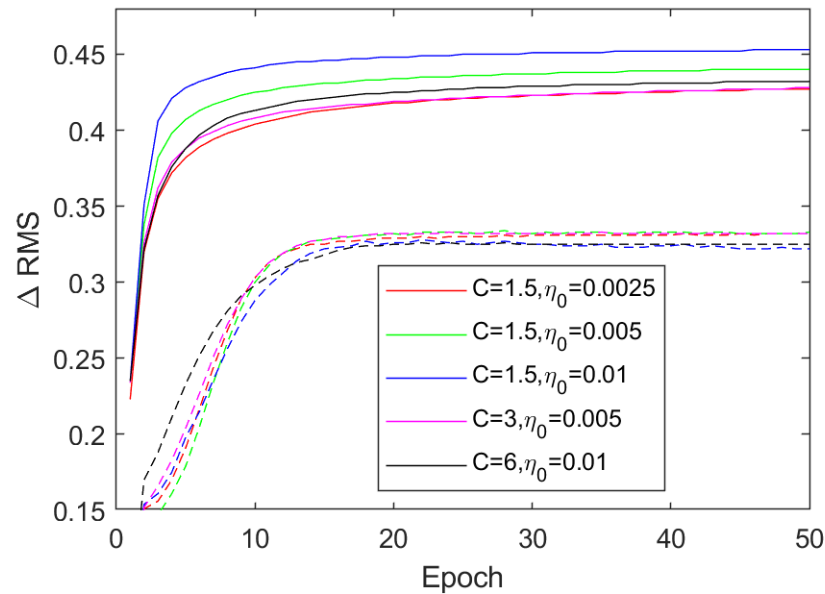


Рисунок 3.1 – Изменение среднеквадратической ошибки ΔRMS (формула (3.3)) прогноза приземной температуры в зависимости от номера эпохи обучения на обучающей выборке Ru13-2017 (сплошные линии) и на валидационной выборке Ru13-2018 (пунктирные линии) при различных значениях параметров η_0 и C стратегии изменения скорости обучения (формула (1.30)) при обучении методом AdamW

На рисунке 3.1 на оси абсцисс отложен номер эпохи обучения t_E , а по оси ординат – изменение среднеквадратической ошибки ΔRMS на обучающей выборке Ru13-2017 (сплошные линии) и на валидационной выборке Ru13-2018 (пунктирные линии).

Рассмотрим теперь различные методы обучения нейронных сетей: стохастический градиентный спуск (SGD), а так же различные варианты адаптивного метода градиентного спуска: RMSProp [229], [148], Adam [171], [132], AdamW [187], RAdam [186] и AdaBelief [252]. Рекуррентные соотношения этих методов приведены в параграфах 1.1.7 и 1.1.8. Так же рассматривались варианты оптимизации с применением центрирования градиентов (gradient centralization, далее GC), формула (1.42), которое было предложено в [248] и совместимо с любым из алгоритмов градиентного спуска.

В таблице 3.2 приведены результаты сравнения оценок прогнозов при использовании различных алгоритмов градиентного спуска, где **жирным** шрифтом отмечены результаты, отличающиеся от наилучших не более чем на 0.001. Корректировались прогнозы модели COSMO-Ru в конфигурации COSMO-Ru6ENA, подвергнутые КСК (глава 2). Видно, что качество прогнозов, прошедших коррекцию, обученную различными алгоритмам градиентного спуска, отличается очень слабо, а наилучшие результаты были достигнуты при использовании алгоритма AdamW.

Таблица 3.2 – Изменение среднего значения функции потерь Хьюбера Δe (формула (3.3)) на обучающей Ru6-2019 и на валидационной Ru6-2020 выборках для различных алгоритмов градиентного спуска при коррекции прогнозов COSMO-Ru6ENA+KCK различных метеорологических параметров

	$T, ^\circ\text{C}$		$Td, ^\circ\text{C}$		$P_0, \text{гПа}$		$ \vec{U} , \text{м/с}$	
Алгоритм	Δe_{train}	Δe_{valid}	Δe_{train}	Δe_{valid}	Δe_{train}	Δe_{valid}	Δe_{train}	Δe_{valid}
RMSProp	0.144	0.104	0.124	0.113	0.045	0.011	0.041	0.039
RMSProp+GC	0.141	0.106	0.122	0.112	0.046	0.010	0.040	0.039
Adam	0.147	0.106	0.128	0.118	0.047	0.010	0.041	0.039
Adam+GC	0.146	0.106	0.125	0.109	0.049	0.013	0.042	0.039
AdaBelief	0.147	0.109	0.126	0.116	0.046	0.012	0.042	0.039
AdaBelief+GC	0.146	0.107	0.124	0.113	0.047	0.010	0.042	0.039
AdamW	0.146	0.110	0.127	0.113	0.047	0.014	0.042	0.039
RAdam	0.144	0.106	0.127	0.108	0.050	0.012	0.042	0.039

По результатам экспериментов, приведенным в таблицах 3.1 и 3.2, везде далее использовался алгоритм градиентного спуска AdamW и следующие параметры стратегии обучения (1.30): $\eta_0 = 0.01$, $C = 1.5$, а число эпох обучения было равно 20.

3.2.2. Оптимизация параметров инициализации и регуляризации

При оптимизации параметров нейронной сети использовалась тихоновская (L_2) регуляризация, формула (1.39), параграф 1.1.8. Тихоновская регуляризация несколько уменьшает среднеквадратичное значение параметров нейронной сети (достигнутое при оптимизации). На среднеквадратичное значение параметров влияет и инициализация этих параметров перед началом обучения.

В данной работе использовалась инициализация Ксавье (1.46) с различными значениями параметра g_0 . Параметр регуляризации l_2 отличался для первой и последующих эпох обучения. Для первой эпохи обучения значение l_2 выбиралось так, чтобы за первую эпоху обучения $t_E = 1$ по методу AdamW при $\nabla_{\theta_k} L \equiv 0$ параметр нейронной сети θ_k изменился в g_1 раз. Таким образом, параметр l_2 вычислялся по формуле:

$$l_2 = \begin{cases} \frac{1 - g_1^{m/M}}{\eta_0}, & t_E = 1; \\ l_2', & t_E > 1. \end{cases} \quad (3.4)$$

Вместо параметра g_1 , можно говорить о параметре l_2 , связанным с ним по формуле (3.4), но тогда будет неудобно сравнивать результаты: при том же l_2 , но другой скорости обучения η_0

или другом размере minibatch m результаты обучения могут сильно отличаться. Если гиперпараметры g_0 и g_1 меньше 1, то они уменьшают L_2 -норму вектора параметров $\vec{\theta}$ нейронной сети: гиперпараметр g_0 при инициализации, а гиперпараметр g_1 во время первой эпохи обучения.

Были проведены численные эксперименты с различными гиперпараметрами l_2' , g_0 и g_1 . Так же как в предыдущем параграфе 3.2.1 эксперименты проводились с обучением трехслойного персептрона с 48-24 нейронами на скрытых слоях и с функциями активации ReLU для коррекции температуры с одной и той же псевдослучайной инициализацией параметров нейронной сети, умноженной на g_0 . Результаты приведены в таблице 3.3.

Таблица 3.3 – Изменение среднеквадратической ошибки ΔRMS (формула (3.3)) прогноза приземной температуры на обучающей Ru13ENA-2017 и на валидационной Ru13ENA-2018 выборках при различных значениях гиперпараметров l_2' , g_0 , g_1

Эпоха обучения			1	5	10	20	1	5	10	20
l_2'	g_0	g_1	ΔRMS_{train}				ΔRMS_{valid}			
0.01	0.3	0.1	0.379	0.458	0.463	0.467	0.070	0.102	0.130	0.311
0.01	0.6	0.1	0.379	0.456	0.461	0.464	0.097	0.100	0.128	0.322
0.01	0.8	0.1	0.377	0.455	0.461	0.465	0.095	0.100	0.134	0.320
0.01	1.0	0.1	0.376	0.458	0.463	0.466	0.096	0.098	0.130	0.327
0.01	0.3	0.2	0.382	0.454	0.459	0.462	-0.100	0.112	0.175	0.328
0.01	0.6	0.2	0.377	0.452	0.458	0.462	0.084	0.105	0.171	0.327
0.01	0.8	0.2	0.377	0.454	0.460	0.463	0.043	0.123	0.215	0.330
0.01	1.0	0.2	0.376	0.454	0.459	0.463	-0.069	0.121	0.183	0.328
0.01	0.3	0.4	0.381	0.451	0.456	0.459	-0.885	0.098	0.199	0.334
0.01	0.6	0.4	0.381	0.452	0.456	0.460	0.097	0.160	0.265	0.328
0.01	0.8	0.4	0.378	0.454	0.458	0.462	-0.087	0.149	0.262	0.333
0.01	1.0	0.4	0.377	0.447	0.452	0.457	-0.074	0.135	0.253	0.336
0.01	0.3	0.7	0.383	0.451	0.456	0.459	-1.506	0.175	0.294	0.327
0.01	0.6	0.7	0.377	0.453	0.458	0.460	-0.227	0.223	0.308	0.326
0.01	0.8	0.7	0.375	0.444	0.451	0.455	-1.708	0.237	0.314	0.330
0.01	1.0	0.7	0.375	0.449	0.454	0.457	-11.366	0.231	0.318	0.338
0.01	0.3	1.0	0.379	0.450	0.454	0.458	-2.159	0.247	0.316	0.329
0.01	0.6	1.0	0.374	0.446	0.453	0.458	-4.297	0.281	0.328	0.327
0.01	0.8	1.0	0.372	0.444	0.454	0.458	-20.166	0.254	0.330	0.332
0.01	1.0	1.0	0.369	0.448	0.454	0.457	-16.708	0.275	0.329	0.336
0	0.90	0.5	0.380	0.451	0.456	0.459	-0.048	0.200	0.294	0.327
0	0.95	0.5	0.380	0.453	0.457	0.460	-2.553	0.221	0.316	0.330
0	0.90	0.8	0.376	0.448	0.453	0.455	-7.504	0.274	0.326	0.327
0	0.95	0.8	0.375	0.444	0.45	0.454	-2.487	0.269	0.324	0.331
0	0.90	0.9	0.374	0.444	0.449	0.453	-2.55	0.250	0.323	0.324
0	0.95	0.9	0.374	0.450	0.455	0.458	-14.347	0.265	0.317	0.326

0.01	0.90	0.5	0.380	0.451	0.456	0.459	-0.048	0.195	0.284	0.331
0.01	0.95	0.5	0.380	0.453	0.457	0.460	-2.553	0.207	0.304	0.333
0.01	0.90	0.8	0.376	0.448	0.453	0.455	-7.504	0.272	0.327	0.331
0.01	0.95	0.8	0.375	0.444	0.451	0.455	-2.487	0.272	0.322	0.334
0.01	0.90	0.9	0.374	0.444	0.449	0.453	-2.55	0.261	0.324	0.326
0.01	0.95	0.9	0.374	0.451	0.455	0.459	-14.347	0.271	0.318	0.331
0.02	0.90	0.5	0.380	0.451	0.456	0.459	-0.048	0.182	0.265	0.332
0.02	0.95	0.5	0.380	0.453	0.457	0.460	-2.553	0.194	0.290	0.334
0.02	0.90	0.8	0.376	0.448	0.453	0.456	-7.504	0.266	0.323	0.332
0.02	0.95	0.8	0.375	0.444	0.451	0.455	-2.487	0.263	0.315	0.336
0.02	0.90	0.9	0.375	0.444	0.453	0.458	-0.949	0.271	0.320	0.337
0.02	0.95	0.9	0.375	0.452	0.458	0.462	-6.145	0.256	0.321	0.336
0.05	0.90	0.5	0.380	0.451	0.457	0.461	-0.048	0.144	0.203	0.324
0.05	0.95	0.5	0.380	0.453	0.457	0.461	-2.553	0.149	0.226	0.324
0.05	0.90	0.8	0.375	0.449	0.454	0.457	-0.725	0.155	0.228	0.329
0.05	0.95	0.8	0.375	0.447	0.453	0.458	-2.983	0.169	0.243	0.322
0.05	0.90	0.9	0.374	0.450	0.456	0.460	-6.242	0.166	0.241	0.329
0.05	0.95	0.9	0.375	0.453	0.458	0.461	-8.78	0.193	0.259	0.333

Из таблицы 3.3 видно, что в данном случае эффект от регуляризации l_2' весьма мал, а параметры g_0 и g_1 весьма сильно влияют на изменение среднеквадратической ошибки ΔRMS в процессе обучения на валидационной выборке Ru13-2018. Видно, что на валидационной выборке лучшие результаты на первой эпохе обучения достигаются при $g_1 = 0.1$, но на 5-10й эпохе обучения – при $g_1 = 1$, а к 20-й эпохе – при $g_0 = 1$, $g_1 \geq 0.4$. При этом на первой эпохе при $g_1 \geq 0.5$ выигрыш от коррекции на валидационной выборке отрицательный. В дальнейшем использовались следующие значения гиперпараметров: $l_2' = 0.01$, $g_0 = 0.95$, $g_1 = 0.5$. Согласно таблице 3.3 такие параметры обеспечивают компромисс между скоростью обучения и достижением положительного эффекта от коррекции на всем протяжении процесса обучения.

3.2.3. Оптимизация архитектуры и количества нейронов нейронной сети

Были проведены численные эксперименты с нейронными сетями с $N_{block} = 2-8$ слоями с различным количеством нейронов следующих архитектур: многослойный перцептрон (далее MLP) [85], Residual Network (далее ResNet) [153], Gated Residual Network (далее GRN) [226]. Соответствующие этим архитектурам графы вычислений представлены на рисунке 3.2.

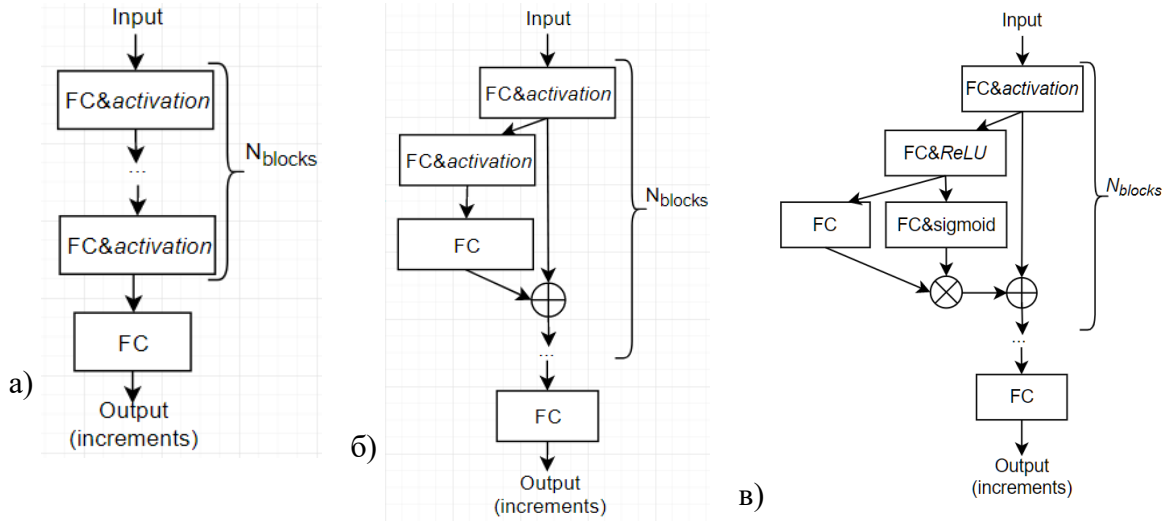


Рисунок 3.2 – Графы вычислений нейронных сетей различных архитектур: а) многослойный перцептрон (MLP); б) Residual Network (ResNet); в) Gated Residual Network (GRN). Тут FC – полносвязный слой

Сети типов ResNet и их улучшенная версия GRN предложены в [153] и [226] соответственно, с целью преодолеть так называемую проблему малых градиентов (параграф 1.1.4, формула (1.22)). Согласно экспериментам в [153] нейронная сеть, содержащая блоки, осуществляющие отображение вида

$$\vec{x} \mapsto \vec{y} = \vec{x} + \vec{R}(\vec{x}, \vec{\theta}), \quad (3.5)$$

где $\vec{R}$ – нейронная сеть, может обучаться быстрее сети, в которой слагаемые $\vec{R}$ были отброшены. Если параметры $\vec{\theta}$ блока (3.5) находятся в общем положении, то в процессе обратного распространения ошибки через блок вида (3.5) градиент ошибки не уменьшается:

$$\|\nabla_{\vec{x}} e\|_2 \geq \|\nabla_{\vec{y}} e\|_2$$

В нейронной сети архитектуры ResNet [153] помимо полносвязных слоев (глава 1, формула (1.19)) есть так же residual блоки типа (3.5), в которых вычисления $\vec{R}(\vec{x}, \vec{\theta})$ происходит по следующим формулам:

$$\begin{aligned} \vec{h} &= \max(\vec{0}, W_h \vec{x} + \vec{b}_h) \\ \vec{R}(\vec{x}, \vec{\theta}) &= W_r \vec{h} + \vec{b}_r \end{aligned} \quad (3.6)$$

В нейронной сети архитектуры GRN [226] в блоках типа (3.5) вычисление $\vec{R}(\vec{x}, \vec{\theta})$ происходит по формулам:

$$\begin{aligned}
\vec{h} &= \max(\vec{0}, W_h \vec{x} + \vec{b}_h) \\
\vec{g} &= W_g \vec{h} + \vec{b}_g \\
\vec{R}(\vec{x}, \vec{\theta}) &= (W_r \vec{h} + \vec{b}_r) \circ \sigma(\vec{g})
\end{aligned} \tag{3.7}$$

В формулах (3.6) и (3.7) вектор параметров $\vec{\theta}_k$ состоит из коэффициентов линейных операторов $W_h, W_g, W_r: \mathbb{R}^{N_k} \rightarrow \mathbb{R}^{N_k}$ и векторов $\vec{b}_{hk}, \vec{b}_{rk}, \vec{b}_{gk} \in \mathbb{R}^{N_k}$; логистическая функция $\sigma(x) = (1 + e^{-x})^{-1}$ и оператор $\max$ применяются покомпонентно. Логика вычислений в (3.6) следующая: сначала выбираем условия $\vec{h}_k > 0$ введения поправок (3.5), а затем определяем величины поправок $\vec{r}_k$. Таким образом поправки являются кусочно-линейными функциями.

В формуле (3.7) поправки дополнительно микшируются нелинейным множителем $\sigma(g) \in (0; 1)$. При больших отрицательных значениях компонент $g_{s,k}, s = 1, \dots, N_k$ значения функции $\sigma(g_{s,k})$ малы, и блок (3.7) слабо меняет значения соответствующих признаков $y_{s,k}$. При $g_{s,k}$ близких к 0 блок (3.7) близок к квадратичным функциям. При больших положительных $g_{s,k}$ – блок (3.7) близок к кусочно-линейным функциям. Таким образом, использование микширующего множителя в блоках (3.7) делает класс моделей МО F , вычисляемых нейронной сетью, существенно более универсальным.

В проведенных диссертантом численных экспериментах по коррекции прогнозов модели COSMO-Ru в конфигурации COSMO-Ru-13ENA суммарное количество параметров нейронных сетей архитектур MLP, ResNet и GRN n_θ варьировалось в пределах от 1288 до 189797. На рисунке 3.3 по оси абсцисс отложено количество параметров нейронной сети n_θ в логарифмическом масштабе, а по оси ординат – выигрыш от коррекции (3.1), выраженный в среднеквадратической погрешности ΔRMS . Видно, что на обучающей выборке зависимость ΔRMS от количества параметров нейронной сети n_θ близка к линейной функции от $\log n_\theta$. При валидации ΔRMS имеет выраженный максимум при $7000 < n_\theta < 20000$. Применение более сложных архитектур нейронной сети (ResNet и GRN) в данном случае не оправдано, так как более простые MLP показывают близкие результаты.

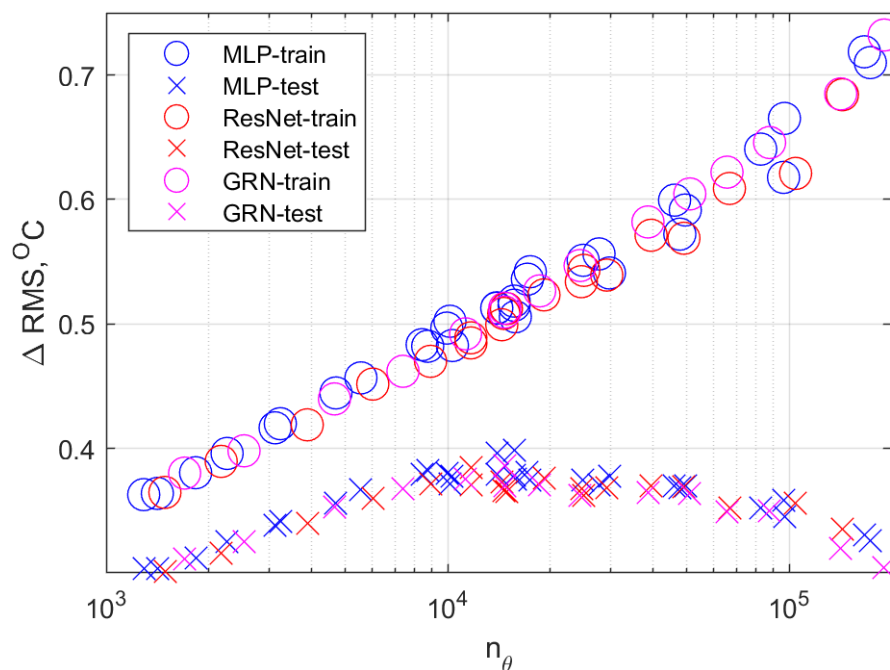


Рисунок 3.3 – Результаты валидации нейронных сетей различной архитектуры с различным количеством слоёв (таблица 3.4)

В таблице 3.4 приведены некоторые результаты валидации нескольких MLP с различным количеством нейронов на скрытых слоях, и с количеством параметров, лежащим в оптимальном согласно рисунку 3.3 диапазоне. Видно, что при заданном количестве нейронов конкретный выбор функции активации (*th*, *ReLU*, *LeakyReLU*) практически не оказывает влияния на оценки.

При использовании функций активации *ReLU* имеется два важных преимущества по сравнению с использованием функции активации *th*: во-первых, вычисления занимают гораздо меньше тактов процессора, а во-вторых, помогает обойти проблему малых градиентов.

Таблица 3.4 – Результаты валидации некоторых многослойных персептронов

№	Количество нейронов N_k	Количество параметров n_θ	Функция активации	ΔRMS_{train}	ΔRMS_{valid}
1	128-64-128	23193	<i>ReLU</i>	0.566	0.398
2	128-64-32	16665	<i>ReLU</i>	0.546	0.391
3	128-64	14553	<i>ReLU</i>	0.537	0.405
4	96-48-96	14345	<i>ReLU</i>	0.538	0.401
5	96-48-24	10601	<i>ReLU</i>	0.523	0.399
6	96-48	9401	<i>ReLU</i>	0.510	0.397
7	64-64-64	11737	<i>ReLU</i>	0.527	0.400
8	64-64	7449	<i>ReLU</i>	0.499	0.401
9	64-32-64	7545	<i>ReLU</i>	0.502	0.395
10	64-32-16	5817	<i>ReLU</i>	0.485	0.400
11	64-32	5273	<i>ReLU</i>	0.481	0.405
12	48-48-48	7289	<i>ReLU</i>	0.503	0.387
13	48-48	4841	<i>ReLU</i>	0.478	0.403

14	48-24-48	4913	<i>ReLU</i>	0.484	0.397
15	48-24-12	3905	<i>ReLU</i>	0.469	0.398
16	48-24	3593	<i>ReLU</i>	0.465	0.399
17	32-32-32	3865	<i>ReLU</i>	0.468	0.395
18	32-32	2745	<i>ReLU</i>	0.449	0.396
19	32-16-32	2793	<i>ReLU</i>	0.447	0.391
20	32-16	2169	<i>ReLU</i>	0.430	0.392
21	64-32-16	5929	<i>tanh</i>	0.487	0.393
22	64-32-16	5817	<i>LeakyReLU</i>	0.482	0.407
23	64-32-16	5817	<i>PReLU</i>	0.489	0.403

Из таблицы 3.4 видно, что из рассмотренных MLP лучшее качество имеют прогнозы, рассчитанные трех- и четырехслойными персептронами. Для дальнейших оценок были использовано среднее арифметическое прогнозов двух MLP с 64-32 нейронами на скрытых слоях.

В случае, когда нейронная сеть корректирует прогнозы, прошедшие систематическую коррекцию КСК или комплексификацию ККП (глава 2), наилучшие результаты на валидационной выборке показала нейронная сеть архитектуры GRN с функцией активации *th* и с 24-16 нейронами на скрытых слоях при коррекции скорости ветра, температуры воздуха и точки росы, и с 12-8 нейронами при коррекции давления.

3.3 Учет локальных параметров с помощью Embedding слоёв

На метеорологические величины в конкретной точке $\vec{x}$ влияют так называемые локальные параметры, прежде всего описывающие подстилающую поверхность. В работе [225] показано, что могут оказаться полезны такие локальные параметры орографии, как угол наклона склона, высота преобладающей вершины и так называемый седловой параметр.

В работе [209] нейронные сети с embedding слоем использовались для коррекции ансамблевых прогнозов приземной температуры воздуха, и было показано преимущество нейронных сетей с embedding слоем, в том числе и над методом случайного леса [114], [195].

Так же, как и в [209], в данной работе embedding слой принимает на вход индекс синоптической станции и возвращает некоторый вектор размерности E , постоянный для данной станции (параграф 1.1.6, формула (1.28)). Получившийся вектор конкатенируется (склеивается) с вектором прочих используемых предикторов. Таким образом оптимизируется локальный параметр (параметры) для каждой синоптической станции.

При использовании такого embedding слоя пропадает возможность применять нейронную сеть для коррекции прогноза в точке, отличной от точки синоптической станции, данные

наблюдений на которой были использованы при обучении. В численных экспериментах рассмотрен набор предикторов при $N_{history} = 1, N_{old} = 1$ в терминах раздела 1 текущей главы. В этих экспериментах корректировались следующие прогнозы:

1. прогнозы по модели COSMO-Ru в конфигурациях COSMO-Ru6ENA, COSMO-RuBy;
2. прогнозы по модели COSMO-Ru в конфигурациях COSMO-Ru6ENA, COSMO-RuBy, подвергнутые процедуре КСК (первый раздел главы 2);
3. комплексифицированные по алгоритму ККП прогнозы (второй раздел главы 2).

Размерность embedding была равна $E=1,2,3,4$. Результаты численных экспериментов показаны в таблице 3.5. Видно, что с ростом параметра E оценки в среднем немного улучшаются.

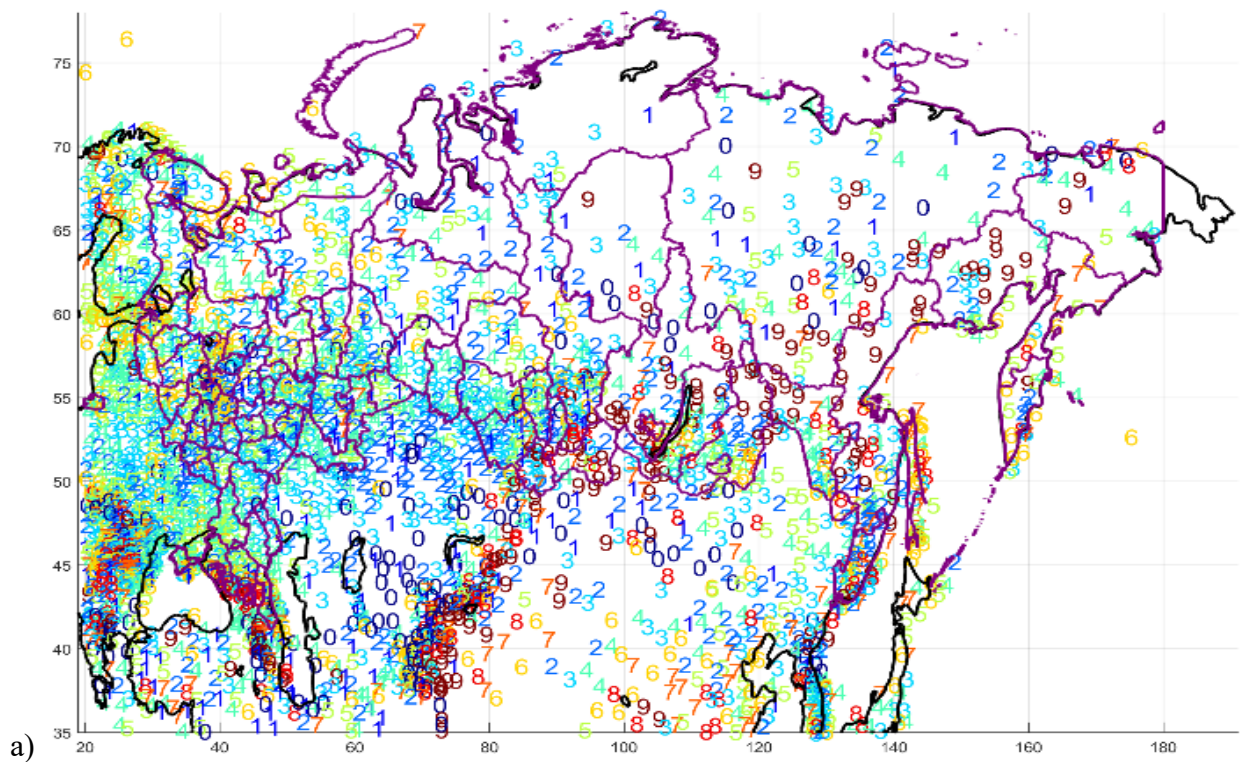
При коррекции прогнозов модели COSMO-Ru начиная с $E=2$ это улучшение становится незначительным. Отметим, что конфигурация метода КНС, не использующая embedding ($E=0$), может применяться к оригинальным прогнозам во всех точках, а не только в точках синоптических станций. Конфигурация метода КНС, использующая embedding ($E>0$), применима только в точках синоптических станций, для которых вычислены значения embedding.

При коррекции прогнозов COSMO-Ru, подвергнутым процедуре КСК, а также к прогнозам, рассчитанным по методу ККП (глава 2) улучшение от использования embedding можно считать незначительным. Согласно таблице 3.5 применение метода КНС к откорректированным КСК и ККП прогнозам давления, приведенного к уровню моря, очень слабо (до 0,015 и 0,041 гПа, соответственно), поэтому в расчетах в оперативном режиме метод КНС к прогнозам давления не применялся.

Таблица 3.5 – Изменение среднеквадратической погрешности ΔRMS на валидационном архиве Ru6-2020 при коррекции прогнозов различных метеорологических параметров X по методу КНС при использовании embedding различной размерности E

COSMO-Ru						COSMO-Ru+КСК				
E	0	1	2	3	4	0	1	2	3	4
$T, ^\circ\text{C}$	0.290	0.315	0.345	0.349	0.364	0.128	0.139	0.142	0.144	0.147
$Td, ^\circ\text{C}$	0.379	0.410	0.438	0.443	0.457	0.120	0.124	0.122	0.125	0.127
$P_0, \text{гПа}$	0.291	0.329	0.334	0.325	0.354	0.012	0.015	0.012	0.012	0.013
$ \vec{U} , \text{м/с}$	0.291	0.374	0.388	0.386	0.390	0.038	0.038	0.039	0.038	0.038
ККП										
$T, ^\circ\text{C}$	0,062	0,071	0,074	0,077	0,078					
$Td, ^\circ\text{C}$	0,084	0,086	0,089	0,091	0,090					
$P_0, \text{гПа}$	0,041	-0,006	0,006	-0,002	-0,011					
$ \vec{U} , \text{м/с}$	0,034	0,035	0,036	0,039	0,037					

На рисунках 3.4 и 3.5 значения embedding при $E=I$ показаны на географической карте. На этом рисунке эти значения с целью визуализации нормированы так, чтобы среднее их значение было равно 5 и чтобы 99% значений укладывались в интервал $[0,10]$. На рисунке 3.4 видно, что значения embedding зависят: от орографии (регион Центральной Азии); от преобладающей атмосферной циркуляции (видна граница между Западной и Восточной Сибирью). Отметим, что для различных синоптических станций Центрального федерального округа (ЦФО) найденный embedding (рисунок 3.5) сильно различается при относительно малых отличиях в орографии и атмосферной циркуляции. Можно предположить, что это причинами этих различий являются в том числе и городские эффекты [236].



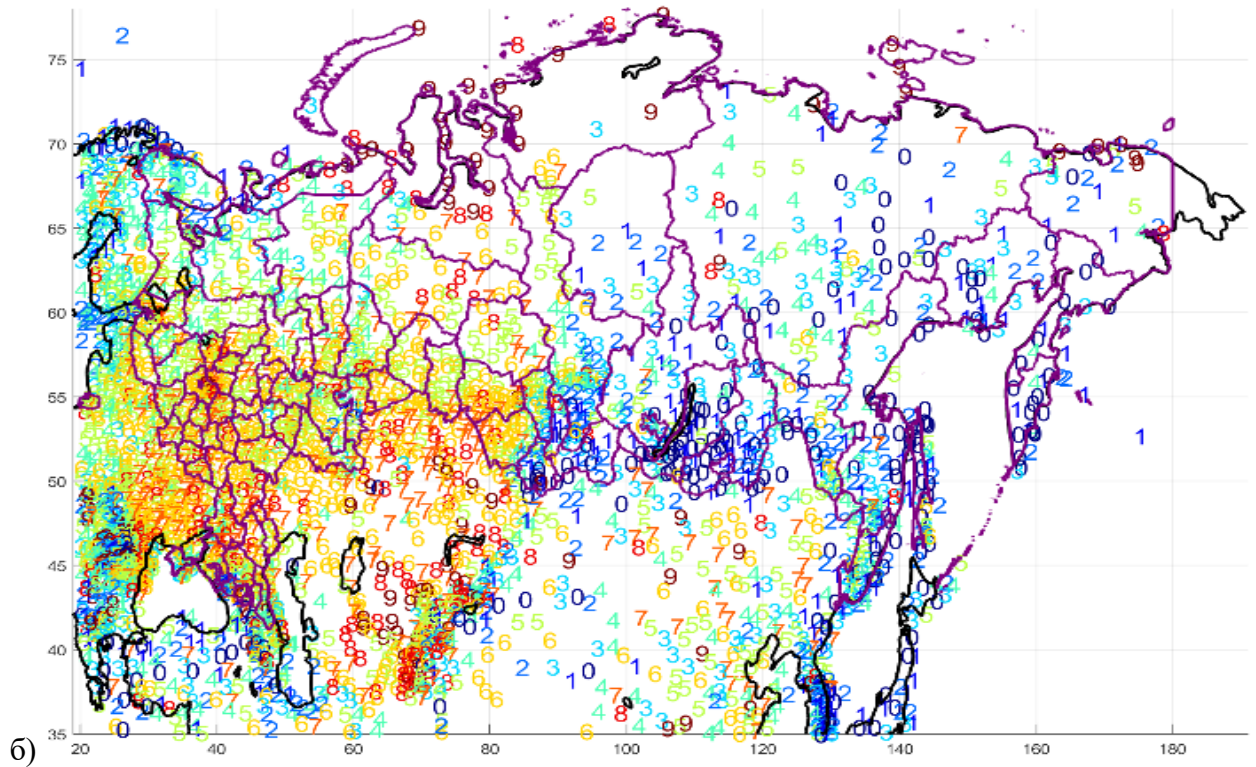
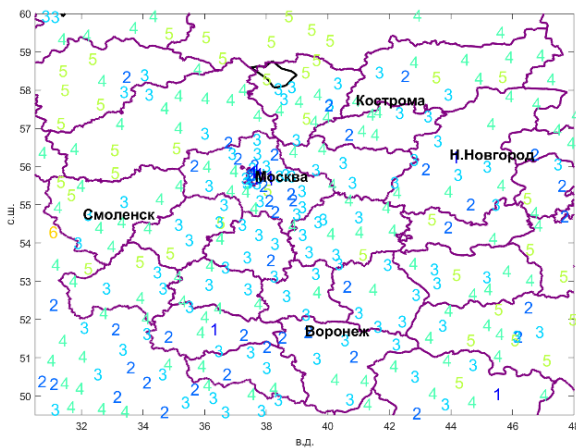


Рисунок 3.4 – Значения embedding размерности $E=I$ для различных синоптических станций на географической карте: а) для задачи коррекции температуры воздуха; б) для задачи коррекции скорости ветра

а)



б)

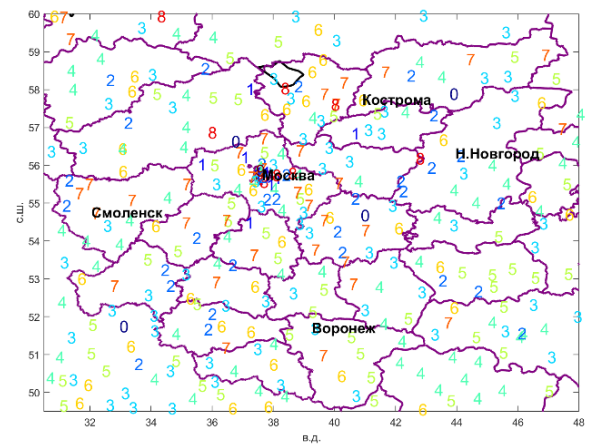


Рисунок 3.5 – То же, что на рисунке 3.4, но на территории ЦФО

3.4 Оценка важности различных предикторов

Под **важностью** (importance) предиктора подразумевают то, насколько ухудшится качество прогнозов метода машинного обучения (МО) в случае, если этот предиктор не использовать. Проводить такие оценки полезно, поскольку возможна ситуация, когда один из предикторов не несёт (или несёт слишком мало) полезной для решаемой задачи информации и

его использование лишь зашумляет ответы, даваемые методом МО. Поэтому отбрасывание такого предиктора может улучшить качество прогноза. В некоторых случаях предикторы отбрасывают из-за того, что трата ресурсов, затраченных на получение их значений, не соответствует получаемому улучшению качества ответов применяемого метода.

Можно рассмотреть сокращенный набор предикторов и провести оптимизацию метода заново, но провести такую процедуру для всех используемых предикторов часто оказывается слишком дорого. Поэтому на практике как правило обучают модель МО F на полном наборе предикторов, а затем применяют различные эвристические методы оценки важности различных предикторов.

Классические задачи оценки важности, например метод class activation map [250], разработаны для задач обработки изображений и предполагают, что значения всех предикторов (различных пикселей изображения) распределены согласно одному и тому же вероятностному закону. Это предположение не верно для различных метеорологических полей (например, вероятностные распределения значений скорости ветра и температуры воздуха принципиально различны).

Для оценки важности различных предикторов используем метод permutation importance, использованный, например, в [209] для задачи коррекции ансамблевых прогнозов приземных метеорологических параметров. Метод permutation importance заключается в следующем. Чтобы оценить важность i -го предиктора Imp_i , возьмем псевдослучайную перестановку $\pi(j), j = 1 \dots M$ рассматриваемого набора из M реализаций входных параметров $X_i, i = 1 \dots N$ и оценим важность Imp_i как изменение функции потерь e (использовалась функция потерь Хьюбера, формула (1.4)) при применении этой перестановки к i -му предиктору, то есть:

$$\text{Imp}_i = \frac{1}{M} \sum_{j=1}^M e\left(F\left(X_{1,j}, \dots, X_{i,\pi(j)}, \dots, X_{N,j}\right), X_{\text{fact},j}\right) - e\left(F\left(\bar{X}_j\right), X_{\text{fact},j}\right). \quad (3.8)$$

Оценка важности по формуле (3.8) требует N применений обученной модели МО F , что существенно дешевле, чем N обучений на уменьшенных наборах предикторов. Важным преимуществом метода permutation importance (3.8) является то, что при применении перестановки к набору значений i -го предиктора не меняется их вероятностное распределение, а значит, можно предположить, что и вероятностные распределения ответов модели МО F тоже изменятся слабо. Оценку по формуле (3.8) можно уточнить, усреднив её по l различным псевдослучайным перестановкам π_k :

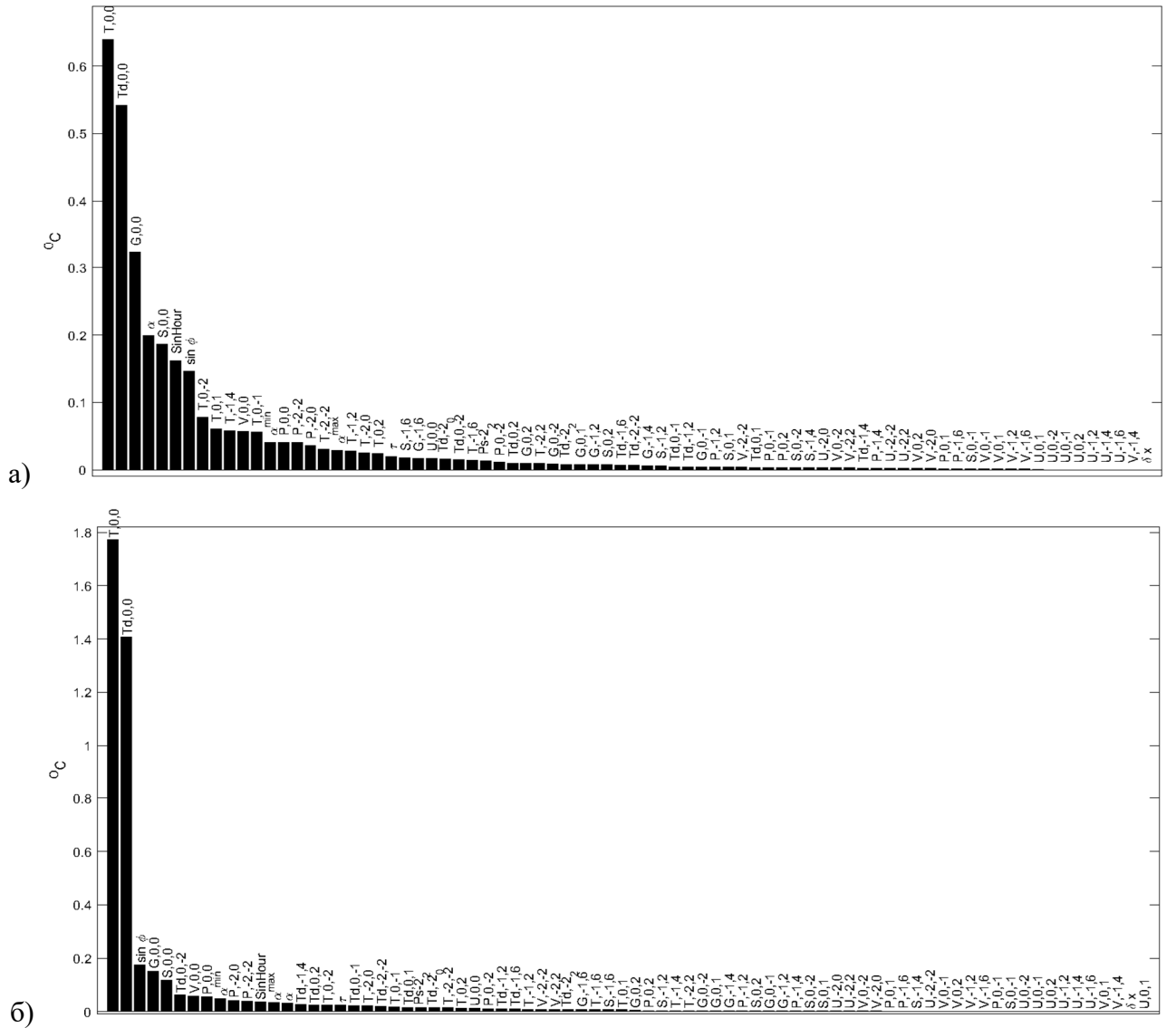
$$\text{Imp}_i = \frac{1}{l} \sum_{k=1}^l \text{Imp}_i(\pi_k). \quad (3.9)$$

Результаты оценки важности по формуле (3.9) при $l = 10$ для случая коррекции откорректированных КСК прогнозов модели COSMO-Ru в конфигурации COSMO-Ru6ENA для случая $N_{history} = 2, N_{old} = 0$ (всего 78 предикторов) приведены на рисунке 3.6 в порядке убывания их важности Imp_i .

На рисунке 3.6 приняты следующие обозначения: за Z, k, s , где $Z = T, Td, P, G, S, U, V$, $k = 0, -1, -2$, $s \in \mathbb{Z}$ обозначен следующий предиктор:

$$Z, k, s = \begin{cases} Z(t, \tau), & \text{при } k = s = 0; \\ Z(t, \tau) - Z(t + 12\tau \cdot k, \tau + 3\tau \cdot s), & \text{иначе.} \end{cases}$$

То есть за $Td, -1, 2$ обозначен предиктор $Td(t, \tau) - Td(t - 12\tau, \tau + 6\tau)$. Обозначения остальных предикторов на рисунке 3.6 приведены в 1 разделе текущей главы.



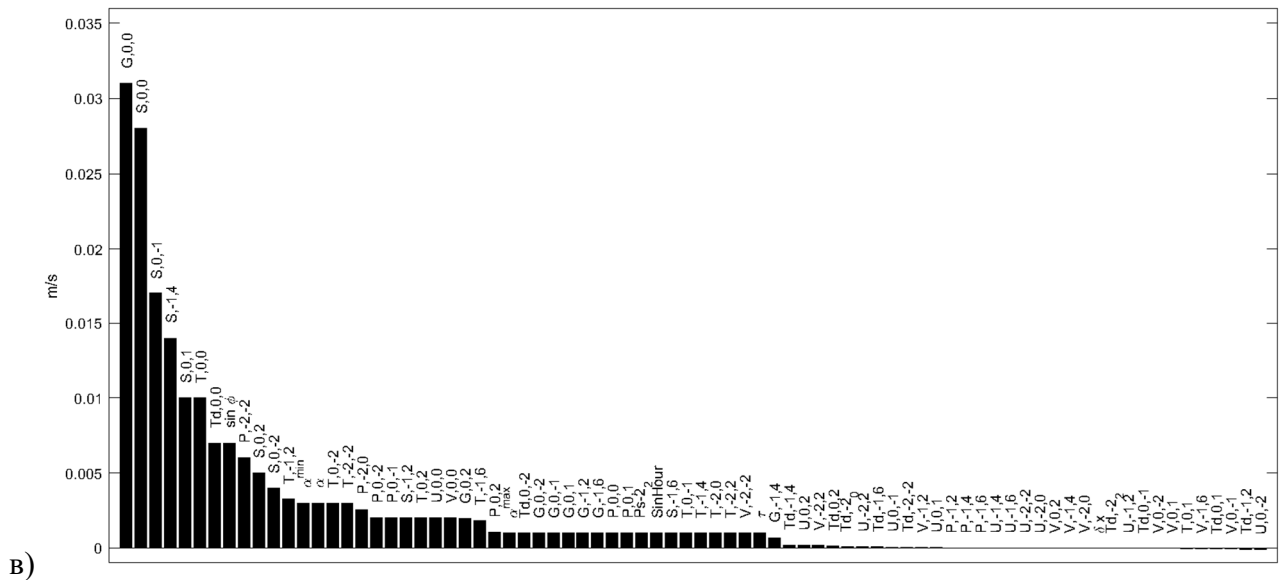


Рисунок 3.6 – Важность различных предикторов метода КНС для коррекции приземных прогнозов COSMO-Ru+КСК: а) при коррекции температуры воздуха T , °C; б) при коррекции точки росы Td , °C; в) при коррекции модуля скорости ветра $|\vec{U}|$, м/с

Из рисунков 3.6а, 3.6б видно, что при коррекции температуры воздуха T и точки росы Td наибольшую важность имеют предикторы $T(t, \tau)$ и $Td(t, \tau)$, а наименьшие важности у предикторов, описывающих вектор ветра $Z = U, V$. При коррекции модуля скорости ветра (рисунок 3.6в) наибольшую важность имеют предикторы, равные модулю скорости $S(t, \tau)$ и скорости порывов ветра $G(t, \tau)$.

Важность оказалась отрицательной лишь в случае коррекции скорости ветра (рисунок 3.6в), причем минимальное значение составляло всего -1.2×10^{-4} м/с. Таким образом, не удалось обнаружить предикторы, которые сколько-нибудь значительно ухудшают качество итогового прогноза, но при этом важность некоторых из предикторов существенно меньше, чем для других. Этот факт косвенно говорит о том, что используемого архива оказалось достаточно для рассмотрения используемого набора предикторов.

Если сбор информации о предикторах с низкой важностью требовал бы много ресурсов, то такие предикторы можно было бы отбросить и провести оптимизацию модели МО заново, однако поскольку в данном случае это не так, то такое отбрасывание не производилось.

Выводы из главы 3

При коррекции прогноза с заблаговременностью τ систематическая коррекция и комплексификация (глава 2) учитывает лишь прогнозы той же заблаговременности τ и того же

метеорологического параметра. Предложенный в главе 3 метод КНС учитывает такие дополнительные предикторы (рассмотренные списки предикторов описаны в **первой части главы**), как прогнозы метеорологических параметров, отличных от корректируемого, а также прогнозы с заблаговременностью, отличной от заблаговременности τ корректируемого прогноза.

Во второй части главы были рассмотрены вопросы подбора оптимальных гиперпараметров (понятие гиперпараметра введено в параграфе 1.1.11): используемого алгоритма градиентного спуска и его параметров (параграф 3.2.1); параметров регуляризации и инициализации (параграф 3.2.2); архитектуры нейронной сети (параграф 3.2.3); количества нейронов на скрытых слоях (параграф 3.2.3).

В третьей части главы рассматривается вопрос использования дополнительных параметров, индивидуальных для каждой синоптической станции. Показано (таблица 3.5), что использование индивидуальных параметров может быть полезно только при коррекции исходных прогнозов ЧМПП, но не при дополнительной коррекции прогнозов, подвергнутых КСК или ККП (глава 2). Таким образом, при предварительном использовании КСК и ККП достаточно иметь архив наблюдений за последние $t_l = 35$ дней, используемый для применения КСК или ККП. Для оптимизации же рассматриваемого в этой главе метода КНС не используется многолетний архив наблюдений на всех синоптических станциях, для которых прогноз будет составляться в оперативном режиме.

В отличие от рассмотренных в главах 2 и 4 методов (являющихся частично интерпретируемыми, то есть «серыми ящиками»), метод КНС является «черным ящиком». Поскольку нейронная сеть является нелинейной функцией, то вопрос оценки важности (**четвертая часть главы**) тех или иных используемых предикторов является нетривиальным. Было показано, что наибольшую важность имеют значения прогностических полей той же заблаговременности, а наименьшую – направление ветра.

ГЛАВА 4. Интерполяция поправок по горизонтальным координатам

Рассмотрим задачу интерполяции (анализа) поправок, вычисленных в предыдущих главах в точки сетки некоторой ЧМПП, прогноз которой будем называть первым приближением. Разумеется, используемое при анализе первое приближение существенно влияет на окончательный результат интерполяции. В метеорологии, как правило, интерполируют не сами поля физических параметров, а их инкременты [120]. Термин инкремент используется в различных математических моделях существенно различными способами. Здесь термин инкремент означает отклонение интересующего нас поля от его первого приближения.

Предлагаемая интерполяция применима для интерполяции любых других двумерных полей (в том числе и при отсутствии полей первого приближения).

Первый раздел главы посвящен истории исследований статистической структуры и корреляционных функций (КФ) метеорологических полей. В параграфе 4.1.1 будет описан метод оптимальной интерполяции, для применения которого необходимо сначала оценить КФ интерполируемого поля. В параграфе 4.1.2 приведены известные параметрические семейства КФ однородных изотропных случайных полей. В параграфах 4.1.3 и 4.1.4 описывается подход выравнивающего отображения (feature mapping), позволяющий свести задачу построения КФ неоднородного неизотропного поля к задаче построения КФ однородного и изотропного поля в пространстве большей размерности.

Во втором разделе главы суммированы результаты предыдущих пунктов и приведена предлагаемая диссертантом квазилинейная неоднородная оптимальная интерполяция (КНОИ). КНОИ использует нейронные сети и приближенно оценивает КФ неоднородного поля по некоторым известным предикторам неоднородности. В параграфах 4.2.1 и 4.2.2 обсуждаются особенности реализации и вычислительные аспекты предлагаемого алгоритма, а в параграфах 4.2.3 и 4.2.4 – результаты численных экспериментов.

4.1. Корреляционные функции и оптимальная интерполяция

При интерполяции геофизических полей по горизонтальным координатам [126], [47], как правило, используют метод оптимальной интерполяции (ОИ) [36], или метод кригинга [54], [190], [217]. Стандартные реализации методов ОИ и кригинга теоретически эквивалентны, отличие заключается в том, что метод оптимальной интерполяции использует оценку положительно определенной корреляционной функции (КФ) интерполируемого поля, а метод кригинга – оценку структурной функции.

Метод ОИ используется в метеорологических задачах более полувека. В частности, он был реализован диссертантом при обработке полей влажности почвы [24], [23], [22].

Существует множество реализаций методов ОИ и кригинга [47]. Эти реализации отличаются способом оценки полей средних (первого приближения), дисперсий и КФ (вариограммы) интерполируемой функции.

При универсальном кригинге [101], [6] (другое название кригинг с трендом) оценивают разложение поля средних по некоторому набору базисных функций. Близкая к универсальному кригингу идея использования нейронных сетей для оценки поля первого приближения предложена в [169], [168] для задач интерполяции данных о черномыльских радиоактивных осадках.

Отметим так же работу [212], в которой результат интерполяции данных о экстремальных температурах в Великобритании искался в виде нейронной сети, зависящей от широты, долготы и географических параметров (таких как параметры орографии, расстояния до побережья, крупных рек).

В диссертации поле средних будет оцениваться с помощью нейронной сети, принимающей на вход дополнительные параметры (предикторы неоднородности), причем если в [169], [168] нейронная сеть использовалась для оценки глобального тренда, то тут нейронная сеть оценивает локальную изменчивость.

КФ описывают статистические связи значений метеорологических полей в двух точках пространства (или времени, или пространства-времени). При оценке КФ необходимо обеспечить положительную определенность полученной оценки. Существует два принципиально разных подхода к оценке статистической структуры метеорологических полей: подход, основанный на изучении данных наблюдений за метеорологическими полями и подход, основанный на ансамблевых ЧМПП.

Система ансамблевого усвоения данных [159], [242], [188] использует ансамблевый прогноз погоды [8] и предполагает, что рассматриваемая ЧМПП адекватно описывает неоднородности метеорологических полей, а ансамблевая система прогнозирования хорошо описывает распределение возможных состояний атмосферы. Классический ансамблевый прогноз погоды использует прогнозы ЧМПП, полученные по нескольким различным возмущенным начальным метеорологическим полям, отличающихся друг от друга в пределах точности системы усвоения данных [77]. Однако этого оказывается недостаточно для качественной оценки распределения значений метеорологических полей, поэтому применяют и другие методы генерации ансамблей, например, рассматривающие в том числе и возмущения параметризаций ЧМПП [94].

Важным преимуществом ансамблевого подхода является возможность оценки неоднородной КФ с учётом локальной синоптической ситуации [159]. Для оценки более крупномасштабных КФ необходимо увеличивать количество членов ансамбля с нескольких десятков до тысяч [231], что, как правило, невозможно. Поэтому на практике в ансамблевом подходе рассматривают лишь локальные КФ.

В неансамблевом подходе при оценке неоднородных КФ, напротив, как правило учитывают лишь глобальную изменчивость (например, оценивают КФ отдельно для нескольких широтных поясов). Очевидно, с ростом числа наблюдений решающее значение будет иметь описание локальных неоднородностей.

Далее будет рассматриваться только подход, основанный на изучении статистической структуры наблюдений за метеорологическими полями.

При изучении статистической структуры метеорологических полей, часто предполагают, что эти поля (или их инкременты) – реализации случайных полей, и что эти случайные поля изотропны и однородны по горизонтальным координатам, то есть совместное распределение значений метеорологического поля в двух точках инвариантно относительно совместного движения по горизонтали двух точек, поведение метеорологических величин в которых исследуется. Для оптимальной интерполяции достаточно более слабого предположения о однородности и изотропности второго момента, то есть инвариантности не совместного распределения, а только КФ. Из предположения о однородности и изотропности следует, что корреляционная функция метеорологических полей в двух исследуемых точках, расположенными на одной высоте (барическом уровне) зависит только от расстояния между двумя этими точками.

Теория КФ стационарных случайных процессов и однородных изотропных случайных полей [87], [53], [243], [38], [92] применялась для изучения статистической структуры метеорологических полей. Теория корреляционных функций однородных изотропных случайных полей помогла развить технологию двумерного объективного анализа, основанную на методе оптимальной интерполяции [35], [22].

При усвоении метеорологических данных в качестве математического ожидания интерполируемого поля лучше использовать не климатические поля, а прогноз небольшой заблаговременности [88]. Для этого исследуется статистическая структура не самих метеорологических полей (их отклонений от климатических полей), а уже полей погрешностей ЧМПП [2].

Исследовалась уже не двумерная структура отдельных метеорологических полей [144], а их трехмерная структура [140], [99] и кросскорреляционные связи различных

метеорологических полей на разных высотах [4]. Трехмерная статистическая структура используется в технологии трехмерного вариационного усвоения данных 3D-Var [106].

При изучении трехмерной структуры метеорологических полей делают дополнительные предположения о том, как именно их КФ K зависит от сдвигов по горизонтали и по вертикали. Например, предполагают, что КФ представима в виде произведения однородных и изотропных КФ по горизонтальным и вертикальной координатам (в качестве вертикальной координаты z может использоваться не только высота, давление или логарифм давления, но и, например, потенциальная температура [211]):

$$K((z_1, \bar{x}_1), (z_2, \bar{x}_2)) = K_z(|z_1 - z_2|) K_x(|\bar{x}_1 - \bar{x}_2|).$$

Все эти предположения, очевидно, являются весьма грубыми [4], [1]. Особенно это заметно для приземных метеорологических полей, поскольку на приземные поля сильно влияет неоднородность подстилающей поверхности Земли: горы, моря, озера [197], леса, города [236] и другие географические объекты. Для метеорологических полей в свободной атмосфере гипотеза изотропности метеорологических полей по горизонтальным переменным в p -системе координат была подвергнута проверке на большом архиве аэрологических наблюдений (в предположении однородности внутри одной широтной зоны). Эта проверка подтвердила сильную анизотропию в некоторых, довольно обширных частях атмосферы Земли [1].

Частично отказавшись от однородности и изотропности метеорологических полей, можно вычислять корреляционные функции не для всевозможных пар точек, а отдельно для лежащих с одной стороны (с разных сторон) от атмосферного фронта. Эти две (вычисленные автором диссертации в работах [3] и [26]) корреляционные функции (рисунок 2.3) заметно отличаются друг от друга. Однако, этих двух корреляционных функций недостаточно для реализации метода оптимальной интерполяции: необходимо гарантировать положительную определенность сложной КФ, учитывающей атмосферные фронты, при любой конфигурации атмосферных фронтов. Кроме того, разрывная КФ будет приводить к разрывным метеорологическим полям.

Отметим, что используемый в некоторых работах индекс Морана (Moran's I, [198], [183]) позволяет оценить неоднородности интерполируемого поля в предположении неподвижности этих неоднородностей (что не выполнено для атмосферных фронтов).

Увеличение количества и качества наблюдений вне заранее установленных сроков (в первую очередь, за счет наблюдений с искусственных спутников Земли, с самолетов и метеорологических радаров, а также за счет увеличения числа наземных автоматических станций с весьма частыми измерениями) потребовало перейти от трехмерного (3D-Var) к четырехмерному (4D-Var) усвоению данных [124].

Чтобы отказаться от однородности и изотропности метеорологических полей, обеспечив положительную определенность КФ, можно оценивать *выравнивающее отображение* (другие названия: деформация, feature mapping) [218], [253], то есть инъекцию рассматриваемой области Ω в пространство H , в котором интерполируемое поле предполагается однородным и изотропным. Задача оценки выравнивающего отображения для реальных метеорологических полей весьма сложна: выравнивающее отображение меняется со временем и зависит от метеорологических параметров.

Известные подходы к оценке выравнивающего отображения для двумерных полей предлагают оптимизировать выравнивающее отображение в виде деформации, чтобы после деформации пространства рассматриваемая реализация поля переходила в как можно более однородное и изотропное поле. Под деформацией понимают [218] квазиконформное отображение, то есть биекцию, сохраняющую углы между кривыми почти во всех точках. В [218] и [97] предложено аппроксимировать выравнивающее отображение в виде сплайнов. В работах [249] и [246] выравнивающее отображение строится с применением методов машинного обучения.

Ниже в диссертации строится методика оценки условного выравнивающего отображения, общего для различных реализаций, но учитывающего дополнительные параметры, кратко описанная автором в [118]. Эти дополнительные параметры описывают причины неоднородностей полей метеорологических параметров (далее *предикторы неоднородности* Θ).

При оценке условного выравнивающего отображения нужно обеспечить его инъективность. Одно из возможных решений этой проблемы: рассматривать выравнивающее отображение, являющееся графиком некоторой функции [112]. В данной работе выравнивающее отображение будем искать в виде графика нейронной сети, принимающей на вход предикторы неоднородности. Поскольку рассмотренные предикторы неоднородности меняются со временем, то это выравнивающее отображение будет различаться в различные моменты времени.

Отметим, что во всех цитированных выше работах, при оценке КФ и выравнивающего отображений оптимизировалась рассматриваемая ковариационная модель (максимизировался логарифм правдоподобия ковариационной модели). При оптимизации ковариационной модели достаточно рассматривать лишь различные пары значений рассматриваемого поля и предположить, что значения интерполируемого случайного поля распределены согласно заранее выбранному вероятностному распределению. Выбор вероятностного распределения может существенно повлиять на качество интерполяции.

В данной же работе будем напрямую минимизировать погрешность интерполяции в точку «пропущенной» станции. Прямая оптимизация погрешности интерполяции существенно сложнее, поскольку при интерполяции рассматриваются наборы значений интерполируемого поля в нескольких точках ближайших синоптических станций, а не только в отдельных парах точек. Но при этом не делается никаких предположений о вероятностном распределении значений интерполируемого случайного поля, что должно позволить достичь лучшей точности полученной интерполяции.

Отметим, что все цитированные выше исследования проводились в предположении, что оптимизируемый оператор интерполяции (аппроксимации) является линейным по известным значениям интерполируемого поля. В данной работе предлагается рассматривать выравнивающее отображение, учитывающее известные значения интерполируемого поля инкремента. Таким образом, рассматриваемый оператор интерполяции будет квазилинейным.

4.1.1. Метод оптимальной интерполяции

Рассмотрим задачу интерполяции некоторого скалярного поля $X(\vec{x})$, значения которого известны в некотором нерегулярном наборе точек $\{\vec{x}_i | i = 1 \dots N\}$ в области $\Omega \subset \mathbb{R}^n$, в некоторый другой набор точек в той же области $\{\vec{y}_j | j = 1 \dots N'\} \subset \Omega \subset \mathbb{R}^n$. В математической метеорологии как правило интерполируют не сами значения поля X , а инкремент: разность его значений $\Delta X(\vec{x}) = X(\vec{x}) - X_0(\vec{x})$ с каким-то его первым приближением $X_0(\vec{x})$. В качестве первого приближения $X_0(\vec{x})$ здесь рассматривались прогнозы по модели COSMO-Ru в конфигурациях COSMO-Ru6ENA и COSMO-RuBy.

В задаче усвоения данных интерполируются погрешности прогноза, то есть разность измеренного поля X_{fact} и прогноза X_0 :

$$\Delta X(\vec{x}) = X_{fact}(\vec{x}) - X_0(\vec{x}). \quad (4.1)$$

В данной работе будем искать поправленное скалярное прогностическое поле. Для вычисления откорректированного поля интерполируем поправки, вычисленные КСК (глава 2) согласно формуле (2.1), то есть интерполируемые инкременты равны:

$$\Delta X(\vec{x}_i) = X_C(\vec{x}_i) - X_0(\vec{x}_i). \quad (4.2)$$

Также рассматривалась задача по интерполяции не откорректированного прогноза X_C , а комплексного прогноза X_{comp} , расчет которого описан во второй части главы 2, тогда интерполируемые инкременты были равны:

$$\Delta X(\vec{x}_i) = X_{comp}(\vec{x}_i) - X_0(\vec{x}_i). \quad (4.3)$$

Допустим, что для каждой точки $\vec{y}$ из набора $\{\vec{y}_j\}$ уже определено множество влияющих точек $I(\vec{y})$ из набора $\{\vec{x}_i\}$. При оптимальной интерполяции ищут оптимальную линейную формулу интерполяции поправок (4.2) или (4.3) вида

$$INC_X(\vec{y}, I) = \mu(\vec{y}) + \sum_{i \in I} w_X(\vec{x}_i, \vec{y}) [\Delta X(\vec{x}_i) - \mu(\vec{x}_i)], \quad (4.4)$$

где $INC_X(\vec{y}, I)$ – поле инкремента (результат интерполяции); w_X – интерполяционные веса; $\mu(\vec{x})$ – математическое ожидание случайного поля $\Delta X(\vec{x})$. Будем искать оптимальные веса w_X и математическое ожидание $\mu(\vec{x})$.

Говоря о линейности интерполяционной формулы (4.4), имеется ввиду линейная зависимость поля инкрементов $INC_X(\vec{y})$ от известных значений инкремента $\Delta X(\vec{x}_i)$, но не от координат $\vec{y}$. Все строящиеся ниже поля инкрементов $INC_X(\vec{y})$ не являются линейными функциями координат $\vec{y}$, поскольку оптимальные интерполяционные веса $w_X(\vec{x}_i, \vec{y})$ – нелинейные функции координат $\vec{y}$.

Далее будет также рассмотрен случай, когда интерполяционные веса w_X зависят от известных значений инкремента $\Delta X(\vec{x}_i)$. В этом случае можно говорить о квазилинейной интерполяционной формуле вида (4.4).

В данной работе использовался простой алгоритм, выбирающий в качестве влияющих 16 ближайших к $\vec{y}$ точек. Число 16 было выбрано экспериментально, а также поскольку на современных компьютерах решение систем линейных уравнений с числом неизвестных, кратным 8, утилизирует большую часть доступной вычислительной производительности.

Существуют более сложные способы выбора влияющих точек. Например, в [41] предлагается «кулоновский» алгоритм выбора с учетом взаимного расположения влияющих точек $\{\vec{x}_i\}$. По результатам экспериментов из [41] «кулоновский» алгоритм в среднем обеспечивает положительность большего количества интерполяционных весов w_X по сравнению с выбором ближайших. Подробнее вопрос возникновения отрицательных весов рассмотрен в параграфе 4.2.1.

Значение поля инкрементов $INC_X(\vec{y})$ может быть вычислено по формуле (4.4) при любом $\vec{y}$. Значения же $\Delta X(\vec{y})$ вычисляются по систематической коррекцией (глава 2), которая применима только в местах расположения синоптических станций. Там, где известны оба значения и $INC_X(\vec{y})$, и $\Delta X(\vec{y})$, они совпадают, если выполнено условие $\lim_{\vec{x} \rightarrow \vec{y}} K(\vec{x}, \vec{y}) = 1$.

Поле $X_{interp}(\vec{y}) = X_0(\vec{y}) + INC_X(\vec{y}, I)$ является оценкой поля $X_{fact}(\vec{y})$, однако некоторые из полей X_{interp} , могут быть не согласованы между собой, то есть описывать физически невозможное состояние системы, например точка росы может при такой интерполяции получиться больше температуры $Td_{interp} > T_{interp}$. Чтобы этого избежать, следуя параграфу 1.1.2, применим согласование, аналогичное формулам (2.2):

$$X_A(\vec{y}) = adapt(X_0(\vec{y}) + INC_X(\vec{y}, I)),$$

где $adapt$ – оператор согласования, формула (1.3). Далее будем оптимизировать функционалы качества (параграф 1.1.1) для прогнозов полей X_A , где $X = P, T, Td, \vec{U}, G$. Интерполяционные веса для компонент ветра и порывов всегда выбирались совпадающими между собой $w_U = w_V = w_{|\vec{U}|} = w_G$.

Будем оптимизировать формулу интерполяции (4.4) по интерполяционным весам w_X и по оценке математического ожидания μ , используя в качестве заранее заданного штрафного функционала среднее значение функции потерь e (используемые функции потерь приведены в параграфах 1.1.1 и 1.1.2), вычисляемую по прогнозу, полученному интерполяцией откорректированного прогноза X_C на «пропущенную» станцию:

$$L(w_X, b_X) = \frac{1}{N} \sum_{i=1}^N e(X_A(\vec{x}_i, I(\vec{x}_i) \setminus \{i\}, w_X, \mu), X_{fact}(\vec{x}_i)) \rightarrow \min_{w_X, \mu}. \quad (4.5)$$

Осреднение в штрафном функционале (4.5) производится по множеству по очереди «пропускаемых» станций.

В случае интерполяции инкрементов, отличных от погрешностей (4.1), например инкрементов (4.2) или (4.3), измеренное значение $X_{fact}(\vec{x}_i)$ неизвестно в момент интерполяции и отличается от интерполируемого результата коррекции $X_C(\vec{x}_i)$ или $X_{comp}(\vec{x}_i)$. Поэтому помимо задачи (4.5), можно рассматривать задачу о взаимном уточнении результатов систематической коррекции для различных наборов соседних синоптических станций:

$$L(w_X, b_X) = \frac{1}{N} \sum_{i=1}^N e(X_A(\vec{x}_i, I_i(\vec{x}_i), w_X, \mu), X_{fact}(\vec{x}_i)) \rightarrow \min_{w_X, \mu}. \quad (4.6)$$

В результате проведенных численных экспериментов по минимизации функционала (4.6) диссертанту не удалось достичь значимого (более чем на 2%) улучшения оценок прогнозов $X_A(\vec{x}_i)$, полученных минимизацией функционала (4.6) по сравнению с оценками прогнозов $X_C(\vec{x}_i)$.

При оптимальной интерполяции [36] предполагаются заранее известными детерминированные функции: средние $\mu(\vec{x})$, дисперсии $\sigma^2(\vec{x})$ и нормированная корреляционная функция (далее НКФ) $R(\vec{x}, \vec{y})$ интерполируемого случайного поля $\Delta X(\vec{x})$:

$$\begin{aligned} E[\Delta X(\vec{x})] &= \mu(\vec{x}); \\ E[\Delta X(\vec{x}) - \mu(\vec{x})]^2 &= \sigma^2(\vec{x}); \\ E[(\Delta X(\vec{x}) - \mu(\vec{x}))(\Delta X(\vec{y}) - \mu(\vec{y}))] &= \sigma(\vec{x})R(\vec{x}, \vec{y})\sigma(\vec{y}). \end{aligned}$$

Здесь E – математическое ожидание. Как правило предполагают, что интерполируемые инкременты ΔX содержат ошибку наблюдений $\zeta(\vec{x}_i)$:

$$\Delta X(\vec{x}_i) = \Delta X'(\vec{x}_i) + \zeta(\vec{x}_i),$$

где $\Delta X'(\vec{x})$ – непрерывное поле истинных инкрементов.

Предположим, что ошибки наблюдений $\zeta(\vec{x}_i)$:

- а) несмещенные, то есть $E\zeta(\vec{x}) \equiv 0$ (напомним, что мы интерполируем поправки постпроцессинга (4.2));
- б) не коррелируют с истинными инкрементами $\Delta X'$:

$$E[\zeta(\vec{x})\Delta X'(\vec{y})] \equiv 0;$$

- в) взятые в различных точках не коррелируют между собой, то есть их КФ $\psi(\vec{x}, \vec{y})$ пропорциональна индикатору нуля I_0 :

$$\psi(\vec{x}, \vec{y}) = E[\zeta(\vec{x})\zeta(\vec{y})] = \varepsilon\sigma(\vec{x})I_0(\vec{x} - \vec{y})\sigma(\vec{y}), \quad (4.7)$$

где $1 > \varepsilon > 0$ – константа, которую будем называть относительной нормой шума.

Из предположений б) и в) вытекает, что НКФ R инкрементов ΔX и НКФ K истинных инкрементов $\Delta X'$ связаны соотношением:

$$R(\vec{x}, \vec{y}) = (1 - \varepsilon)K(\vec{x}, \vec{y}) + \varepsilon I_0(\vec{x} - \vec{y}). \quad (4.8)$$

Рассмотрим квадратичную функцию потерь $e(X_{fact}, X_{pred}) = (X_{fact} - X_{pred})^2$.

Минимизируя (4.5), оптимальные интерполяционные веса w_X находим по формулам

многомерной линейной регрессии, решая СЛАУ с симметричной положительно определенной матрицей R_{xx} :

$$w_X(\{\vec{x}\}, \vec{y}) = \sigma_x^{-1} \circ R_{xx}^{-1} K_{xy} \sigma(\vec{y}) = \sigma_x^{-1} \circ \bar{w}_X \sigma(\vec{y}), \quad (4.9)$$

где K_{xy} – столбец, состоящий из значений КФ $K(\vec{x}_i, \vec{y})$; σ_x^{-1} – столбец, состоящий из чисел, обратных к $\sigma(\vec{x}_i)$; R_{xx} – корреляционная матрица значений поля ΔX в точках множества $\{\vec{x}_i\}$, которая согласно формуле (4.8) равна

$$R_{xx} = K_{xx} + \varepsilon I,$$

где K_{xx} – корреляционная матрица значений поля $\Delta X'$ в точках множества $\{\vec{x}_i\}$; εI – корреляционная матрица ошибок наблюдений, I – единичная матрица. Нормированные интерполяционные веса $\bar{w}_X(\vec{x}_i, \vec{y})$ вычисляются по формуле:

$$\bar{w}_X(\{\vec{x}\}, \vec{y}) = w_X(\{\vec{x}\}, \vec{y}) \sigma_x / \sigma(\vec{y}) = R_{xx}^{-1} K_{xy}.$$

Чрезвычайно важно, чтобы используемая при оптимальной интерполяции используемая оценка НКФ $K(\vec{x}, \vec{y})$ была ядром положительно определенного линейного оператора. Действительно, если это не так, то при некотором наборе влияющих станций $\{\vec{x}_i\}$ корреляционная матрица R_{xx} может оказаться сингулярной, что или сделает невозможным вычисление интерполяционных весов w_X по формуле (4.9), или приведет к большой погрешности интерполяции.

Обычно спектр линейных операторов K_{xx} хотя и положителен, но накапливается к нулевому значению [172]. Такое накопление может приводить к тому, что корреляционная матрица, составленная из недостаточно точно вычисленных значений корреляционной функции, не будет положительно определенной. В монографии [45] предложен алгоритм регуляризации матриц, которые «немного утратили свою положительную определенность». Минимальными поправками алгоритм превращает её в положительно определенную с обусловленностью не хуже заранее заданного числа. В статьях [4], [26] этот алгоритм, основанный на теории возмущений эрмитовых операторов, был обобщен для регуляризации КФ и применен к оценкам КФ аэрологических полей.

На практике, чтобы обеспечить положительную определенность НКФ $K(\vec{x}, \vec{y})$, её можно искать среди некоторого заранее выбранного семейства положительно определенных НКФ $K(\vec{x}, \vec{y}, \vec{v})$, где $\vec{v}$ – оптимизируемые параметры семейства.

4.1.2. Корреляционные функции однородного изотропного случайного поля

Как правило, при использовании оптимальной интерполяции предполагают однородность и изотропность интерполируемого случайного поля $\Delta X(\vec{x})$ по горизонтальным переменным [59], [35], [140]. При таком предположении НКФ $K(\vec{x}, \vec{y})$ зависит лишь от расстояния $r = |\vec{x} - \vec{y}|$ между точками:

$$K(\vec{x}, \vec{y}) = K(|\vec{x} - \vec{y}|) = K(r).$$

Обсудим сначала как связаны гладкость и непрерывность случайного поля инкрементов $INC_X(\vec{y})$ и гладкость и непрерывность функции $K(r)$. А затем укажем некоторые семейства положительно определенных НКФ $K(r)$.

Поле инкремента $INC_X(\vec{y})$ будет C^{2k} -гладко зависеть от $\vec{y}$, если и только если 1) $K \in C^{2k}[0, +\infty)$ и 2) первые k нечетных правых производных по r в нуле равны нулю:

$$\lim_{r \rightarrow +0} K'(r) = \lim_{r \rightarrow +0} K'''(r) = \dots = \lim_{r \rightarrow +0} K^{(2k-1)}(r) = 0.$$

Второе требование связано с тем, что полярные координаты имеют особенность в начале координат (переход от декартовых координат не является взаимно однозначным отображением). Если второе требование не выполнено, то при $\vec{y} \in \{\vec{x}_i\}$ поле инкремента $INC_X(\vec{y})$ не будет дифференцируемым. Действительно, устремляя точку $\vec{y}$ к точке $\vec{x}_i$ по противоположным направлениям, получим ненулевые значения производной противоположных знаков. Если случайное поле однородно, но не изотропно, то условия дифференцируемости функции в начале координат более сложное [44]. Условия гладкости КФ, не скалярного, а векторного поля на плоскости, приведены в [44], [43].

Отметим, что шум наблюдений $\varepsilon > 0$ можно интерпретировать как параметр регуляризации по Тихонову [79]. Действительно, в формуле (4.9) при решении плохо обусловленной системы линейных уравнений с матрицей K_{xx} мы заменили её на матрицу $R_{xx} = K_{xx} + \varepsilon I$.

Поскольку НКФ K неотрицательно определена, то минимальное собственное значение матрицы R_{xx} не меньше ε . Таким образом, при $\varepsilon > 0$ матрица R_{xx} в (4.9) строго положительно определена, даже если среди влияющих точек $\{\vec{x}_i\}$ есть повторяющиеся.

Согласно теореме о кругах Гершгорина [43], максимально возможное собственное значение матрицы R_{xx} не превосходит $m - (m-1)\varepsilon$, где m – количество влияющих точек

(размер квадратной матрицы R_{xx}). Это значение достигается только в случае, когда все точки $\bar{x}_i$ совпадают. Таким образом, число обусловленности матрицы R_{xx} не превосходит $m-1+m\varepsilon^{-1}$ при любых возможных наборах влияющих точек $\{\bar{x}_i\}$.

Опишем теперь способы представления НКФ однородного изотропного поля.

Согласно первой теореме Шёнберга [219] непрерывная функция $K(r)$, $r \in [0, +\infty)$ является НКФ изотропного однородного непрерывного скалярного случайного поля в пространстве $\mathbb{R}^d$ (принадлежит классу Φ_d) тогда и только тогда, когда она представима в виде

$$K(r) = \int_0^{+\infty} \left(\frac{2}{rt}\right)^{d/2-1} \Gamma\left(\frac{d}{2}\right) J_{d/2-1}(rt) d\xi(t), \quad (4.10)$$

где $\xi(t)$ – некоторая вероятностная мера (функция вероятностного распределения); $J_{d/2-1}$ – функция Бесселя первого рода $d/2-1$ порядка; Γ – гамма-функция Эйлера.

В этом случае, используя тождество

$$J_{-1/2}(x) = \cos x \sqrt{\frac{2}{\pi x}},$$

формулу (4.10) при $d=1$ можно сильно упростить и получить её из теоремы Бохнера – Хинчина [110], [87]:

$$K(r) = \int_0^{+\infty} \cos(rt) d\xi(t).$$

Согласно второй теореме Шёнберга [219] непрерывная функция $K(r)$, $r \in [0, +\infty)$ принадлежит классу $\Phi_\infty = \bigcap_{d=1}^{\infty} \Phi_d$ тогда и только тогда, когда она представима в виде

$$K(r) = \int_0^{+\infty} \exp(-r^2 t^2) d\xi(t), \quad (4.11)$$

где $\xi(t)$ – некоторая вероятностная мера.

Таким образом, для любого натурального d класс Φ_d содержит в том числе и знакопеременные функции, но все элементы Φ_∞ являются положительными убывающими функциями.

Классическим примером НКФ из $\Phi_\infty \cap C^\infty$, все нечетные производные которой в начале координат равны нулю является гауссовская КФ [141]:

$$K(r) = \exp(-r^2/2). \quad (4.12)$$

Отметим, что на практике часто вместо гауссовской НКФ (4.12) рассматривают её C^2 -гладкое приближение с ограниченным носителем. В работе [140] построена следующая положительно определенная НКФ $K_{GC}(r)$ из $\Phi_1 \cap \Phi_2 \cap \Phi_3$, приближающая гауссовскую (4.12):

$$K_{GC}(r) = \begin{cases} 1 - \frac{5}{3}r^2 + \frac{5}{8}r^3 + \frac{1}{4}r^4 - \frac{1}{4}r^5, & 0 \leq r < 1, \\ 4 - 5r + \frac{5}{3}r^2 + \frac{5}{8}r^3 - \frac{1}{2}r^4 + \frac{1}{12}r^5 - \frac{2}{3r}, & 1 \leq r < 2, \\ 0, & 2 \leq r. \end{cases} \quad (4.13)$$

Ограниченность носителя функции K_{GC} гарантирует, что влияние каждого инкремента $\Delta X(\vec{x}_i)$ на $INC_X(\vec{y})$ ограничено некоторым кругом радиуса $R_{\max}$ вокруг $\vec{x}_i$, то есть $w_X(\vec{x}_i, \vec{y}) \neq 0$ тогда и только тогда, когда $|\vec{x}_i - \vec{y}| < R_{\max} = 2$. Этот факт положительно влияет (ускоряет и стабилизирует) на вычисление интерполяционных весов.

Отметим, что на современных компьютерах вычисление экспоненты занимает в примерно 25 раз больше времени, чем умножение, поэтому вычисление K_{GC} (формула (4.13)) оказывается примерно в 3-4 раза быстрее вычисления гауссовской КФ.

Другим примером КФ из пересечения классов функций $\Phi_\infty \cap C^\infty$ служит НКФ [92], [105]:

$$K(r) = (1+r)\exp(-r). \quad (4.14)$$

Некоторые другие примеры НКФ из $\Phi_d, d \geq 2$ приведены в [144].

При минимизации погрешности интерполяции (4.5) удобно рассматривать в качестве КФ функцию $K(r)$ из заранее заданного конечномерного семейства положительно определённых НКФ $K(r, \vec{v})$, где $\vec{v}$ – параметры семейства НКФ. Параметры $\vec{v}$ будем оптимизировать, минимизируя среднюю погрешность интерполяции (4.5). Достигнутая наименьшая погрешность интерполяции (4.5) будет, очевидно, зависеть в том числе и от заранее выбранного семейства НКФ $K(r, \vec{v})$.

Поскольку вычисление значений НКФ, представленной в виде (4.10) вычислительно трудоёмко (требуется многократно вычислять значения функции Бесселя), что существенно увеличивает вычислительную сложность оптимизации погрешности интерполяции (4.5) по сравнению с НКФ, представленной в виде (4.11), то предпочтительным оказывается выбор НКФ из класса Φ_∞ . Вычислительная сложность вычисления НКФ весьма важна: для вычисления

значения поля инкрементов $INC_X(\vec{y})$ в одной точке $\vec{y}$ вычисляется $m(m+1)/2$ различных значений НКФ, где m – количество влияющих точек.

Укажем некоторые очевидные свойства НКФ, позволяющие легко конструировать семейства НКФ:

А) если $K(r)$ – НКФ и $R > 0$, то $K(r/R)$ – НКФ;

Б) если $K_1(r), K_2(r)$ – НКФ, $\alpha, \beta > 0, \alpha + \beta = 1$, то $\alpha K_1(r) + \beta K_2(r)$ – НКФ.

Используя свойства А и Б (которые легко вывести из теорем Шёнберга [219]), сконструируем параметрическое семейство НКФ, в котором будем искать НКФ интерполируемого однородного и изотропного поля.

Возьмем две НКФ: гауссовскую (4.12) и функцию (4.14) и с некоторыми весами сконструируем из этих двух НКФ трехпараметрическое семейство НКФ:

$$K(r, \vec{v}_0) = (1 - \beta)(1 + r/R_1) \exp(-r/R_1) + \beta \exp(-r^2/2R_2^2), \quad (4.15)$$

где $\vec{v}_0 = (\beta, R_1, R_2)$ – трехмерный вектор параметров семейства НКФ, оптимизируемых в смысле (4.5); $R_1, R_2 > 0$ – радиусы корреляции; $0 \leq \beta \leq 1$. Учитывая формулы (4.8) и (4.15), приходим к следующему представлению для четырехпараметрического семейства НКФ R :

$$R(r, \vec{v}) = \varepsilon I_0(r) + (1 - \varepsilon) \left[(1 - \beta)(1 + r/R_1) \exp(-r/R_1) + \beta \exp(-r^2/2R_2^2) \right], \quad (4.16)$$

где $\vec{v} = (\varepsilon, \beta, R_1, R_2)$ – четырехмерный вектор параметров семейства НКФ R . Перед стартом оптимизации погрешности интерполяции (4.5) использовались следующие начальные значения параметров семейства (4.140): $\beta = 0.8$, $R_1 = 100 \text{ км}$, $R_2 = 200 \text{ км}$, $\varepsilon = 1/3$.

Причины того, что рассматриваются только НКФ из класса Φ_∞ будут пояснены ниже, в следующем параграфе 4.1.3.

4.1.3. Выравнивающее отображение

Реальные метеорологические поля, особенно приземные, на которые заметное влияние оказывает подстилающая поверхность, не однородны и не изотропны. Некоторые аэрологические поля в некоторых регионах также не изотропны [1], [3], [26].

Согласно теореме Мерсера [196], непрерывное ядро $K(x, y)$, где $x, y \in [0, 1]$, любого положительно определенного, интегрируемого в квадрате (класса L^2) линейного оператора, представимо в виде:

$$K(x, y) = \sum_{j=1}^{\infty} \lambda_j e_j(x) e_j(y),$$

где ряд сходится абсолютно и равномерно почти во всех точках; e_j – некоторый ортонормированный базис, состоящих из функций из $C \cap L^2[0;1]$; $\lambda_j \geq 0$ – собственные значения.

Теорема Мерсера обобщается на случаи:

- а. когда $\vec{x}, \vec{y} \in \Omega$ – любой связный компакт в банаховом пространстве (например, в $\mathbb{R}^n$) [222];
- б. векторзначных операторов [119];
- в. матрицезначных операторов [129].

Обобщения б) и в) должны позволить обобщить предлагаемую в данной работе интерполяцию на случай совместной интерполяции нескольких метеорологических параметров.

Рассмотрим отображение $g: \Omega \rightarrow H$ в бесконечномерное гильбертово пространство H , вычисляемое согласно формуле

$$g(\vec{x}) = (e_1(\vec{x})\sqrt{\lambda_1}, e_2(\vec{x})\sqrt{\lambda_2}, \dots), \quad (4.17)$$

где e_j и $\lambda_j \geq 0$ взяты из теоремы Мерсера. Это отображение переводит компакт Ω в бесконечномерное гильбертово пространство H . Для отображения g выполнено равенство:

$$K(\vec{x}, \vec{y}) = \langle g(\vec{x}), g(\vec{y}) \rangle_H, \quad (4.18)$$

где $\langle \cdot, \cdot \rangle_H$ – скалярное произведение в H .

Таким образом, из теоремы Мерсера следует, что для положительно определенного ядра K существует отображение $g: \Omega \rightarrow H$ в некоторое гильбертово пространство H , такое что после применения отображения g ядро рассматриваемого положительно определенного оператора переходит в скалярное произведение в *воспроизводящем ядро K гильбертовом пространстве H* [7] (*reproducing kernel Hilbert space* [102], [222]). Отображение g называют *выравнивающим отображением (feature mapping)*.

Верно и обратное: если существует непрерывное отображение $g: \Omega \rightarrow H$ в некоторое гильбертово пространство H , то $K(\vec{x}, \vec{y})$ из (4.18) – непрерывное ядро неотрицательно определенного оператора. Действительно, корреляционная матрица K_{xx} равна неотрицательно определенной матрице Грама системы векторов $\{g(\vec{x}_i)\}$.

В случае, когда ядро $K(\vec{x}, \vec{y})$ положительно определенного оператора является НКФ случайного поля, то для всех $\vec{x} \in \Omega$ выполнено:

$$1 = K(\vec{x}, \vec{x}) = \langle g(\vec{x}), g(\vec{x}) \rangle_H = \|g(\vec{x})\|_H,$$

то есть образы $g(\vec{x})$ всех точек $\vec{x} \in \Omega$ лежат на единичной сфере в $\mathbb{S}^\infty \subset H$. Используя тождество параллелограмма, получаем:

$$\begin{aligned} K(\vec{x}, \vec{y}) &= \langle g(\vec{x}), g(\vec{y}) \rangle_H = \frac{1}{2} [\|g(\vec{x})\|_H + \|g(\vec{y})\|_H - \|g(\vec{x}) - g(\vec{y})\|_H] = \\ &= 1 - \frac{1}{2} \|g(\vec{x}) - g(\vec{y})\|_H, \end{aligned}$$

то есть выравнивающее отображение g переводит рассматриваемое случайное поле с НКФ $K(\vec{x}, \vec{y})$ в однородное и изотропное поле f^* на $g(\Omega)$.

Назовем нормированным случайным полем $f_\xi(\vec{x})$ случайное поле удовлетворяющее равенствам $Ef_\xi(\vec{x}) = 0$, $Ef_\xi(\vec{x})^2 = 1$ для всех рассматриваемых $\vec{x}$. Выше продемонстрировано, что для нормированных случайных полей справедлива следующая лемма:

Лемма. Для любого нормированного непрерывного случайного поля $f_\xi(\vec{x})$ на компакте $\vec{x} \in \Omega$ в банаховом пространстве с непрерывной корреляционной функцией $K(\vec{x}, \vec{y})$, существует непрерывное выравнивающее отображение $g(\vec{x})$, вкладывающее компакт Ω в некоторое гильбертово пространство H , в котором случайное поле $f_\xi^*(g(\vec{x})) = f_\xi(\vec{x})$ есть однородное и изотропное.

Утверждение леммы можно записать в виде коммутативной диаграммы:

$$\begin{array}{ccc} \Omega & \xrightarrow{f} & C(\Omega) \\ g \downarrow & & \uparrow g^* \\ H & \xrightarrow{f^*} & C(H). \end{array}$$

Отметим, что в приведенном выше примере (4.17) выравнивающее отображение g вкладывает компакт Ω в сферу $\mathbb{S}^\infty \subset H$, а структурная функция поля f^* есть метрика в H .

Далее будем искать выравнивающее отображение g , переводящее рассматриваемое поле в однородное и изотропное в пространстве H , а от требования, что структурная функция поля f^* есть метрика в H откажемся.

На практике вычислить отображение с бесконечномерным образом H невозможно и необходимо приблизить истинное выравнивающее отображение отображением с

конечномерным образом в $\mathbb{R}^{2+d}$, $d \geq 0$. В работе [137] было доказано, что если ядро $K(\vec{x}, \vec{y})$ симметричного положительно определенного оператора на n -мерной сфере липшицево, тогда собственные значения λ_j таковы, что при некотором $\delta > 0$ последовательность $j^{1+1/n+\delta} \lambda_j$ ограничена. Отсюда следует, что ряд в теореме Мерсера сходится достаточно быстро, и липшицево ядро $K(\vec{x}, \vec{y})$ можно приблизить, используя выравнивающее отображение с конечномерным образом.

Поскольку рассматриваемое выравнивающее отображение g с конечномерным образом $\mathbb{R}^{2+d}$ является лишь аппроксимацией отображения с бесконечномерным образом, поэтому предположение о принадлежности рассматриваемой НКФ (4.15) к классу Φ_∞ , а не Φ_{2+d} (параграф 4.1.2) весьма логично. Однако, очевидно, что если истинная НКФ поля f принимает в том числе и отрицательные значения, то мы не сможем её приблизить предлагаемым методом. Если же поле f^* можно непрерывно продлить с $g(\Omega)$ на все пространство H , то НКФ поля f^* неотрицательна и принадлежит классу Φ_∞ .

Задача оценки выравнивающего отображения для данного случайного поля весьма сложна: истинная КФ неизвестна, а известны лишь данные наблюдений за значениями рассматриваемого поля. В таких условиях заранее определить минимально возможную размерность гильбертова пространства H , при отображении в которое наше случайное поле станет перейдет в однородное и изотропное с некоторой заранее заданной погрешностью не представляется возможным.

Теорема Мерсера [196] несколько опередила своё время: ещё не была достаточно развита теория гильбертовых пространств. Первые исследователи теории КФ [110], [219], [87], [53], [243] не рассматривали выравнивающее отображение. Теория воспроизводящих ядро гильбертовых пространств появилась несколько позже [102]. История исследования ядер положительно определенных операторов и их приложений рассказана в [136].

Подход оценки КФ неоднородного анизотропного поля с помощью выравнивающего отображения не нов. В работах [218] и [97] для $\vec{x} \in \mathbb{R}^2$ были предложены способы оценки некоторой «оптимальной» деформации $g: \mathbb{R}^2 \mapsto \mathbb{R}^2$, такой, что после применения деформации рассматриваемая реализация поля $X(\vec{x})$ как можно более однородна и изотропна, то есть:

$$K(\vec{x}, \vec{y}) \approx K(|g(\vec{x}) - g(\vec{y})|) = K(r). \quad (4.19)$$

В работах [249] и [246] рассматриваются способы построения биекции $g: \Omega \leftrightarrow \Omega' \subset \mathbb{R}^2$, переводящей данную реализацию случайного поля в как можно более однородное и изотропное

гауссовское поле. В [249] для данных дистанционного зондирования Земли деформация области Ω строится методами машинного обучения в виде последовательности (нескольких десятков и сотен) конформных преобразований плоскости $\mathbb{R}^2$.

Упомянем также работу [165], в которой теория воспроизводящих ядер используется для оценки временной корреляционной функции для системы стохастических ОДУ.

4.1.4. Выравнивающее отображение как график нейронной сети

Будем приближать выравнивающее отображение $g: \Omega \rightarrow \mathbb{R}^{2+d}$ с использованием нейронной сети. Если не налагать никаких дополнительных условий на нейронную сеть, вычисляющую деформацию g , то отображение g может не быть инъекцией. То есть образы некоторых из влияющих точек могут совпадать: $g(\vec{x}_i) = g(\vec{x}_k)$ при $k \neq i$. Тогда рассматриваемое поле инкрементов после деформации неоднозначно и не является функцией на образе $g(\Omega)$. Хотя это делает необратимой матрицу K_{xx} , но из-за учёта шума наблюдений (при $\varepsilon > 0$) матрица R_{xx} остаётся обратимой, а оптимальная интерполяция (параграф 4.1.2) – применимой. Однако, этот факт может существенно усложнить оптимизацию выравнивающего отображения.

Поскольку при оптимизации нейронных сетей используется метод градиентного спуска, то выравнивающее отображение меняется малыми шагами в предположении малого количества локальных минимумов оптимизируемого функционала. В то же время, переход между топологически неэквивалентными образами $g(\Omega)$ сопровождается немонотонным изменением расстояний между влияющими точками, и немонотонным изменением оптимизируемого функционала. Поэтому предпочтительно использовать выравнивающее отображение, переводящее рассматриваемую область Ω в топологически эквивалентную ей.

При вычислении деформации g желательно учитывать известные параметры $\Theta(\vec{x})$, которые влияют на неоднородность рассматриваемого случайного поля, например, параметры, описывающие топографию, время суток на которое даётся прогноз, время года, подстилающую поверхность, тип землепользования и любые другие параметры, известные на модельной сетке. Далее параметры $\Theta(\vec{x})$, конкретные списки которых для каждого из интерполируемых метеорологических параметров X приведены в параграфе 4.2.2, будем называть **предикторами неоднородности**.

Значения предикторов неоднородности $\Theta(t, \vec{x})$ вообще говоря могут отличаться для различных реализаций интерполируемого поля $\Delta X(t, \vec{x})$. Во всех приведенных выше работах рассматривался случай деформации $g: \Omega \mapsto \Omega'$, где $\Omega, \Omega' \subset \mathbb{R}^2$. При учете же предикторов неоднородности отображение $\vec{x} \mapsto g(\vec{x}, \Theta(t, \vec{x})) \in \mathbb{R}^2$, как правило, не будет инъекцией. Поэтому, для учёта предикторов неоднородности $\Theta(\vec{x})$, будем искать вложение (инъекцию) $g: \Omega \mapsto \Omega \times \mathbb{R}^d, d \geq 1$, отображающее Ω в график некоторой функции ϖ от предикторов неоднородности $\Theta(t, \vec{x})$, то есть

$$g(t, \vec{x}) = \langle \vec{x}, \varpi(\Theta(t, \vec{x})) \rangle. \quad (4.20)$$

Оценивать отображение ϖ будем с помощью нейронной сети (далее сеть ϖ будем называть сетью вложения), которую будем оптимизировать, минимизируя итоговую погрешность интерполяции (4.5). Конкретные аргументы и вид нейронной сети ϖ , будут указаны ниже в параграфах 4.2.2 и 4.2.3.

В работе [112] вложение g (как и диссертации) строилось в виде графика некоторой функции $f(\vec{x})$, то есть вложение $g(\vec{x}) = \langle \vec{x}, f(\vec{x}) \rangle$. В [112] при построении $f(\vec{x})$ не использовались нейронные сети и не привлекались дополнительные параметры (предикторы неоднородности) $\Theta(\vec{x})$.

В данной работе предлагается свести задачу интерполяции неоднородного анизотропного поля в области $\Omega \subset \mathbb{R}^2$ к задаче интерполяции однородного и изотропного поля в области $\Omega \times \mathbb{R}^d \subset \mathbb{R}^{2+d}$ в пространстве на $d \geq 1$ большей размерности (далее использовался $d = 1 \dots 4$). Метод оптимальной интерполяции полей инкрементов $INC_X(\vec{y})$ по формуле (4.4) с весами, вычисленными по формуле (4.9), требует оценки КФ. Будем конструировать КФ K неоднородного анизотропного поля в виде (4.19), используя:

1. вложение $g: \Omega \mapsto \Omega \times \mathbb{R}^d$, оцениваемое как график нейронной сети ϖ , принимающей на вход предикторы неоднородности $\Theta(t, \vec{x})$, формула (4.20);

2. НКФ $K(r)$ однородного и изотропного поля в $2+d$ -мерном пространстве из семейства $K(r, \vec{v})$, формула (4.15), принадлежащему классу Φ_∞ .

Эквивалентная формулировка: применяем оптимальную интерполяцию в **расширенном** $2+d$ -мерном пространстве с НКФ $K(r)$.

Параметры семейства НКФ $\vec{v}$ в данной работе оптимизировались одновременно с отображением g , полем средних $\mu(\vec{x})$ и полем стандартных отклонений $\sigma(\vec{x})$. Критерий качества: минимум погрешности (4.5) прогноза (в «пропущенной» точке):

$$L(\vec{\theta}, \vec{v}) = L\left(INC_X\left(\varpi(\vec{\theta}), \mu(\vec{\theta}), \sigma(\vec{\theta}), \vec{v}\right)\right) \rightarrow \min_{\vec{\theta}, \vec{v}}, \quad (4.21)$$

где $\vec{\theta}$ – оптимизируемые параметры трех нейронных сетей ϖ, μ, σ . Нейронные сети ϖ, μ, σ принимают на вход предикторы неоднородности $\Theta(t, \vec{x})$, то есть значения полей, описывающих неоднородность интерполируемого поля.

В силу формулы (4.9) интерполяционные веса w , а значит и поле инкрементов $INC_X(\vec{y})$ зависят только от «относительных значений» $\sigma(\vec{x})/\sigma(\vec{y})$, но не от абсолютных $\sigma(\vec{x})$. Если сумма интерполяционных весов $\sum_i w(\vec{x}_i)$ равна 1, то инкременты $INC_X(\vec{y})$ зависят только от $\mu(\vec{x}) - \mu(\vec{y})$, но не от $\mu(\vec{x})$. Гистограмма сумм интерполяционных весов $\sum_i w(\vec{x}_i)$ показана на рисунке 4.1. Видно, что эта сумма в 95% случаев лежит в интервале от 0.98 до 1.04, то есть в абсолютном большинстве случаев погрешность интерполяции почти не зависит от абсолютных значений $\mu(\vec{x})$, а зависит только от относительных значений $\mu(\vec{x}) - \mu(\vec{y})$.

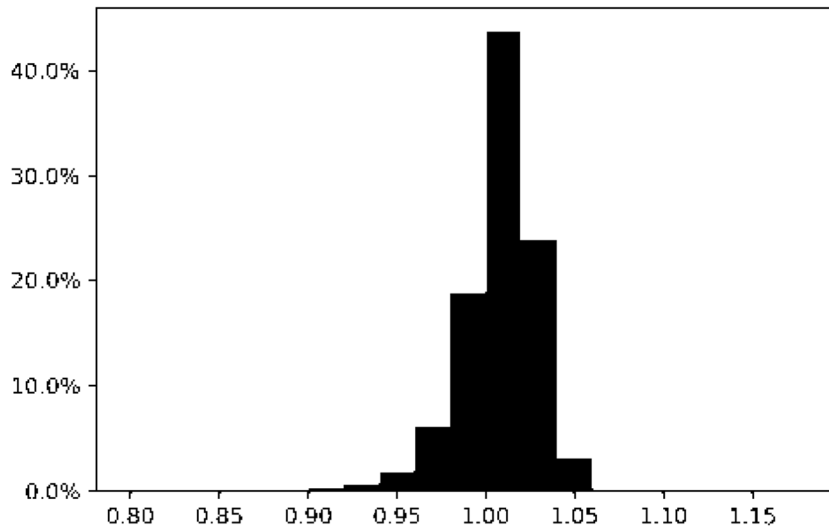


Рисунок 4.1 – Гистограмма сумм интерполяционных весов $\sum_i w(\vec{x}_i)$ при интерполяции поправок к температуре воздуха T

Поэтому оптимальные в силу минимальной погрешности (4.5) проинтерполированного прогноза

$$X_A(\bar{y}) = \text{adapt}(X_0(\bar{y}) + \text{INC}_X(\bar{y}, I))$$

нейронные сети $\mu(\bar{x})$ и $\sigma(\bar{x})$ определены с точностью до аддитивной константы $\bar{\mu}$ и до мультипликативной константы $\bar{\sigma}$ соответственно. Для различных реализаций интерполируемого поля инкрементов константы $\bar{\mu}$ и $\bar{\sigma}$ могут быть различными, поэтому задача оценки констант $\bar{\mu}$ и $\bar{\sigma}$ не является тривиальной при известных нейронных сетях μ и σ , оптимальных в силу (4.21).

Оценить абсолютные значения средних $\mu(\bar{x})$, стандартного отклонения $\sigma(\bar{x})$ можно, если предположить, что инкременты ΔX распределены согласно какому-то распределению, например гауссовскому. Плотность m -мерного гауссовского распределения с нулевыми средними, единичными дисперсиями и корреляционной матрицей R_{xx} равна [55]:

$$\rho(\bar{x}) = \frac{1}{(2\pi)^{m/2} \sqrt{\det R_{xx}}} \exp\left(-\bar{x}^T R_{xx} \bar{x} / 2\right), \quad (4.22)$$

где $\det R_{xx}$ – определитель матрицы R_{xx} .

Для оценки абсолютных значений средних $\mu(\bar{x})$ и стандартных отклонений $\sigma(\bar{x})$ вместо минимизации погрешности (4.5) прогноза $X_A(\bar{y})$ можно максимизировать логарифм правдоподобия того, что инкременты ΔX выбраны из гауссовского распределения со средним $\mu(\bar{x})$, дисперсией $\sigma^2(\bar{x})$ и корреляционной матрицей R_{xx} . Обозначим нормированные инкременты $\overline{\Delta X} = (\Delta X - \mu)/\sigma$, тогда за функцию потерь выберем минус логарифм правдоподобия ковариационной модели (4.22):

$$e(\bar{\theta}, \bar{v}) = \frac{1}{2} \left[\overline{\Delta X}^T R_{xx}^{-1} \overline{\Delta X} + \ln \det R_{xx} + m \ln 2\pi \right]. \quad (4.23)$$

По результатам численных экспериментов по оптимизации правдоподобия (4.23) погрешности интерполяции (4.5) оказывались заметно выше, чем при прямой оптимизации (4.5). Это вполне естественно: при прямой оптимизации погрешности интерполяции (4.5) не использовалось дополнительное предположение о том, что инкременты ΔX выбраны из гауссовского распределения. По-видимому, для оценки абсолютных значений средних μ и стандартных отклонений σ требуется учесть дополнительные предикторы неоднородности $\Theta(\bar{x})$, которые являются входными аргументами нейронных сетей μ и σ .

Во всех известных автору работах, оптимизирующих выравнивающее отображение [218], [97], [112], [249], [246] оптимизировался логарифм правдоподобия рассматриваемой ковариационной модели (например, (4.23)), а не погрешность интерполяции (4.5). Однако

максимизация правдоподобия рассматриваемой ковариационной модели не гарантирует минимизации погрешности интерполяции.

Во-видимому, прямая оптимизация погрешности интерполяции ранее никем не рассматривалась, поскольку она заметно сложнее оптимизации ковариационной модели. Действительно, при оптимизации ковариационной модели достаточно рассматривать только пары точек и при вычислении функции потерь (4.23) обращать корреляционную матрицу размера 2×2 . При прямой же оптимизации погрешности интерполяции необходимо обращать матрицу существенно большего размера (равному количеству влияющих станций m). Вычислительные сложности, возникающие при решении СЛАУ в градиентном машинном обучении, подробнее рассмотрены далее в параграфе 4.2.1.

В настоящей работе никакого существенного уменьшения (более чем на 1%) средней погрешности интерполяции (4.5) за счёт оптимизации нейронной сети σ достичь не удалось, поэтому нигде далее нейронная сеть σ не оптимизировалась и полагалось $\sigma(\vec{x}) \equiv 1$. Можно предположить, что поля стандартных отклонений σ рассмотренных метеорологических параметров являются более крупномасштабными чем типичные расстояния между станциями и их сложно оценить по используемым данным о значениях инкремента ΔX на соседних синоптических станциях.

Отметим, что в некоторых случаях рассматривается задача интерполяции не только по пространству, но и по времени. Предположение об однородности в пространстве-времени, может оказаться весьма грубым. Часто предполагают, что НКФ в пространстве-времени представима в некотором виде, например, в виде произведения НКФ в зависимости от сдвига по пространству $K_x(r)$ и НКФ в зависимости от сдвига по времени по времени $K_t(t_1 - t_2)$:

$$K((t_1, \vec{x}), (t_2, \vec{y})) = K_t(t_1 - t_2) K_x(|\vec{x} - \vec{y}|).$$

Более сложные конструкции, позволяющие получить из НКФ K_x и K_t семейства НКФ K в пространстве-времени, предложены в [125] и [99]. В данной работе не рассматривались НКФ, зависящие от сдвига по времени, но предлагаемый подход позволяет оценить НКФ неоднородного анизотропного случайного поля в пространстве-времени, ища вложение в несколько отличном от (4.20) виде, например в виде:

$$g(t, \vec{x}) = \langle \vec{x}, \varpi(t, \Theta(t, \vec{x})) \rangle$$

или в виде

$$g(t, \vec{x}) = \langle \vec{x}, \varpi_t(t), \varpi(t, \Theta(t, \vec{x})) \rangle,$$

где ϖ_t – некоторая нейронная сеть.

4.2. Квазилинейная неоднородная оптимальная интерполяция инкрементов

Предположим, что при интерполяции учитывается набор из n_Θ полей предикторов $\bar{\Theta}(t, \vec{x})$, влияющих на неоднородность и анизотропность интерполируемого поля $\Delta X(t, \vec{x})$ по пространству. Этот набор предикторов может включать любые известные поля, например: данные о землепользовании; данные об орографии; данные о растительности и типах почв; данные о времени суток и времени года. Значения предикторов неоднородности $\bar{\Theta}(t, \vec{x})$ могут различаться для различных реализаций (моментов времени t_j) случайного поля $\Delta X(t_j, \vec{x})$.

Применение гладкой инъекции $g: \Omega \mapsto \Omega \times \mathbb{R}^d$, формула (4.20), с гладкими по $\vec{x}$ предикторами неоднородности $\Theta(\vec{x})$ использование НКФ из семейства (4.15) гарантирует гладкость результата интерполяции $INC_X(\vec{y})$ по $\vec{y}$ в силу бесконечной гладкости всех используемых при расчётах формул (4.4), (4.9), (4.15).

При минимизации погрешности (4.5) прогноза $X_A(\vec{y})$ будем искать, формула (4.21), следующие параметры:

1. параметры семейства НКФ $\vec{v}$, формула (4.16);
2. параметры двух нейронных сетей: сети вложения ϖ и сети средних μ .

Сеть вложения ϖ , возвращает d аргументов и позволяет вычислить вложение $g: \Omega \rightarrow \Omega \times \mathbb{R}^d$, области $\Omega \subset \mathbb{R}^n$ в пространство на d большей размерности (далее будем называть это пространство *расширенным*):

$$g(\vec{x}) = g(\Theta(\vec{x}), X_0(\vec{x}), \Delta X(\vec{x})) = (\vec{x}, \varpi(\Theta(\vec{x}), X_0(\vec{x}), \Delta X(\vec{x}))). \quad (4.24)$$

Сеть средних $\mu(\Theta(\vec{x}), X_0(\vec{x}))$ оценивает относительные средние значения интерполируемого поля.

В данной работе предполагается, что параметр нормы шума ε , используемый в формуле (4.16), является константой (то есть, не зависит от $\vec{x}$), однако, в дальнейшем от этого предположения можно отказаться и вычислять этот параметр, например, нейронной сетью.

При вычислении нейронной сети вложения ϖ в точках $\vec{y}_j$, в которые интерполировалось поле, согласно формуле (4.24) используются значения $\Delta X(\vec{y}_j)$, которые неизвестны. Вместо $\Delta X(\vec{y}_j)$ подставим результат интерполяции $\Delta X(\vec{x}_i)$ в предположении однородности,

изотропности интерполируемого поля инкрементов и линейности интерполяционной формулы с НКФ $K(r)$ из семейства (4.15). Тогда, как и в случае линейной по ΔX оптимальной интерполяции, при малом шуме наблюдений $\varepsilon \rightarrow 0$ выполняется $INC_X(\bar{x}_i) \rightarrow \Delta X(\bar{x}_i)$, то есть результат интерполяции в точку станции при $\varepsilon \rightarrow 0$ стремится к известному значению.

Вложение g , очевидно, инъективно. В приведённых ниже результатах экспериментов рассматривались двумерные ($n=2$) реальные метеорологические поля, а размерность образа g лежала в пространстве на $d=1...4$ больше.

Итак, алгоритм интерполяции состоит из следующих шагов:

1. Отображение всех влияющих точек $\{\bar{x}_i\}$ и точек в которые производится интерполяция $\{\bar{y}_j\}$, в расширенное пространство согласно формуле (4.24) с использованием нейронной сети вложения ϖ ;
2. Выбор множества влияющих точек $I(\bar{y}_j) \subseteq \{\bar{x}_i\}$ для каждого $\bar{y}_j$;
3. Вычисление корреляционных матриц $R_{xx} = K_{xx} + \varepsilon I$ и K_{xy} согласно формуле (4.19) и выбранной НКФ в расширенном пространстве из семейства (4.15);
4. Оценка средних μ нейронной сетью;
5. Вычисление интерполяционных весов w по формуле (4.9);
6. Вычисление значений поля инкрементов $INC_X(\bar{y}_j)$ по формуле (4.4).

Приведенный алгоритм интерполяции инкрементов назовем квазилинейной неоднородной оптимальной интерполяцией (КНОИ). Напомним, что в пункте 2 в качестве влияющих точек выбирались ближайшие $m=16$ точек в расширенном пространстве.

Обсудим некоторые ожидаемые свойства КНОИ.

В областях, где поле $\vec{\varpi}(\bar{x})$ меняется слабо, КНОИ с g , вычисляемым по формуле (4.24) близка к оптимальной интерполяции однородного и изотропного случайного поля. КНОИ будет неоднородной и анизотропной в тех областях, где поле $\vec{\varpi}(\bar{x})$ – а значит и какое-то из полей $\vec{\Theta}(\bar{x})$ – имеет большой градиент. Если сеть вложения ϖ зависит от первого приближения $X_0(\bar{x})$, то КНОИ будет иметь способность уменьшать радиус корреляции в областях больших градиентов первого приближения X_0 (по сравнению с областями со слабо меняющимся первым приближением). В зонах больших градиентов предположение об однородности и изотропности полей может нарушаться. В [26] и [3] были получены оценки: существенно различные НКФ метеорологических полей для двух пар точек, лежащих по одну и по разные стороны от

атмосферного фронта, рисунок 2.3. По результатам этих оценок было показано, что во втором случае радиус корреляции примерно вдвое меньше.

Если сеть вложения ϖ не зависит от инкремента $\Delta X(\vec{x})$, то результат интерполяции INC линейно зависит от известных инкрементов $\Delta X(\vec{x})$, формула (4.4).

4.2.1. Вычисление интерполяционных весов

Предлагаемая КНОИ метеорологических полей имеет параметры: $\vec{\theta}$ – параметры нейронных сетей; $\vec{v} = (\varepsilon, \beta, R_1, R_2)$ – параметры семейства НКФ (4.16). Оптимизацию параметров $\vec{\theta}$ и $\vec{v}$ будем осуществлять методом градиентного спуска (параграф 1.1.7) в силу минимизации погрешности интерполяции (4.5). Для применения метода градиентного спуска необходимо вычислить градиенты $\nabla_{\vec{\theta}} e$ и $\nabla_{\vec{v}} e$ функции потерь e . Вычислять эти градиенты $\nabla_{\vec{\theta}} e$ и $\nabla_{\vec{v}} e$ будем методом обратного распространения ошибки [34], параграф 1.1.4.

Для вычисления коэффициентов нормированных интерполяционных весов $\bar{w}_X$ по формуле (4.9) решается система линейных алгебраических уравнений (СЛАУ) $R_{xx} \bar{w}_X = K_{xy}$. Обратное распространение ошибки для вычисления нормированных интерполяционных весов $\bar{w}_X$ можно представить в виде формул:

$$\begin{aligned} \nabla_{K_{xy}} e &= R_{xx}^{-1} \nabla_{\bar{w}_X} e; \\ \nabla_{R_{xx}} e &= -R_{xx}^{-1} K_{xy} \nabla_{\bar{w}_X}^T e R_{xx}^{-1} = -\bar{w}_X \nabla_{K_{xy}}^T e, \end{aligned} \quad (4.25)$$

где $\nabla_{R_{xx}} e$ – матрица того же размера, что и матрица R_{xx} , составленная из градиентов функции потерь e по соответствующим элементам матрицы R_{xx} .

Большие величины градиентов $\nabla_{R_{xx}} e = \nabla_{K_{xx}} e$ и $\nabla_{K_{xy}} e$ функции потерь e увеличивают [63], [37] количество итераций метода градиентного спуска. При использовании метода градиентного спуска первого порядка требуется порядка $O(\|\nabla_{\varpi} e\|) = O(\|\nabla_{\varpi} K_{xx} \otimes \nabla_{K_{xx}} e\|)$ итераций.

Поскольку использовалась функция потерь Хьюбера [162], формула (1.4), градиент которой не превосходит 2, градиент $\nabla_{\bar{w}_X} e$ можно оценить сверху как

$$\|\nabla_{\bar{w}_X} e\|_2 \leq C_B = 2 \max_{\vec{x}} |BIAS_X(\vec{x})|.$$

Из формулы (4.25) следует, что отношение модулей градиентов $\nabla_{K_{xx}}e$ и $\nabla_{\bar{w}_X}e$ можно оценить как $\|K_{xy}\| \|R_{xx}^{-1}\|^2$ или как $\|\bar{w}_X\|_2 \|R_{xx}^{-1}\|$. Таким образом, если норма вектора интерполяционных весов удовлетворяет условию $\|\bar{w}_X\|_2 < C$, то оценить $\|\nabla_{K_{xx}}e\|$ сверху можно величиной $C_B C \|R_{xx}^{-1}\|$. Оценить градиент $\|\nabla_{K_{xy}}e\|$ сверху можно величиной $C_B \|R_{xx}^{-1}\|$.

Согласно параграфу 4.1.2, норма $\|R_{xx}^{-1}\|$ не превосходит ε^{-1} , где ε – один из оптимизируемых параметров $\bar{u}$ (формула (4.7)) КНОИ. Таким образом, если в процессе оптимизации параметр ε станет очень малым, то сходимость метода градиентного спуска сильно замедлится: в силу формул (4.9) и (4.25) градиент $\|\nabla_{K_{xx}}e\|$ можно оценить как $O(\varepsilon^{-2})$.

Чтобы предотвратить подобное замедление сходимости при оптимизации параметров $\bar{\theta}$ и $\bar{u}$ КНОИ, предлагается рассматривать только случаи, в которых выполняется неравенство

$$\sum_{i \in I(\bar{y})} \bar{w}_X(\bar{x}_i, \bar{y})^2 < C^2, \quad (4.26)$$

где использовалась константа C , равная 1.5. Если неравенство (4.26) не выполнено (а в 84% случаев неравенство (4.26) выполнялось), то соответствующее слагаемое не включалось в оптимизируемый штрафной функционал (4.5). Тогда в силу (4.25) градиент $\|\nabla_{K_{xx}}e\|$, влияющий на сходимость метода градиентного спуска, можно оценить сверху величиной $C_B C \|R_{xx}^{-1}\|$.

Можно надеяться, что при выполнении неравенства (4.26) значения градиента $\nabla_{g(x)}e$ будут ограничены и метод градиентного спуска будет сходиться. К этому есть следующие предпосылки:

1. При $\|\bar{w}_X\|_2 = \|R_{xx}^{-1} K_{xy}\|_2 < C$ норма градиента $\|\nabla_{K_{xy}}e\|_2 = \|R_{xx}^{-1} \nabla_{\bar{w}_X}e\|_2$ как правило соизмерима с нормой градиента $\nabla_{\bar{w}_X}e$;
2. Для большинства рассматриваемых корреляционных матриц K_{xx} норма $\|R_{xx}^{-1}\|$ существенно меньше ε^{-1} и лишь для некоторых «плохих» конфигураций расположения ближайших станций она близка к ε^{-1} ;
3. Поскольку матрица K_{xx} часто содержит много малых значений КФ (4.15), при вычислении которой используются экспоненты, то при дальнейшем обратном распространении ошибки градиент $\nabla_{K_{xx}}e$ умножится на малый градиент КФ $\nabla_{g(x)}K_{xx}$.

Такая оптимизация с отбором «хороших» случаев согласно неравенству (4.26) в таблице 4.1 назван «неявная с выбором», а без отбора случаев – «неявная». Из таблицы 4.1 видно, что предлагаемый отбор «хороших» случаев согласно неравенству (4.26), заметно улучшает точность получаемой интерполяции.

В отличие от обучения, при применении КНОИ ограничение (4.26) не используется.

Вычисление коэффициентов оптимальной интерполяции по формуле (4.9) и градиентов $\nabla_{K_{xy}}e$ по формуле (4.25) требует решения системы линейных алгебраических уравнений (СЛАУ) с положительно определенной симметричной матрицей R_{xx} , что является сравнительно дорогой процедурой, требующей $O(m^3)$ умножений и сложений, где m – количество влияющих точек.

Решать СЛАУ с матрицей R_{xx} в (4.9) и (4.25) будем итеративным методом сопряженных градиентов, предназначенного для решения СЛАУ с симметричной положительно определенной матрицей [233]. Использовался критерий остановки: либо достигнуто m итераций, либо невязка $|R_{xx}\bar{w}_X - K_{xy}|$ не превосходит 10^{-6} . Использование итеративного метода сопряженных градиентов в некоторых случаях позволяет найти приближенное решение для таких СЛАУ, численное решение которой явными методами (например, методами квадратного корня, Гаусса, LU -разложения) невозможно [19] по причине плохой обусловленности СЛАУ.

Для применения итеративного метода решения СЛАУ необходимо задать первое приближение для решения СЛАУ. При вычислении градиентов $\nabla_{K_{xy}}e$ по первой из формул (4.25) использовалось первое приближение, равное нулю.

Приведем первое приближение для $\tilde{w}$, используемое при решении уравнения в (4.9) и удовлетворяющее неравенству (4.26) с константой $C=1$.

Предлагаем при оптимизации погрешности интерполяции (4.5) вместо коэффициентов $\bar{w}$, вычисленных по формуле (4.9), использовать другую интерполяцию (4.4) поля инкрементов ΔX с коэффициентами, найденными по приближенной явной формуле, не требующей решения системы линейных уравнений, а именно:

$$\bar{w}_X(\vec{x}_i, \vec{y}) \approx \tilde{w}_X(\vec{x}_i, \vec{y}) = \frac{K(\vec{x}_i, \vec{y})}{\mathcal{G}_i} \bigg/ \max \left(1, \sum_{i \in I(\vec{y})} \frac{K(\vec{x}_i, \vec{y})}{\mathcal{G}_i} \right), \quad (4.27)$$

где $\mathcal{G}_i = \sum_{k \in I(\vec{y})} R(\vec{x}_k, \vec{x}_i)$ можно интерпретировать как штраф за высокую плотность наблюдений

в окрестности влияющей точки $\vec{x}_i$. Подобный штраф вводят при так называемой декластеризации данных [166], [47], проводимой перед применением ОИ (кригинга).

Поскольку в параграфе 4.1.2 было выбрано семейство строго положительных НКФ $R(r)$, формула (4.16), то выполнено неравенство $\mathcal{G}_i \geq R(\vec{x}_i, \vec{x}_i) = 1$, а значит и неравенства:

$$0 < \tilde{w}_X(\vec{x}_i, \vec{y}) \leq \frac{K(\vec{x}_i, \vec{y})}{\mathcal{G}_i} \leq K(\vec{x}_i, \vec{y}).$$

Поэтому относительное влияние отдельного наблюдения $\Delta X(\vec{x}_i)$ на результат интерполяции $INC_X(\vec{y})$ не превосходит $K(\vec{x}_i, \vec{y})$, то есть оценки корреляции инкрементов в точках $\vec{x}_i$ и $\vec{y}$. Интерполяционный вес $\tilde{w}_X(\vec{x}_i, \vec{y})$ близок к $K(\vec{x}_i, \vec{y})$, если все остальные наблюдения производились в точках, как очень далеких от точки $\vec{x}_i$ и $\max$ в (4.27) равен 1.

Рассмотрим предельный случай, когда параметр $\varepsilon = 0$ и все $\vec{x}_i \rightarrow \vec{x}_1 \neq \vec{y}$. Тогда, очевидно, веса $\tilde{w}_X$ стремятся к константе: $\tilde{w}_X(\vec{x}_i, \vec{y}) \rightarrow 1/m$. В тоже время матрица R_{xx} стремится к необратимой матрице K_{xx} (полностью состоящей из единиц) и предельные значения весов $\bar{w}_X(\vec{x}_i, \vec{y})$ не определены. При этом формула (4.27) вычисляет одно из решений СЛАУ $R_{xx}\bar{w} = K_{xy}$, а итерации метода сопряженных градиентов не изменяют его.

При постоянном разбросе $\sigma(\vec{x}) = const$ различные значения функции максимума в формуле (4.27) можно интерпретировать как различные варианты задачи интерполяции. В задаче экстраполяции плотность наблюдений (с учётом штрафов $\mathcal{G}_i$) в окрестности $\vec{y}$ мала и $\max = 1$, а

сумма весов $\sum_{j=1}^m \tilde{w}(\vec{x}_j, \vec{y}) < 1$. В задаче интерполяции плотность наблюдений достаточна ($\max > 1$)

и сумма весов равна единице: $\sum_{j=1}^m \tilde{w}(\vec{x}_j, \vec{y}) = 1$. При $\max \gg 1$ можно говорить задаче сглаживания

рассматриваемого поля.

Отметим два важных свойства, делающих оптимизацию, использующую формулу (4.27), более быстрой по сравнению с весами ОИ (4.9).

Во-первых, все веса $\tilde{w}_X$ положительны в то время, как веса $\bar{w}_X$ могут принимать и отрицательные значения. Отрицательные значения весов $\bar{w}_X$ часто возникают при экстраполяции, когда точка $\vec{y}$, в которую интерполируется поле ΔX , окружена влияющими точками не со всех сторон.

Отметим, что при использовании выравнивающего отображения в расширенное пространство размерности $2+d$, отрицательные веса интерполяции неизбежно возникают гораздо чаще, чем в предположении однородности и изотропности интерполируемого поля.

Действительно, задача интерполяции в физическом пространстве при переходе в расширенное пространство уже может оказаться задачей экстраполяции.

Отрицательные значения весов $\bar{w}_X$ часто приводят к существенной погрешности интерполяции [36], [41]. Таким образом, погрешность интерполяции с весами $\bar{w}_X$ зависит от выбора влияющих точек гораздо сильнее, чем при интерполяции с весами $\tilde{w}_X$.

Во-вторых, компоненты градиента $\nabla_{K_{xx}} \tilde{w}$ по модулю не превосходят 1, в отличие от неявной формулы (4.9), для которой градиент $\nabla_{K_{xx}} \bar{w}$ ограничен числом обусловленности корреляционной матрицы R_{xx} . Благодаря этому, сходимость в процессе оптимизации (4.21) при использовании явной аппроксимации (4.27) происходит существенно быстрее, чем при использовании неявной формулы (4.9).

Вычисление по формуле (4.27) требует порядка $O(m^2)$ сложений и $O(m)$ умножений против $O(m^3)$ умножений и $O(m^3)$ сложений при вычислении по формулам (4.9). Для КНОИ время, затраченное на вычисление интерполяционных весов по формуле (4.9) весьма близко к времени для вычисления нейронных сетей ϖ, μ, σ , а так же вычислению компонент корреляционных матриц R_{xx} и K_{xy} при известных $g(\bar{x}_i)$. Поэтому, согласно численным экспериментам, замена формулы (4.9) на более экономичную формулу (4.27) приводит к ускорению вычислений лишь на 5-15%.

Пример сравнения результатов интерполяции с коэффициентами, вычисленными по формулам (4.9) и (4.27), показан на рисунке 4.2. Видно, что разница между результатами двух версий интерполяции относительно невелика, но при использовании явного приближения (4.27) градиент результирующего поля существенно меньше, особенно в областях с высокой плотностью наблюдений. Таким образом, можно предположить, что вычисления по формуле (4.27) требуют существенно меньшей точности вычислений по сравнению с неявной формулой (4.9).

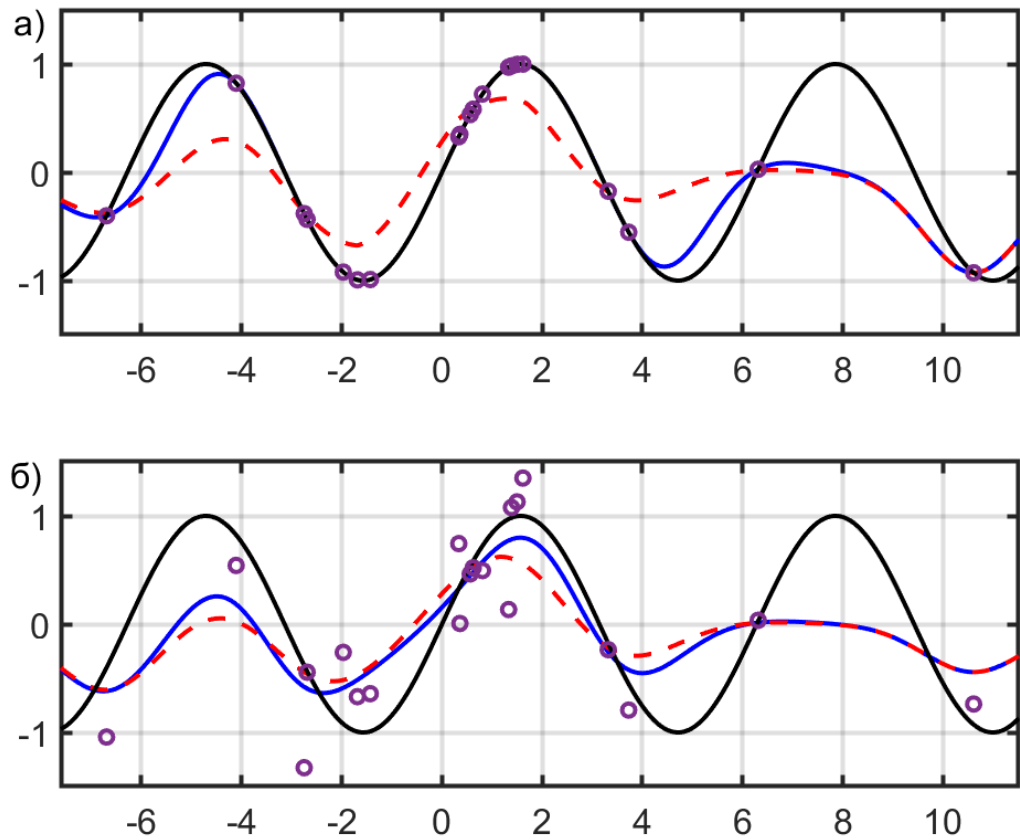


Рисунок 4.2 – Сравнение результатов интерполяции функции $f(x) = \sin x$ (черные линии): оптимальная интерполяция (4.9) – синие линии; интерполяция с весами, вычисленными по формуле (4.27) – красные пунктиры. а) при $\varepsilon = 0$; б) при $\varepsilon = 0.4$. Значения инкрементов, которые использовались при интерполяции, показаны кружками

Сравнение точности интерполяции функции одной переменной $f(x) = \sin x$ на отрезке $[-\pi; \pi]$ для различной амплитуды шума ε и количества точек случайным образом выбранных точек показано на рисунке 4.3. Видно, что явная формула (4.9) даёт несколько лучшую точность, но с ростом амплитуды шума ε различие в точности результатов интерполяции по двум формулам быстро убывает и при $\varepsilon \geq 0.4$ его можно считать несущественным.

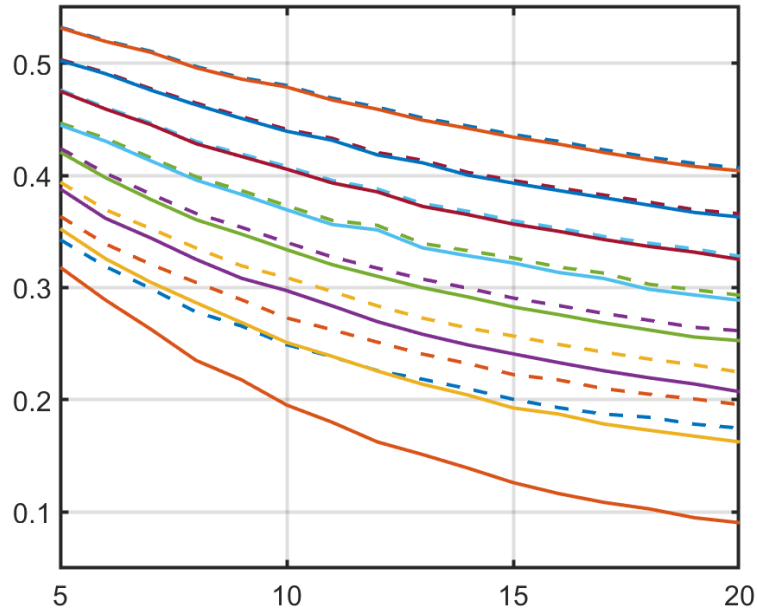


Рисунок 4.3 – Сравнение среднеквадратической погрешности интерполяции функции $f(x) = \sin x$ на отрезке $[-\pi; \pi]$ в зависимости от количества влияющих точек (ось абсцисс): для оптимальной интерполяции (4.9) – сплошные линии; для интерполяции с весами, вычисленными по формуле (4.27) – пунктиры. Снизу вверх амплитуда шума $\varepsilon = 0, 0.1, \dots, 0.7$

При оптимизации (4.21) начальное значение ε далее выбиралось равным $1/3$, то есть согласно рисунку 4.3 лежало примерно на границе областей с относительно низкой и относительно высокой ошибкой, вносимой неточностью формулы (4.27). Полученные в результате оптимизации значения параметра ε приведены в таблице 4.2.

Для вычисления интерполяционных весов w_X можно использовать как явную формулу (4.27), так и неявную формулу (4.9). При этом при оптимизации нейронных сетей ϖ, μ, σ в силу (4.21) и при их оперативном применении не обязательно использовать одну и ту же формулу для вычисления интерполяционных весов w_X .

В таблице 4.1 и на рисунке 4.4 приведено сравнение интерполяции на реальных данных при интерполяции поправок (4.2), вычисленных КСК (глава 2), с коэффициентами, найденными по явной (4.9) и по неявной (4.27) формулам при $d = 2$. Приведено увеличение Δe среднего значения функции потерь e (использовалась функция потерь Хьюбера, формула (1.4)) для прогноза на пропущенной станции относительно прогноза X_C , вычисленному КСК:

$$\Delta e = \frac{1}{N} \sum_{i=1}^N e(X_A(\bar{x}_i, I \setminus \{i\}), X_{fact}(\bar{x}_i)) - e(X_C(\bar{x}_i), X_{fact}(\bar{x}_i)) \quad (4.28)$$

Согласно результатам численных экспериментов, приведенных в таблице 4.1, для всех рассмотренных метеорологических параметров погрешность интерполяции минимальна, если при обучении использовать явную формулу (4.27), а при оперативном применении – неявную

(4.9). Для модуля скорости ветра при оперативном применении (при использовании для обучения явной формулы) явная и неявная формулы дают эквивалентную погрешность интерполяции. На втором месте результаты, полученные при использовании только явной формулы. Наихудшие результаты получаются, если при обучении использовать неявную формулу, а при применении – явную.

Из полученных оценок можно сделать вывод о том, что при минимизации погрешности интерполяции (4.21) факт того, что компоненты градиента $\nabla_{K_{xx}} \tilde{w}_X$ не превосходит 1, оказался важнее теоретической (в предположении верности используемой модели ковариаций) неоптимальности интерполяционных весов $\tilde{w}_X$ по сравнению с весами w_X .

При оптимизации (4.5) рассматривался так же гибридный вариант вычисления весов w_X , сочетающий веса, вычисляемые по формулам (4.9) и (4.27), а именно:

$$\hat{w}_X(\vec{x}_i, \vec{y}) = \begin{cases} w_X(\vec{x}_i, \vec{y}), & \text{если } \sum_{i \in I(\vec{y})} |w_X(\vec{x}_i, \vec{y})|^2 < C^2; \\ \tilde{w}_X(\vec{x}_i, \vec{y}), & \text{иначе.} \end{cases} \quad (4.29)$$

Таким образом, модуль вектора интерполяционных весов $\hat{w}_X$ ограничен $C=1.5$, а значит и градиент $\nabla_{K_X} \hat{w}_X$ ограничен, что должно улучшить сходимость градиентного спуска. В среднем в 84% случаев в формуле (4.29) использовался первый случай $\hat{w}_X = w_X$, то есть в абсолютном большинстве случаев использовались веса, обеспечивающие теоретически наименьшую погрешность. При оперативном применении использование (4.29) не рассматривалось, поскольку оно приводит к разрывности поля инкрементов $INC_X(\vec{y})$.

На рисунке 4.4 приведена динамика изменения Δe (4.28) при интерполяции инкрементов температуры воздуха ($X=T$) в зависимости от шага метода градиентного спуска на наборе данных Ru6-2019. Параметры нейронных сетей инициализировались одинаково, а minibatch-и совпадали для всех способов вычисления интерполяционных весов. По оси ординат отложено значение Δe на обучающей выборке за последних 50 шагов градиентного спуска. Приведенные графики не монотонны, поскольку оценки на различных шагах получены при интерполяции различных реализаций прогностических полей температуры воздуха.

Из рисунка 4.4 видно, что при использовании неявной формулы вычисления интерполяционных весов (4.9), для того чтобы обогнать по качеству явную формулу вычисления интерполяционных весов (4.27) потребовалось около 1560 шагов градиентного спуска, а при использовании гибридной формулы (4.29) – всего 480 шагов, то есть более чем в 3 раза меньше. Заметим, что при использовании явной формулы, обучение на первых шагах градиентного спуска

происходит заметно быстрее. Следует ожидать, что явная формула может быть эффективнее неявной при использовании наборов данных существенно меньшего размера.

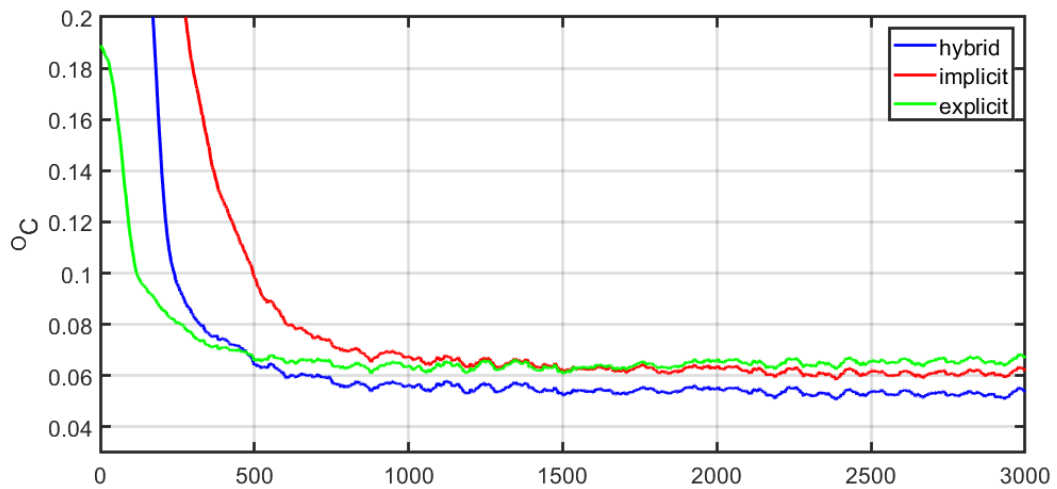


Рисунок 4.4 – Динамика увеличения (ось ординат) функции потерь $\Delta \epsilon$, формула (4.28) при интерполяции инкрементов температуры воздуха, вычисленных КСК, в зависимости от шага метода градиентного спуска (ось абсцисс) для различных формул вычисления интерполяционных весов: явного (4.27) – зеленая линия; неявного (4.9) – красная; гибридного (4.29) – синяя

Таблица 4.1 – Увеличение функции потерь $\Delta \epsilon$, формула (4.28), для интерполяции инкрементов, вычисленных КСК, при $d = 2$ по результатам тестирования на независимом архиве прогнозов COSMO-RuBENA за январь 2021 г. При вычислении интерполяционных коэффициентов по явной (4.27), по неявной (4.9), по неявной (4.9) с выбором (4.26) и по гибридной (4.29) формулам

Обучение	Неявная	Неявная с выбором	Явная	Гибридная	Неявная	Неявная с выбором	Явная	Гибридная
Тестирование	Неявная	Неявная	Явная	Неявная	Явная	Явная	Явная	Явная
$T, ^\circ C$	0,202	0,102	0,146	0,089	0,329	0,164	0,166	0,164
$Td, ^\circ C$	0,173	0,087	0,117	0,064	0,296	0,203	0,146	0,183
P_0 , гПа	0,076	0,052	0,060	0,049	0,107	0,101	0,090	0,099
$ \vec{U} $, м/с	0,213	0,143	0,175	0,127	0,252	0,167	0,175	0,153

Из таблицы 4.1 видно, что наименьшую погрешность демонстрирует интерполяция, обученная с использованием гибридной формулы (4.29), а на втором месте – обученная с использованием неявной с выбором формулы (4.26), а на третьем – явной формулы (4.27). Таким образом, можно сделать следующие выводы. Во-первых, ограниченность весов $\hat{w}_X$ в процессе обучения интерполяции сильно улучшает качество получаемой интерполяции. Во-вторых,

случаи, отброшенные согласно критерию (4.26), все же несут некоторую полезную информацию для оптимизации параметров интерполяции.

4.2.2. Предикторы неоднородности

В качестве предикторов неоднородности $\Theta(t, \vec{x})$, являющихся аргументами нейронных сетей вложения ϖ и смещения μ , будем использовать следующие параметры: первое приближение, время суток и время года, а также высоту поверхности Земли над уровнем моря. Напомним, что в данной работе оптимизация стандартного отклонения σ поля инкрементов ΔX не производилась и при интерполяции полагалось, что $\sigma \equiv 1$. Кроме того, при интерполяции давления, приведенного к уровню моря, средний инкремент полагался нулевым: $\mu_P \equiv 0$.

Опишем используемые наборы предикторов неоднородности при интерполяции инкрементов различных метеорологических параметров подробнее. Для всех метеорологических параметров в качестве предикторов неоднородности $\Theta(t, \vec{x})$ использовались следующие 4 параметра:

1. $\alpha_{sun}(t + \tau)$ – высота Солнца над горизонтом (напомним, что τ – заблаговременность прогноза);
2. $\alpha_{sun}^{\max}, \alpha_{sun}^{\min}$ – высота Солнца над горизонтом в зените и надире за текущие сутки (2 аргумента);
3. $\sin(2\pi t_J/365)$ – синус времени года, где t_J – юлианской день.

При интерполяции инкрементов давления, приведенного к уровню моря, ($X = P_0$) набор предикторов неоднородности $\Theta(t, \vec{x})$, помимо указанных выше 4 величин включал, также следующие 2 величины:

4. $P_0(t, \tau, \vec{x})$ – прогностическое значение давления;
5. $\Delta P(t, \tau, \vec{x})$ – интерполируемый инкремент.

При интерполяции инкрементов температуры воздуха и точки росы ($X = T, Td$) образом набор предикторов неоднородности $\Theta(t, \vec{x})$, помимо указанных выше 4 величин, включал также следующие 5 величин:

4. $T_0(t, \tau, \vec{x}), Td_0(t, \tau, \vec{x})$ – прогностические значения температуры воздуха и точки росы;
5. $H_0(\vec{x})$ – высота над уровнем моря.

6. $\Delta T(t, \tau, \vec{x}), \Delta Td(t, \tau, \vec{x})$ – интерполируемые инкременты.

При интерполяции инкрементов скорости ветра (случай $X = \vec{U}$) набор предикторов неоднородности $\Theta(t, \vec{x})$, помимо указанных выше 4 величин, включал также следующие 6 величин:

4. $|\vec{U}_0(t, \tau, \vec{x})|$ – модуль скорости ветра;

5. $\vec{U}_0(t, \tau, \vec{x})$ – вектор скорости ветра (2 компоненты);

6. $H_0(\vec{x})$ – высота над уровнем моря;

7. $\Delta S(t, \tau, \vec{x}) = |\vec{U}_0(t, \tau, \vec{x}) + \Delta \vec{U}(t, \tau, \vec{x})| - |\vec{U}_0(t, \tau, \vec{x})|$ – инкремент модуля скорости ветра;

8. $|\Delta \vec{U}(t, \tau, \vec{x})|$ – модуль инкремента вектора скорости ветра.

При минимизации погрешности (4.21) проинтерполированного прогноза $X_A(\vec{y})$ в первую эпоху обучения количество влияющих точек m увеличивалась с 5 до 16, а в остальные эпохи было равно 16. В качестве влияющих точек выбирались ближайшие в расширенном пространстве.

В некоторых современных ЧМПП для описания подстилающей поверхности используется так называемые тайловый (tile) и мозаичный подходы [232], [95], [236]. В тайловом подходе внутри каждого квадрата (треугольника) модельной сетки некоторые доли площади заняты различными объектами (водной поверхностью, лесом, городом и так далее). В мозаичном подходе подстилающая поверхность рассматривается на сетке более высокого разрешения. Предлагаемая КНОИ позволяет учитывать тайлы (мозаику) при интерполяции приземных метеорологических полей: для этого достаточно в список предикторов неоднородности добавить параметры, описывающие эти тайлы (мозаику).

4.2.3. Результаты оптимизации

Были проведены численные эксперименты по оптимизации КНОИ, применяемого для интерполяции инкрементов (4.2), вычисляемых КСК (2.2) при различном количестве дополнительных размерностей $d = 1...4$. Чтобы рассматривать более многомерные вложения желательно увеличить количество предикторов неоднородности $\Theta(t, \vec{x})$, равное, согласно параграфу 4.2.2, от 6 до 12. Функция ϖ искалась в виде трехслойного персептрона с 24-12 нейронами на скрытых слоях и функцией активации:

$$ELU(x) = \begin{cases} x, & x \geq 0; \\ e^x - 1, & x < 0. \end{cases} \quad (4.30)$$

Использовалась функция активации (4.30), а не более стандартные $ReLU$ и $tanh$, поскольку первая не обратима: «склеивает» точки, что ухудшит число обусловленности корреляционной матриц K_{xx} и R_{xx} , а вторая ограничена, то есть образ нашего пространства будет ограничен и значения НКФ K будут ограничены снизу, что опять же ухудшит число обусловленности корреляционных матриц K_{xx} и R_{xx} .

Вместе с оптимизацией нейронных сетей ϖ и μ оптимизировалась так же параметры $\vec{U}$ семейства НКФ, (4.16). Далее рассматривается так же случай $d = 0$, соответствующий предположению однородности и изотропности интерполируемого поля – далее эта интерполяция называется классической оптимальной интерполяцией (ОИ). В этом случае нейронные сети ϖ и μ не рассматривались, а оптимизировались только параметры семейства НКФ (4.16).

Результаты оптимизации параметра ε , описывающего амплитуду шума и стоящего в (4.7) и (4.15) при индикаторе нуля, приведены в таблице 4.2. Производилась оптимизация КНОИ для инкрементов, вычисленных КСК для прогнозов модели COSMO-Ru в конфигурации COSMO-Ru6ENA. Из таблицы 4.2 видно, что при вычислении интерполяционных весов $\tilde{w}_X$ по явной формуле (4.27) оптимальные константы ε существенно меньше, чем при вычислении интерполяционных весов по неявной (4.9) и по гибридной (4.29) формулам. Меньшие значения означают, что корреляционная модель объясняет большую часть дисперсии инкрементов ΔX .

Таблица 4.2 – Найденные значения относительной нормы шума ε для КНОИ инкрементов коррекции прогнозов модели COSMO-Ru в конфигурации COSMO-Ru6ENA, оптимальные при интерполяции инкрементов различных приземных метеорологических параметров при различной размерности d выхода сети вложения ϖ и различных интерполяционных весах

X	Веса	$d = 0$	$d = 1$	$d = 2$	$d = 3$	$d = 4$
T	Гибридные $\hat{w}_X$	0.946	0.627	0.600	0.572	0.543
	Неявные $\bar{w}_X$	0.946	0.941	0.684	0.671	0.647
	Явные $\tilde{w}_X$	0.641	0.679	0.629	0.653	0.681
Td	Гибридные $\hat{w}_X$	0.946	0.785	0.726	0.615	0.593
	Неявные $\bar{w}_X$	0.946	0.948	0.604	0.610	0.599
	Явные $\tilde{w}_X$	0.636	0.654	0.620	0.651	0.629
P_0	Гибридные $\hat{w}_X$	0.484	0.304	0.294	0.266	0.259
	Неявные $\bar{w}_X$	0.485	0.344	0.331	0.294	0.286
	Явные $\tilde{w}_X$	0.612	0.004	0.005	0.005	0.005
$ \vec{U} $	Гибридные $\hat{w}_X$	0.648	0.372	0.348	0.313	0.305
	Неявные $\bar{w}_X$	0.668	0.343	0.308	0.252	0.249
	Явные $\tilde{w}_X$	0.713	0.011	0.011	0.009	0.006

Графики оптимальных НКФ при $d=0$ и $d=2$ приведены на рисунке 4.5. Производилась оптимизация КНОИ для инкрементов ККП (формула (4.3)) с первым приближением по модели COSMO-Ru в конфигурации COSMO-Ru6ENA. Отметим, что отложенное на рисунке 4.5 по горизонтальной оси расстояние при $d=2$ измеряется в расширенном пространстве и реализуется в физическом пространстве только при постоянных предикторах неоднородности $\Theta(\vec{x})$, что в реальности невозможно. При использовании выравнивающего отображения область значений КФ существенно увеличивается, что говорит о том, что выравнивающее отображение помогает более эффективно отбирать влияющие станции.

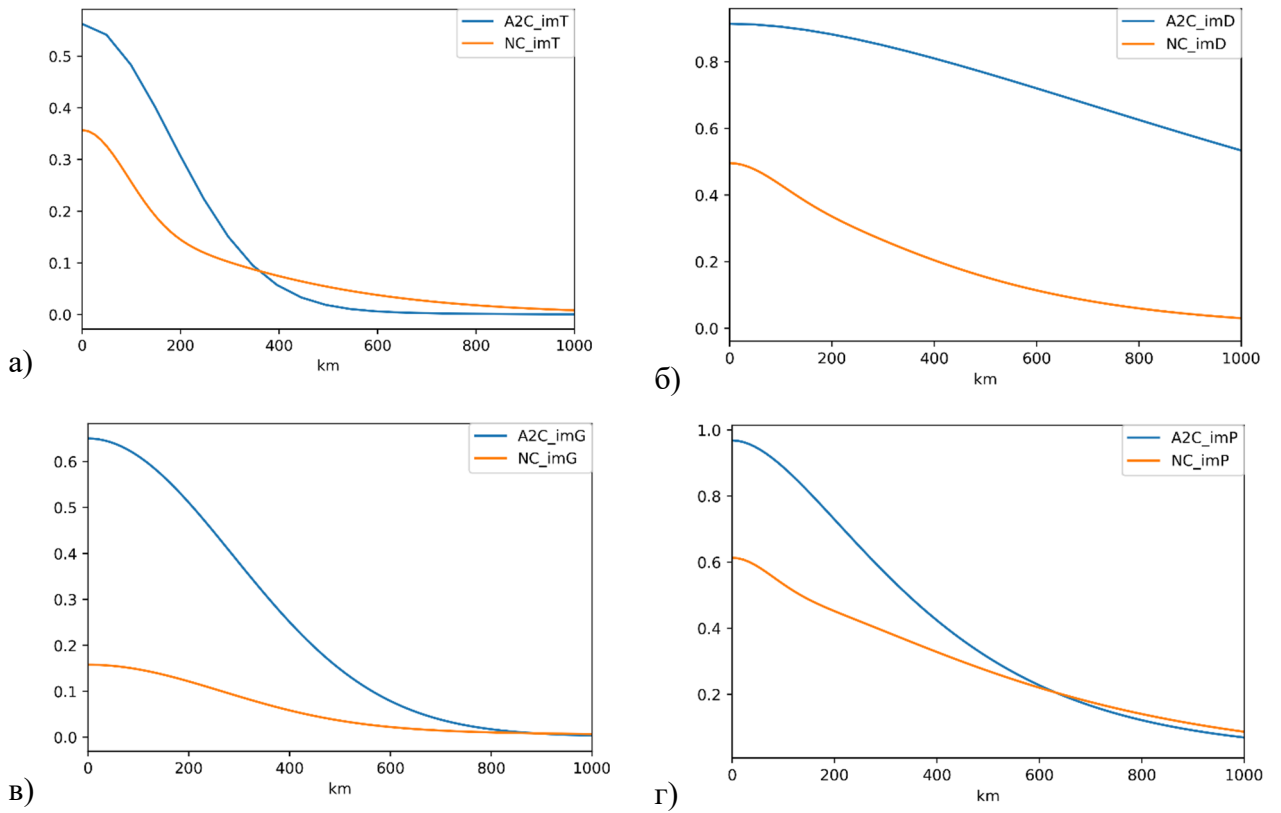


Рисунок 4.5 – Оптимальные корреляционные функции $R(r)$ из семейства (4.15) в случае $d=2$ (синие кривые) в зависимости от расстояния в расширенном пространстве и в предположении однородности и изотропности (оранжевые) в зависимости от расстояния для интерполяции комплексного прогноза ККП (глава 2) с первым приближением по модели COSMO-Ru в конфигурации COSMO-Ru6ENA различных приземных метеорологических параметров:

а) температуры воздуха T_{2m} ; б) точки росы Td_{2m} ; в) модуля скорости ветра $|\vec{U}_{10m}|$;

г) давления, приведенного к уровню моря P_0

Из таблицы 4.2 и рисунка 4.5 видно, что при большей размерности вложения оптимальным является меньшее ε и большие корреляции для точек, расположенных на том же расстоянии. Этот результат ожидаем: в пространстве большей размерности плотность облака точек-образов влияющих станций в среднем меньше, а значит для регуляризации в (4.9) будет достаточно

меньшего ε , а то же среднее значение корреляции будет достигаться при большем радиусе корреляции.

Отметим, что в предположении однородности и изотропности поля (при $d=0$) и значение ε получается заметно больше, чем при $d \geq 1$, таким образом, использование выравнивающего отображения g_X позволяет объяснить существенно большую часть дисперсии рассматриваемого метеорологического поля X .

На рисунке 4.6 продемонстрирован пример оцененных с использованием нейронной сети вложения ϖ_T (при $d=2$) значений КФ $K(\bar{x}, \bar{y})$ инкрементов температуры воздуха в отмеченной точке $\bar{x}$ (фиолетовая «X») и в точке $\bar{y}$, пробегающей по рассматриваемой области вблизи холодного атмосферного фронта. Видно, что найденная оценка КФ является КФ сильно анизотропного поля: вдоль фронта корреляции заметно больше, чем поперек фронта, что подтверждает результаты из работы [3], продемонстрированные на рисунке 2.3. Причем в данном случае теплая воздушная масса заметно более однородна (внутри нее корреляции заметно больше), чем холодная. Последний факт подтверждается изолиниями поля температуры.

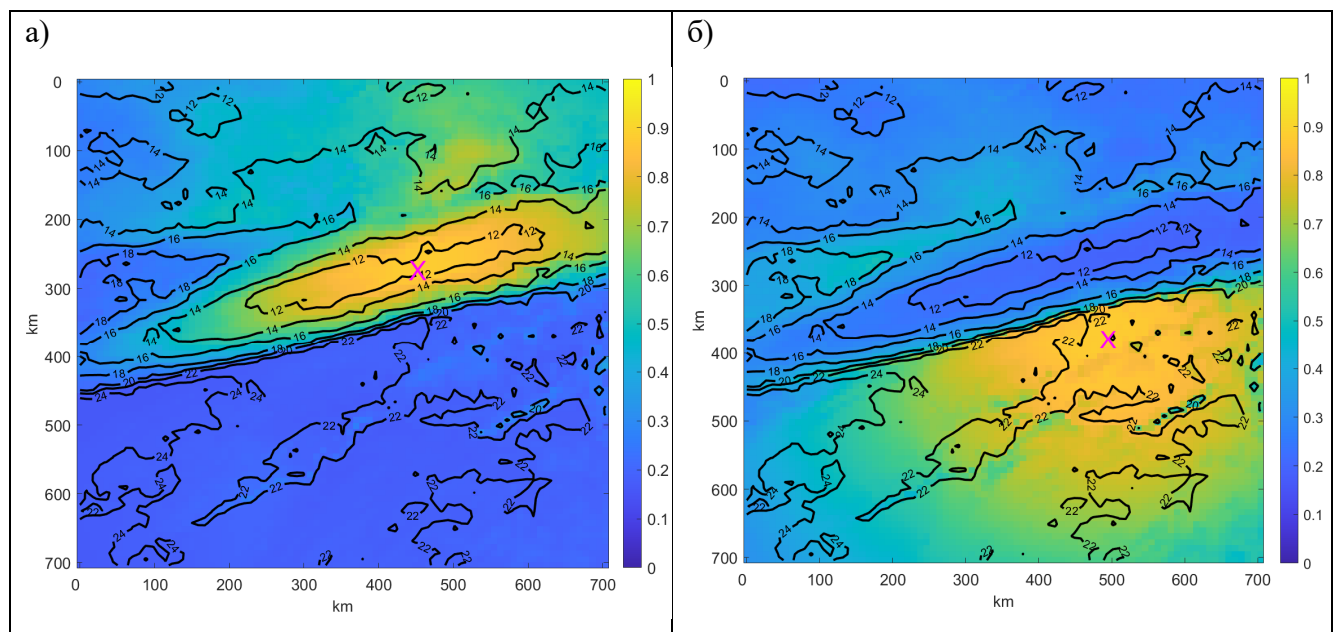
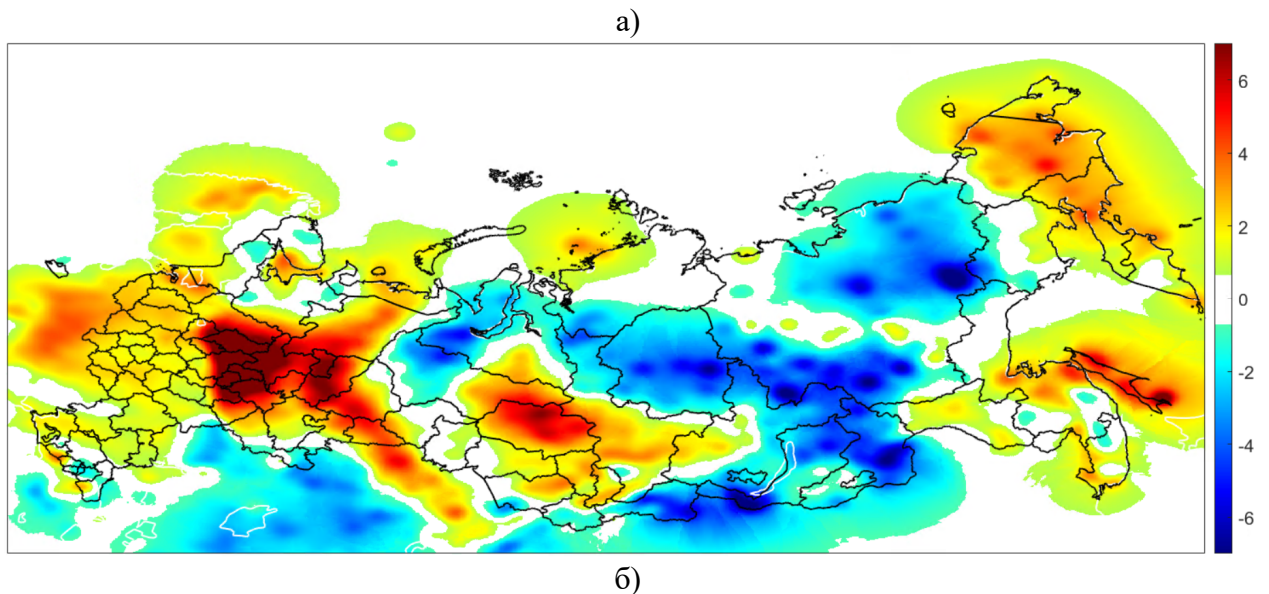


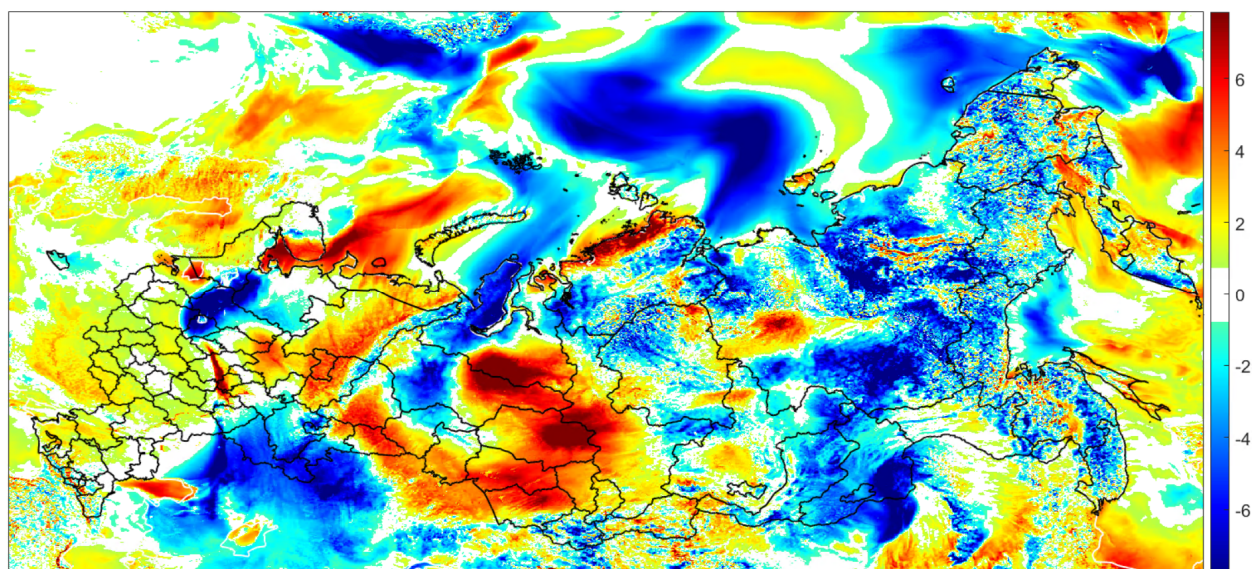
Рисунок 4.6 – Значения неоднородной КФ $K(\bar{x}, \bar{y})$ поля инкрементов температуры воздуха (раскраска) вблизи холодного атмосферного фронта, где точка $\bar{y}$ пробегает по области, а точка $\bar{x}$ отмечена фиолетовой буквой «X»: а) для $\bar{x}$ внутри холодной воздушной массы; б) для $\bar{x}$ внутри теплой воздушной массы. Черным нанесены изолинии температуры

На рисунке PP4.7 приведен пример интерполяции инкрементов КСК для прогностического поля T_{2m} модели COSMO-Ru в конфигурации COSMO-Ru6ENA с заблаговременностью 72ч от начального срока 00:00 ВCB 24 февраля 2021 г. на сетку

конфигурации COSMO-RuBENA (далее инкремент коррекции). Получившиеся поля инкрементов (рисунки РР4.7а, РР4.7в) сравниваются с полем анализа на срок прогноза (от 00ч 27 февраля 2021 г.) из которого вычтен рассматриваемый прогноз (далее инкремент анализа, рисунок РР4.7б). Видно, что получившиеся поля инкремента коррекции сильно коррелируют с полем инкремента анализа. В данном случае корреляция этих двух полей составила 32.5% и 36.1% для классической ОИ и КНОИ соответственно. При этом, инкременты коррекции вычислялись только для синоптических станций Российской Федерации (РФ) и стран Содружества независимых государств (СНГ). Если рассматривать только территорию РФ, то корреляция возрастает до 50.1% и 55.2% соответственно.

Согласно рисунку 4.5, КФ полей инкрементов температуры, скорости ветра и давления убывают существенно быстрее при учёте выравнивающего отображения, что приводит к более детализированным полям инкрементов (пример инкрементов поля температуры показан на рисунке 4.7). По-видимому, одной из причин более быстрого убывания КФ является в том числе и то, что оптимизировалась погрешность интерполяции, что позволило существенно уменьшить завышение оценки корреляции, вызываемую смещенной оценкой поля средних μ .





в)

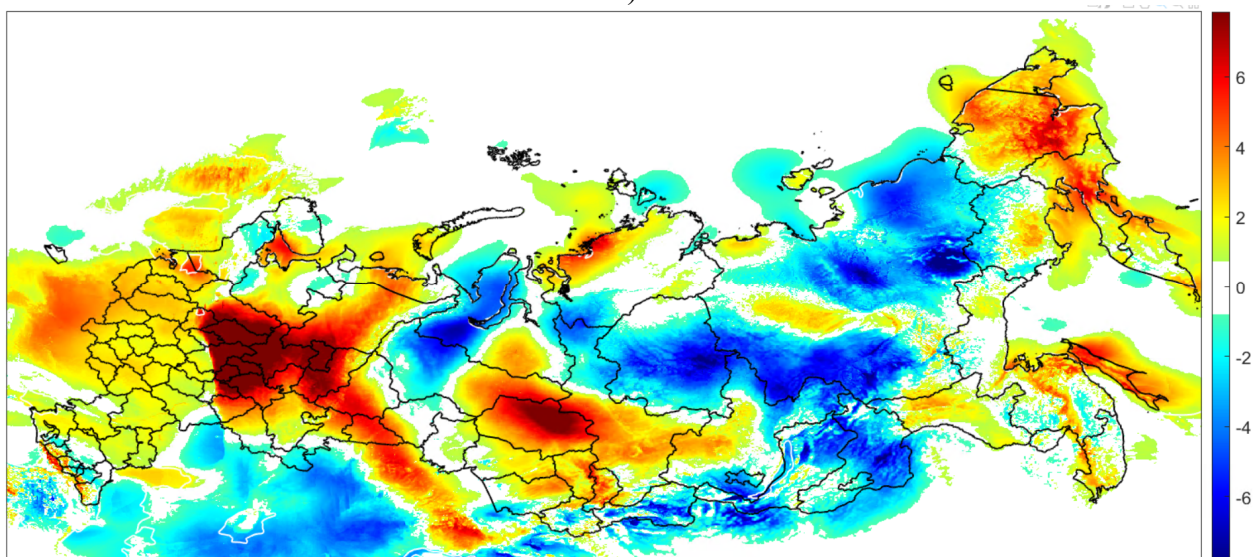


Рисунок РР4.7 – Пример полей инкремента коррекции для прогноза поля T_{2m} по модели COSMO-Ru в конфигурации COSMO-Ru6ENA с заблаговременностью 72 ч от начального срока 00:00 ВСВ 24 февраля 2021 г., вычисленного:
а) классической ОИ; б) анализом от 00:00 ВСВ 27 февраля 2021 г.;
в) предлагаемой КНОИ

Таблица 4.3 – Величина изменения функции потерь Δe , формула (4.28) для различных приземных метеорологических параметров при интерполяции инкрементов КСК различной размерности d выхода сети вложения ϖ_X при тестировании на независимом архиве прогнозов COSMO-Ru6ENA за январь 2021 г

X	$d = 0$	$d = 1$	$d = 2$	$d = 3$	$d = 4$
$T, ^\circ\text{C}$	0,216	0,162	0,146	0,142	0,141
$Td, ^\circ\text{C}$	0,207	0,133	0,117	0,111	0,113
$P_0, \text{гПа}$	0,063	0,061	0,060	0,059	0,059
$ \vec{U} , \text{м/с}$	0,221	0,184	0,175	0,173	0,174

В таблице 4.3 показано насколько в среднем увеличивается функции потерь Δe , формула (4.28) при интерполяции инкрементов, вычисленных КСК на соседних станциях (без использования наблюдений на данной станции) по сравнению с прогнозами КСК, формула (4.28). Видно, что при интерполяции поля давления можно предположить однородность и изотропность ($d=0$), а при интерполяции остальных рассмотренных параметров оптимальной размерностью вложения является $d=2$.

4.2.4. Сравнительные оценки качества интерполяции

В качестве полей первого приближения использовались прогнозы X_0 по модели COSMO-Ru в конфигурации COSMO-Ru6ENA. Численный эксперимент по оценке качества интерполяции состоял из следующих шагов:

1. Согласно КСК (глава 2) и КНС (глава 3) в точках синоптических станций вычислялись инкременты $\Delta X = X_N - X_0$, где X_N – результат коррекции, формула (3.1), вычисленный в оперативном режиме (через 05:12 после начального срока, таблица 1.9);
2. Вычислялись X_A – результат интерполяции инкрементов ΔX в точки «пропущенных» станций;
3. Оценивалось качество прогнозов X_A .

Сравнивались средние абсолютные погрешности прогноза X_A метеорологических параметров $X = T, Td, P, \vec{U}$, полученные следующими 4 способами:

1. Исходный прогноз COSMO-Ru6ENA;
2. Прогноз X_C – прогноз COSMO-Ru6ENA, откорректированный КСК и КНС (с использованием измерений на данной станции);
3. Прогноз X_A при $d=2$ (далее в этом параграфе «КНОИ»);
4. Прогноз X_A при $d=0$, то есть в предположении однородности и изотропности поля $BIAS_X$ инкрементов коррекции КСК+КНС. Далее в этом параграфе «оптимальная интерполяция» (ОИ).

Результаты сравнения прогнозов метеорологических параметров $X = T, Td, P, \vec{U}$, за июль 2020 и февраль 2021 по Европейской и Азиатской частям России приведены на рисунках 4.8-4.11 в зависимости от заблаговременности прогноза. За 0 по оси ординат на рисунках взята погрешность прогноза X_N , рассчитанного КСК+КНС, то есть по оси абсцисс отложена величина

$$\Delta MAE = \frac{1}{N} \sum_{j=1}^N |X_{A,j} - X_{fact,j}| - |X_{N,j} - X_{fact,j}|. \quad (4.31)$$

Таким образом, на рисунках 4.8-4.11 показано насколько в среднем абсолютная погрешность откорректированного прогноза больше в точке, данные измерений в которой недоступны при постпроцессинге по сравнению с точкой, в которой измерения проводились регулярно и использовались при постпроцессинге.

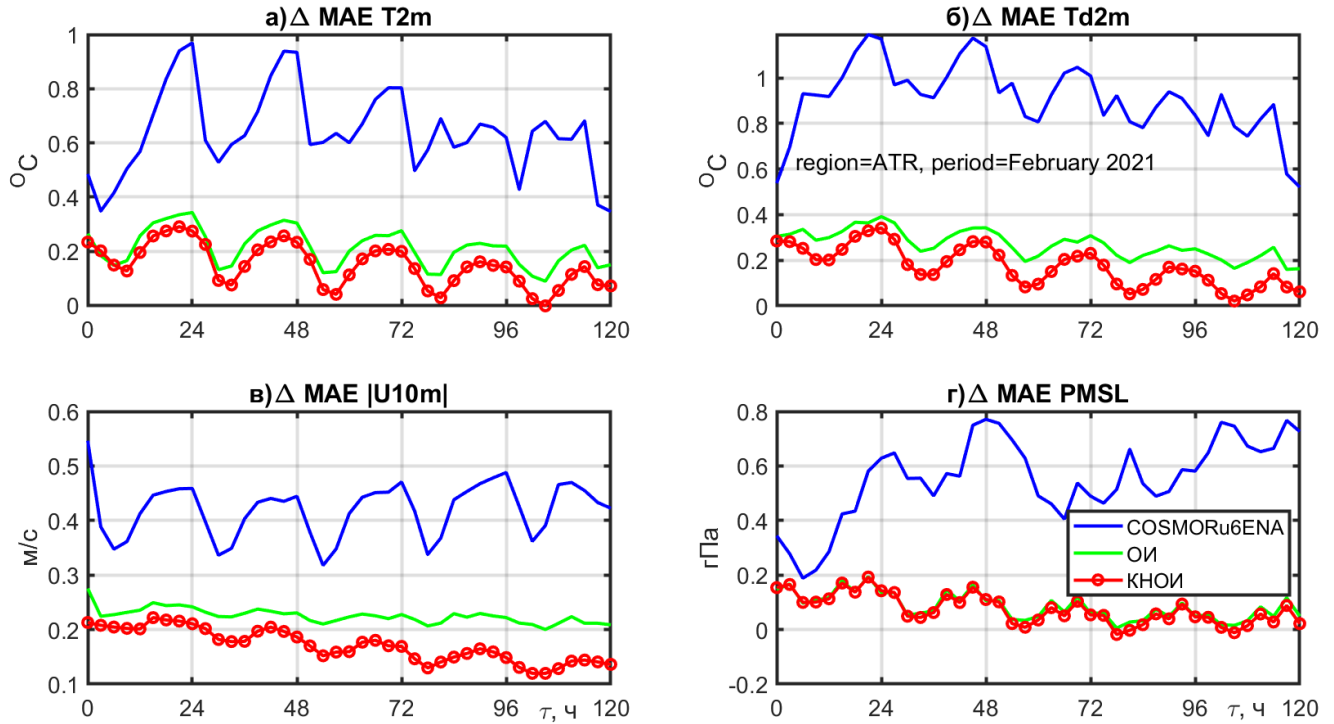


Рисунок РР4.8 – Величина ΔMAE , формула (4.31) по архиву за февраль 2021 г. по АЧР: синяя линия – исходный прогноз COSMORu-6-ENA; зеленая – оптимальная интерполяция; красная – КНОИ. Тут за 0 по оси ординат взята погрешность прогноза X_N , рассчитанного КСК+КНС.

Для прогнозов различных приземных метеорологических параметров:

а) температуры воздуха T_{2m} ; б) точки росы Td_{2m} ;

в) модуля скорости ветра $|\vec{U}_{10m}|$; г) давления, приведенного к уровню моря P_0

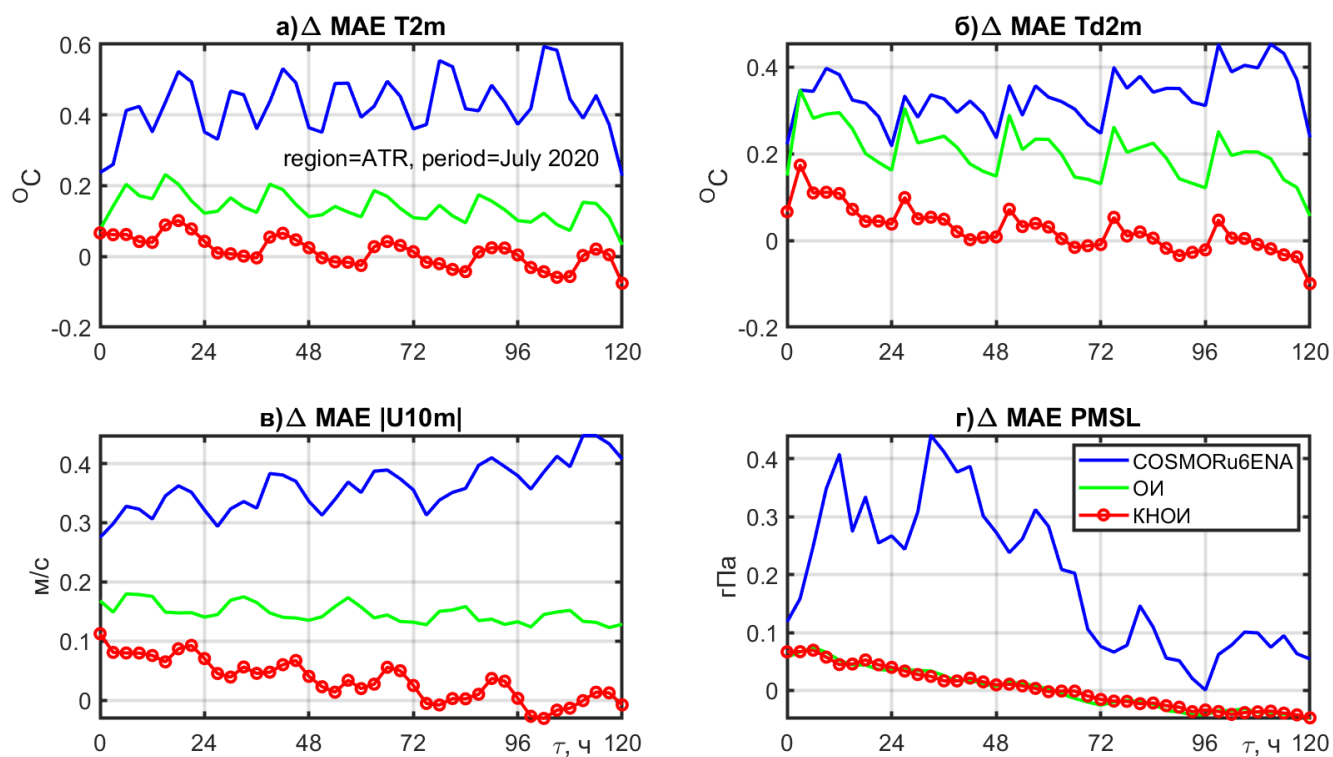


Рисунок PP4.9 – То же, что на рисунке PP4.8, но за июль 2020 г

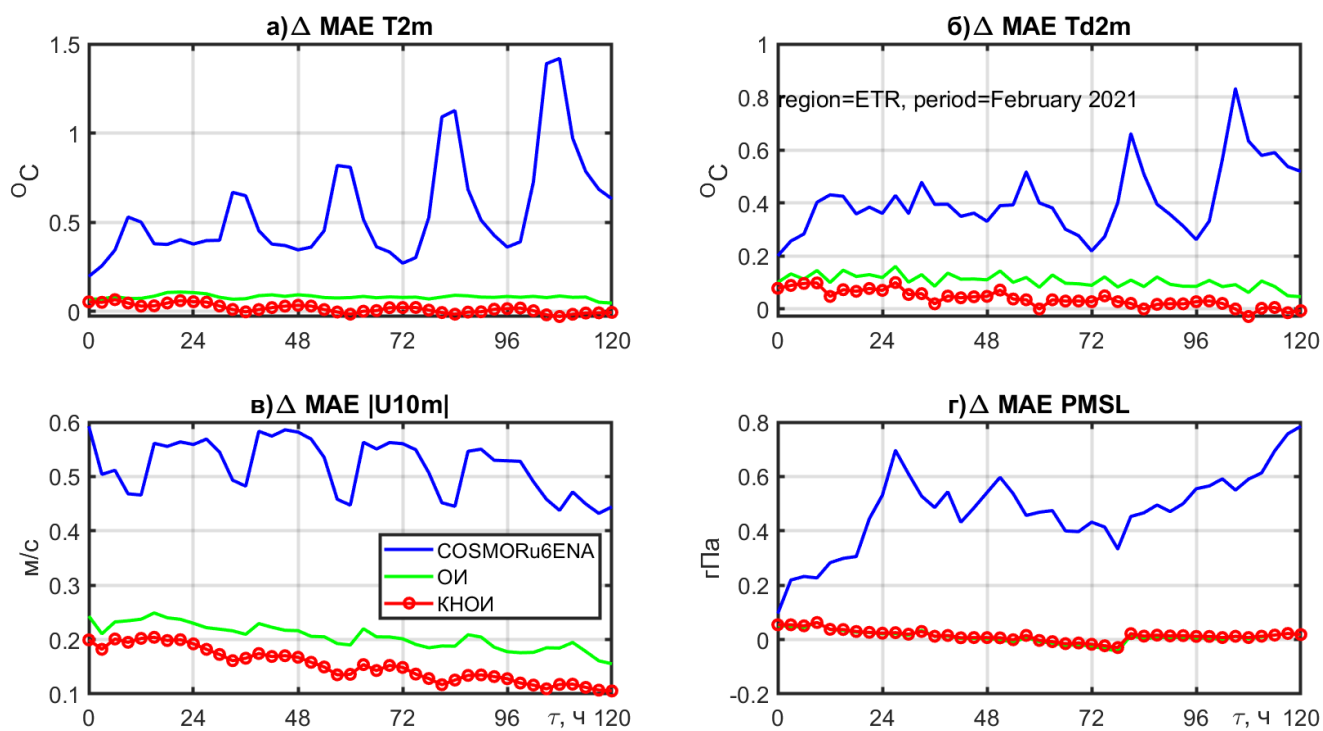


Рисунок PP4.10 – То же, что на рисунке PP4.8, но за февраль 2021 г. по ЕЧР

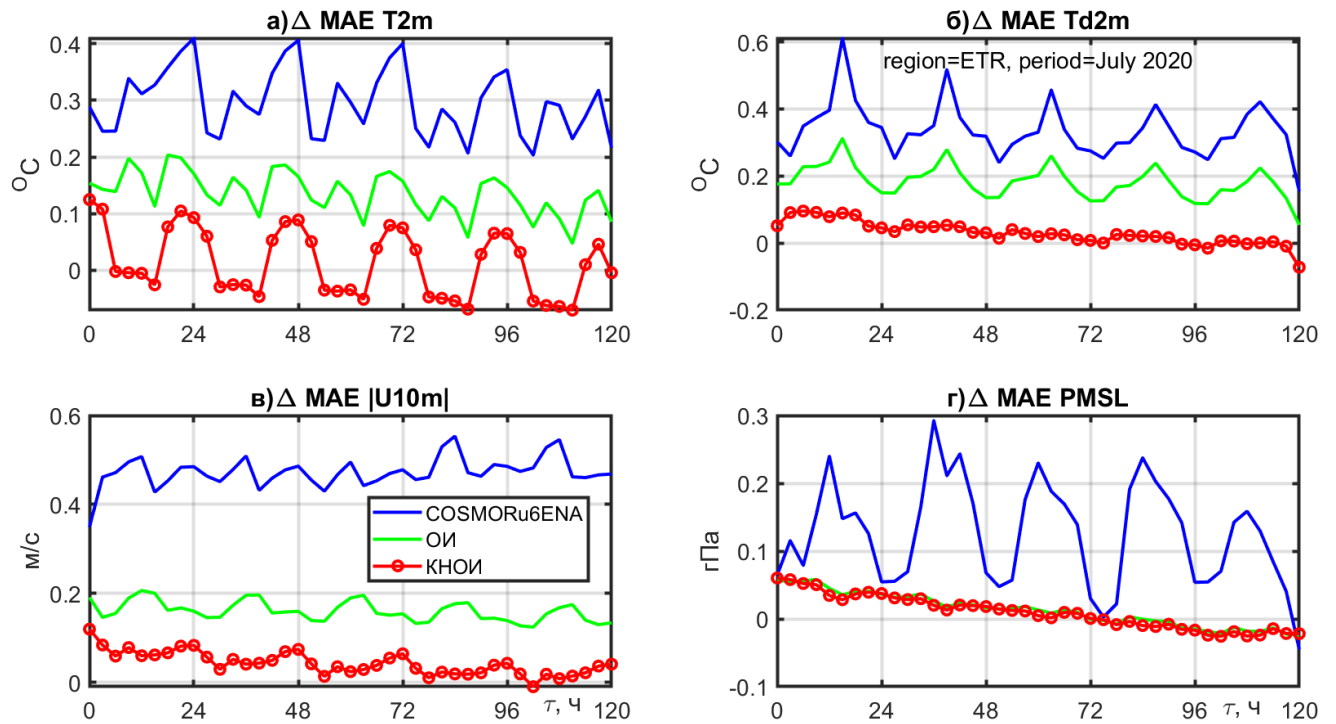


Рисунок PP4.11 – То же, что на рисунке PP4.8, но за июль 2020 г. по ЕЧР

Видно, что при интерполяции давления, приведенного к уровню моря, P качество прогнозов, проинтерполированных ОИ и КНОИ, весьма близко (средняя абсолютная погрешность отличается не более 0.05 гПа), то есть при интерполяции давления P можно предполагать однородность и изотропность.

Для остальных рассмотренных метеорологических параметров ($X = T, Td, \vec{U}$) в феврале 2021 г. (рисунки PP4.8 и PP4.10) средняя абсолютная погрешность прогноза, проинтерполированного ОИ немного больше погрешности прогноза, проинтерполированного КНОИ: абсолютная погрешность прогноза температуры больше на 0.03-0.07 °C; точки росы на 0.04-0.17 °C; скорости ветра на 0.02-0.08 м/с. В июле (рисунки PP4.9 и PP4.11) погрешность прогноза, проинтерполированного ОИ заметно больше прогноза, проинтерполированного КНОИ: абсолютная погрешность прогноза температуры больше на 0.08-0.19 °C; точки росы на 0.14-0.24 °C; скорости ветра на 0.07-0.18 м/с.

Согласно рисункам PP4.9б и 4.11б при интерполяции поля точки росы Td_{2m} в июле предположение об однородности и изотропности является очень грубым и погрешность прогноза по методу ОИ лишь немного лучше, чем исходный прогноз по модели COSMO-Ru6ENA.

При некоторых заглаженностях, больших 48ч (а именно: для T_{2m} и Td_{2m} в дневное время и для P_0 в любое время суток) прогноз, проинтерполированный КНОИ на «пропущенную» станцию оказывается лучше прогноза КСК+КНС. То есть результат постпроцессинга без использования измерений на данной станции (с использованием измерений на соседних

станциях) оказывается лучше, чем результат постпроцессинга с использованием измерений на данной станции (и только на ней). Вероятная причина: при интерполяции используются данные с нескольких метеорологических станций, что подавляет шумы, вызванные ошибками контроля качества данных наблюдений.

Выводы из главы 4

В главе 4 описана квазилинейная неоднородная интерполяция (КНОИ) и полученные с её помощью прогнозы.

В классической оптимальной интерполяции [36] используются существенные предположения, влияющие на качество интерполяции:

- однородность и изотропность интерполируемого поля по горизонтальным координатам;
- распределение значений интерполируемого поля согласно определенному вероятностному распределению, например, гауссовскому;
- линейность оператора интерполяции.

КНОИ реализует существенно более общий случай оптимальной интерполяции, в котором эти три существенных предположения заменены на два существенно более слабых:

- неоднородности интерполируемого поля могут быть описаны некоторыми заранее заданными (быть может, меняющимися со временем) предикторами неоднородности;
- оптимальный оператор интерполяции представим в некоторой квазилинейной форме.

При оптимизации параметров классической оптимальной интерполяции максимизируют правдоподобие рассматриваемой ковариационной модели, а при оптимизации параметров КНОИ минимизировалась погрешность интерполяции в точку «пропущенной» станции. Такой подход позволил отказаться от предположения о том, что интерполируемое поле является гауссовским.

Применение КНОИ улучшает качество интерполяции метеорологических полей, не являющихся однородным и (или) значения которых распределены согласно негауссовскому вероятностному закону. Действительно, согласно рисунку 4.5 КНОИ позволяет объяснить существенно (в 1.5 и более раз) большую часть дисперсии интерполируемого поля. КНОИ вырабатывает гораздо более детализированные поля, чем ОИ.

КНОИ применена для интерполяции по горизонтальным координатам поправок к прогнозам приземных метеорологических параметров, вычисленным постпроцессингом, предложенным в главах 2 и 3.

Было продемонстрировано (параграф 4.2.4), что для целей интерполяции поля поправок можно считать однородными и изотропными в случае поправок к полю давления, приведенного

к уровню моря, а для поправок к остальным приземным метеорологическим полям (температуры воздуха, точки росы, скорости ветра) – только на равнинных территориях в зимний период времени.

ГЛАВА 5. Сравнительные оценки качества прогнозов

Во всех приведенных ниже оценках участвовали прогнозы ЧМПП, поступившие в базы данных Гидрометцентра России до 05:00 ВСВ соответствующих суток. Все оцениваемые поправленные прогнозы КСК+КНС и квазилинейный комплексный прогноз с нейросетевой коррекцией (ККП+КНС, далее в этой главе «комплексный прогноз») рассчитаны в оперативном режиме и записаны в базы данных Гидрометцентра России до 05:15 ВСВ (17:15 ВСВ для начального срока 12:00 ВСВ) соответствующих суток.

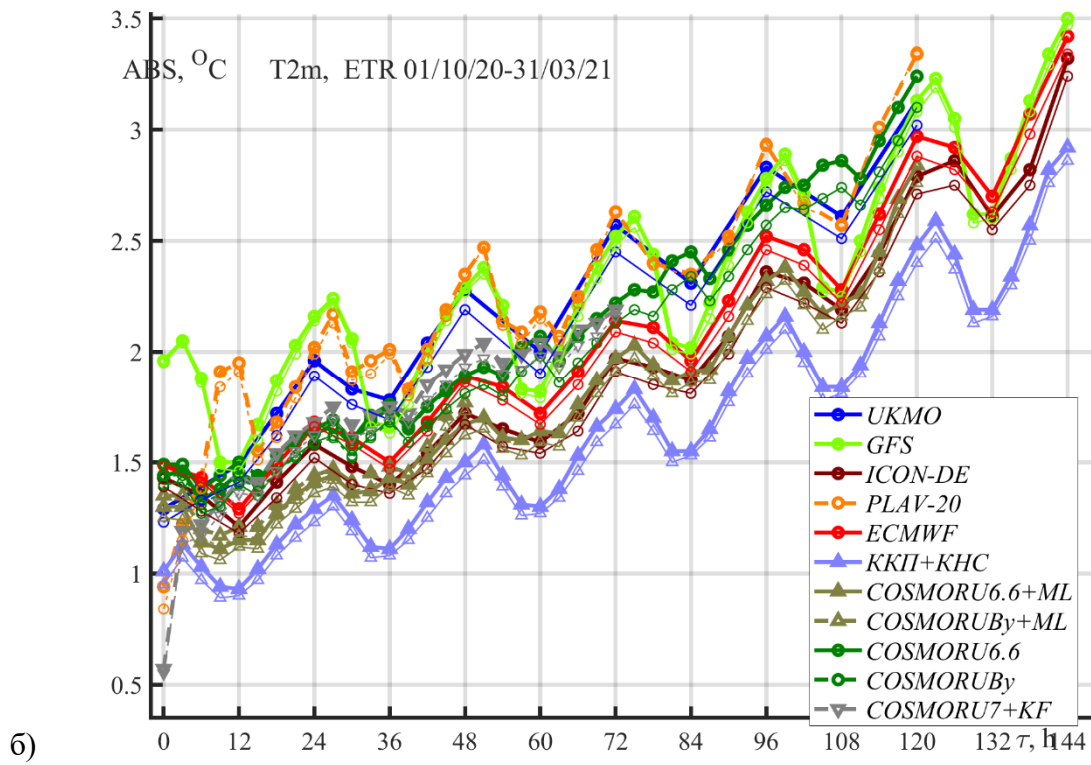
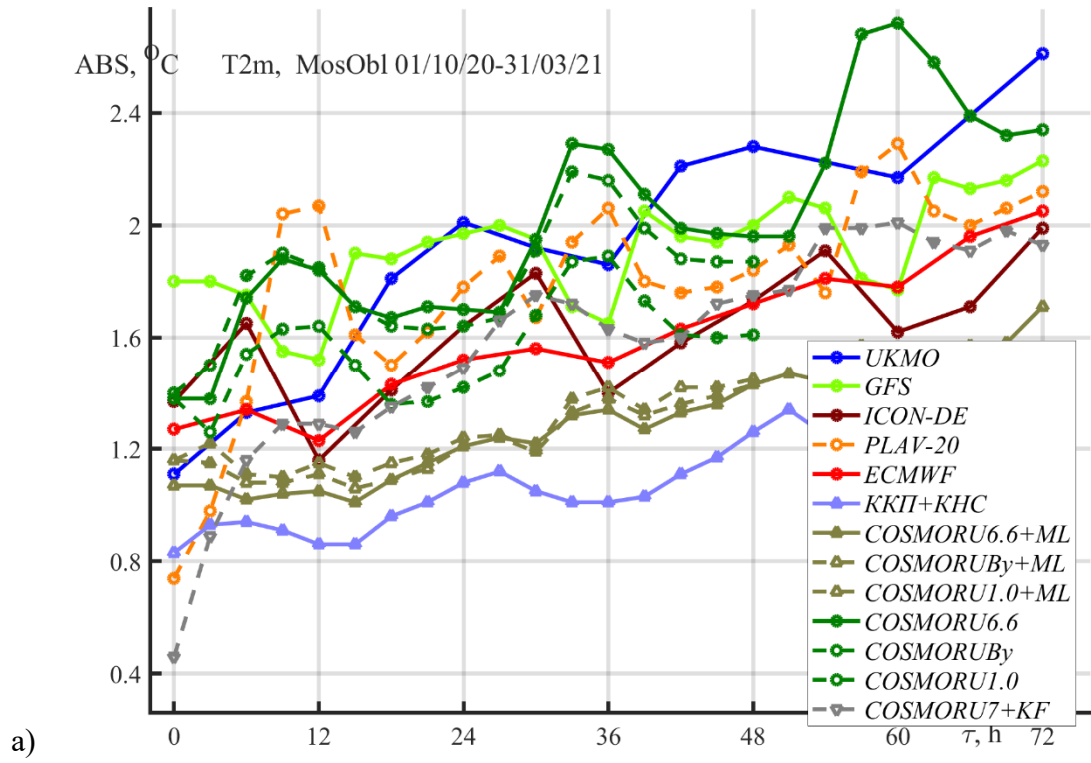
Поправленные прогнозы с заблаговременностью 0 и 3 часа не имеют практического значения, поскольку: на момент записи соответствующих прогностических значений уже были доступны фактически наблюдаемые на данной синоптической станции значения метеорологических параметров и предлагаемый постпроцессинг вполне мог бы использовать эти значения (и использует, но лишь для прогнозов с заблаговременностями 24, 27, 48, 51 и так далее часов).

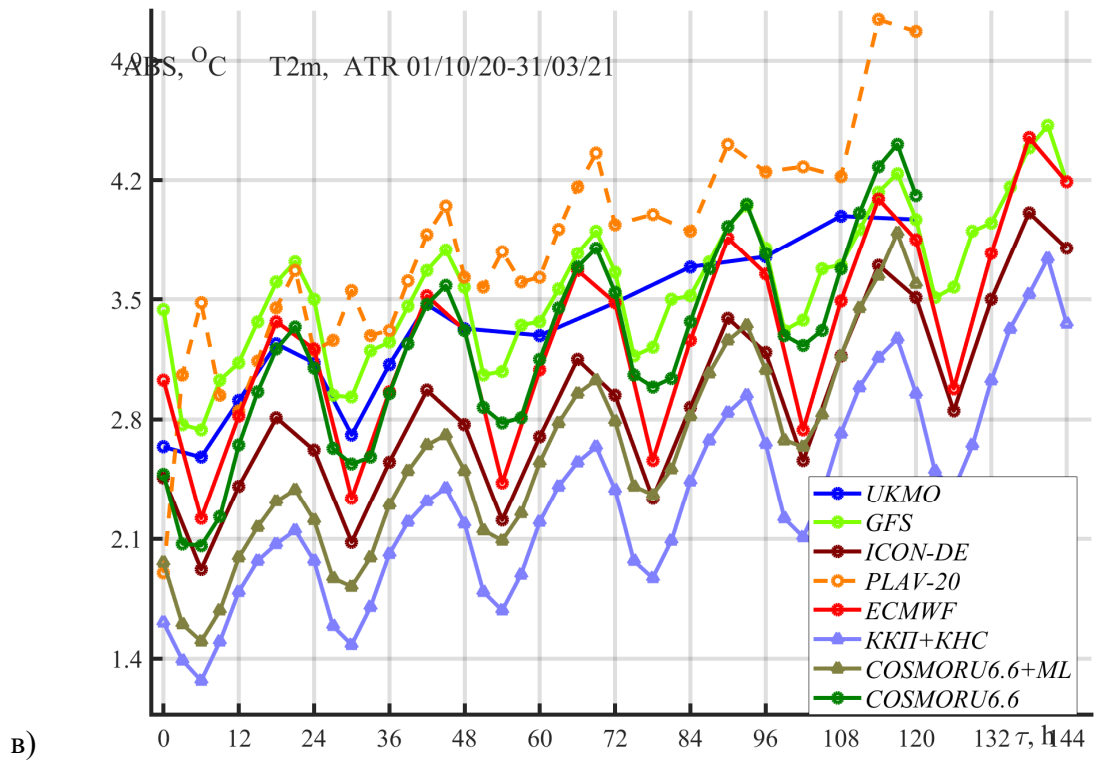
В этой главе приведены результаты сравнения качества различных прогнозов от начального срока 00:00 ВСВ за период с 1 октября 2020 г по 31 марта 2021 г (182 суток). Аналогичные оценки, но от начального срока 12:00 ВСВ приведены в приложении А.

5.1. Сравнительные оценки качества прогнозов температуры воздуха и точки росы

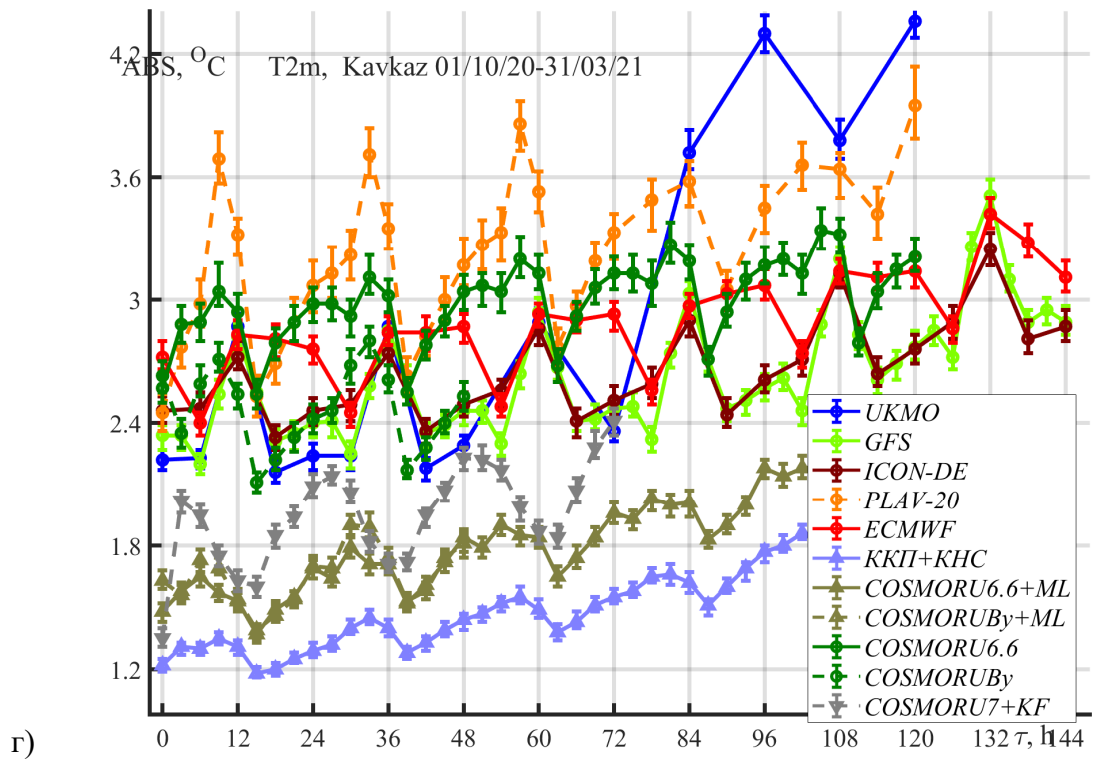
На рисунке 5.1 приведены сравнительные оценки качества прогнозов температуры воздуха на высоте 2 м по различным ЧМПП, а также предлагаемому постпроцессингу: комплексному прогнозу ККП+КНС, COSMO-Ru6ENA+КСК+КНС, COSMO-RuBy+КСК+КНС, COSMO-Ru1Msk+КСК+КНС (обозначены как COSMORU6.6+ML, COSMORU2.2+ML и COSMORU1.0+ML соответственно) и откорректированными с помощью фильтра Калмана прогнозами модели COSMO-Ru конфигурации Cosmo-Ru7-ETR [5] – обозначены как COSMORU7+KF.

На рисунке 5.1 показаны оценки отдельно для следующих 5 территорий: отдельно для Московского региона, Европейской части России (ЕЧР), Азиатской части России (АЧР), Кавказа и территории Средней Азии. Границы рассматриваемых областей показаны на рисунке 1.3, параграф 1.2.2. Аналогичные оценки для начального срока 12:00 ВСВ приведены в приложении А на рисунке А.1.





B)



Г)

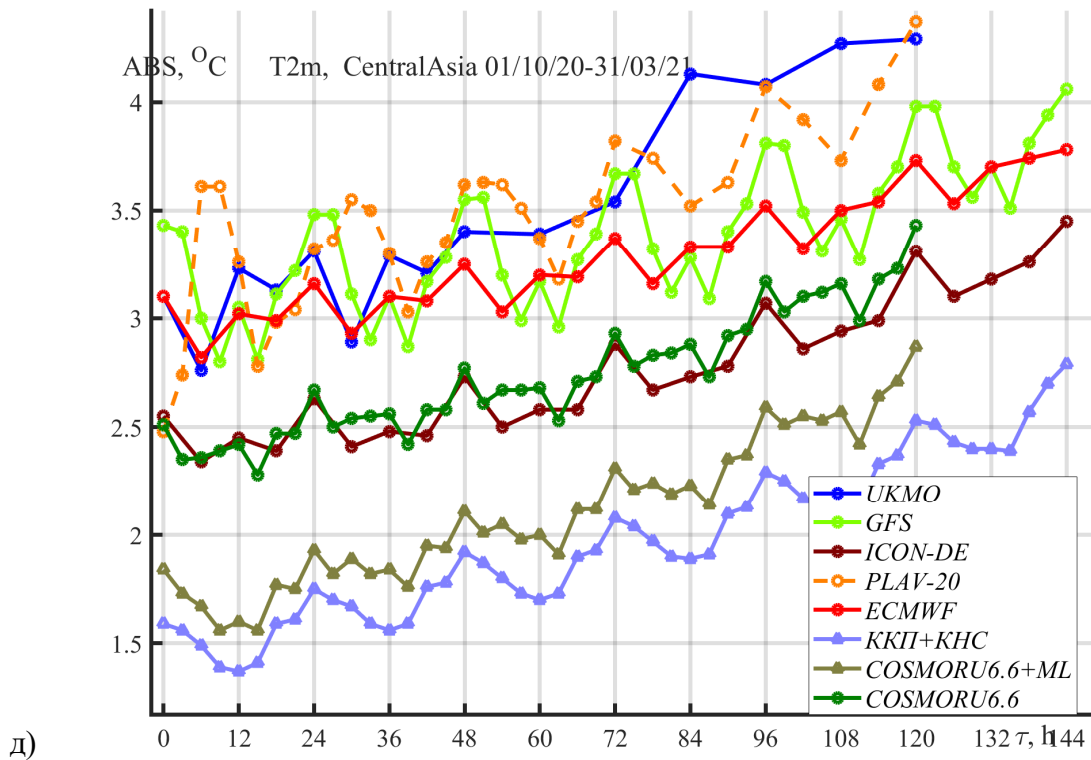


Рисунок 5.1 – Средняя абсолютная ошибка прогнозов (ось абсцисс) температуры воздуха на высоте 2 м от начального срока 00:00 ВСВ по различным методам за период с 1 октября 2020 г. по 31 марта 2021 г. по различным регионам в зависимости от заблаговременности (ось ординат): а) Московский регион; б) ЕЧР; в) АЧР; г) Кавказ; д) Средняя Азия

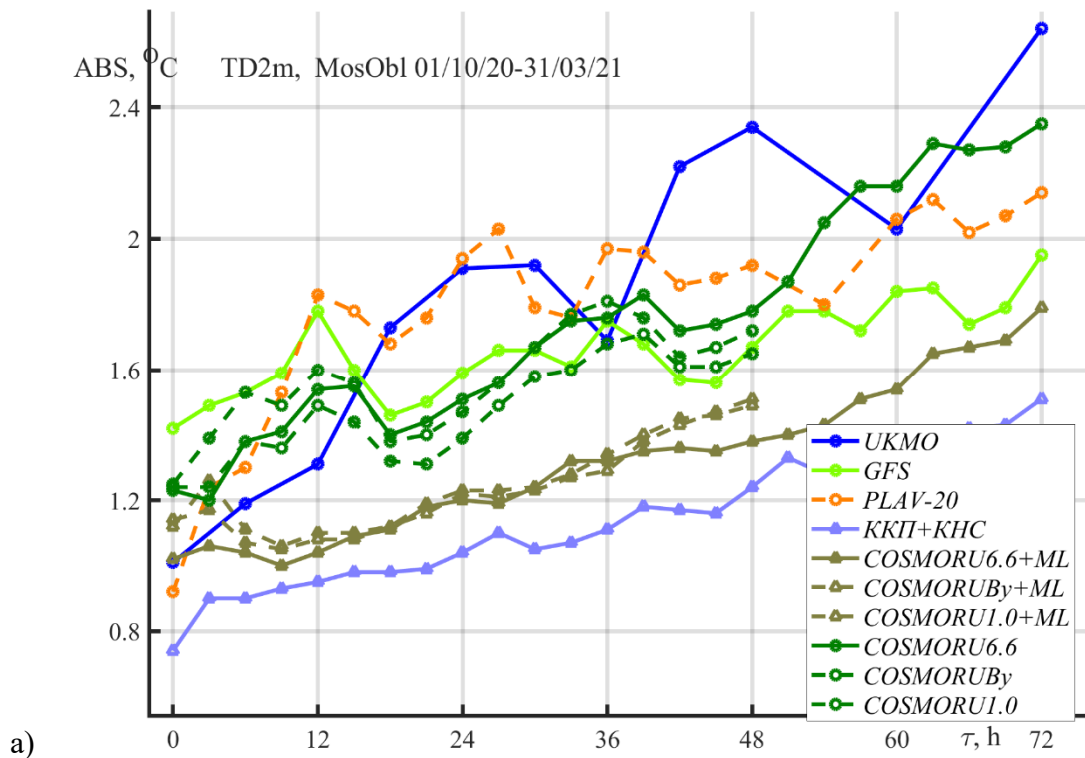
Видно, что наибольшие погрешности и наибольшая абсолютная величина эффекта от коррекции КНС наблюдается для территории Средней Азии, а наименьшие погрешности и наименьший эффект от коррекции – для территории Московского региона и ЕЧР.

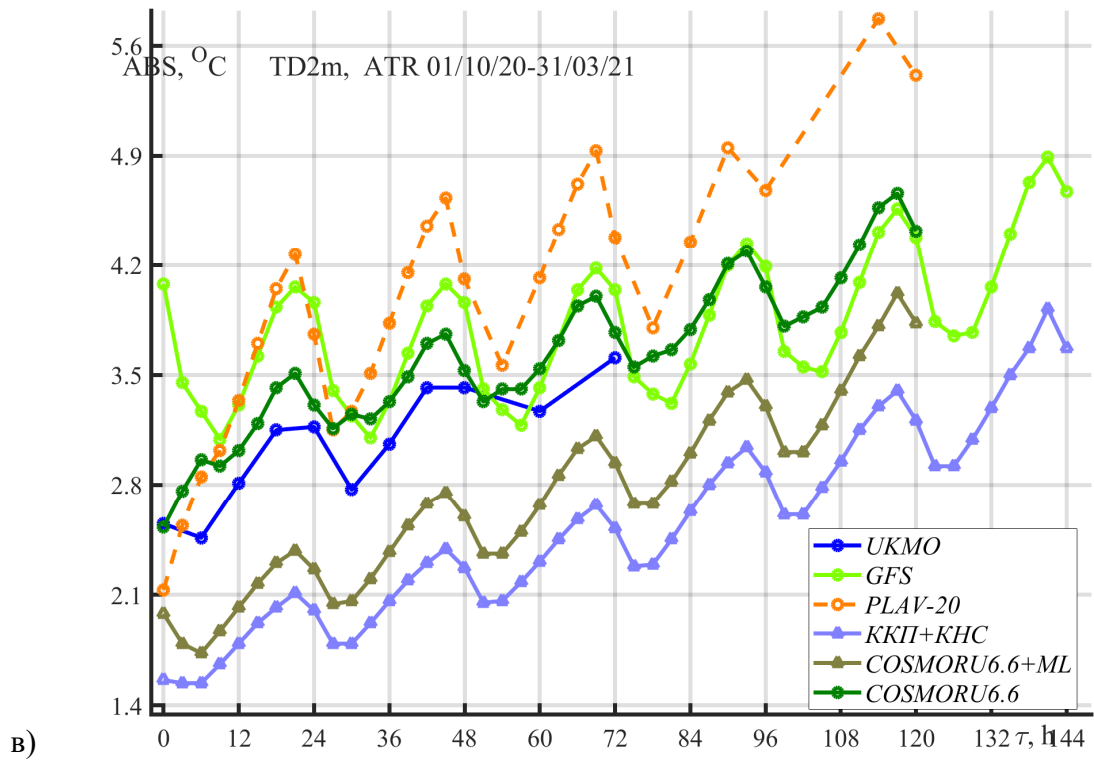
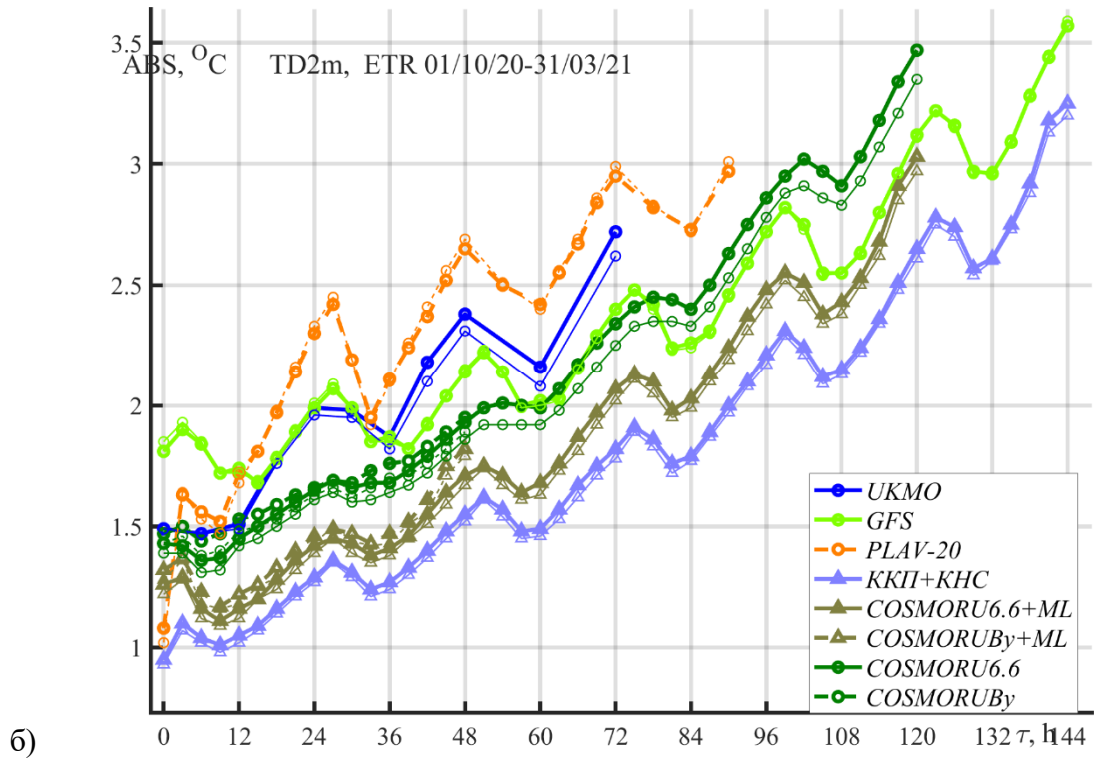
Анализируя результаты для территории Московского региона (рисунок 5.1а), видим, что качество рассмотренных прогнозов модели COSMO-Ru в конфигурации COSMO-Ru1Msk, использующей параметризацию городских процессов существенно лучше, чем в конфигурации COSMO-Ru6ENA. Однако, после применения поправок качество прогнозов COSMO-Ru6ENA+КСК+КНС в среднем оказывается чуть-чуть лучше (разница в абсолютной погрешности MAE не более 0.06°C) качества прогнозов COSMO-Ru1Msk+КСК+КНС. Можно предположить, что это связано с тем, что объем используемого при обучении архива синхронизированных прогнозов и наблюдений по конфигурации COSMO-Ru6ENA примерно в 200 раз больше соответствующего объема архива по конфигурации COSMO-Ru1Msk. Аналогичный эффект наблюдается при сравнении качества прогнозов COSMO-Ru6ENA+КСК+КНС и COSMO-RuBy+КСК+КНС на ЕЧР (рисунок 5.1б) – для конфигурации COSMO-RuBy объем соответствующего архива был примерно в 10 раз меньше, чем для конфигурации COSMO-Ru6ENA.

Предлагаемый постпроцессинг КСК+КНС оказался эффективен для всех рассмотренных конфигураций COSMO-Ru на всех рассмотренных территориях для всех заблаговременностей.

Суммировать полученные оценки постпроцессинга прогнозов по модели COSMO-Ru в конфигурации COSMO-Ru6ENA можно следующим образом: постпроцессинг позволяет превзойти (абсолютная погрешность MAE меньше на $0.06-0.18^{\circ}\text{C}$ в ЕЧР и на $0.15-0.42^{\circ}\text{C}$ в АЧР) качество лучшей ЧМПП (ICON-De) в первые 3 суток, а для последующих суток даёт близкие по качеству (различие в MAE не более 0.05°C в ЕЧР и не более 0.13°C в АЧР) к ICON-De прогнозы. Исключением являются горные регионы (Кавказ и Средняя Азия), где откорректированные прогнозы сильно лучше (MAE меньше на $0.35-0.85^{\circ}\text{C}$) любых неоткорректированных, что связано, в том числе, и с тем, что при интерполяции данных ЧМПП, поступающих через сеть международного обмена (таблица 1.6), не учитывалась разница высот (а из-за более грубых сеток эта разница высот для иностранных ЧМПП существенно больше, чем для рассмотренных конфигураций модели COSMO-Ru).

Средняя абсолютная погрешность комплексного прогноза температуры воздуха меньше соответствующей погрешности прогнозов COSMO-Ru6ENA+КСК+КНС на $0.13-0.41^{\circ}\text{C}$ в ЕЧР и на $0.18-0.64^{\circ}\text{C}$ в АЧР, причем разница растет с ростом заблаговременности и соответствует уменьшению заблаговременности COSMO-Ru6ENA+КСК+КНС примерно на 1 сутки.





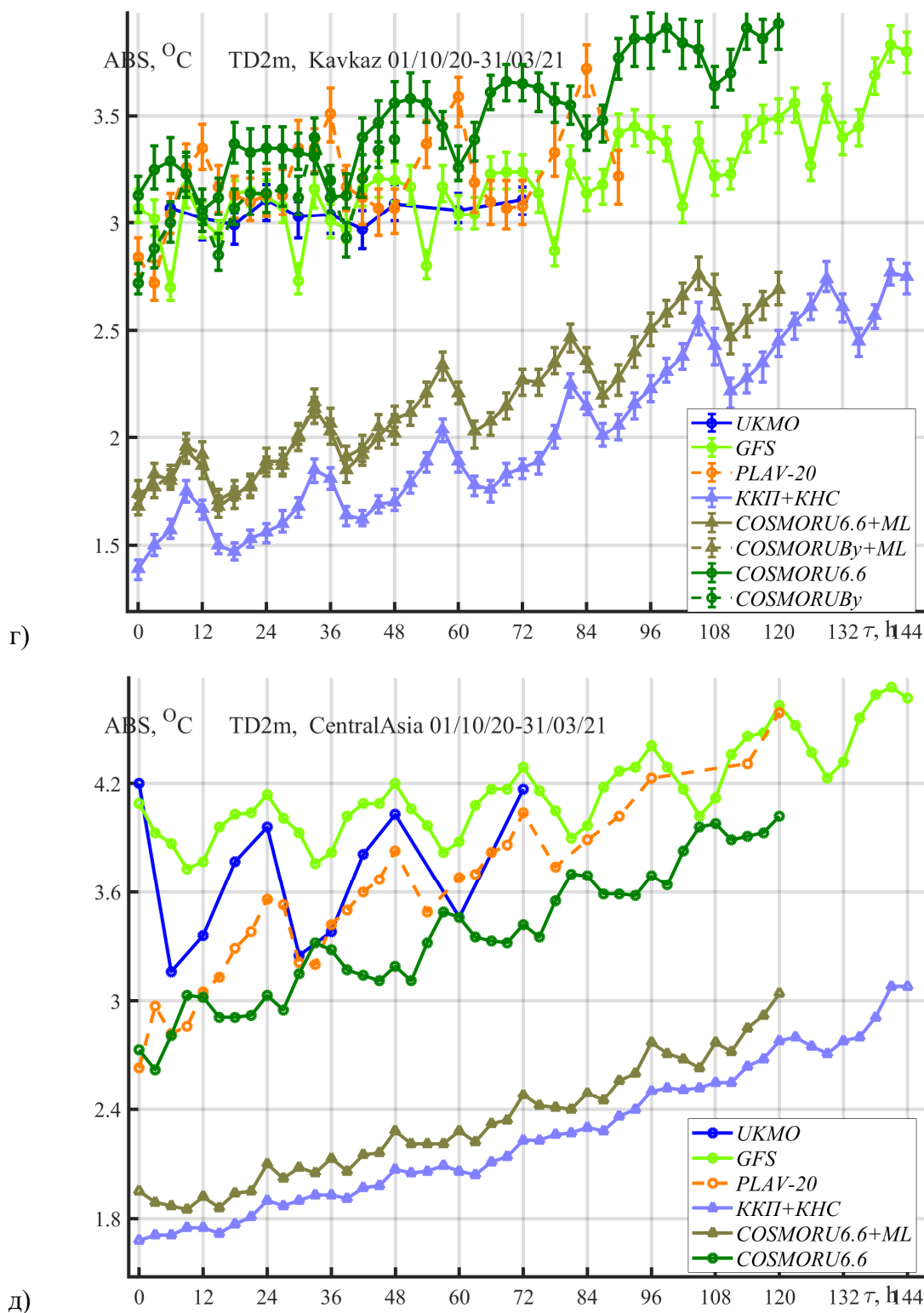


Рисунок 5.2 – Средняя абсолютная ошибка прогнозов (ось абсцисс) точки росы на высоте 2 м от начального срока 00:00 ВСВ по различным методам за период с 1 октября 2020 г. по 31 марта 2021 г. по различным регионам в зависимости от заблаговременности (ось ординат): а) Московский регион; б) ЕЧР; в) АЧР; г) Кавказ; д) Средняя Азия

Оценки средней абсолютной ошибки *MAE* прогнозов точки росы на высоте 2 м приведены на рисунке 5.2. Аналогичные оценки для начального срока 12:00 ВСВ приведены в приложении А на рисунке А.2. Результаты коррекции с помощью фильтра Калмана отсутствуют, так как авторы алгоритма [5] используют его только для прогнозов температуры воздуха. Прогнозы точки росы по моделям ICON-De и ECMWF не оценивались, поскольку они не поступают оперативно в базы данных Гидрометцентра России (таблица 1.6).

Средняя абсолютная погрешность *MAE* прогнозов точки росы по предлагаемому постпроцессингу модели COSMO-Ru в конфигурации COSMO-Ru6ENA на ЕЧР на 0.06-0.59°C меньше погрешностей лучшей из иностранных моделей (различие в среднем убывает с ростом заблаговременности) и на 0.28-0.75°C меньше абсолютной погрешности исходных прогнозов COSMO-Ru6ENA. На АЧР преимущество прогнозов COSMO-Ru6ENA+KCK+KHC над лучшей иностранной моделью увеличивается до 0.27-1.13°C и 0.51-0.89°C соответственно.

Средняя абсолютная погрешность комплексного прогноза точки росы меньше соответствующей погрешности прогнозов COSMO-Ru6ENA+KCK+KHC на 0.09-0.41°C в ЕЧР и на 0.19-0.62°C в АЧР, причем разница растет с ростом заблаговременности.

Погрешности прогнозов температуры воздуха и точки росы имеют сильный суточный ход: в ночное время погрешности прогнозов по предлагаемым методам на 15-30% больше.

Оценки доверительных интервалов для Кавказского региона получены методом бутстрепа (bootstrap) [134]. На рисунках 5.1г и 5.2г приведены 5% и 95% квантили оценок по методу бутстрепа. Для остальных регионов такие оценки не приводятся, поскольку доверительные интервалы получились очень узкими (например, $\pm 0.02^\circ\text{C}$ для оценок средней абсолютной ошибки температуры и точки росы в ЕЧР).

Для ЕЧР оценки вычислялись двумя способами: помимо применяемого для остальных регионов усреднения абсолютных ошибок прогнозов по отдельным станциям, так же применялось предварительное усреднение по квадратам $1^\circ \times 1^\circ$ сетки. Полученные таким образом оценки прогнозов на $1^\circ \times 1^\circ$ сетке показаны на рисунках 5.1б и 5.2б тонкими линиями. Видно, что такое осреднение по квадратам $1^\circ \times 1^\circ$ сетки улучшает все оценки, сохраняя при этом все сделанные выводы. Это улучшение весьма логично: сходимость методов решения уравнений в частных производных, используемых моделями прогноза погоды, обеспечивается в сеточной норме (теорема Лакса - Рябенского [75], [56], [43]).

Из рисунков 5.1 и 5.2 видно, что качество комплексного прогноза приземной температуры воздуха и точки росы по всем рассмотренным регионам в среднем эквивалентно качеству поправленных прогнозов COSMO-Ru6ENA+KCK+KHC с на 1-1.5 суток меньшей заблаговременностью. Отметим, что модели, прогнозы которых имеют сравнительно большую

погрешность (PLAV, UKMO) использованы при расчетах весовой функции (вторая часть главы 3) с наибольшими весами (параграф 2.2.2, рисунки 2.12б и 2.12в).

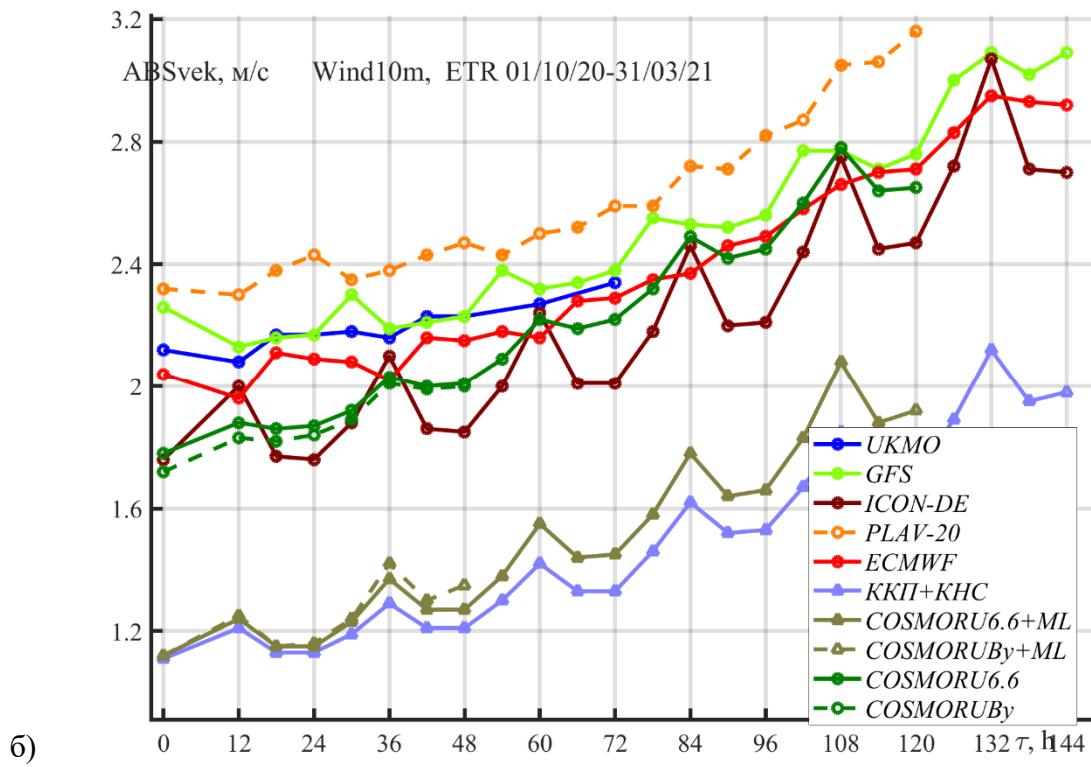
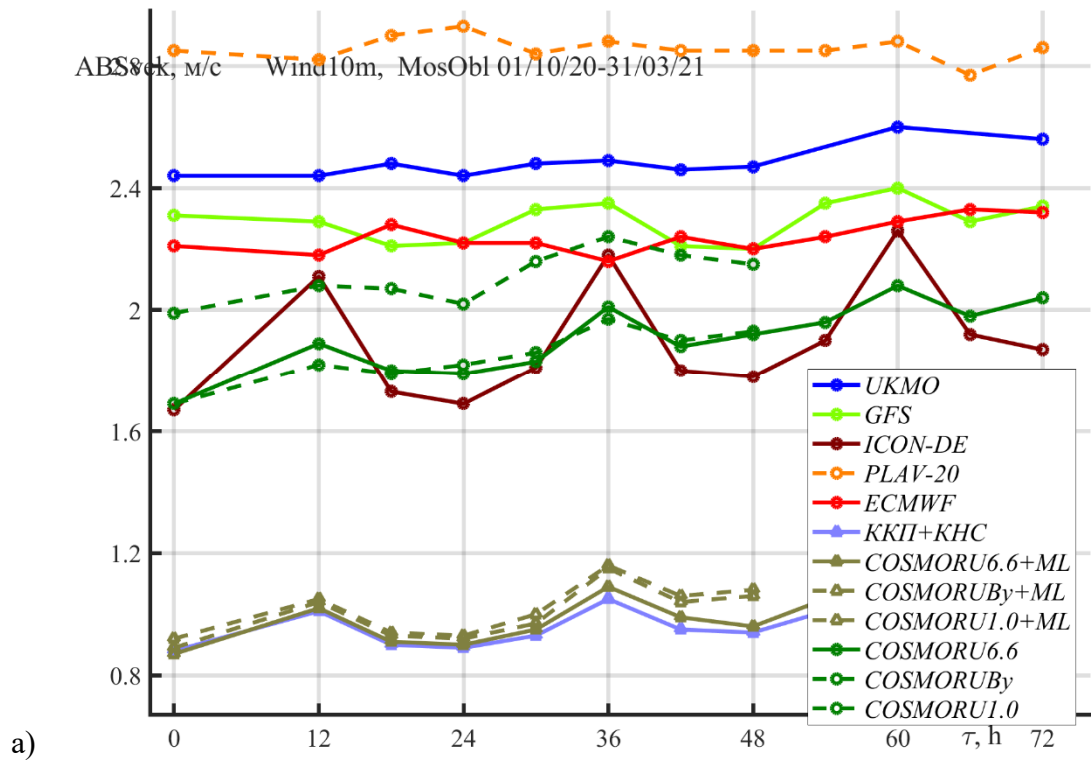
5.2. Сравнительные оценки качества прогнозов скорости ветра

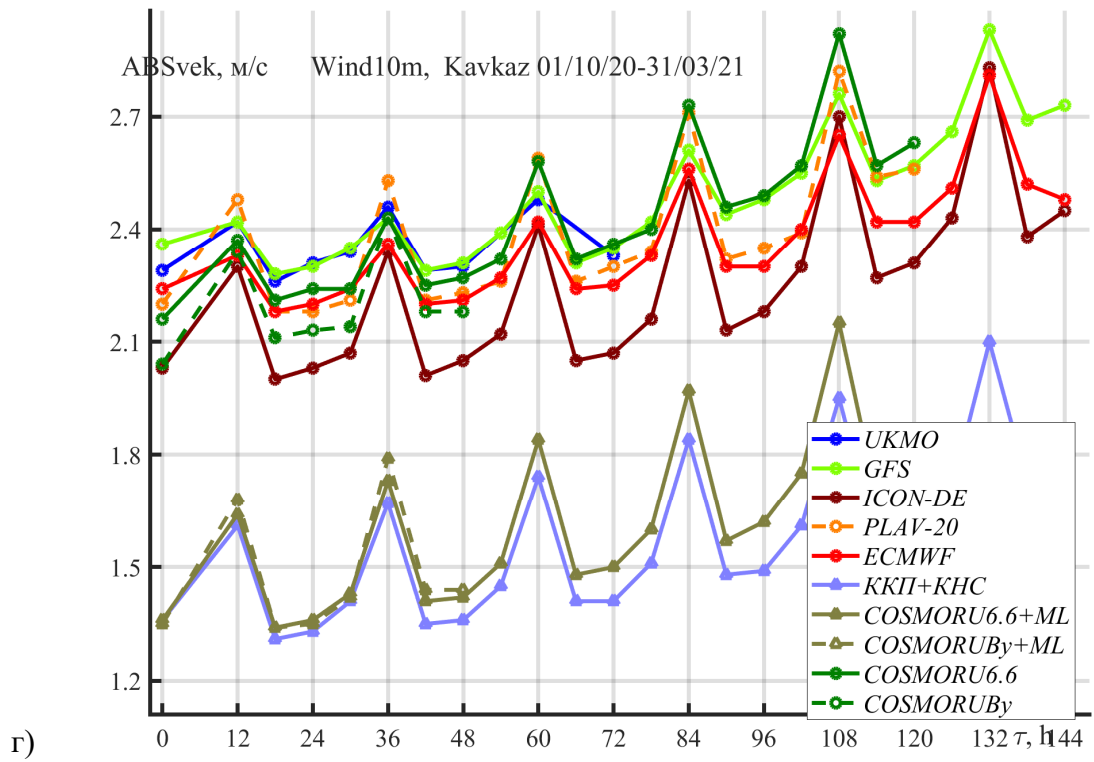
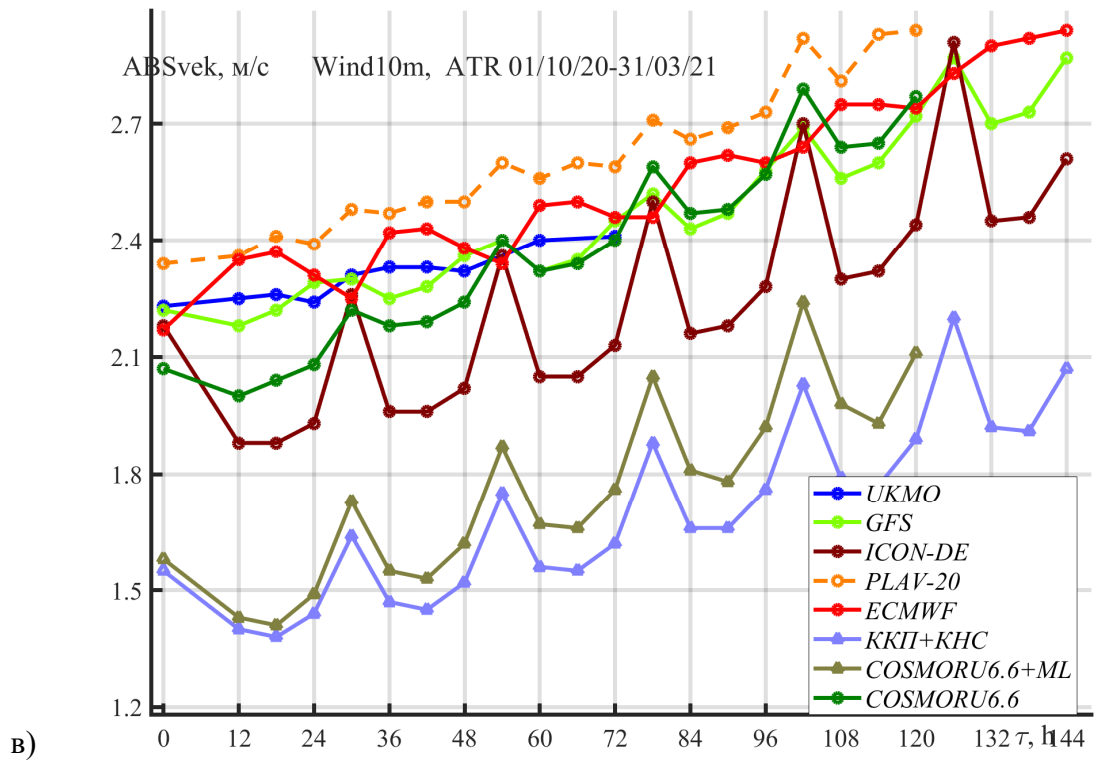
На рисунке 5.3 приведены сравнительные оценки среднего модуля векторной ошибки прогнозов скорости ветра на высоте 10 м по различным ЧМПП, а также по предлагаемым постпроцессингом: комплексным прогнозом и коррекцией КСК+КНС для прогнозов модели COSMO-Ru в различных конфигурациях. Аналогичные оценки для начального срока 12:00 ВСВ приведены в приложении А на рисунке А.4.

Погрешности прогнозов скорости ветра на высоте 10 м имеют сильный суточный ход: днём погрешности прогнозов по предлагаемым методам примерно на 25% больше. Этот суточный ход связан в том числе с суточным ходом абсолютных скоростей ветра в природе.

Средняя абсолютная погрешность прогнозов по предлагаемым методам постпроцессинга на 0.5-0.7 м/с меньше погрешности исходных прогнозов, что соответствует уменьшению заблаговременности прогноза на 4 и более суток.

Из рисунка 5.3 видно, что качество комплексного прогноза скорости ветра по всем рассмотренным регионам в среднем эквивалентно качеству поправленных прогнозов COSMO-Ru6ENA+КСК+КНС с на 0.5 суток меньшей заблаговременностью, а при той же заблаговременности средняя абсолютная векторная ошибка меньше на 0.03-0.22 м/с.





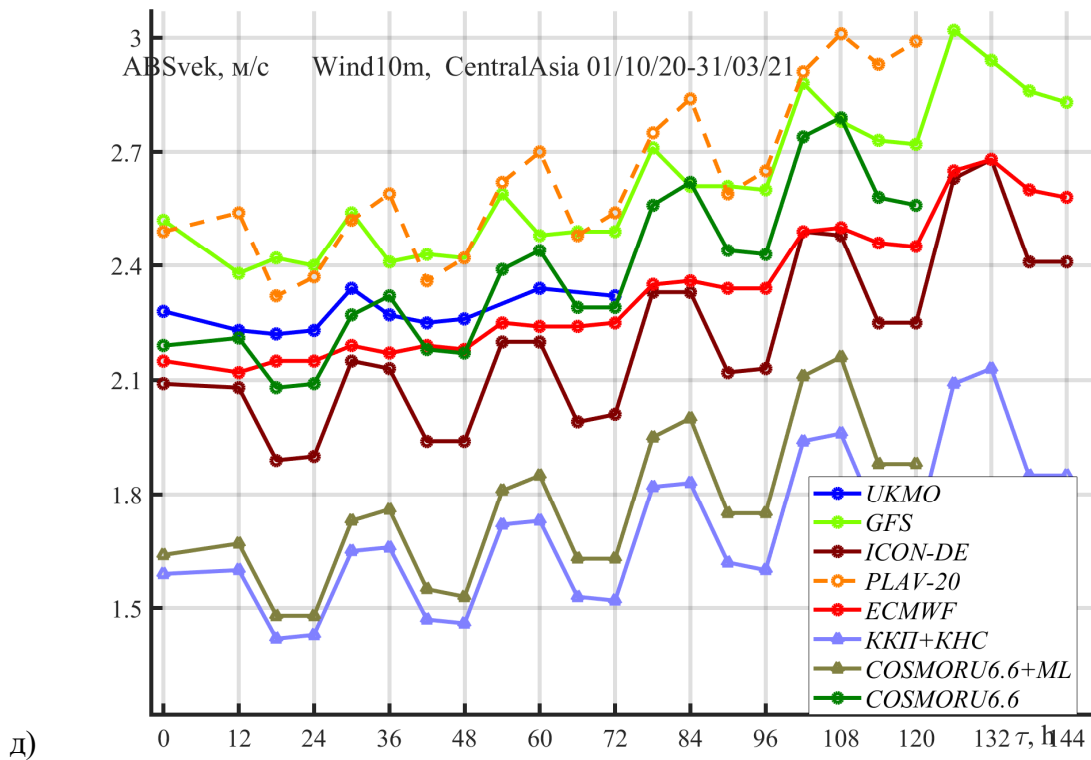


Рисунок 5.3 – Средний модуль векторной ошибки (ось абсцисс) прогнозов скорости ветра на высоте 10 м от начального срока 00:00 ВСВ по различным методам за период с 1 октября 2020 г. по 31 марта 2021 г. по различным регионам в зависимости от заблаговременности (ось ординат): а) Московский регион; б) ЕЧР; в) АЧР; г) Кавказ; д) Средняя Азия

В таблицах 5.1 и 5.2 приведены таблицы сопряженности и оценки качества (параграф 1.1.2): предупреденность Pr , доля ложных тревог FAR (false alarm rate) и критерий Багрова – Хайдке HSS (формула (1.14), [131], [155], [18]) для различных методов прогноза порывов ветра по двум градациям: более 12 и более 18 м/с. Аналогичные оценки для начального срока 12:00 ВСВ приведены в приложении А в таблицах А.1 и А.2. Видно, что комплексный прогноз обеспечивает заметно большие (а значит, и лучшие) значения критерия Багрова – Хайдке и предупреденности.

Предупреденность комплексного прогноза для градации 12 м/с превосходит предупреденность прогнозов лучшей из рассмотренных моделей (COSMO-Ru6.6ENA, ICON-De) на 4-5% (68% против 73% и 61% против 65%), а для градации 18 м/с разница в предупреденности гораздо больше (56% против 34% и 58% против 38% для заблаговременностей 12 и 24ч, соответственно).

Таблица 5.1 – Показатели успешности и таблицы сопряженности (таблица 1.1) прогнозов порывов ветра более 12 и более 18 м/с с заблаговременностью 12 ч от начального срока 00:00 ВСВ по различным методам за период с 1 октября 2020 г по 31 марта 2021 г по ЕЧР

Более 12 м/с			Более 18 м/с		
GFS	Pr=59% FAR=54%	HSS=0.46	GFS	Pr=17% FAR=72%	HSS=0.21
5229	6252	11481	76	200	276
3588	91136	94724	375	105554	105929
8817	97388	106205	451	105754	106205
ICON-De	Pr=68% FAR=51%	HSS=0.52	ICON-De	Pr=34% FAR=66%	HSS=0.34
6069	6354	12426	156	303	459
2803	91564	94369	302	106029	106331
8872	97918	106790	458	106332	106790
COSMO-Ru6.6ENA	Pr=66% FAR=50%	HSS=0.52	COSMO-Ru6.6ENA	Pr=34% FAR=63%	HSS=0.35
5519	5536	11055	153	261	414
2801	86930	89731	291	100081	100381
8320	92466	100786	444	100342	100786
COSMO-Ru6.6ENA + KCK+KHC	Pr=63% FAR=65%	HSS=0.55	COSMO-Ru6.6ENA + KCK+KHC	Pr=47% FAR=59%	HSS=0.43
5258	4190	6448	210	306	516
3062	88276	91338	234	100036	100270
8320	92466	100786	444	100342	100786
COSMO-RuBy	Pr=53% FAR=44%	HSS=0.51	COSMO-RuBy	Pr=26% FAR=47%	HSS=0.35
4375	3433	7808	115	103	218
3843	87963	91806	328	99068	99396
8218	91396	99614	443	99171	99614
COSMO-RuBy + KCK+KHC	Pr=55% FAR=38%	HSS=0.55	COSMO-RuBy + KCK+KHC	Pr=37% FAR=52%	HSS=0.42
4512	2732	7244	164	178	342
3706	88664	92370	279	98993	99272
8218	91396	99614	443	99171	99614
Комплексный прогноз	Pr=73% FAR=49%	HSS=0.56	Комплексный прогноз	Pr=56% FAR=67%	HSS=0.41
6450	6152	12602	253	506	759
2363	90928	93291	200	104934	105134
8813	97080	105893	453	105440	105893

Постпроцессинг прогнозов модели COSMO-Ru в конфигурациях COSMO-Ru6.6ENA и COSMO-RuBy показывает более высокие значения критерия Багрова – Хайдке (0.50-0.52 для градации 12 м/с и 0.43-0.44 для градации 18 м/с), чем исходные прогнозы той же конфигурации (0.46 для градации 12 м/с и 0.37-0.40 для градации 18 м/с). Предупрежденность поправленных прогнозов чуть выше (39-41% против 35-38% у исходных прогнозов) для градации более 18 м/с и чуть-чуть ниже (48-53% против 52-61% у исходных прогнозов) для градации 12 м/с.

Таблица 5.2 – То же, что в таблице 5.1, но с заблаговременностью 24 ч

Более 12 м/с			Более 18 м/с		
GFS	Pr=55% FAR=76%	HSS=0.31	GFS	Pr=22% FAR=85%	HSS=0.18
2628	11455	14983	54	301	355
1212	90010	91222	190	105660	105850
4740	101465	106205	244	105961	106205
ICON-De	Pr=52% FAR=51%	HSS=0.48	ICON-De	Pr=28% FAR=55%	HSS=0.34
2465	2582	5047	68	83	151
2298	99445	101743	176	106463	106639
4763	102027	106790	244	106546	106790
COSMO-Ru6.6ENA	Pr=61% FAR=59%	HSS=0.46	COSMO-Ru6.6ENA	Pr=38% FAR=63%	HSS=0.37
2777	3989	6766	93	158	251
1808	92212	94020	151	100384	100535
4585	96201	100786	244	100542	100786
COSMO-Ru6.6ENA + KCK+KHC	Pr=53% FAR=45%	HSS=0.52	COSMO-Ru6.6ENA + KCK+KHC	Pr=41% FAR=54%	HSS=0.43
2434	1988	4422	99	114	213
2151	94213	96364	145	100428	100573
4585	96201	100786	244	100542	100786
COSMO-RuBy	Pr=52% FAR=54%	HSS=0.46	COSMO-RuBy	Pr=35% FAR=54%	HSS=0.40
2315	2761	5076	77	89	166
2112	91840	93952	145	98717	98862
4427	94601	99028	222	98806	99028
COSMO-RuBy + KCK+KHC	Pr=48% FAR=43%	HSS=0.50	COSMO-RuBy + KCK+KHC	Pr=39% FAR=49%	HSS=0.44
2148	1606	3753	87	83	170
2280	92995	95275	135	98723	98858
4427	94601	99028	222	98806	99028
Комплексный прогноз	Pr=65% FAR=52%	HSS=0.53	Комплексный прогноз	Pr=58% FAR=61%	HSS=0.47
3098	3309	6407	142	220	362
1640	97658	99298	101	105242	105343
4738	100967	105705	243	1054621	105705

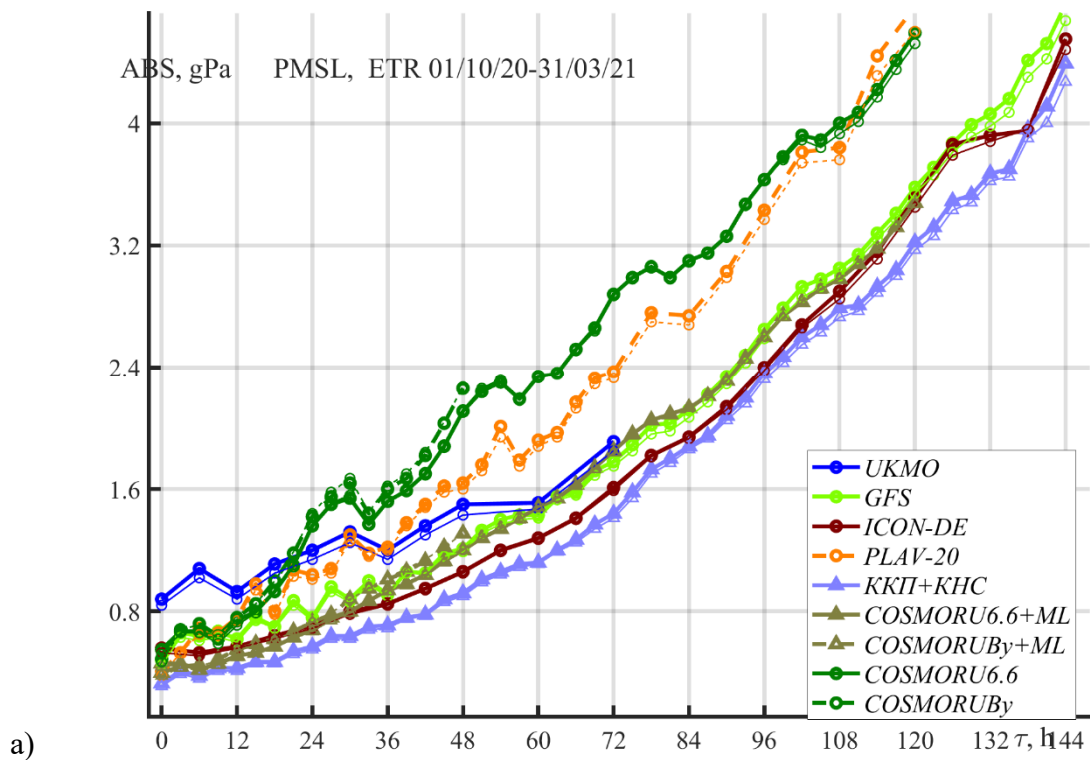
5.3. Сравнительные оценки качества прогнозов давления

На рисунке 5.4 приведены сравнительные оценки качества прогнозов давления, приведенного к уровню моря, P_0 по различным ЧМПП, а также по предлагаемым постпроцессингом: комплексным прогнозом и коррекцией КСК+КНС для прогнозов модели COSMO-Ru в различных конфигурациях COSMO-Ru6ENA и COSMO-RuBy. Аналогичные оценки для начального срока 12:00 ВСВ приведены в приложении А на рисунке А.4.

Из рисунка 5.4 видно, что качество комплексного прогноза P_0 по всем рассмотренным регионам, в среднем эквивалентно качеству прогнозов лучшей из рассмотренных моделей (ICON-De и GFS) с на 0.5 суток меньшей заблаговременностью.

Средняя абсолютная погрешность поправленных прогнозов COSMO-Ru6ENA меньше погрешности исходных прогнозов на 0.23-0.95 гПа на ЕЧР и на 0.28-1.28 гПа на АЧР. При сравнении с прогнозами лучших из доступных иностранных моделей MAE в первые 2 суток немного меньше (до 0.22 гПа на ЕЧР и до 0.41 гПа на АЧР), а в последующие сутки немного (до 0.20 гПа на ЕЧР и до 0.42 гПа на АЧР) больше погрешности моделей ICON-De и GFS.

Отметим, что суточный ход погрешностей прогнозов давления гораздо меньше (не более 5%) суточного хода (15-30%) рассмотренных выше погрешностей прогнозов температуры, точки росы и скорости ветра.



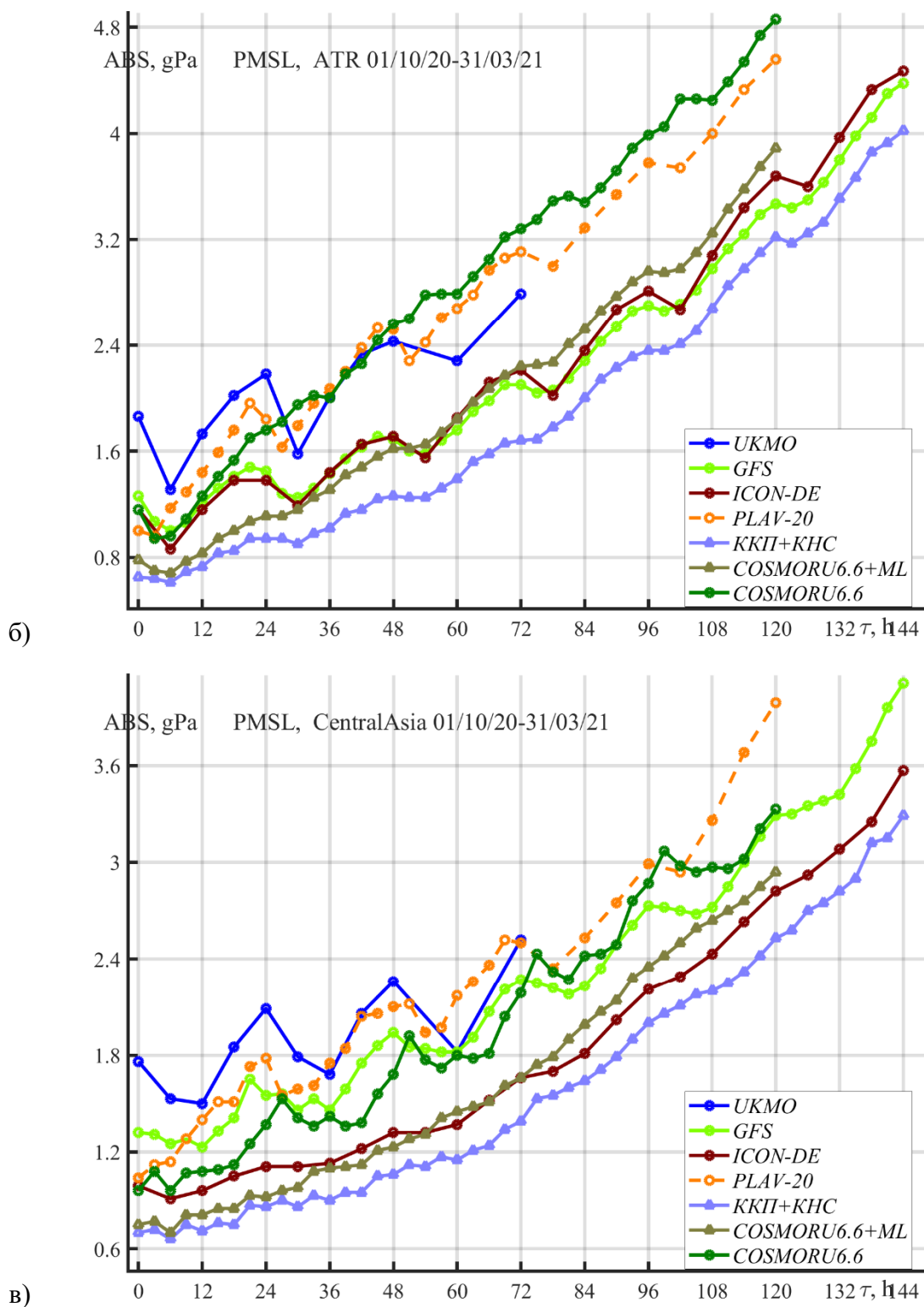


Рисунок 5.4 – Средняя абсолютная ошибка прогнозов (ось абсцисс) давления, приведенного к уровню моря, от начального срока 00:00 ВСВ по различным методам за период с 1 октября 2020 г. по 31 марта 2021 г. по различным регионам в зависимости от заблаговременности (ось ординат):
а) ЕЧР; б) АЧР; в) Средняя Азия

5.4. Сравнительные оценки качества прогнозов экстремальных температур воздуха

Результаты сравнительных оценок качества прогнозов минимальной ночной и максимальной дневной температур воздуха в 86 центрах субъектов РФ синоптиков УГМС, передаваемые в ФГБУ «Гидрометцентр России» в коде КП-68, комплексного прогноза и метода РЭП (расчет элементов погоды, [30], [29], [31]) приведены в таблице 5.3. Результаты сравнения из таблицы 5.3 за период с 1 октября 2020 г. по 31 марта 2021 г. по ЕЧР и АЧР предоставлены диссертанту А.Н. Багровым.

Таблица 5.3 – Сравнение качества различных прогнозов минимальной ночной и максимальной дневной температур воздуха по центрам субъектов РФ за период с 1 октября 2020 г. по 31 марта 2021 г.: оправдываемость, % / средняя ошибка МЕ, °С / средняя абсолютная ошибка МАЕ, °С

Минимальная ночная, ЕЧР	1 сутки	2 сутки	3 сутки
ККП+КНС	96.0/-0.1/1.17	95.1/-0.1/1.31	92.7/-0.1/1.46
РЭП	94.6/ 0.1 /1.30	92.0/ 0.1 /1.53	89.9/ 0.0 /1.64
УГМС (код КП-68)	94.8/-0.3/1.37	92.3/-0.3/1.57	89.8/-0.3/1.74
Максимальная дневная, ЕЧР	1 сутки	2 сутки	3 сутки
ККП+КНС	97.8/ 0.0/1.00	96.7/ 0.0/1.14	94.0/ 0.1/1.35
РЭП	96.6/ 0.0 /1.14	94.9/ 0.0 /1.32	92.9/ 0.0 /1.45
УГМС (код КП-68)	96.3/-0.2/1.24	94.3/-0.2/1.40	92.3/-0.1/1.53
Минимальная ночная, АЧР	1 сутки	2 сутки	3 сутки
ККП+КНС	92.8/-0.1/1.39	90.6/-0.1/1.57	89.0/-0.2/1.74
РЭП	90.4/ 0.1 /1.58	88.5/ 0.1 /1.70	86.1/ 0.0 /1.86
УГМС (код КП-68)	89.3/- 0.1 /1.66	86.3/- 0.1 /1.85	83.5/-0.2/2.05
Максимальная дневная, АЧР	1 сутки	2 сутки	3 сутки
ККП+КНС	96.6/ 0.0/1.15	94.8/ 0.0/1.28	93.4/ 0.0/1.42
РЭП	94.6/ 0.0 /1.33	93.2/ 0.0 /1.48	91.0/ 0.0 /1.58
УГМС (код КП-68)	93.4/-0.2/1.46	90.4/-0.3/1.6	88.2/-0.2/1.76

Из таблицы 5.3 видно, что по обеим рассмотренным территориям и для всех рассмотренных заблаговременностей первое место (выделено жирным) среди рассмотренных прогнозов занимает комплексный прогноз, на втором месте прогноз РЭП, а на последнем – прогноз синоптиков. За рассмотренный период качество комплексного прогноза в среднем лучше, чем прогноз синоптиков с заблаговременностью на 1-2 суток меньшей и близко по качеству к прогнозу РЭП с заблаговременностью на 1 сутки меньшей.

Отметим, что синоптикам УГМС на ЕЧР в момент составления прогноза, вообще говоря, были доступны оцениваемые тут комплексные прогнозы от 00:00 ВСВ, размещаемые на сайте методического кабинета Гидрометцентра России [9] в оперативном режиме. Синоптикам УГМС

на АЧР в момент составления прогноза были доступны лишь прогнозы от 12:00 ВСВ предыдущих суток. По-видимому, это может объяснить, почему отставание качества прогнозов УГМС от качества комплексных прогнозов в АЧР примерно на 0.5 суток больше чем в ЕЧР.

Таблица 5.4 – То же, что в таблице 5.3, но сравнение производилось только для случаев, когда фактическое межсуточное изменение прогнозируемой экстремальной температуры составляло более 5 градусов

Минимальная ночная, ЕЧР	1 сутки N=546	2 сутки N=508	3 сутки N=513
ККП+КНС	91.7/-0.1/1.54	89.7/-0.2/1.64	83.6/-0.3/1.94
РЭП	88.1/ 0.0 /1.70	86.3/ 0.0 /1.87	80.8/ -0.2 /2.11
УГМС (код КП-68)	88.8/-0.4/1.83	84.8/-0.3/2.06	79.7/-0.5/2.34
Максимальная дневная, ЕЧР	1 сутки N=695	2 сутки N=620	3 сутки N=636
ККП+КНС	92.5/ 0.0/1.43	89.8/ 0.0/1.63	84.6/ 0.0/1.94
РЭП	86.6/ 0.0 /1.76	83.9/ 0.0 /1.96	80.3/ 0.0 /2.26
УГМС (код КП-68)	87.4/-0.2/1.85	85.0/-0.3/1.99	81.5/-0.2/2.15
Минимальная ночная, АЧР	1 сутки N=421	2 сутки N=348	3 сутки N=344
ККП+КНС	88.6/ 0.0/1.67	84.9/-0.1/1.94	82.7/-0.3/2.13
РЭП	83.5/ 0.1/1.95	79.8/ 0.0 /2.15	77.3/ -0.1 /2.38
УГМС (код КП-68)	83.2/-0.1/2.05	80.0/-0.3/2.25	75.2/-0.3/2.52
Максимальная дневная, АЧР	1 сутки N=577	2 сутки N=484	3 сутки N=475
ККП+КНС	95.0/-0.1/1.37	90.2/-0.3/1.60	86.6/-0.3/1.84
РЭП	89.1/ -0.1 /1.75	88.5/ -0.3 /1.91	82.0/ -0.3 /2.17
УГМС (код КП-68)	84.8/-0.5/2.03	81.0/-0.7/2.24	79.1/-0.6/2.46

В таблице 5.4 отдельно приведены результаты сравнительных оценок прогноза экстремальных температур для тех случаев, когда фактическое межсуточное изменение этой экстремальной температуры по сравнению с предыдущими сутками составило более 5°C. Естественно, средние оценки качества всех прогнозов в таблице 5.4 несколько хуже, чем в таблице 5.3, однако качество комплексного прогноза максимальной дневной температуры снизилось заметно меньше: например, на ЕЧР оправдываемость при прогнозе на 1, 2 и 3 суток уменьшилась на 5.3%, 6.9% и 9.4% соответственно, в то время как оправдываемость прогнозов РЭП уменьшилась на 10.0%, 11.0% и 12.6%, а прогнозов УГМС уменьшилась на 11.9%, 9.7% и 10.8%.

На рисунке 5.5 приведены результаты сравнения качества комплексного прогноза минимальной ночной и максимальной дневной температур воздуха с прогнозами, представленными на интернет-сайтах meteoinfo.ru, gismeteo.ru, yandex.ru, foreca.com, rp5.ru, meteovesti.ru, weather.com за период с 1 октября 2020 по 31 марта 2021. Сайт weather.com оценивался начиная с 14 октября 2020. Результаты сравнения предоставлены Н.А.Светловой.

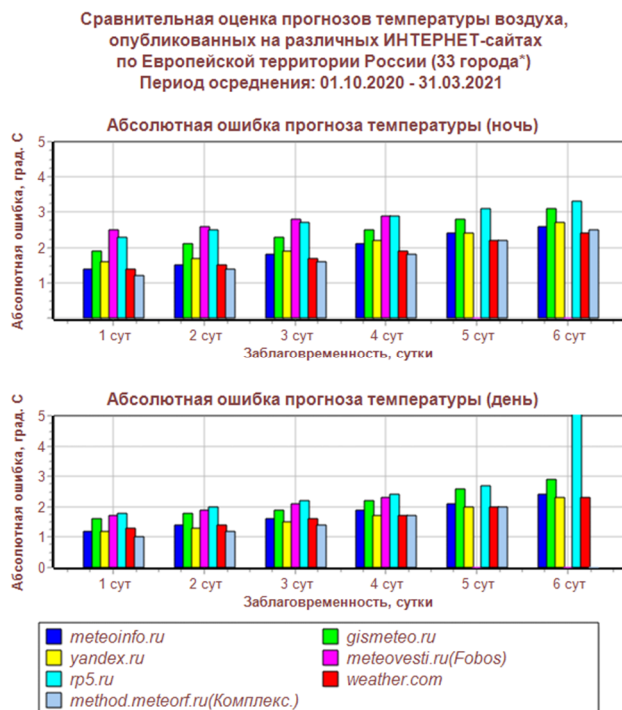
Производилось сравнение прогнозов по 53 крупным городам РФ: по 33 городам на ЕЧР и по 20 городам на АЧР. Прогнозы скачивались с сайтов около 09:00 ВСВ, а комплексный прогноз вычислялся в оперативном режиме счета в 05:15 ВСВ.

Отметим, что прогнозы на официальном сайте Гидрометцентра России meteoinfo.ru рассчитываются по методу РЭП. Однако, автор метода РЭП уточняет свои прогнозы после поступления прогностической информации из ECMWF, что приводит к тому, что прогнозы с сайта meteoinfo.ru (и их оценки качества) отличаются от прогнозов по методу РЭП, которые выше сравнивались с прогнозами синоптиков УГМС.

Прогнозы на сайте yandex.ru вычисляются с использованием машинного обучения по методу Матрикснет (фирменной реализацией метода ансамбля деревьев решений) и основаны на прогнозах по моделям GFS и WRF (Weather Research and Forecasting) [52].

Из рисунка 5.5 видно, что комплексный прогноз занимает среди рассмотренных сайтов первое место при заблаговременности прогноза до 3 суток, при заблаговременности 4-5 суток делит первое и второе место с сайтом weather.com.

а)



б)

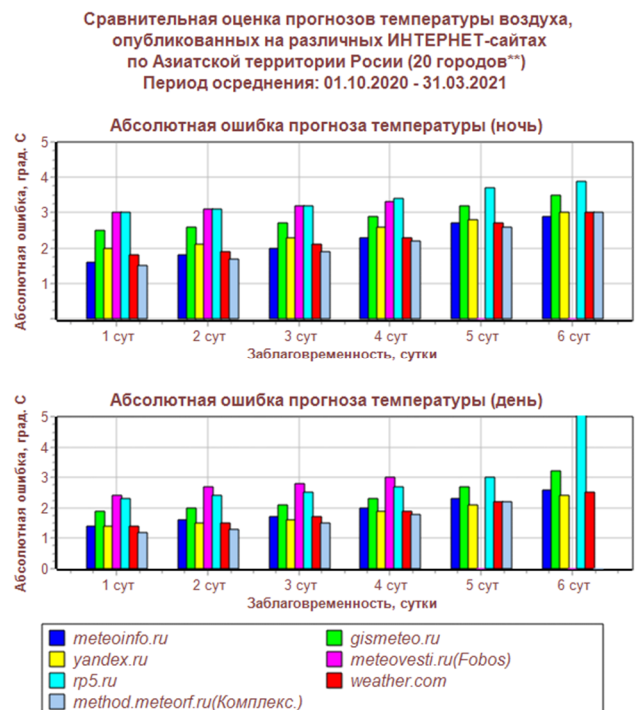


Рисунок 5.5 – Результаты сравнительной оценки средней абсолютной ошибки комплексного прогноза минимальной ночной и максимальной дневной температур воздуха с прогнозами, представленными на сайтах сети интернет:

а) по 33 городам на ЕЧР; б) по 20 городам на АЧР

Выводы из главы 5

В главе 5 представлены результаты сравнительных оценок качества прогнозов, вычисленным постпроцессингом, предложенным в главах 2 и 3.

Продemonстрировано, что качество поправленных КСК+КНС прогнозов модели COSMO-Ru в конфигурации COSMO-Ru6ENA (далее COSMO-Ru6ENA+КСК+КНС) эквивалентно качеству исходных прогнозов COSMO-Ru6ENA с на 2-2.5 суток меньшей заблаговременности.

Качество комплексного прогноза приземной температуры воздуха и точки росы в среднем эквивалентно качеству прогнозов COSMO-Ru6ENA+КСК+КНС с на 1-1.5 суток меньшей заблаговременностью, которое близко или чуть лучше, чем качество прогнозов лучшей из доступных численных моделей.

Качество комплексного прогноза скорости ветра весьма близко к качеству поправленных прогнозов COSMO-Ru6ENA+КСК+КНС. Однако, качество комплексного прогноза порывов ветра несколько превосходит качество поправленных прогнозов COSMO-Ru6ENA+КСК+КНС, особенно при прогнозе порывов более 18 м/с.

При сравнении качества комплексного прогноза с прогнозами синоптиков УГМС и прогнозам РЭП выяснилось, что качество комплексного прогноза близко к качеству прогнозов УГМС с на 1-2 суток меньшей заблаговременностью и близко по качеству к прогнозу РЭП с на 1 сутки меньшей заблаговременностью.

Среди 7 рассмотренных погодных сайтов качество прогнозов по территории РФ за период с 1 октября 2020 г по 31 марта 2021 г на сайте комплексного прогноза занимает первое место при заблаговременности прогноза до 3 суток, а при заблаговременности 4-5 суток делит первое и второе место с сайтом weather.com.

Заключение

Основные результаты

В диссертации предложен новый нейросетевой комплекс постпроцессинга (коррекции) численных моделей прогноза погоды (ЧМПП), которые были обучены, валидированы и протестированы на больших архивах (объемом порядка 10^6 - 10^8 пар из прогностических и фактических значений) прогнозов ЧМПП и синхронных им измерений на синоптических метеорологических станциях на территории Российской Федерации и сопредельных государств.

Предложенный комплекс постпроцессинга:

- 1) Использует прогнозы приземных метеорологических параметров одной или нескольких различных ЧМПП;
- 2) Учитывает последние (за период длительностью 35 дней) измерения на синоптических станциях;
- 3) Применим также и в случаях частичного отсутствия прогностической информации, и (или) в случаях отсутствия (отбраковки алгоритмом контроля качества) части данных наблюдений;
- 4) Выдаёт прогнозы как в точках синоптических станций, так и в виде полей на модельной сетке;
- 5) Существенно улучшает качество прогнозов;
- 6) Применим в оперативном режиме счета на существующей инфраструктуре (для подготовки прогнозов в виде полей на сетке высокого разрешения необходимо использование GPU).

По результатам тестирования в холодный период года в оперативном режиме счета качество прогнозов приземных метеорологических параметров (температуры воздуха (срочной, максимальной дневной и минимальной ночной), точки росы, скорости ветра, порывов ветра, давления, приведенного к уровню моря) по предлагаемому постпроцессингу близко к качеству используемых исходных прогнозов ЧМПП с на 2-4 суток меньшей заблаговременностью.

Примеры продукции

Примеры продукции, использующей прогнозы, рассчитанные предложенным в диссертации постпроцессингом, представлены в приложении Б. Комплексный прогноз доступен оперативно:

- для населенных пунктов России и Беларуси демонстрируется на сайте методического кабинета Гидрометцентра России <http://method.meteorf.ru/ansambl/ansambl.html> [9], где прогнозы с заблаговременностью до 96ч доступны для всех регионов РФ в виде карт (пример карты показан на рисунке Б.1 в приложении Б, авторы Н.А.Светлова Е.Н.Шакотько), а для некоторых населенных пунктов – в виде метеограмм (пример показан на рисунке Б.2 в приложении Б);
- для населенных пунктов Центральной Азии в виде метеограмм (внешний вид аналогичен метеограммам для станций РФ) на сайте проекта по прогнозированию суровых погодных условий в Центральной Азии <http://swfdp-ca.meteoinfo.ru/prognozy/mmforecasts>;
- для населенных пунктов России на сайте «О погоде для специалистов» Гидрометцентра России <https://special.meteoinfo.ru/> в разделах «метеограммы» (дублируются метеограммы с сайта методического кабинета) и «прогнозы по пунктам» (пример приведен на рисунке Б.3 в приложении Б).

Прогнозы, рассчитанные КСК+КНС для модели COSMO-Ru в конфигурациях COSMO-Ru6ENA и COSMO-RuBy демонстрируются оперативно на сайтах:

- Численных прогнозов погоды для метеорологического обеспечения <http://u2019.meteoinfo.ru/services>: на метеограммах демонстрируются графики поправленных прогнозов температуры и точки росы на высоте 2 м (пример метеограммы показан на рисунке Б.4 приложения Б);
- «О погоде для специалистов» Гидрометцентра России <https://special.meteoinfo.ru/> в разделе «метеограммы» (дублируются метеограммы с <http://u2019.meteoinfo.ru/services>).

Комплексный прогноз максимальной дневной и минимальной ночной температур воздуха оценивается оперативно и эти оценки демонстрируются на сайте методического кабинета в виде:

- Сравнения средних абсолютных ошибок прогнозов, размещенных на некоторых сайтах сети интернет (пример приведен на рисунке Б.5 Приложения Б). Оперативно доступны сравнения за прошедший месяц, за последние 10 дней, для отдельного города и средние ошибки по 33 городам ЕЧР и 20 городам АЧР;
- Сравнения качества прогнозов (пример приведен на рисунке Б.6 Приложения Б) с прогнозами УГМС, передаваемых в коде КП-68 и прогнозами по методу РЭП [29];
- Сравнение прогнозов на ближайшую ночь и следующий день с фактическими значениями в виде графиков прогнозов за последние 15 суток. Пример сравнения приведен на рисунке Б.7 Приложения Б.

Сравнение средних и средних абсолютных ошибок комплексного прогноза (отдельно для почти всех синоптических станций РФ) оперативно демонстрируется на сайте «О погоде для специалистов». Пример такого сравнения показан на рисунке Б.8 приложения Б.

Использование результатов диссертации

Разработанный комплекс постпроцессинга позволяет выпускать продукцию, а именно откорректированные прогнозы приземных метеорологических параметров в виде метеограмм для точек синоптических станций, а также в виде синоптических карт.

Подготавливаемые прогнозы могут быть использованы непосредственно потребителем, так и быть использованы для дальнейшего постпроцессинга. Например, автором были разработаны методики прогноза потребления электроэнергии [25], [27] и нагрузки на службу «Скорой помощи» [117], [42], использующие прогнозы приземных метеорологических параметров, рассчитанные предлагаемым постпроцессингом.

Рассчитываемая по предлагаемым постпроцессингом продукция (примеры продукции приведены в приложении Б) размещается на интернет-сайтах ФГБУ «Гидрометцентр России»:

- методического кабинета <http://method.meteorf.ru/ansambl/ansambl.html>;
- проекта по прогнозированию суровых погодных условий в Центральной Азии <http://swfdp-ca.meteoinfo.ru/prognozy/mmforecasts>;
- Численных прогнозов погоды для метеорологического обеспечения <http://u2019.meteoinfo.ru/services>;
- «О погоде для специалистов» <https://special.meteoinfo.ru/>.

В приложении В приведены отзывы синоптиков региональных УГМС и ЦГМС о продукции, рассчитываемой по предлагаемым постпроцессингом и предоставляемой на сайте методического кабинета Гидрометцентра России.

Перспективы дальнейших исследований

Для уменьшения погрешностей моделирования результаты предложенного постпроцессинга могут быть использованы путём «усвоения» полученных поправок численной моделью, например, методом nudging [103]. В методе nudging усвоение происходит путем добавления в правую часть решаемых моделью дифференциальных уравнений в частных производных форсинга, притягивающего модельные переменные к измеренным (в данном случае

откорректированным) значениям. Пример усвоения поправок для поля H500 (высоты поверхности 500 гПа), вычисляемых методами машинного обучения, предложен в работе [240].

Дальнейшее развитие разработанного постпроцессинга может происходить в нескольких направлениях:

1. добавление дополнительных предикторов в список аргументов нейронных сетей (feature engineering);
2. поиск более совершенных реализаций используемых нейронных сетей;
3. применение к другим метеорологическим полям, например, к аэрологическим;
4. замена используемых предположений на более общие.

Развитие в направлениях 1 и 2 может улучшить качество продукции постпроцессинга. Работа в направлении 1 может осуществляться как методом перебора дополнительных предикторов, так как и путем тестирования индексов, разработанных в других работах. Работа в направлении 2 может заключаться:

- а. в применении других архитектур нейронной сети, например: свёрточных нейронных сетей (CNN, [181]); рекуррентных нейронных сетей (RNN [157], [147]); архитектур, найденных алгоритмами поиска архитектуры [107], [130], [203];
- б. ансамблировании нескольких нейронных сетей;
- в. совершенствовании методов обучения, например, использовании так называемого активного обучения, когда выбор minibatch делается неслучайным [230].

Остановимся подробнее на возможном развитии в направлениях 3 и 4 для каждой из предложенных моделей МО.

Аэрологические поля существенно менее чувствительны к подстилающей поверхности, поэтому можно предположить, что гипотеза локальности погрешностей ЧМПП для них заметно нарушается. Аналогичная гипотеза нарушается и для погрешностей долгосрочных прогнозов. При коррекции долгосрочных прогнозов погоды (например, [80]) используются дальние корреляционные связи, поэтому развитие КСК, ККП и КНС в направлении 3 вероятно будет тесно связано с применением CNN к прогностическим полям (направление 2) и (или) использованием индексов циркуляции атмосферы (направление 1). Причем индексы циркуляции могут вычисляться CNN.

КСК и ККП использовали существенное предположение квазилинейности, отказаться от которого можно, используя нейронные сети с механизмом внимания [237]. Учет кросскорреляционных связей смещений различных ЧМПП может улучшить качество прогнозов, рассчитанных ККП (направление 4). Для учёта этих кросскорреляционных связей потребуется решать СЛАУ с корреляционной матрицей в левой части. Методы преодоления трудностей,

возникающих при решении СЛАУ в модели МО, использующей нейронные сети, были проработаны в параграфе 4.2.1 настоящей работы.

При развитии в направлении 3 КНОИ может быть применена для интерполяции любых других двумерных полей, в том числе в системе усвоения данных метеорологических наблюдений. Его применение должно существенно улучшить качество интерполяции для тех метеорологических полей, которые не являются однородными, но причины (предикторы) неоднородности их известны. Возможно не непосредственное применение КНОИ, а применение полученных при оптимизации КНОИ оценок неоднородных корреляционных функций (КФ) в существующих системах усвоения данных, использующих, например, метод вариационного усвоения данных [106], фильтр Калмана [159].

При развитии КНОИ в направлении 4 можно попытаться отказаться от предположения о представлении выравнивающего отображения в виде графика функции. Этот отказ приведет к тому, что максимальная возможная корреляция для данного расстояния между точками станет непостоянной по всей рассматриваемой области.

Из предложенных моделей МО наиболее практически значимым, пожалуй, является КНОИ, который может получить дальнейшее развитие не только в постпроцессинге, но и найти применения в других областях математического моделирования в геофизике.

Во-первых, представляется возможным обобщить КНОИ на случаи: совместной интерполяции нескольких метеорологических полей; интерполяции трех- и четырехмерных метеорологических полей.

Во-вторых, в чуть более отдаленной перспективе КНОИ может быть применена в системе усвоения данных. Отметим, что в системе усвоения данных используются в том числе и данные косвенных наблюдений, например, с искусственных спутников Земли, для усвоения которых используется так называемый оператор наблюдения. Оператор наблюдения (и (или) обратный к нему [22]) строят в том числе и методами МО [179], [245], [213], [238].

При использовании КНОИ для усвоения данных косвенных наблюдений можно проводить совместную оптимизацию параметров интерполяции и параметров оператора наблюдений. Такая совместная оптимизация: а) будет использовать метод градиентного спуска в направлении градиента, вычисленного методом обратного распространения ошибки [34], что может дать синергетический эффект; б) позволит при оценке оператора наблюдений использовать все наблюдения за тот же срок, а не только колоцированные пары из данных прямых и косвенных наблюдений. Колоцированными называют пары наблюдений, произведенным в достаточно близких (для данной задачи) точках пространства-времени. При классическом поэтапном подходе оптимизация оператора наблюдений и оценка КФ

осуществляются независимо, и их совместная оптимизация может осуществляться только методами нулевого порядка: в классическом подходе критерии качества, используемые в задаче оценки оператора наблюдений и в задаче оценки КФ различны.

Более отдаленной перспективой является создание полностью нейросетевой системы численного прогноза погоды. Существующие системы прогноза погоды состоят из трех относительно плохо согласованных блоков: препроцессинга, физической модели прогноза и постпроцессинга. Предложенная в работе КНОИ после дополнительной настройки может быть применен так же и в блоке препроцессинга ЧМПП. Таким образом, предложенные алгоритмы и идеи при соответствующем развитии геофизических нейросетевых моделей [170] открывают путь к созданию физически обоснованной полностью нейросетевой системы прогноза погоды, все параметры которой могут быть оптимизированы одновременно.

Список сокращений и условных обозначений

АССОИ – автоматизированная система сбора и обработки информации ФГБУ «Гидрометцентр России»

АЧР – Азиатская часть России

БД – базы данных «Прогноз» ФГБУ «Гидрометцентр России»

ВСВ – всемирное согласованное время

г – год

ЕЧР – Европейская часть России

ЕОС – естественные ортогональные составляющие

КНОИ – квазилинейная неоднородная анизотропная интерполяция

КНС – коррекция с использованием нейронных сетей

ККП – квазилинейный комплексных прогноз

КСК – квазилинейная систематическая коррекция

КФ – корреляционная функция

ЛВФ – логарифм весовой функции

МО – машинное обучение

НКФ – нормированная корреляционная функция

ОДУ – обыкновенное дифференциальное уравнение

ОИ – оптимальная интерполяция

ОС – операционная система

ПЛАВ – полулагранжева модель, основанная на уравнении абсолютного вихря

Росгидромет – Федеральная служба по гидрометеорологии и мониторингу окружающей среды

РФ – Российская Федерация

РЭП – расчет элементов погоды

СЕАКЦ – Северо-Евразийский климатический центр

СибНИГМИ – Сибирский региональный научно-исследовательский гидрометеорологический институт

СНГ – содружество независимых государств

СЛАУ – система линейных алгебраических уравнений

УГМС – управление по гидрометеорологии и мониторингу окружающей среды Росгидромета

ФГБУ – Федеральное государственное бюджетное учреждение

ЦГМС – Центр по гидрометеорологии и мониторингу окружающей среды

ЦФО – Центральный федеральный округ

ч – час

ЧМПП – численная модель прогноза погоды

CNN – convolutional neural network (свёрточные нейронные сети)

COSMO – Consortium for small-scale modeling (Консорциум мелкомасштабного моделирования)

CPU – central processing unit (центральное процессорное устройство)

DAG – directed acyclic graph (ориентированный граф без циклов)

DWD – Deutsche wetterdienst (Немецкая метеорологическая служба)

ECMWF – European centre for medium-range weather forecasts (Европейский центр среднесрочных прогнозов погоды)

ETS – equitable threat score (критический индекс успешности)

FAR – false alarm rate (доля ложных тревог)

FC – fully-connected layer (полносвязный слой)

GC – gradient centralization (центрирование градиентов)

GFS – global forecast system (глобальная система прогнозирования)

GPU – graphics processing unit (графическое процессорное устройство)

GRIB – general regularly-distributed information in binary form (общая регулярно распространяемая информация в двоичной форме)

HSS – Heidke skill score (критерий Багрова – Хайдке)

ICON – icosahedral nonhydrostatic model (икосаэдрическая негидростатическая модель)

JMA – Japan meteorological agency (Японское метеорологическое агентство)

MAE – mean absolute error (средняя абсолютная погрешность)

ME – mean error (средняя погрешность)

MLP – multilayer perceptron (многослойный перцептрон)

RMS – root-mean square error (среднеквадратичная погрешность)

RNN – recurrent neural network (рекуррентные нейронные сети)

UKMO – United Kingdom meteorological office (метеорологическое бюро Соединенного Королевства)

$\nabla_{\vec{x}} F$ – градиент функции F по вектору $\vec{x}$

$\|\cdot\|_2$ – евклидова норма

$I_A(x) = \begin{cases} 1, & x \in A, \\ 0, & x \notin A, \end{cases}$ – индикаторная функция множества A

$\langle \vec{x}, \vec{y} \rangle$ – конкатенация (склеивание) векторов $\vec{x}$ и $\vec{y}$

$\sigma(x) = (1 + \exp(-x))^{-1} \in (0, 1)$ – логистическая функция [93]

EX – математическое ожидание величины X

$\nabla_A F$ – матрица (тензор) того же размера, что матрица (тензор) A , составленная из градиентов функции F по соответствующим элементам матрицы (тензора) A

A^T – матрица, транспонированная к матрице A

$\# A$ – мощность (количество элементов) множества A

$\|A\| = \sup_x \frac{\|Ax\|}{x}$ – норма оператора

$\circ$ – поэлементное умножение (произведение Адамара)

$\delta_j^i = \begin{cases} 1, & i = j, \\ 0, & i \neq j, \end{cases}$ – символ Кронекера

$p(A|B)$ – условная вероятность события A , если произошло событие B

$\text{ReLU}(x) = \max(0, x) = x_+$ – функция ReLU (rectified linear unit)

Список литературы

1. Алдухов О. А., Гордин В. А.. Оценки анизотропии корреляционной структуры полей метеорологических величин по наблюдениям глобальной аэрологической сети // Изв. РАН. Серия «Физика атмосферы и океана», т. 41, № 3, 2005. с. 399-409.
2. Алдухов О.А., Багров А.Н., Гордин В.А.. Статистические характеристики прогностических метеорологических полей и их использование для объективного анализа // Метеорология и гидрология, т. 10, 2002. с. 18-33.

3. Алдухов О.А., Быков Ф.Л., Гордин В.А.. Крупномасштабные трехмерные корреляционные функции для атмосферы Земли // Ярославский педагогический вестник, т. 4, 2011. с. 36-43.
4. Алдухов О.А., Гордин В.А. Трехмерные корреляционные функции основных аэрологических величин // Изв. РАН. Физика атмосферы и океана, т. 37, № 1, 2001. с. 3-23.
5. Алферов Ю.В., Климова Е.Г.. Опыт использования фильтра Калмана для коррекции численного прогноза приземной температуры воздуха // Гидрометеорологические исследования и прогнозы, т. 378, № 4, 2020. с. 28-42.
6. Армстронг М. Основы линейной геостатистики. М.: Недра, 1998. 149 с.
7. Ароншайн Н. Теория воспроизводящих ядер // Математика, т. 7, № 2, 1963. с. 67-130.
8. Астахова Е.Д. Построение ансамблей начальных полей для системы кратко- и среднесрочного ансамблевого прогнозирования погоды // Труды Гидрометцентра России, № 342, 2008. с. 98-117.
9. Багров А.Н., Быков Ф.Л., Гордин В.А., Светлова Н.А., Пурина И.Э.. Комплексный прогноз по данным различных атмосферных моделей для городов России и Республики Беларусь [Электронный ресурс] // Методический кабинет Гидрометцентра России: [сайт]. URL: <http://method.meteorf.ru/ansambl/ansambl.html> (дата обращения: 04.03.2021).
10. Багров А.Н., Быков Ф.Л., Гордин В.А.. Комплексный прогноз приземных метеоэлементов // Метеорология и гидрология, т. 5, 2014. с. 5-16.
11. Багров А.Н., Быков Ф.Л., Гордин В.А.. Схема оперативного краткосрочного комплексного прогноза ветра // Метеорология и гидрология, т. 7, 2018. с. 19-26.
12. Багров А.Н., Быков Ф.Л., Гордин В.А.. Схема оперативного краткосрочного комплексного прогноза приземной температуры воздуха и влажности // Метеорология и гидрология, т. 8, 2018. с. 5-18.
13. Багров А.Н., Быков Ф.Л., Гордин В.А. Сборник тезисов Международной научно-практической конференции, посвященной 90-летию Российского государственного гидрометеорологического университета // Схема комплексного прогноза погоды по территории стран СНГ на срок до 6 суток. 2020. с. 122-123.
14. Багров А.Н., Гордин В.А., Локтионова Е.А., Очан Н.Ю.. Контроль и архивация в Росгидрометцентре глобальных данных о приземной температуре воздуха // Метеорология и гидрология, т. 2, 1993. с. 18-26.
15. Багров А.Н., Гордин В.А., Халявин А.В.. Ансамблевый прогноз приземной температуры воздуха и количества осадков // Метеоспектр, № 4, 2009. с. 113-115.

16. Багров Н.А. Аналитическое представление последовательности метеорологических полей посредством естественных ортогональных составляющих // Труды ЦИП, № 74, 1959. с. 3-74.
17. Багров Н.А. К вопросу об оценке гидрометеорологических прогнозов // Метеорология и гидрология, № 6, 1953. с. 13-16.
18. Багров Н.А. О статистических свойствах некоторых оценок прогнозов // Труды ММЦ, № 9, 1966. с. 61-69.
19. Бахвалов Н.С., Жидков Н.П., Кобельков Г.Г.. Численные методы. М.: Бином, 2008. 636 с.
20. Блинов Д.В., Ривин Г.С.. Система краткосрочной негидростатического прогноза погоды COSMO-Ru: технологическая линия // Труды Гидрометцентра России, № 365, 2017. с. 142-162.
21. Борщ С.В., Симонов Ю.А. Оперативная система краткосрочных гидрологических прогнозов расхода воды на реках бассейна Кубани // Труды Гидрометцентра России, № 349, 2013. с. 58-79.
22. Быков Ф.Л., Василенко Е.В., Гордин В.А., Тарасова Л.Л.. Статистическая структура поля влажности верхнего слоя почвы по данным наземных и спутниковых наблюдений // Метеорология и гидрология, № 6, 2017. с. 68-84.
23. Быков Ф.Л., Василенко Е.В., Гордин В.А., Тарасова Л.Л. Материалы II Всероссийской научной конференции с международным участием «Применение средств дистанционного зондирования Земли в сельском хозяйстве // Совместный оперативный анализ станционных и спутниковых данных о влагозапасе почвы. СПб. 2018. с. 279-284.
24. Быков Ф.Л., Василенко Е.В., Гордин В.А., Тарасова Л.Л. Сборник тезисов докладов четырнадцатой всероссийской открытой конференции "Современные проблемы дистанционного зондирования Земли из космоса // Анализ влажности почвы по данным наземной сети и дистанционного спутникового зондирования. 2016. Р. 334.
25. Быков Ф.Л., Гордин В.А.. Краткосрочный прогноз часового потребления электроэнергии с учетом погоды для субъектов РФ // Известия РАН: Энергетика, т. 5, 2017. с. 47-56.
26. Быков Ф.Л., Гордин В.А.. Трехмерный объективный анализ структуры атмосферных фронтов // Известия РАН. Физика атмосферы и океана, т. 48, № 2, 2012. с. 172-188.
27. Быков Ф.Л. Межвузовская научно-техническая конференция студентов, аспирантов и молодых специалистов им. Е.В. Арменского // Краткосрочный прогноз потребления

- электроэнергии с учетом температуры воздуха и кластеризацией потребителей. 2019. с. 12-13.
28. Быков Ф.Л. Статистическая коррекция прогнозов погоды по модели COSMO с помощью нейронных сетей // Метеорология и гидрология, т. 3, 2020. с. 5-20.
 29. Васильев П.П., Васильева Е.В., Горлач И.А.. Метод прогноза экстремальной температуры воздуха до 3 суток по административным центрам субъектов Российской Федерации на основе технологии РЭП (автор - П.П. Васильев) и результаты его испытания // Результаты испытания новых и усовершенствованных технологий, моделей и методов гидрометеорологических прогнозов, № 37, 2010. с. 53-59.
 30. Васильев П.П., Васильева Е.Л., Горлач И.А.. Среднесрочный прогноз температуры воздуха и результаты его испытания // Результаты испытания новых и усовершенствованных технологий, моделей и методов гидрометеорологических прогнозов, № 37, 2010. с. 3-14.
 31. Вильфанд Р.М., Васильев П.П., Васильева Е.Л.. Развитие методов прогноза погоды на основе статистической интерпретации гидродинамических моделей по технологии Гидрометцентра России // In: 80 лет Гидрометцентру России. 2010. с. 313-335.
 32. Вильфанд Р.М., Мартазинова В.Ф., Цепелев В.Ю., Хан В.М., Мироничева Н.П., Елисеев Г.В., Иванова Е.К., Тищенко В.А., Уткузова Д.Н.. Комплексирование синоптико-статистических и гидродинамических прогнозов температуры воздуха на месяц // Метеорология и гидрология, т. 8, 2017. с. 5-17.
 33. Витушкин А.Г. К тринадцатой проблеме Гильберта // Доклады АН СССР, т. 96, № 4, 1954. с. 701-704.
 34. Галушкин А. И. Синтез многослойных систем распознавания образов. М: Энергия, 1974. 368 с.
 35. Гандин Л. С., Каган Р. Л. Статистические методы интерпретации метеорологических данных. Л: Гидрометеиздат, 1976. 359 с.
 36. Гандин Л.С. Объективный анализ метеорологических полей. Л: Гидрометеиздат, 1963.
 37. Гасников А.В. Современные численные методы оптимизации. Метод универсального градиентного спуска. 2nd ed. М.: МФТИ, 2018. 291 с.
 38. Гельфанд И.М., Виленкин Н.Я. Некоторые применения гармонического анализа. Оснащенные гильбертовы пространства. М: Физматгиз, 1961. 472 с.

39. Георгиеский Ю.М. Краткосрочные гидрологические прогнозы. Учебное пособие. Л.: ЛГМИ, 1982. 100 с.
40. Горбань А.Н. Обобщенная аппроксимационная теорема и точное представление многочленов нескольких переменных суперпозициями многочленов от одного переменного // Известия высших учебных заведений. Математика, т. 5, № 432, 1998. с. 6-9.
41. Гордин В.А. "Кулоновский" алгоритм выбора влияющих станций // Метеорология и гидрология, № 12, 2003. с. 100-105.
42. Гордин В.А., Багров А.Н., Быков Ф.Л. Международная конференция и школа молодых ученых по измерениям, моделированию и информационным системам для изучения окружающей среды: ENVIROMIS-2018 // Оперативная схема краткосрочного прогноза погоды и ей приложения. Томск. 2018. с. 93-94.
43. Гордин В.А. Математика, компьютер, прогноз погоды и другие сценарии математической физики. Физматлит, 2010. 733 с.
44. Гордин В.А. Математические задачи гидродинамического прогноза погоды. Аналитические аспекты. Л.: Гидрометеиздат, 1987. 256 с.
45. Гордин В.А. Математические задачи гидродинамического прогноза погоды. Вычислительные аспекты. Л.: Гидрометиздат, 1987. 264 с.
46. Груза Г. В., Рейтенбах Р. Г. Статистика и анализ гидрометеорологических данных. Л.: Гидрометеиздат, 1982. 216 с.
47. Демьянов В.В., Савельева Е.А.. Гео статистика: теория и практика. Наука, 2010. 327 с.
48. Дзядык В.К. Введение в теорию равномерного приближения функций полиномами. М.: Наука, 1977. 512 с.
49. Здерева М.Я., Токарев В.М., Виноградова М.В.. Автоматизированный прогноз температуры воздуха с обучением по методу группового учета аргументов // Труды СибНИГМИ, № 106, 2011. с. 143-151.
50. Здерева М.Я., Торубарова Г.П., Шустова Г.А.. Физико-статистическая схема прогноза экстремальной температуры воздуха по станциям Новосибирской области на 1—5 суток // Труды СибНИГМИ, № 105, 2006. с. 40-46.
51. Ивахненко А.Г., Лапа В.Г.. Кибернетические предсказывающие устройства. Киев: Наукова думка, 1965. 216 с.
52. Как Яндекс прогнозирует погоду [Электронный ресурс] // Yandex: [сайт]. URL: <https://yandex.ru/company/technologies/meteum/> (дата обращения: 03.04.2021).

53. Колмогоров А.Н. Интерполирование и экстраполирование стационарных случайных последовательностей // Изв. АН СССР. Сер. математическая, т. 5, № 1, 1941.
54. Криге Д.Г. Роль математической статистики в методах уточненной оценки промышленного оруднения на рудниках Южной Африки. 1968. с. 252-271.
55. Лоэв М. Теория вероятностей. М: ИЛ, 1962. 720 с.
56. Марчук Г.И. Численные методы в прогнозе погоды. Л. 1967. 356 с.
57. Мещерская А.В., Руховец Л.В., Юдин М.И., Яковлева Н.И. Естественные составляющие метеорологических полей. Л.: Гидрометеиздат, 1970. 200 с.
58. Минский М., Пейперт С.. Персептроны. М: Мир, 1971. 264 с.
59. Монин А.С., Яглом А.М. Статистическая гидромеханика. Vol 1. М: Наука, 1965.
60. Наставление по краткосрочным прогнозам погоды общего назначения. Руководящий документ РД 52.27.724-2009. Обнинск: ИГ-СОЦИН, 2009. 50 с.
61. Наставление по краткосрочным прогнозам погоды общего назначения. Руководящий документ РД 52.27.724-2019. Москва: ФБГУ «Гидрометцентр России», 2019. 65 с.
62. Нестеров Ю.Е. Метод решения задачи выпуклого программирования со скоростью сходимости $O(1/k^2)$ // Доклады академии наук СССР, т. 269, № 3, 1983. с. 543-547.
63. Нестеров Ю.Е. Методы выпуклой оптимизации. М.: МЦНМО, 2010. 281 с.
64. Николенко С., Кадурич А., Архангельская Е.. Глубокое обучение. Погружение в мир нейронных сетей. СПб: Питер, 2018. 480 с.
65. Обухов А.М. О статистически ортогональных разложениях эмпирических функций // Изв. АН СССР. Сер. геофиз., № 3, 1960. с. 432-439.
66. Отчет ФБГУ "СибНИГМИ". Создание технологии глобальных прогнозов метеорологических полей на срок до 10 суток с шагом сетки не грубее 40км для детерминированных и 70 км для вероятностных прогнозов. Новосибирск. 2013.
67. Пановский Г. А., Брайер Г. В. Статистические методы в метеорологии. Л: Гидрометеиздат, 1972. 209 с.
68. Поляк Б.Т. Введение в оптимизацию. М.: Наука, 1983. 384 с.
69. Ривин Г.С., Вильфанд Р.М., Киктев Д.Б., Розинкина И.А., Тудрий К.О., Блинов Д.В., Варенцов М.И., Самсонов Т.Е., Бундель А.Ю., Кирсанов А.А., Захарченко Д.И.. Система численного прогнозирования явлений погоды, включая опасные, для московского мегаполиса: разработка прототипа // Метеорология и гидрология, № 11, 2019. с. 33-45.

70. Ривин Г.С., Розинкина И.А., Астахова Е.Д., Блинов Д.В., Бундель А.Ю., Кирсанов А.А., Шатунова М.В., Чубарова Н.Е., Алферов Д.Ю., Варенцов М.И., Захарченко Д.И., Копейкин В.В., Никитин М.А., Полюхов А.А., Ревокатова А.П., Татаринович Е.В., Чурюлин Е.В.. Система краткосрочного численного прогноза высокой детализации COSMO-Ru, ее развитие и приложения // Гидрометеорологические исследования и прогнозы, т. 4, № 374, 2019. с. 37-53.
71. Ривин Г.С., Розинкина И.А., Вильфанд Р.М., Алферов Д.Ю., Астахова Е.Д., Блинов Д.В., Бундель А.Ю., Казакова Е.В., Кирсанов А.А., Никитин М.А., Перов В.Л., Суркова Г.В., Ревокатова А.П., Шатунова М.В., Чумаков М.М.. Система COSMO-Ru негидростатического мезомасштабного краткосрочного прогноза погоды Гидрометцентра России: второй этап реализации и развития // Метеорология и гидрология, т. 6, 2015. с. 58-70.
72. Ривин Г.С., Розинкина И.А., Вильфанд Р.М., Киктев Д.Б., Тудрий К.О., Блинов Д.В., Варенцов М.И., Захарченко Д.И., Самсонов Т.Е., Репина И.А., Артамонов А.Ю.. Разработка оперативной системы численного прогноза погоды и условий возникновения опасных явлений с высокой детализацией для московского мегаполиса // Метеорология и гидрология, т. 7, 2020. с. 5-19.
73. Розенблатт Ф. Принципы нейродинамики: Персептроны и теория механизмов мозга. М: Мир, 1965. 480 с.
74. Романов А.В., Скрибцов П.В., Червоненкис М.А.. Решение обратных задач русловой гидравлики с использованием нелинейных математических моделей // Труды Гидрометцентра России, т. 349, 2013. с. 142-160.
75. Рябенский В.С. Введение в вычислительную математику. М.: Физматлит, 2000. 296 с.
76. СЕАКЦ. Северо-Евразийский климатический центр [Электронный ресурс] // Северо-Евразийский климатический центр: [сайт]. URL: <http://seakc.meteoinfo.ru/>
77. Ситников И.Г., Полякова И.В.. Практическое применение ансамблей гидродинамических прогнозов метеорологических полей // Метеорология и гидрология, т. 8, 1997. с. 113-118.
78. Страшная А.И., Береза О.В., Тарасова Л.Л., Максименкова Т.А., Шмльгин И.А., Пурина И.Э., Чекулаева Т.С. Современное состояние и проблемы агрометеорологического обеспечения сельского хозяйства России // Гидрометеорологические исследования и прогнозы, т. 4, № 374, 2019. с. 219-240.
79. Тихонов А.Н., Арсенин В.Я.. Методы решения некорректных задач. Наука, 1986. 286 с.

80. Тищенко В.А., Хан В.М., Круглова Е.Н., Куликова И.А.. Применение статистической коррекции детерминистских прогнозов температуры воздуха и осадков по модели ПЛАВ для арктического региона // Труды гидрометеорологического научно-исследовательского центра Российской Федерации, № 361, 2016. с. 47-65.
81. Толстых М.А., Фадеев Р.Ю., Шашкин В.В., Травова С.В., Ройман Г.С., Мизяк В.Г., Рогутов В.С., Шляева А.В., Юрова А.Ю.. Развитие глобальной полулагранжевой модели атмосферы ПЛАВ в 2009–2019 гг. // Гидрометеорологические исследования и прогнозы, т. 4, № 374, 2019. с. 77-91.
82. Толстых М.А. Глобальная полулагранжева модель численного прогноза погоды. М., Обнинск,: ОАО ФОР, 2010. 111 с.
83. Уланова Е.С. Методы агрометеорологических прогнозов. Л. 1959. 280 с.
84. Фахрутдинова Н.П. (отв. редактор). Код для оперативной передачи данных приземных метеорологических наблюдений с сети станций Росгидромета (KH-01 SYNOP). Росгидромет, 2012.
85. Хайкин С. Нейронные сети. М: Вильямс, 2006. 1104 с.
86. Хан В.М., Крыжов В.Н., Вильфанд Р.М., Тищенко В.А., Бундель А.Ю.. Мультимодельный подход при составлении прогнозов погоды на сезон, т. 1, 2011. с. 19-29.
87. Хинчин А.Я. Теория корреляции стационарных стохастических процессов // УМН, № 5, 1938. с. 42-51.
88. Цырульников М.Д., Толстых М.А., Багров А.Н., Зарипов Р.Б.. Развитие глобальной системы усвоения данных с переменным разрешением // Метеорология и гидрология, т. 4, 2003. с. 5-24.
89. Чеботарёв Н.Г. О проблеме резольвент // Учен. зап. Казан. гос. ун-та, т. 114, 1954. с. 189-193.
90. Шакина Н.П., Иванова А.Р.. Прогнозирование метеорологических условий для авиации. М.: Триада лтд, 2016. 312 с.
91. Шакина Н.П., Хоменко И.А., Иванова А.Р., Скрипунова Е.Н.. Образование и прогнозирование замерзающих осадков: обзор литературы и некоторые новые результаты // Труды Гиromетцентра России, № 348, 2012. с. 130-161.
92. Яглом А.М. Корреляционная теория стационарных случайных процессов. Л: Гидрометеиздат, 1981. 263 с.

93. Aldrich J.H., Nelson F.D.. Linear probability, logit, and probit models. London: Sage, 1984. 96 pp.
94. Alferov D., Astakhova E.. Experiments with stochastic perturbation of physical tendencies in COSMO-Ru2-EPS 2017. № 17.
95. Ament F., Simmer C. Improved representation of land-surface heterogeneity in a non-hydrostatic numerical weather prediction model // *Boundary-Layer Meteorol*, Vol. 121, 2006. pp. 153-174.
96. Anaconda. Announcing Anaconda Distribution 2019.10 [Электронный ресурс] // Anaconda: [сайт]. [2019]. URL: <https://www.anaconda.com/blog/anaconda-distribution-2019-10> (дата обращения: 05.03.2020).
97. Anderes E. B., Stein M. L.. Estimating deformations of isotropic gaussian random fields on the plane // *Annals of Statistics*, 2008. pp. 719-741.
98. Anonymous. ICLR // Don't use large mini-batches, use local SGD. 2020.
99. Apanasovich T. V., Genton M. G.. Cross-covariance functions for multivariate random fields based on latent dimensions // *Biometrika*, Vol. 97, № 1, 2010. pp. 15-30.
100. Arcomano T., Szunyogh I., Pathak J., Wikner A., Hunt B., Ott R. E.. A Machine Learning-Based Global Atmospheric Forecast Model // *Geophysical Research Letters*, Vol. 47, № 9, 2020.
101. Armstrong M. Problems with Universal kriging // *Mathematical geology*. 1984. Vol. 16. № 1. pp. 101-108.
102. Aronszajn N. Theory of reproducing kernels // *Trans. Amer. Math. Soc.*, Vol. 68, № 3, 1950. pp. 337-404.
103. Auroux D., Blum J.. A nudging-based data assimilation method: the Back and Forth Nudging (BFN) algorithm // *Nonlin. Processes Geophys.*, Vol. 15, 2008. pp. 305-319.
104. Ba J.L., Kiros J.R., Hinton G.E.. Layer Normalization // *arXiv*. 2016. URL: <https://arxiv.org/pdf/1607.06450.pdf> (дата обращения: 05.03.2021).
105. Balgovind R., Dalcher A., Ghil M., Kalnay E.. A stochastic-dynamic model for the spectral structure of forecast errors // *Mon. Weather Rev.*, Vol. 111, 1983. pp. 701-721.
106. Barker D.M., Huang W., Guo Y-R., Bourgeois A.J., Xioa Q.N.. A Three-Dimensional Variational Data Assimilation System for MM5: Implementation and Initial Results // *Monthly Weather Review*, Vol. 132, № 4, 2004. pp. 897-914.
107. Barret Z., Quoc L.V.. Neural architecture search with reinforcement learning // *arXiv*. 2016. URL: <https://arxiv.org/pdf/1611.01578.pdf> (дата обращения: 05.03.2021).
108. Bengio Y., Goodfellow I., Courville A.. *Deep Learning*. MIT Press, 2015. 800 pp.

109. Bengio Y., Lamblin P., Popovici D., Larochelle H.. Greedy layer-wise training of deep networks // Neural Information Processing Systems. 2006.
110. Bochner S. A theorem on Fourier-Stieltjes integrals // Bull. Amer. Math. Soc., Vol. 40, № 4, 1934. pp. 271-276.
111. Bocquet M., Brajard J., Carrassi A., Bertino L.. Data assimilation as a learning tool to infer ordinary differential equation representations of dynamical models // Nonlin. Processes Geophys., Vol. 26, 2019. pp. 143-162.
112. Bornn L., Shaddick G., Zidek J.V.. Modeling Nonstationary Processes Through Dimension Expansion // Journal of the American Statistical Association, Vol. 107, № 497, 2012. pp. 281-289.
113. Bourlard H., Kamp Y.. Auto-association by multilayer perceptrons and singular value decomposition // Biological Cybernetics, № 59, 1988. pp. 291-294.
114. Breiman L. Random forests // Machine Learning, Vol. 45, 2001. pp. 5-32.
115. Broadcasting [Электронный ресурс] // NumPy: [сайт]. [2008-2020]. URL: <https://numpy.org/doc/stable/user/basics.broadcasting.html> (дата обращения: 09.02.2020).
116. Bykov Ph. L. ICCARUS 2020 Book of Abstracts // Neural networks based statistical correction of COSMO-Ru model surface weather forecasts. Offenbach, Germany. 2020. pp. 47-48.
117. Bykov Ph.L., Gordin V.A.. Forecasting Moscow Ambulance Trips // Higher School of Economics Research Paper, Series: Science, Technology and Innovation. № WP BRP 36/STI/2015. 2015. URL: <http://www.hse.ru/data/2015/03/27/1096061578/36STI2015.pdf> (дата обращения: 08.02.2016).
118. Bykov Ph.L. Optimal interpolation of inhomogeneous fields using neural networks // Research activities in Earth system modelling. Working Group on Numerical Experimentation. 2021. pp. 1-03.
119. Carmeli C., De Vito E. , Toigo A., Umanita V.. Vector valued reproducing kernel Hilbert spaces and universality // Analysis and Applications, Vol. 8, 2011. pp. 377-408.
120. Carrassi A., Bocquet M., Bertino L., Evensen G.. Data assimilation in the geosciences: An overview of methods, issues, and perspectives // WIREs Clim Change, Vol. 9, № e535, 2018. pp. 1-50.
121. Chen R.T.Q., Rubanova Yu., Bettencourt J., Duvenaud D.K.. NIPS Proceedings // Neural ordinary differential equations. 2018.

122. cosmo-model.org. Consortium for Small-scale Modeling [Электронный ресурс] // Consortium for Small-scale Modeling: [сайт]. URL: <http://www.cosmo-model.org/>
123. Cotter A., Shamir O., Srebro N., Sridharan K.. Better mini-batch algorithms via accelerated gradient methods // *Advances in Neural Information Processing Systems*, Vol. 24, 2011. pp. 1647-1655.
124. Courtier P., Thepaut J.-N., Hollingsworth A.. A strategy for operational implementation of 4D-Var, using an incremental approach // *Q. J. R. Meteorol. Soc.*, Vol. 120, 1994. pp. 1367-1387.
125. Cressie N., Huang H.-C.. Classes of nonseparable, spatio-temporal stationary covariance functions // *J. Am. Statist. Assoc.*, Vol. 94, 1999. pp. 1330-1340.
126. Cressie N. *Statistics for spatial data*. New York: John Wiley & Sons, 1991. 900 pp.
127. Cubuk E. D., Zoph B., Mane D., Vasudevan V., Le Q. V.. Autoaugment: Learning augmentation policies from data // *arXiv*. 2018. URL: <https://arxiv.org/pdf/1805.09501.pdf> (дата обращения: 03.05.2021).
128. Cybenko G.V. Approximation by Superpositions of a Sigmoidal function // *Mathematics of Control Signals and Systems*, Vol. 2, № 4, 1989. pp. 303-314.
129. De Vito E., Umanita V., Villa S.. An extension of Mercer theorem to matrix-valued measurable kernels // *Applied and Computational Harmonic Analysis*, Vol. 34, 2013. pp. 339-351.
130. Ding Z., Chen Y., Li N., Zhao D., Chen C.L.Ph.. Efficient Neural Architecture Search: A Broad Version // *arXiv*. 2020. URL: <https://arxiv.org/abs/2001.06679> (дата обращения: 05.03.2021).
131. Doolittle M. H. Association ratios // *Bull. Philos. Soc.*, Vol. 7, 1888. pp. 122-127.
132. Dozat T. Incorporating Nesterov Momentum into Adam // *ICLR Workshop*. 2016. Vol. 1:2013-2016.
133. Duchi J., Hazan E., Singer Y.. Adaptive subgradient methods for online learning and stochastic optimization // *Journal of Machine Learning Research*, Vol. 12, 2011. pp. 2121-2159.
134. Efron B. Bootstrap Methods: Another Look at the Jackknife // *Annals of Statistics*, Vol. 7, № 1, 1979. pp. 1-26.
135. Evensen G. *Data Assimilation: The Ensemble Kalman Filter*. 2nd ed. N.Y.: Springer, 2009. 272 pp.
136. Fasshauer G. Positive Definite Kernels: Past, Present and Future // *Dolomite Res. Notes Approx*, № 4, 2011.
137. Ferreira J.C., Menegatto V.A., Peron A.P.. Integral operators on the sphere generated by positive definite smooth kernels // *Journal of Complexity*, Vol. 24, 2008. pp. 632-647.

138. Fisher R.A. On the mathematical foundations of theoretical statistics, A // Philosophical Transactions of the Royal Society, № 222, 1922. pp. 309-368.
139. Gagne D.J., McGovern A., Xue M.. Machine Learning Enhancement of Storm-Scale Ensemble Probabilistic Quantitative Precipitation Forecasts // Weather and Forecasting, Vol. 29, № 4, 2014. pp. 1024-1043.
140. Gaspari G., Cohn S.E.. Construction of correlation functions in two and three dimensions // Quarterly Journal of the Royal Meteorological Society, Vol. 125, 1999. pp. 723-757.
141. Gauss C. F. Theoria motus corporum coelestium in sectionibus conicis solem ambientium. Vol VII. Hamburg: Friedrich Perthes and I.H. Besser, 1809, 1906.
142. Geer A.J. Learning earth system models from observations: machine learning or data assimilation? // Phil. Trans. R., Vol. 379, № 20200089, 2021.
143. Glorot X., Bengio Y.. Understanding the difficulty of training deep feedforward neural networks // In Proc. AISTATS, Vol. 9, 2010. pp. 249-256.
144. Gneiting T. Correlation functions for atmospheric data analysis // Q. J. R. Meteorol. Soc., Vol. 125, 1999. pp. 2449-2464.
145. Google. Google Colaboratory [Электронный ресурс] // Google Colaboratory: [сайт]. URL: <https://colab.research.google.com/> (дата обращения: 04.02.2021).
146. Goyal P., Dollár P., Girshick R., Noordhuis P., Wesolowski L., Kyrola A., Tulloch A., Jia Ya., He K.. Accurate, Large Minibatch SGD: Training ImageNet in 1 Hour // arXiv. 2017. URL: <https://arxiv.org/pdf/1706.02677.pdf> (дата обращения: 05.03.2021).
147. Graves A., Schmidhuber J.. Framewise phoneme classification with bidirectional LSTM and other neural network architectures // Neural Networks, Vol. 18, № 5-6, 2005. pp. 602-610.
148. Graves A. Generating sequences with recurrent neural networks // arXiv. 2013. URL: <https://arxiv.org/pdf/1308.0850.pdf> (дата обращения: 05.03.2021).
149. Greydanus S., Dzamba M., Yosinski J.. Neural Information Processing Systems // Hamiltonian neural networks. 2019. pp. 15379-15389.
150. Gronquist P., Yao C., Ben-Nun T., Dryden N., Dueben P., Li S., Hoefler T.. Deep Learning for Post-Processing Ensemble Weather Forecasts // arXiv. 2020. URL: <https://arxiv.org/pdf/2005.08748v1.pdf> (дата обращения: 05.03.2021).
151. Guen, V. L., Yin, Y., Dona, J., Ayed, I., de Bézenac, E., Thome, N., Gallinari, P.. Augmenting physical models with deep networks for complex dynamics forecasting 2020. URL: <https://arxiv.org/pdf/1911.05180> (дата обращения: 05.03.2021).

152. Gulrajani I., Ahmed F., Arjovsky M., Dumoulin V., Courville A. C.. Advances in neural information processing systems // Improved training of wasserstein gans. 2017. pp. 5767-5777.
153. He K., Zhang X., Ren S., Sun J.. Deep Residual Learning for Image Recognition // arXiv. 2015. URL: <https://arxiv.org/abs/1512.03385>
154. He K., Zhang X., Ren S., Sun J.. Delving deep into rectifiers: Surpassing human-level performance on ImageNet classification // arXiv. 2015. URL: <https://arxiv.org/pdf/1502.01852.pdf> (дата обращения: 05.03.2021).
155. Heidke P. Berechnung Des Erfolges Und Der Güte Der Windstärkevorhersagen Im Sturmwarnungsdienst // Geografiska Annaler , Vol. 8, 1926. pp. 301-349.
156. Hinton G. E., Salakhutdinov R.. Reducing the dimensionality of data with neural networks // Science, № 313(5786), 2006. pp. 504-507.
157. Hochreiter S., Schmidhuber J.. Long short-term memory // Neural computation, Vol. 9, 1997. pp. 1735-1780.
158. Hogan R.J., Ferro C.A.T., Jolliffe I.T., Stephenson D.B.. Equitability Revisited: Why the “Equitable Threat Score” Is Not Equitable // Wea. and Forecasting, Vol. 25, 2010. pp. 710-726.
159. Houtekamer P.L., Matchell H.L.. Data Assimilation Using an Ensemble Kalman Filter Technique // Monthly Weather Review, Vol. 126, № 3, 1996. pp. 796-811.
160. Huang G., Li Y., Pleiss G., Liu Z., Hopcroft J.E. , Weinberger K.Q.. ICLR // Snapshot Ensembles: Train 1, get M for free. 2017.
161. Huang L., Zhou Y., Zhu F., Liu L., Shao L.. CVPR // Iterative Normalization: Beyond Standardization towards Efficient Whitening. 2019. pp. 4875-4883.
162. Huber P.J. Robust Estimation of a Location Parameter // Ann. Math. Statist., Vol. 35, № 1, 1964. pp. 73-101.
163. Ioffe S., Szegedy C.. Batch Normalization: Accelerating Deep Network Training by Reducing Internal Covariate Shift // arXiv. 2015. URL: <https://arxiv.org/pdf/1502.03167> (дата обращения: 05.03.2021).
164. Jakobsson M. Пурыгrib [Электронный ресурс] // pypi.org: [сайт]. URL: <https://pypi.org/project/пурегриб/> (дата обращения: 08.04.2020).
165. Jiang S.W., Harlim J. Modeling of missing dynamical systems: deriving parametric models using a nonparametric framework // Res Math Sci, Vol. 7, № 16, 2020.
166. Journel A.G. Nonparametric estimation of spatial distributions // Mathematical Geology, Vol. 15, 1983. pp. 445-468.

167. Kalnay E. Atmospheric modeling, data assimilation and predictability. Cambridge university press, 2003. 367 pp.
168. Kanevski M., Demyanov V., Chernov S., Savelieva E., Timonin V.. Neural Network Residual Kriging Application for Climatic Data // The J. of Geographic Information and Decision Analysis, Vol. 2, № 2, 1998. pp. 215-232.
169. Kanevsky M., Arutyunyan R. , Bolshov L., Demyanov V., Maignan M.. Artificial neural networks and spatial estimations of Chernobyl fallout // Geoinformatics, Vol. 7, № 1-2, 1995. pp. 5-11.
170. Kashinath K., Mustafa M., Albert A., Wu J-L., Jiang C., Esmaeilzadeh S., Azizzadenesheli K., Wang R., Chattopadhyay A., Singh A., Manepalli A., Chirila D., Yu R., Walters R., White B., Xiao H., Tchelepi H.A., Marcus P., Anandkumar A., Hassanzadeh P., Prab. Physics-informed machine learning: case studies for weather and climate modelling // Phil. Trans. R. Soc. A, Vol. 379, № 20200093, 2021.
171. Kingma D., Ba J.. Adam: A method for stochastic optimization // arXiv. 2014. URL: <https://arxiv.org/pdf/1412.6980.pdf> (дата обращения: 05.03.2021).
172. Konig H. Eigenvalue Distribution of Compact Operators. Springer Basel AG, 1986. 262 pp.
173. Koza J.R. Genetic Programming – On the Programming of Computers by Means of Natural Selection. MIT Press, 1992. 819 pp.
174. Koza J.R. Genetic Programming II: Automatic Discovery of Reusable Programs. Cambridge, Massachusetts: MIT Pres, 1994. 320 pp.
175. Krasnopolsky V.M., Fox-Rabinovitz M.S., Belochitski A.. Decadal climate simulations using accurate and fast neural network emulation of full, long- and short wave, radiation // Mon Weather Rev, № 136, 2008. pp. 3683-3695.
176. Krasnopolsky V.M., Fox-Rabinovitz M.S., Hou Y.T., Lord S.J., Belochitski A.. Accurate and fast neural network emulations of model radiation for the NCEP coupled climate forecast system: climate simulations and seasonal predictions // Mon Weather Rev, № 138, 2010. pp. 1822-1842.
177. Krasnopolsky V.M., Fox-Rabinovitz M.S.. A new synergetic paradigm in environmental numerical modeling: hybrid models combining deterministic and machine learning components // Ecological Modelling, Vol. 191, № 1, 2006. pp. 5-18.

178. Krasnopolsky V.M., Lin Y.. A neural network nonlinear multimodel ensemble to improve precipitation forecasts over continental US // *Advances in Meteorology*, № Article ID 649450, 2012. P. 11.
179. Krasnopolsky V.M., Schiller H.. Some neural network applications in environmental sciences part I: Forward and inverse problems in satellite remote sensing // *Neural Netw*, Vol. 16, 2003. pp. 321-334.
180. Krasnopolsky V.M. *The Application of Neural Networks in the Earth System Sciences: Neural Network Emulations for Complex Multidimensional Mappings*. Springer, 2013. 206 pp.
181. Krizhevsky A., Sutskever I., Hinton G. E.. ImageNet classification with deep convolutional neural networks // *Neural Information Processing Systems*. 2012. pp. 1106-1114.
182. Lee K., Parish E. J.. Parameterized Neural Ordinary Differential Equations: Applications to Computational Physics Problems // *arXiv*. 2020. URL: <https://arxiv.org/pdf/2010.14685> (дата обращения: 05.03.2021).
183. Li H., Calder C. A., Cressie N.. Beyond Moran's I: testing for spatial dependence based on the spatial autoregressive model // *Geographical Analysis*, Vol. 39, № 4, 2007. pp. 357-375.
184. LightGBM. LightGBM [Электронный ресурс] // LightGBM: [сайт]. URL: <https://lightgbm.readthedocs.io/en/latest/> (дата обращения: 15.09.2019).
185. Lin H.-T., Li L.. Large-margin thresholded ensembles for ordinal regression: Theory and practice // In: *Algorithmic Learning Theory*. Springer, 2006.
186. Liu L., Jiang H., He P., Chen W., Liu X., Gao J., Han J.. ICLR // On the variance of the adaptive learning rate and beyond. 2020.
187. Loshchilov I., Hutter F.. ICLR // Decoupled weight decay regularization. 2019.
188. Lussana C., Seierstad I.A., Nipen T.N., Cantarello L.. Spatial interpolation of two-metre temperature over Norway // *Q J R Meteorol Soc*, Vol. 145, 2019. pp. 3626-3643.
189. Mariotti A., Baggett C., Barnes E. A., Becker E., Butler A., Collins D. C., Albers J.. Windows of opportunity for skillful forecasts subseasonal to seasonal and beyond // *Bulletin of the American Meteorological Society*, Vol. 101, 2020. pp. E608-625.
190. Matheron G. *Principles of Geostatistics* // *Economic Geology*, Vol. 58, 1963. pp. 1246-1266.
191. Mayer K.J., Barnes E.A.. Subseasonal Forecasts of Opportunity Identified by an // *Geophysical Research Letters*, Vol. 48, № e2020GL092092, 2021.
192. Mayoraz L., Ambühl J.. Automatic Gale Warning Proposals for Swiss Lakes and Regional Aerodromes, *Scientific Report MeteoSwiss*, 102, 2016. 70 pp.

193. McCullagh P. Regression models for ordinal data // Journal of the Royal Statistical Society. Series B (Methodological), Vol. 42, № 2, 1980. pp. 109-142.
194. McGovern A., Elmore K.L., Gagne D.J., Haupt S.E., Karstens C.D., Lagerquist R., Shith T., Williams J.K.. Using Artificial Intelligence to Improve Real-Time Decision-Making for High-Impact Weather // Bulletin of the American Meteorological Society, Vol. 98, № 10, 2017. pp. 2073-2090.
195. Meinshausen N. Quantile regression forests // Journal of Machine Learning Research, № 7, 2006. pp. 983-999.
196. Mercer J. Functions of positive and negative type and their connection with the theory of integral equations // Philosophical Transactions of the Royal Society A, Vol. 89, № 559, 1909. pp. 415-446.
197. Mironov D.V. Technical Report. Parameterization of Lakes in Numerical Weather Prediction. Description of a Lake Model. Vol 11. 2008.
198. Moran P.A.P. Notes on Continuous Stochastic Phenomena // Biometrika, Vol. 37, № 1, 1950. pp. 17-23.
199. Nielsen M. A. Neural networks and deep learning. San Francisco: Determination press, 2015, 2018. 224 pp.
200. Oriol V, Toshev A, Bengio S, Erhan D. Show and tell: A neural image caption generator 2015.
201. pangeo-data. WeatherBench [Электронный ресурс] // Github: [сайт]. URL: <https://github.com/pangeo-data/WeatherBench> (дата обращения: 18.01.2021).
202. Pedregosa F., Bach F., Gramfort A. On the Consistency of Ordinal Regression Methods // Journal of Machine Learning Research, Vol. 18, 2017. pp. 1-35.
203. Pham H., Guan M.Y., Zoph B., Le Q.V., Dean J. Efficient Neural Architecture Search via Parameter Sharing // arXiv. 2018. URL: <https://arxiv.org/abs/1802.03268> (дата обращения: 05.03.2021).
204. Python. Python [Электронный ресурс] URL: <https://www.python.org/> (дата обращения: 18.10.2018).
205. PyTorch. PyTorch [Электронный ресурс] // PyTorch: [сайт]. URL: <https://pytorch.org/> (дата обращения: 20.12.2019).
206. Quinlan J.R. Induction of Decision Trees // Mach Learn, Vol. 1, 1986. pp. 81-106.

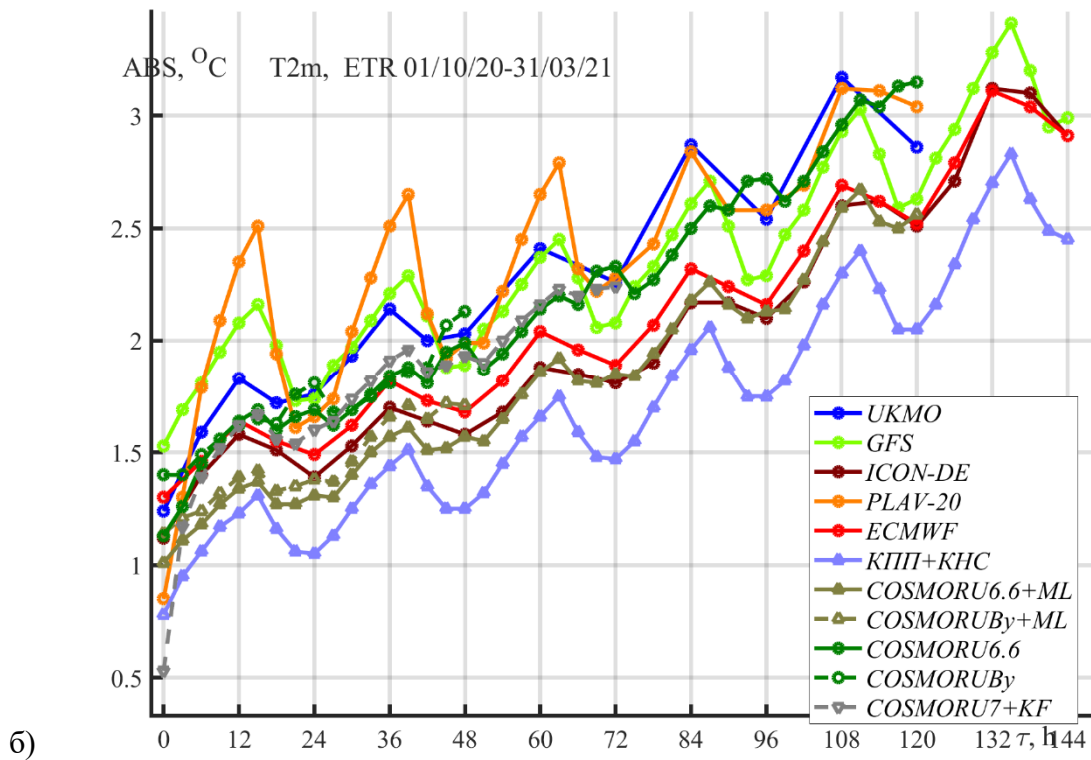
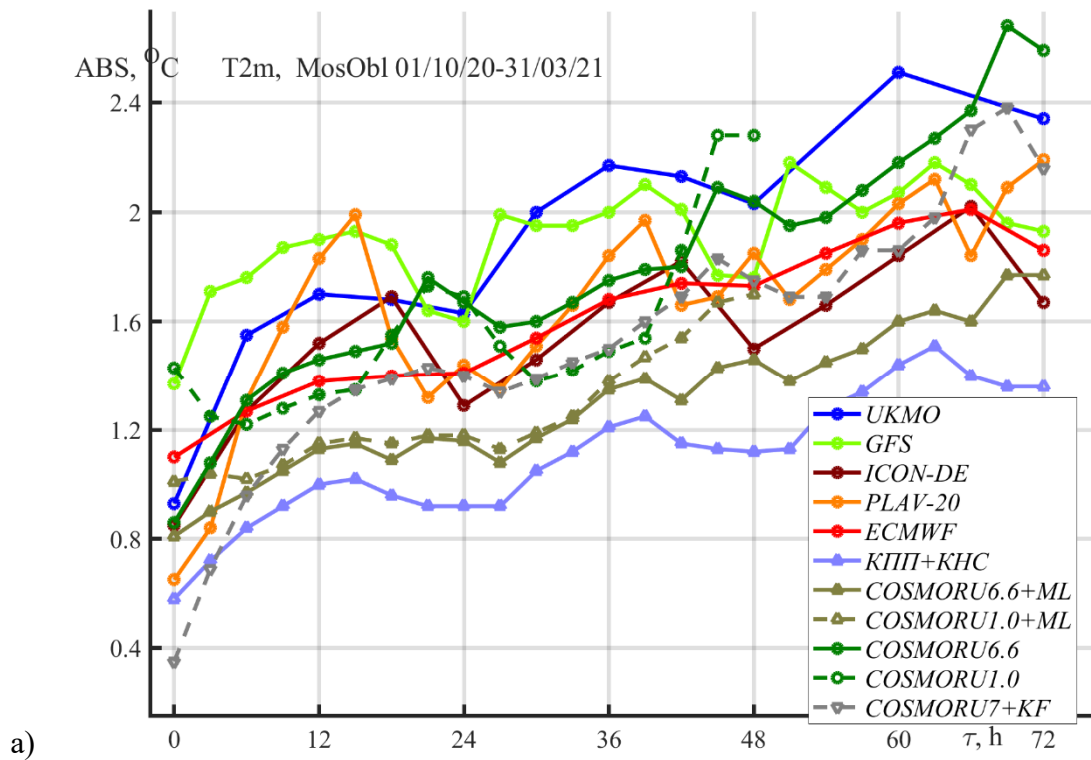
207. Raissi M., Perdikaris P., Karniadakis G.E.. Physics-informed neural networks: A deep learning framework for solving forward and inverse problems involving nonlinear partial differential equations // *Journal of Computational Physics*, Vol. 378, 2019. pp. 686-707.
208. Ranzato M., Boureau Y., LeCun Y. Sparse feature learning for deep belief networks // *Neural Information Processing Systems*. 2007.
209. Rasp S., Lerch S.. Neural networks for postprocessing ensemble weather forecasts // *Monthly Weather Review*, Vol. 146, № 11, 2018. pp. 3885-3900.
210. Rasp S., Thuerey N.. Data-driven medium-range weather prediction with a Resnet pretrained on climate simulations: A new model for WeatherBench // *arXiv*. 2020. URL: <https://arxiv.org/pdf/2008.08626.pdf> (дата обращения: 05.03.2021).
211. Rhodin A., Lange H., Potthast R., Janjic-Pfander T.. Documentation of the DWD Data Assimilation System. 2016. 446 pp.
212. Rigol J.P., Jarvis C.H., Stuart N.. Artificial neural networks as a tool for spatial interpolation // *International Journal of Geographical Information Science*, Vol. 15, № 4, 2001. pp. 323-343.
213. Roberts B., Clayson C.A., Robertson F.R., Jackson D.. Predicting near-surface atmospheric variables from SSM/I using neural networks with a first guess approach // *Geophys Res*, Vol. 115, 2010.
214. Ronneberger O., Fischer Ph., Brox T.. U-Net: Convolutional Networks for Biomedical Image Segmentation // *MICCAI: Lecture Notes in Computer Science*. 2015. Vol. 9351.
215. Rosenblatt F. The perseptron: a probabilistic model for information storage and organization in the brain // *Psychol. Rev.*, Vol. 65, 1958. P. 386.
216. Rumelhart D.E., Hinton G.E., Williams R.J.. Learning Internal Representations by Error Propagation. In: *Parallel Distributed Processing*. Vol 1. Cambridge: MIT Press, 1986. pp. 318-362.
217. Sagar D.B.S., Cheng Q., Agterberg F.. *Handbook of Mathematical Geosciences. Fifty Years of IAMG*. 2018. 918 pp.
218. Sampson P., Guttorp P.. Nonparametric estimation of nonstationary spatial covariance structure // *J. Amer. Statist. Assoc.*, Vol. 87, 1992. pp. 108–119.
219. Schoenberg I.J. Metric spaces and completely monotone functions // *Ann. Math.*, Vol. 39, 1938. pp. 811-841.
220. Smith L. WACV // *Cyclical Learning Rates for Training Neural Networks*. 2017. pp. 464-472.

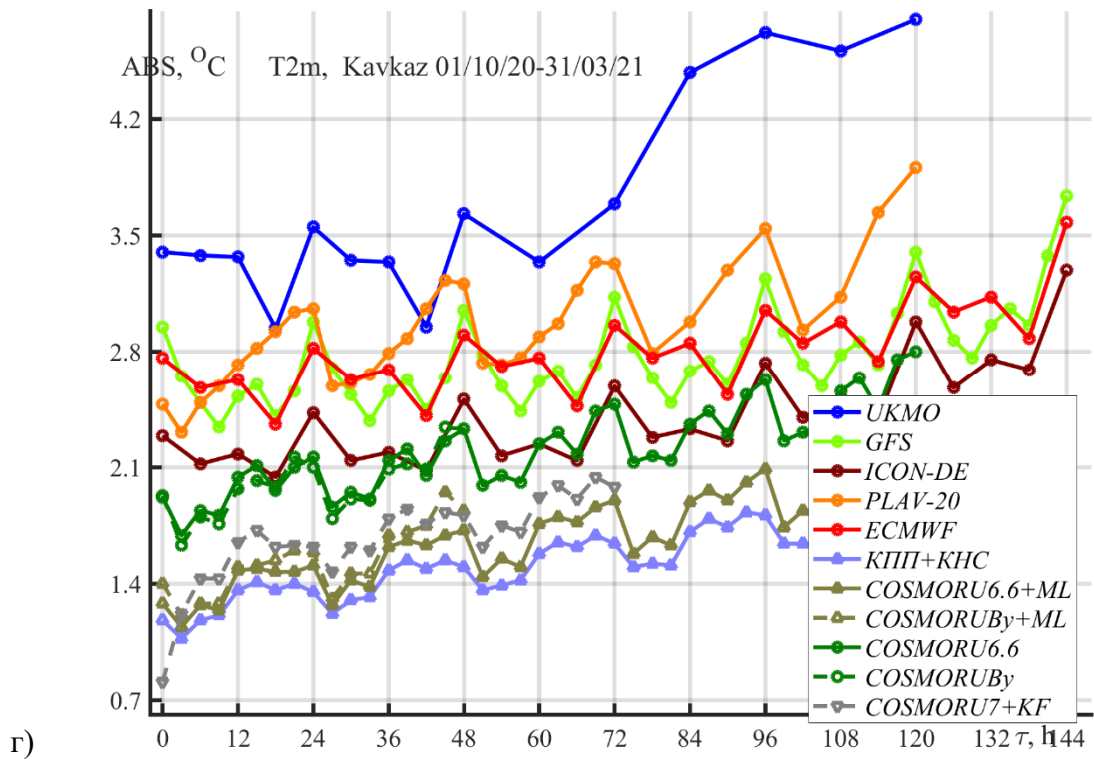
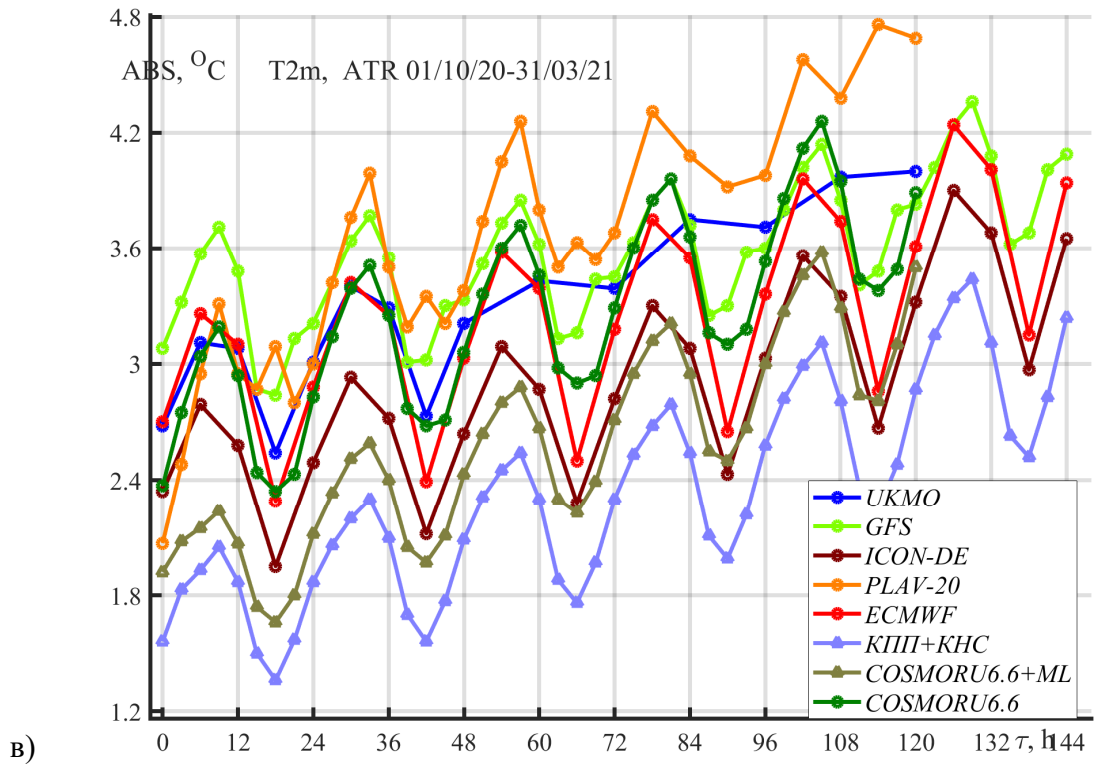
221. Smith S.L., Kindermans P.-J., Ying C., Le Q.V.. ICLR // Don't Decay the Learning Rate, Increase the Batch Size. 2018.
222. Steinwart I., Scovel C. Mercer's Theorem on General Domains: On the Interaction between Measures, Kernels, and RKHSs // Constr Approx, Vol. 35, 2012. pp. 363–417.
223. Stone M.N. The generalized Weierstrass approximation theorem // Math. Mag., Vol. 21, 1948. pp. 167-183.
224. Sutskever I., Martens J., Dahl G., Hinton G.. International conference on machine learning // On the importance of initialization and momentum in deep learning. 2013. pp. 1139-1147.
225. Taillardat M., Mestre O.. From research to applications – examples of operational ensemble // Nonlin. Processes Geophys., Vol. 27, № 2, 2020. pp. 329-347.
226. Tan K., Chen J., Wang D.. IEEE/ACM Transactions on Audio, Speech, and Language Processing // Gated Residual Networks With Dilated Convolutions for Monaural Speech Enhancement. 2019. Vol. 27(1). pp. 189-198.
227. Tan P.-N., Steinbach M.M., Kumar V.. Introduction to Data Mining. 2012. 165 pp.
228. TensorFlow. TensorFlow [Электронный ресурс] // TensorFlow: [сайт]. URL: <https://www.tensorflow.org/> (дата обращения: 18.07.2019).
229. Tieleman T., Hinton G.. Lecture 6.5 - RMSProp, COURSERA: Neural Networks for Machine Learning, 2012.
230. Tsymbalov E., Makarychev S., Shapeev A., Panov M.. IJCAI // Deeper Connections between Neural Networks and Gaussian Processes Speed-up. 2019.
231. Tsyrlunikov M. Is the Local Ensemble Transform Kalman Filter suitable for operational data assimilation? // COSMO News Letter. 2010. № 10.
232. van den Hurk B.J.J.M., Viterbo P., Beljaars A.C.M., Betts A.K. Offline validation of the ERA40 surface scheme // ECMWF. 2000. URL: <https://www.ecmwf.int/node/12900> (дата обращения: 05.03.2021).
233. Van der Vorst H.A. Iterative Krylov Methods for Large Linear System. Cambridge University Press, 2003. 221 pp.
234. van Dyk D.A., Meng X.-L.. The Art of Data Augmentation // Journal of Computational and Graphical Statistics, Vol. 10, 2001. pp. 1-50.
235. Vannitsem S., Bremnes J.B., Demaeyer J., Evans G.R., Flowerdew J., Hemri S., Lerch S., Roberts N., Theis S., Atencia A., Bouallègue Z.B., Bhend J., Dabernig M., De Cruz L., Hieta L., Mestre O., Moret L., Plenkovic I. O., Schmeits M., Taill M.. Statistical Postprocessing for

- Weather Forecasts – Review, Challenges and Avenues in a Big Data World // Bulletin of the American Meteorological Society, Vol. 11, 2020. pp. 1-44.
236. Varentsov M., Samsonov T., Demuzere M.. Impact of Urban Canopy Parameters on a Megacity's Modelled Thermal Environment // Atmosphere, Vol. 11, № 12, 2020. P. 1349.
 237. Vaswani A., Shazeer N., Parmar N., Uszkoreit J., Jones L., Gomez A.N., Kaiser L., Polosukhin I.. Advances in Neural Information Processing Systems // Attention Is All You Need. 2017. pp. 6000-6010.
 238. Wang Y., Wang X., Jian J.. Remote Sensing Landslide Recognition Based on Convolutional Neural Network // Mathematical Problems in Engineering, № Article ID 8389368, 2019.
 239. Warner T. T. Numerical weather and climate prediction. Cambridge University Press, 2010. 526 pp.
 240. Watt-Meyer O., Brenowitz N.D., Clark S.K., Henn B., McGribbon J.J., Perkins W.A., Bretherton C.S.. Correcting weather and climate models by machine learning nudged historical simulations // Geophysical Research Letters, 2021. P. 13.
 241. Weyn J.A., Durran D.R., Caruana R.. Improving data-driven global weather prediction using deep convolutional neural networks on a cubed sphere // arXiv. 2020. URL: <https://arxiv.org/pdf/2003.11927.pdf> (дата обращения: 05.03.2021).
 242. Whitaker J.S., Hamill T.M.. Ensemble data assimilation without perturbed observations // Mon. Wea. Rev., № 130, 2002. pp. 1913-1924.
 243. Wiener N. Extrapolation, interpolation and smoothing of stationary time series. N.Y.: MIT Press, 1949. 163 pp.
 244. Wilks D.S. Statistical Methods in the Atmospheric Sciences. Elsevier, 2006. 704 pp.
 245. Wu A., Hsieh W.W., Tang B.. Neural network forecasts of the tropical Pacific sea surface temperatures // Neural Networks, № 19, 2006. pp. 145-154.
 246. Xue H., Wu Z.-F., Sun W.-X.. IJCAI-19 // Deep Spectral Kernel Learning. 2019.
 247. Ye C., Yang Y., Fermüller C., Aloimonos Y.. On the Importance of Consistency in Training Deep Neural Networks // arXiv. 2017. URL: <https://arxiv.org/pdf/1708.00631.pdf> (дата обращения: 05.03.2021).
 248. Yong H., Huang J., Hua X., Zhang L.. Gradient centralization: A new optimization technique for deep neural networks // arXiv. 2020. URL: <https://arxiv.org/pdf/2004.01461.pdf> (дата обращения: 05.03.2021).

249. Zammit-Mangion A., Tin L. J. Ng, Vu Q., Filippone M.. Deep Compositional Spatial Models // arXiv. 2019. URL: <https://arxiv.org/pdf/1906.02840.pdf> (дата обращения: 05.03.2021).
250. Zhou B., Khosla A., Lapedriza A., Oliva A., Torralba A.. CVPR // Learning Deep Features for Discriminative Localization. 2016. pp. 2921-2929.
251. Zhou K., Zheng Y., Li B., Dong W., Zhang X.. Forecasting Different Types of Convective Weather: A Deep Learning Approach // Journal of Meteorological Research, Vol. 33, 2019. pp. 797-809.
252. Zhuang J., Tang T., Ding Y., Tatikonda S.C., Dvornek N., Papademetris X., Duncan J.. NeurIPS // AdaBelief Optimizer: Adapting Stepsizes by the Belief in Observed Gradients. 2020.
253. Zimmerman D.L. Another Look at Anisotropy in Geostatistics // Mathematical Geology, Vol. 25, № 4, 1993. pp. 453-470.
254. Zurada J.M. Introduction to artificial neural systems. New York: PWS, 1992. 785 pp.

Приложение А. Оценки прогнозов от начального срока 12:00 ВСВ





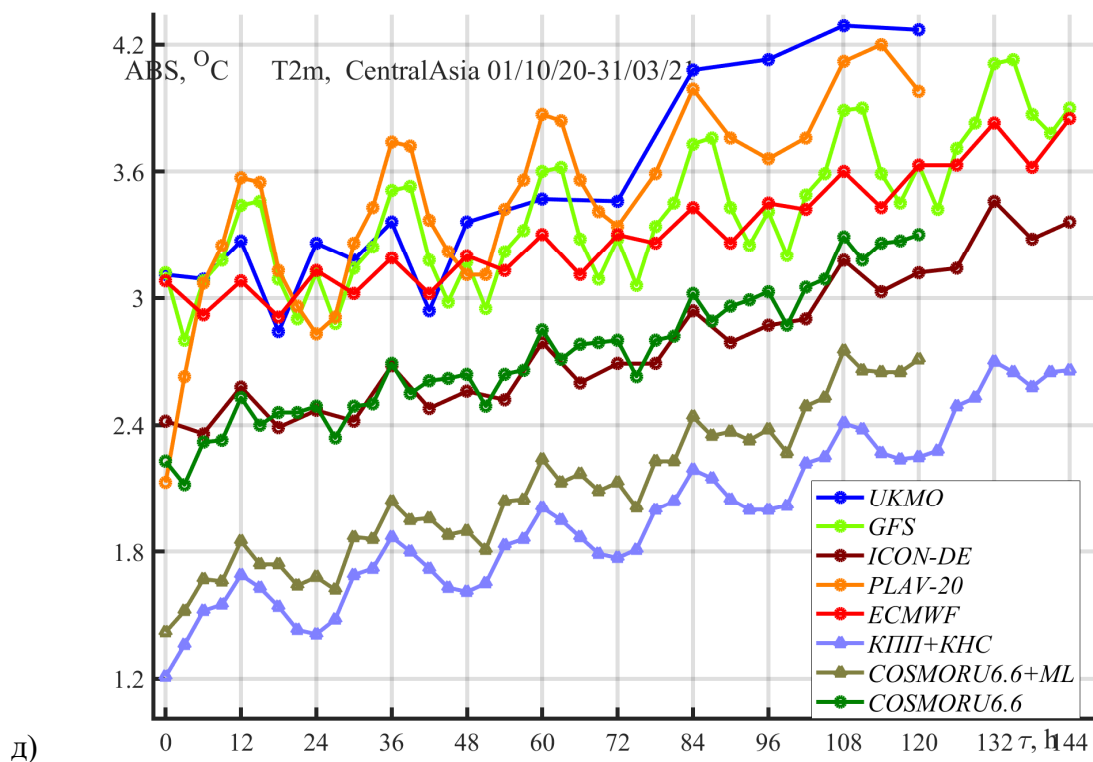
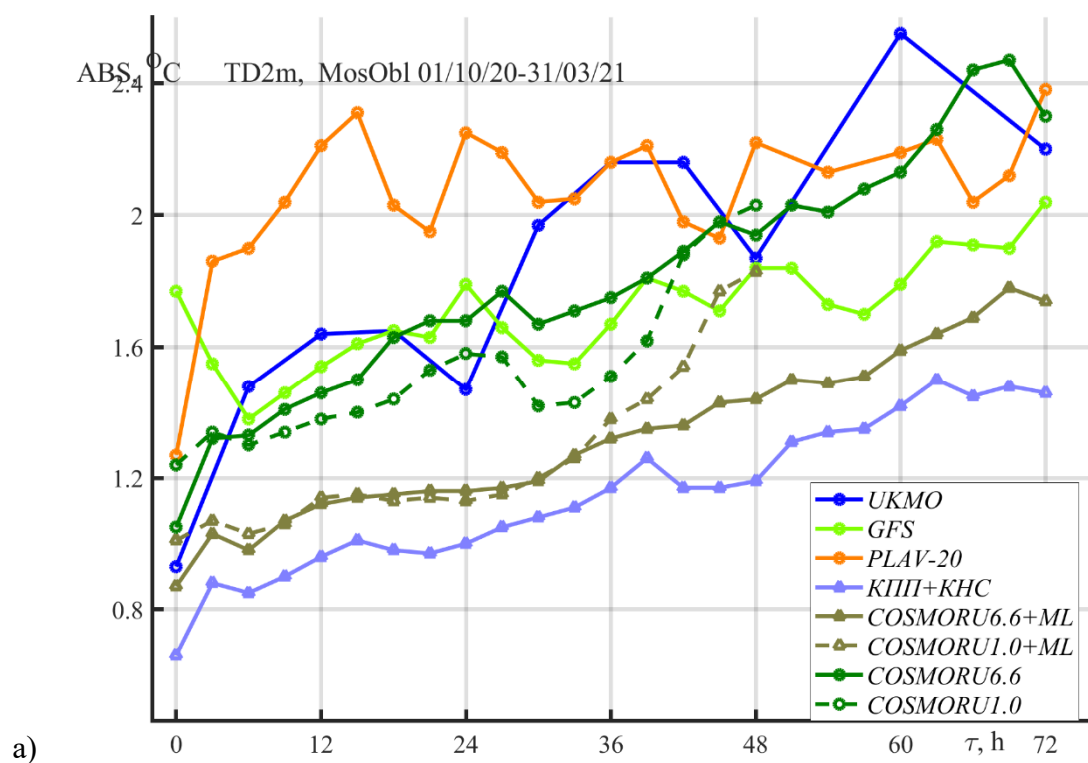
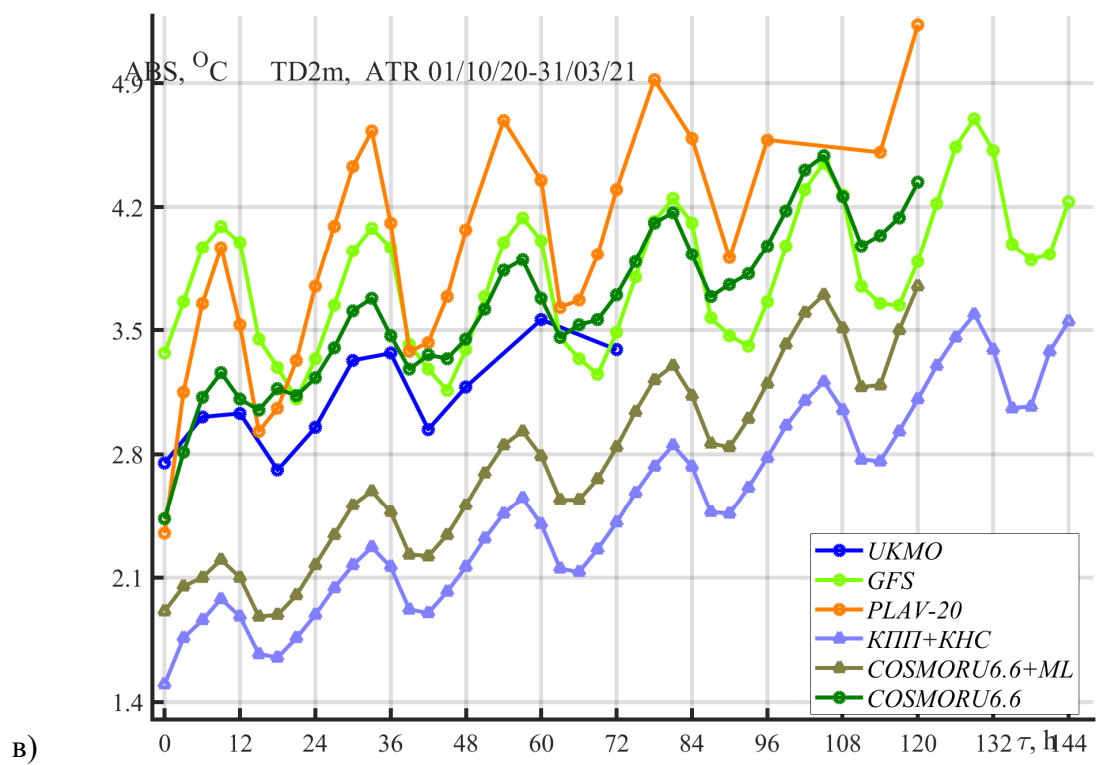
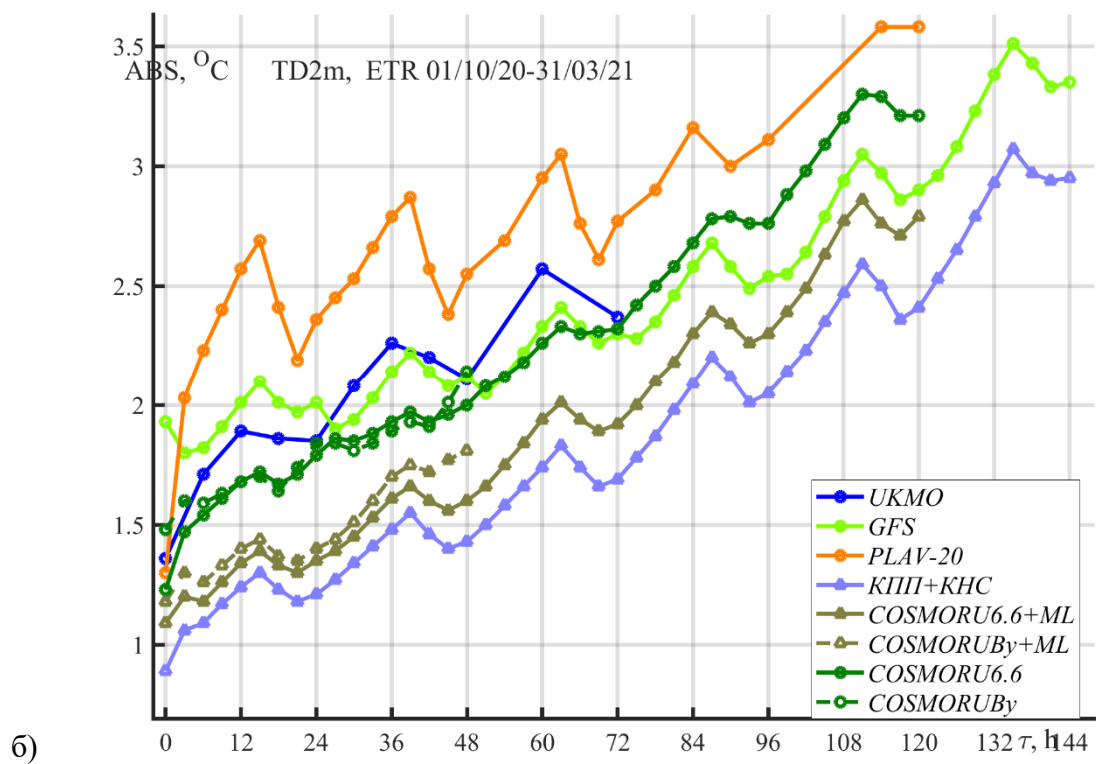


Рисунок А.1 – Средняя абсолютная ошибка (ось абсцисс) прогнозов температуры воздуха на высоте 2 м от начального срока 12:00 ВСУ по различным методам за период с 1 октября 2020 г. по 31 марта 2021 г. по различным регионам в зависимости от заблаговременности (ось ординат): а) Московский регион; б) ЕЧР; в) АЧР; г) Кавказ; д) Средняя Азия





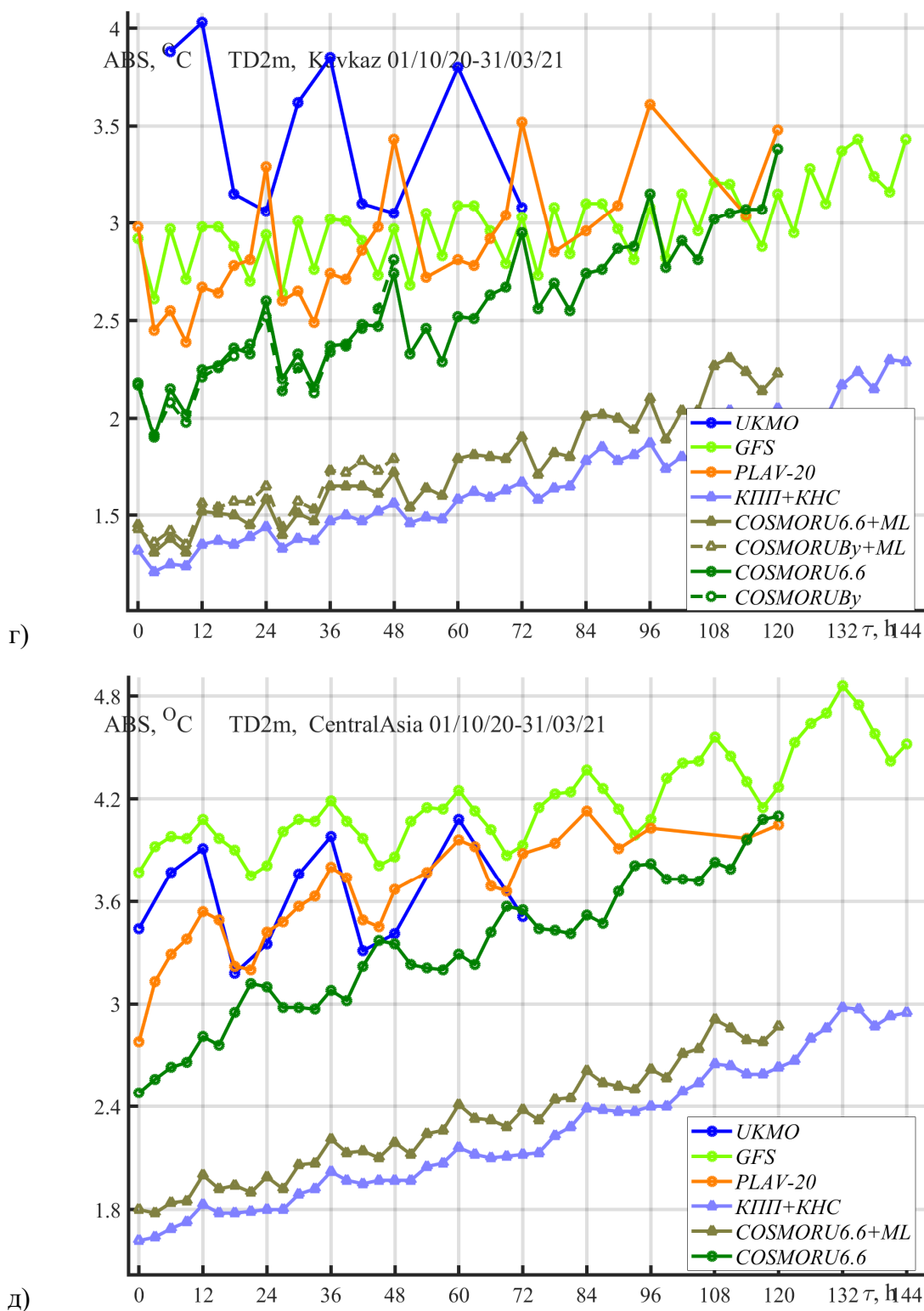
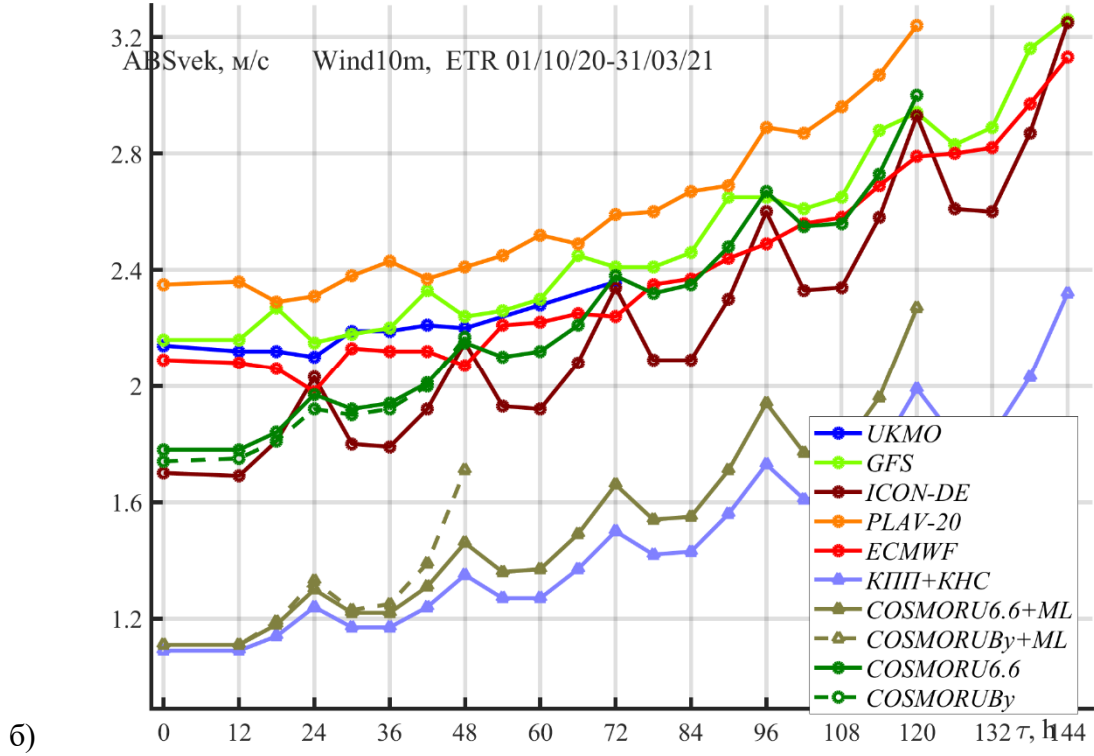
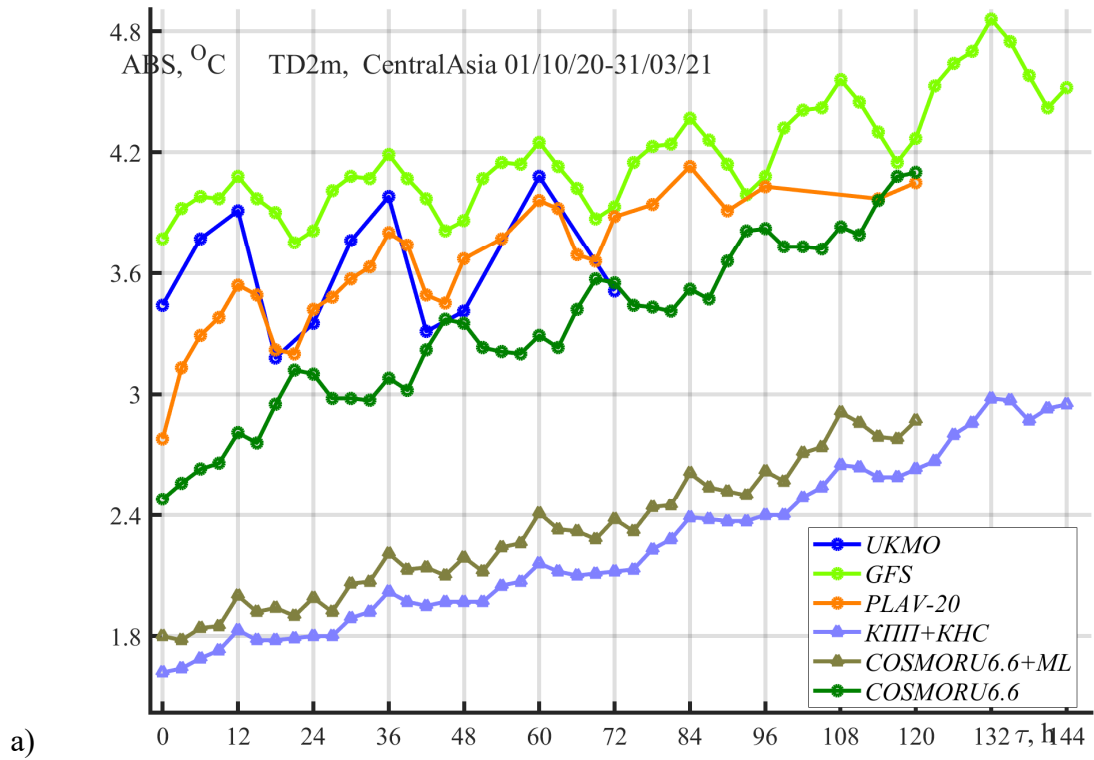
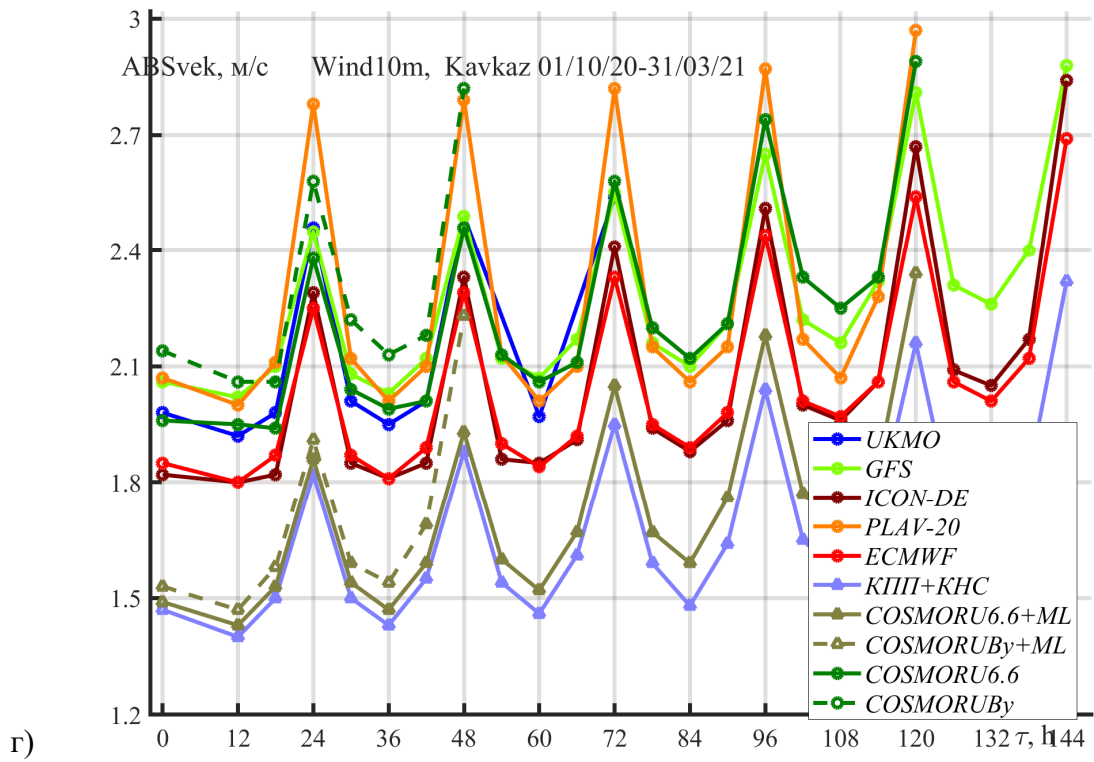
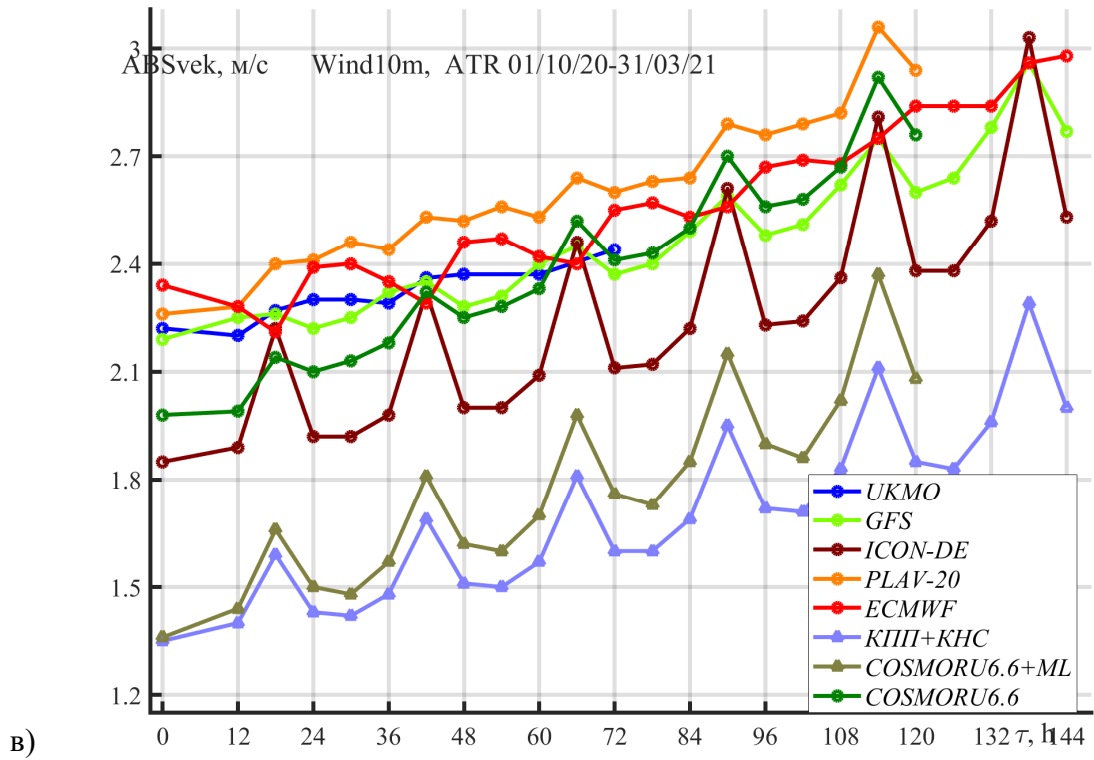


Рисунок А.2 – Средняя абсолютная ошибка (ось абсцисс) прогнозов точки росы на высоте 2 м от начального срока 12:00 ВСВ по различным методам за период с 1 октября 2020 г. по 31 марта 2021 г. по различным регионам в зависимости от заблаговременности (ось ординат): а) Московский регион; б) ЕЧР; в) АЧР; г) Кавказ; д) Средняя Азия





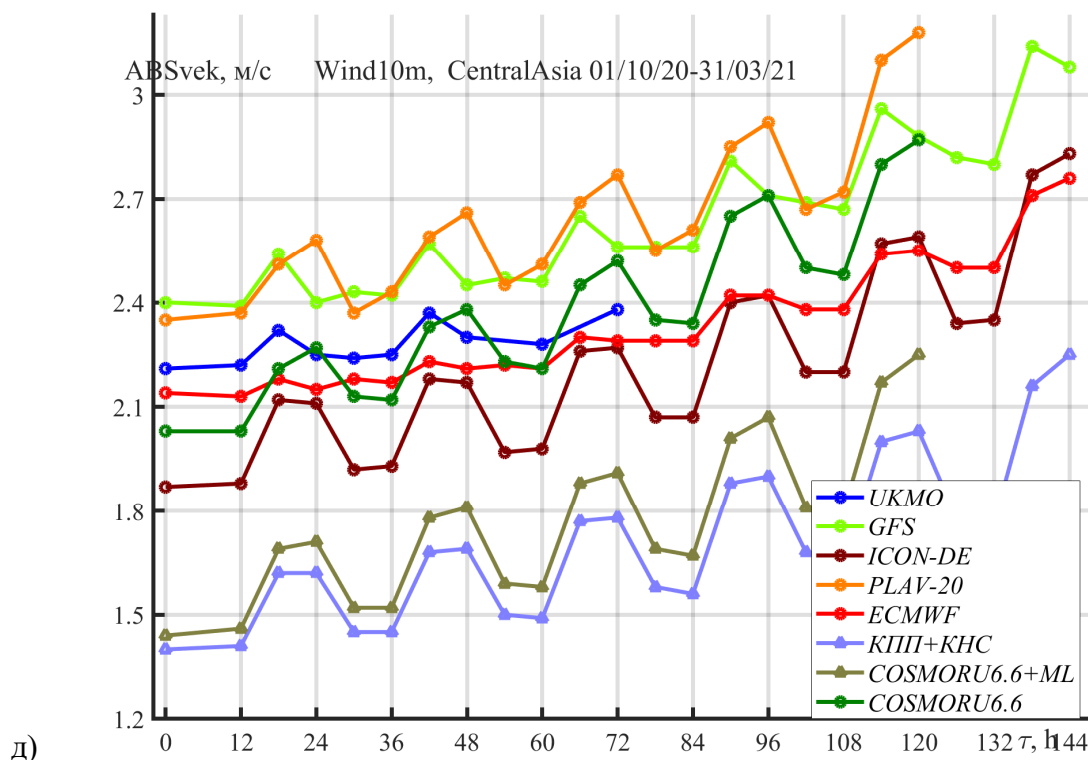


Рисунок А.4 – Средний модуль векторной ошибки (ось абсцисс) прогнозов скорости ветра на высоте 10 м от начального срока 12:00 ВСВ по различным методам за период с 1 октября 2020 г. по 31 марта 2021 г. по различным регионам в зависимости от заблаговременности (ось ординат): а) Московский регион; б) ЕЧР; в) АЧР; г) Кавказ; д) Средняя Азия

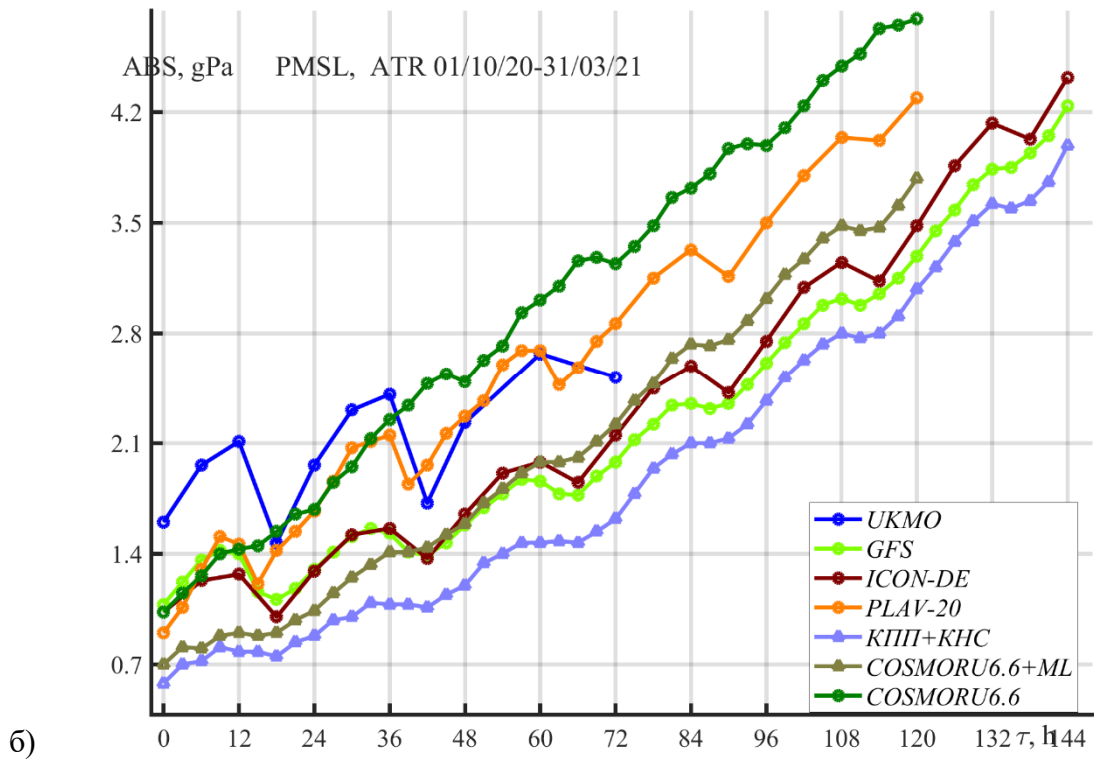
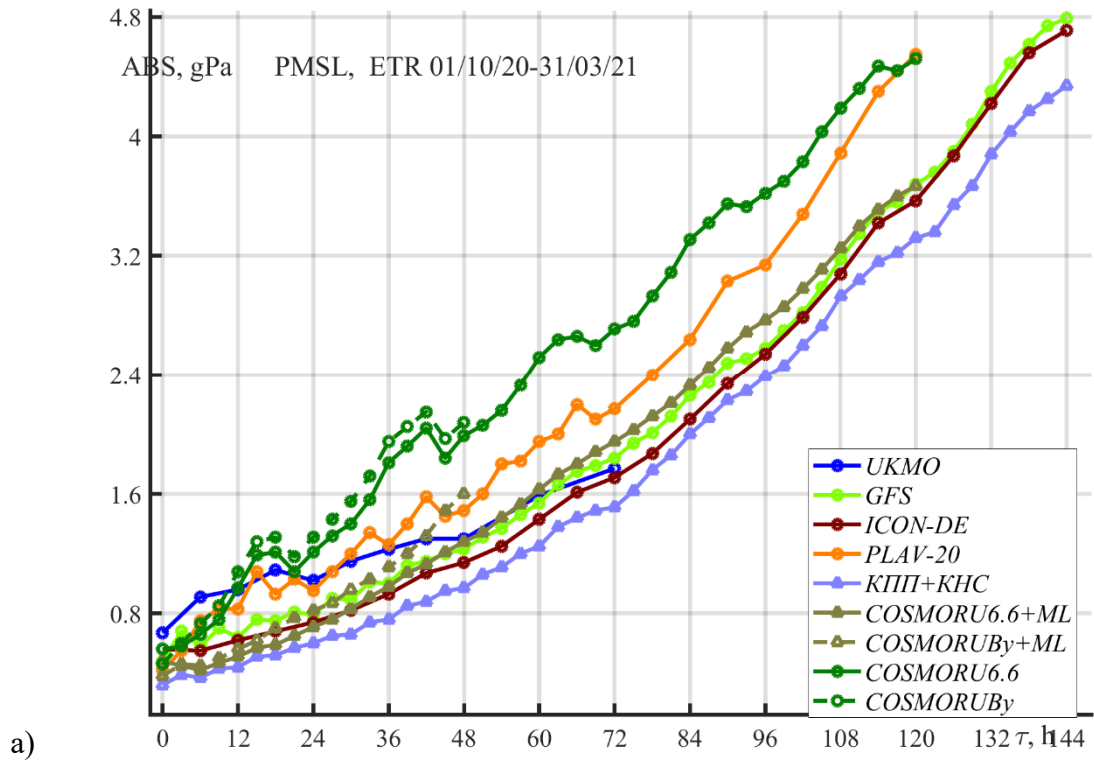
Таблица А.1 – Показатели успешности прогнозов порывов ветра более 12 и более 18 м/с с заблаговременностью 12ч от начального срока 12:00 ВСВ по различным методам за период с 1 октября 2020 г. по 31 марта 2021 г. по ЕЧР

Более 12 м/с			Более 18 м/с		
GFS	Pr=59%	HSS=0.46	GFS	Pr=17%	HSS=0.21
5229	6252	11481	76	200	276
3588	91136	94724	375	105554	105929
8817	97388	106205	451	105754	106205
ICON-De	Pr=68%	HSS=0.52	ICON-De	Pr=34%	HSS=0.34
6069	6354	12423	156	303	459
2803	91564	94367	302	106029	106331
8872	97918	106790	458	106332	106790
COSMO-Ru6.6ENA	Pr=66%	HSS=0.52	COSMO-Ru6.6ENA	Pr=34%	HSS=0.35
5519	5536	11055	153	261	414
2801	86930	89731	291	100081	100372
8320	92466	100786	444	100342	100786
COSMO-Ru6.6ENA + KCK+KHC	Pr=63%	HSS=0.55	COSMO-Ru6.6ENA + KCK+KHC	Pr=47%	HSS=0.43
5258	4190	9448	210	306	516
3062	88276	91338	234	100036	100270
8320	92466	100786	444	100342	100786
COSMO-RuBy	Pr=53%	HSS=0.51	COSMO-RuBy	Pr=26%	HSS=0.35
4375	3433	7808	115	103	218

3843	87963	91806	328	99068	99396
8218	91396	99614	443	99171	99618
COSMO-RuBy + KCK+KHC	Pr=55%	HSS=0.55	COSMO-RuBy + KCK+KHC	Pr=37%	HSS=0.42
4512	2732	7244	164	178	342
3706	88664	92370	279	98993	99272
8218	91396	99614	443	99171	99614
Комплексный прогноз	Pr=73%	HSS=0.56	Комплексный прогноз	Pr=56%	HSS=0.41
6450	6152	12602	253	506	759
2383	90928	93291	200	104934	105134
8813	97080	105893	453	105440	105893

Таблица А.2 – То же, что в таблице А.1, но с заблаговременностью 24 ч

Более 12 м/с			Более 18 м/с		
GFS	Pr=74%	HSS=0.31	GFS	Pr=22%	HSS=0.18
3528	11455	14983	54	301	355
1212	90010	91222	190	105660	105850
4740	101465	106205	244	105961	106205
ICON-De	Pr=52%	HSS=0.48	ICON-De	Pr=28%	HSS=0.34
2465	2582	5047	68	83	151
2298	99445	101743	176	106463	106639
4763	102027	106790	244	106546	106790
COSMO-Ru6.6ENA	Pr=61%	HSS=0.46	COSMO-Ru6.6ENA	Pr=38%	HSS=0.37
2777	3989	6766	93	158	251
1808	92212	94020	151	100384	100535
4585	96201	100786	244	100542	100786
COSMO-Ru6.6ENA + KCK+KHC	Pr=53%	HSS=0.52	COSMO-Ru6.6ENA + KCK+KHC	Pr=41%	HSS=0.43
2434	1988	4422	99	114	213
2151	94213	96364	145	100428	100573
4585	96201	100786	244	100542	100786
COSMO-RuBy	Pr=52%	HSS=0.46	COSMO-RuBy	Pr=35%	HSS=0.40
2315	2761	5076	77	89	166
2112	91840	93952	145	98717	98862
4427	94601	99028	222	98806	99028
COSMO-RuBy + KCK+KHC	Pr=48%	HSS=0.50	COSMO-RuBy + KCK+KHC	Pr=38%	HSS=0.44
2147	1606	3753	89	83	170
2280	92995	95274	135	98723	98858
4427	94601	99028	222	98806	99028
Комплексный прогноз	Pr=65%	HSS=0.53	Комплексный прогноз	Pr=58%	HSS=0.47
3098	3309	6407	142	220	362
1640	97658	99298	101	105242	105343
4738	100967	105705	243	105462	105705



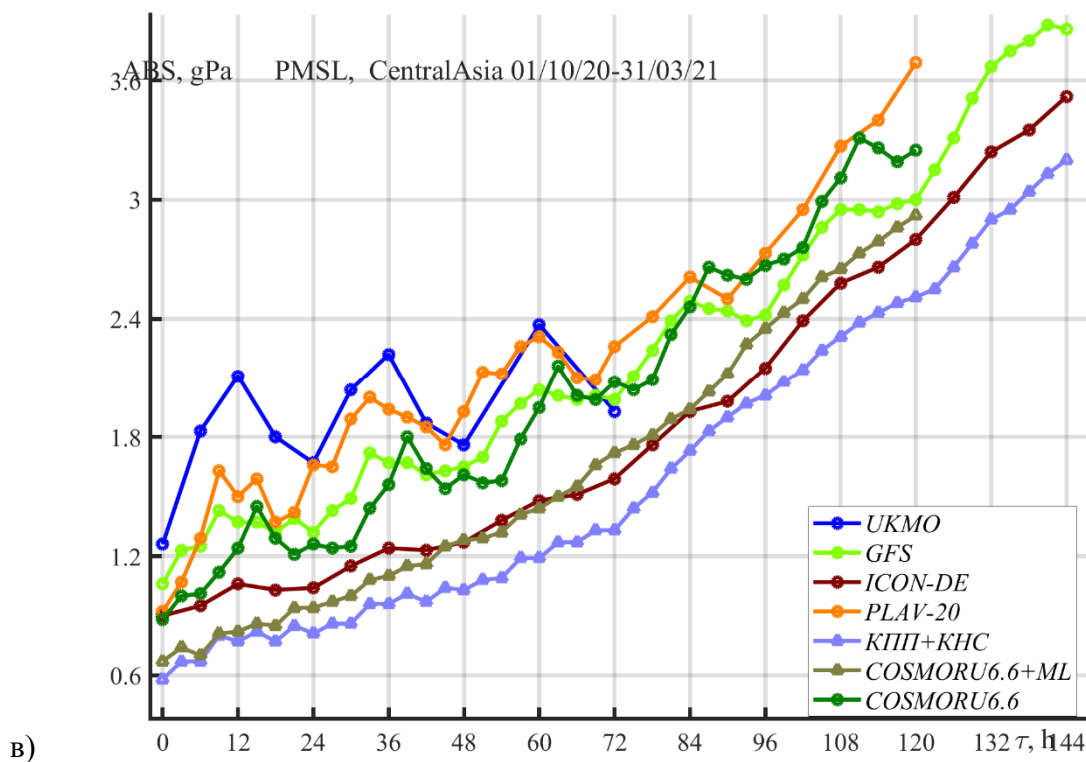


Рисунок А.4 – Средняя абсолютная ошибка прогнозов (ось абсцисс) давления, приведенного к уровню моря от начального срока 12:00 ВСВ по различным методам за период с 1 октября 2020 г. по 31 марта 2021 г. по различным регионам в зависимости от заблаговременности (ось ординат):
а) ЕЧР; б) АЧР; в) Средняя Азия

Приложение Б. Примеры продукции, использующей предлагаемый постпроцессинг

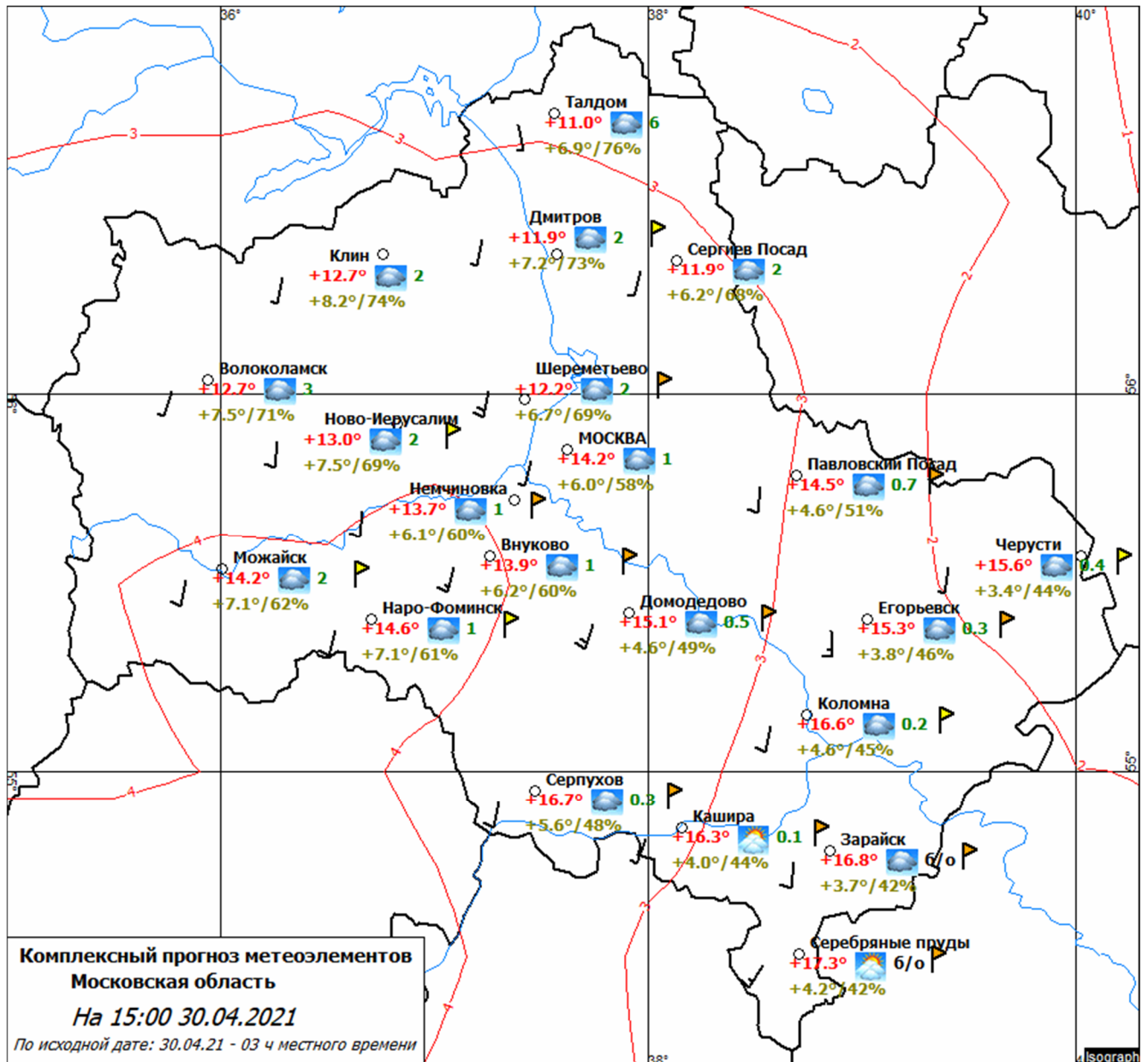


Рисунок Б.1 – Пример прогностической карты с сайта методического кабинета ФГБУ «Гидрометцентр России» [9] с прогнозами, рассчитанных предлагаемым комплексным прогнозом (авторы визуализации Н.А. Светлова, Е.Н.Шакотько)

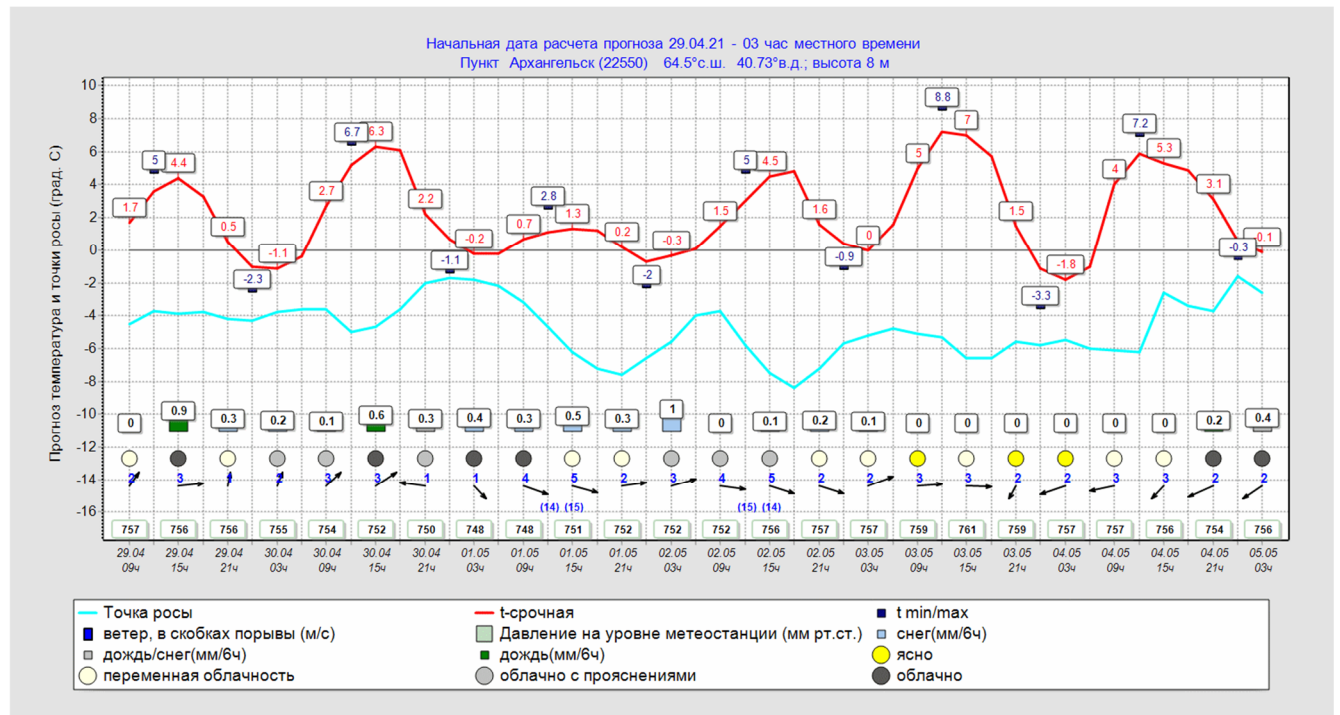


Рисунок Б.2 – Пример метеограммы с сайта методического кабинета
ФГБУ «Гидрометцентр России» [9] с прогнозами,
рассчитанных предлагаемым комплексным прогнозом

Мультимодельный прогноз

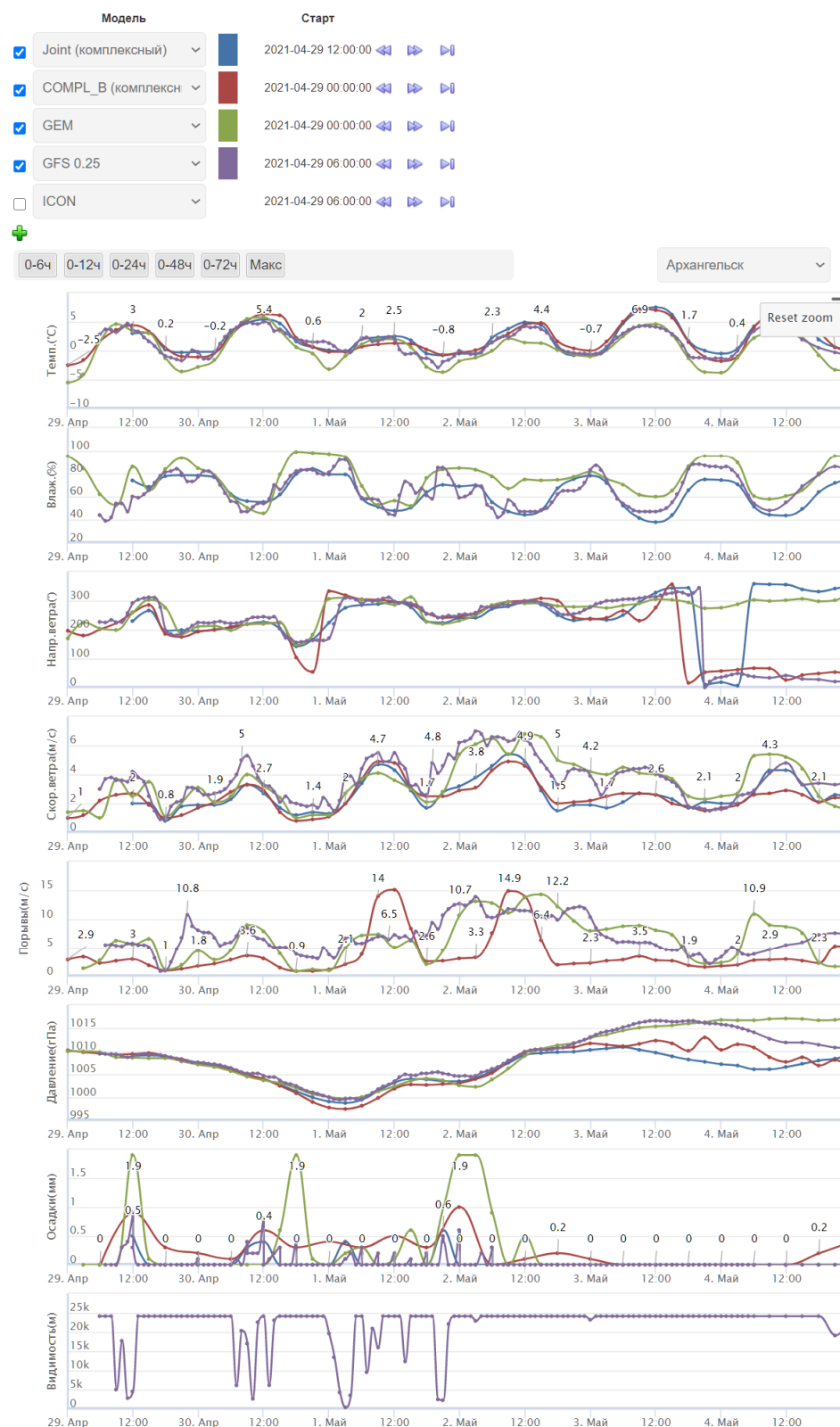


Рисунок Б.3 – Пример сравнительных метеограмм, содержащий прогнозы по предлагаемому комплексному прогнозу (обозначен как COMPLEX, красные кривые), с сайта «Погода для специалистов» <https://special.meteoinfo.ru/> в разделе «Прогнозы по пунктам» (администратор сайта А.В.Смирнов)

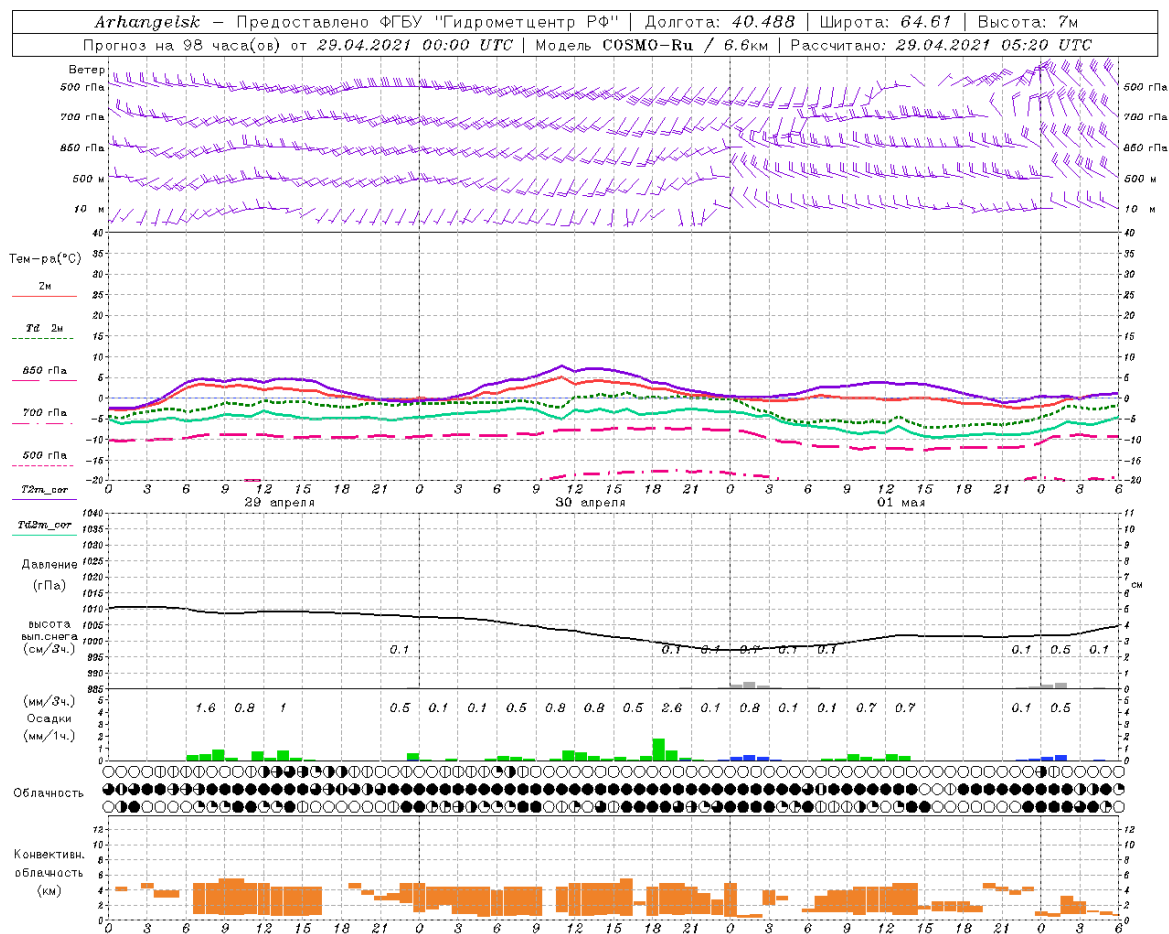


Рисунок Б.4 – Пример метеограммы по модели COSMO-Ru в конфигурации COSMO-Ru6ENA с сайта численных прогнозов погоды для метеорологического обеспечения <http://u2019.meteoinfo.ru/services> с поправленными КСК+КНС прогнозами температуры (фиолетовая кривая) и точки росы (зеленая сплошная кривая)

Сравнительная оценка прогнозов температуры воздуха,
опубликованных на различных ИНТЕРНЕТ-сайтах
по территории России (53 города*)
Период осреднения: 01.03.2021 - 31.03.2021

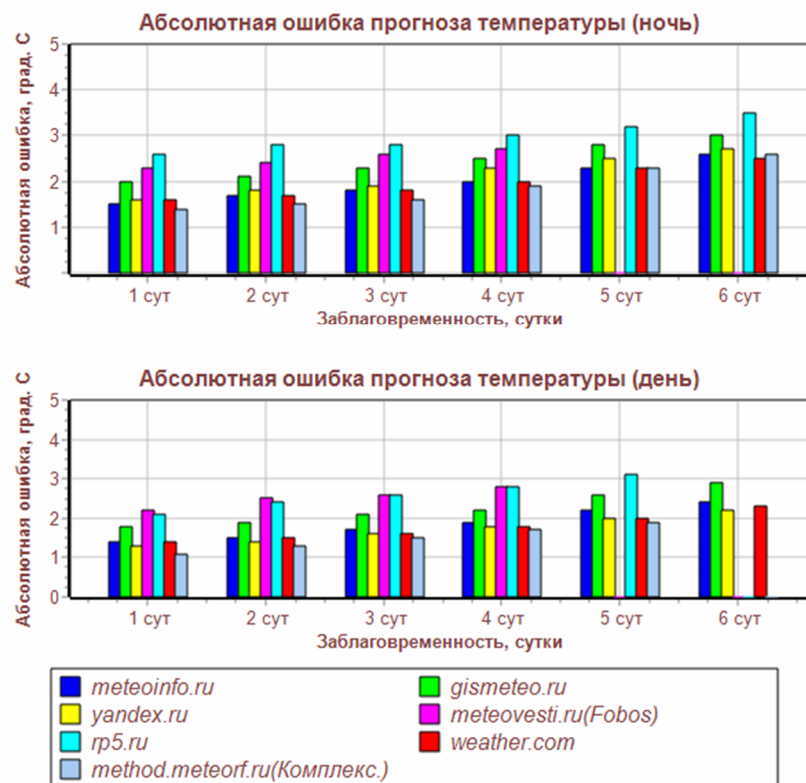


Рисунок Б.5 – Сравнительные оценки качества прогнозов с некоторых сайтов (в том числе по предлагаемому комплексному прогнозу с сайта [9]), ежемесячно публикуемые на сайте методического кабинета ФГБУ «Гидрометцентр России» по адресу <http://193.7.160.230/web/osio/meteogram/53-stations-inet-month.html> (оценки сделаны Н.А. Светловой)

Сравнительная оценка различных методов прогнозов температуры воздуха по административным центрам РФ Период оценки: Март 2021 Мурманск (22113)														
Метод прогноза	Прогнозы на ночь (Tmin)					Прогнозы на день (Tmax)					Прогноз осадков на ночь		Прогноз осадков на день	
	N	OPR	BIAS	ABS	OTN	N	OPR	BIAS	ABS	OTN	N	OPR	N	OPR
1-е сутки														
Синоптический (КП-68)	31	97.	-0.7	1.4	0.37	31	100.	-1.1	1.2	0.25	31	81.	31	86.
РЭП	31	100.	-0.2	1.1	0.30	31	100.	-0.5	0.9	0.19	–	–	–	–
Комплексный прогноз	31	100.	-0.5	0.9	0.25	31	100.	-0.2	0.7	0.14	31	86.	31	89.
2-е сутки														
Синоптический (КП-68)	20	95.	-1.0	1.3	0.29	20	100.	-1.1	1.3	0.26	20	72.	20	86.
РЭП	20	90.	-0.6	1.3	0.28	20	100.	-0.4	1.0	0.19	–	–	–	–
Комплексный прогноз	20	100.	-0.6	1.1	0.24	20	100.	-0.3	0.9	0.17	20	81.	20	85.
3-и сутки														
Синоптический (КП-68)	20	85.	-0.9	1.7	0.30	20	100.	-1.3	1.3	0.20	20	86.	20	90.
РЭП	20	80.	-0.9	1.8	0.32	20	100.	-0.9	1.2	0.18	–	–	–	–
Комплексный прогноз	20	90.	-1.0	1.7	0.30	20	100.	-0.4	0.9	0.14	20	76.	20	88.

N - количество случаев оценки

OPR - оправдываемость,% (по "Наставлению...")

BIAS - средняя арифметическая ошибка температуры,°C

ABS - средняя абсолютная ошибка температуры,°C

OTN - относительная ошибка

Рисунок Б.6 – Сравнительные оценки качества прогнозов по методам УГМС (код КП-68), РЭП и комплексного прогноза, ежемесячно публикуемые для центров регионов РФ на сайте комплексного прогноза [9] в разделе «сравнительная оценка»

Сопоставление фактических и прогностических значений температуры
и осадков при прогнозе на ближайшую ночь и следующий день
(за последние 15 дней) в пункте: Красноярск
Исходный срок прогноза: 07 час местного времени

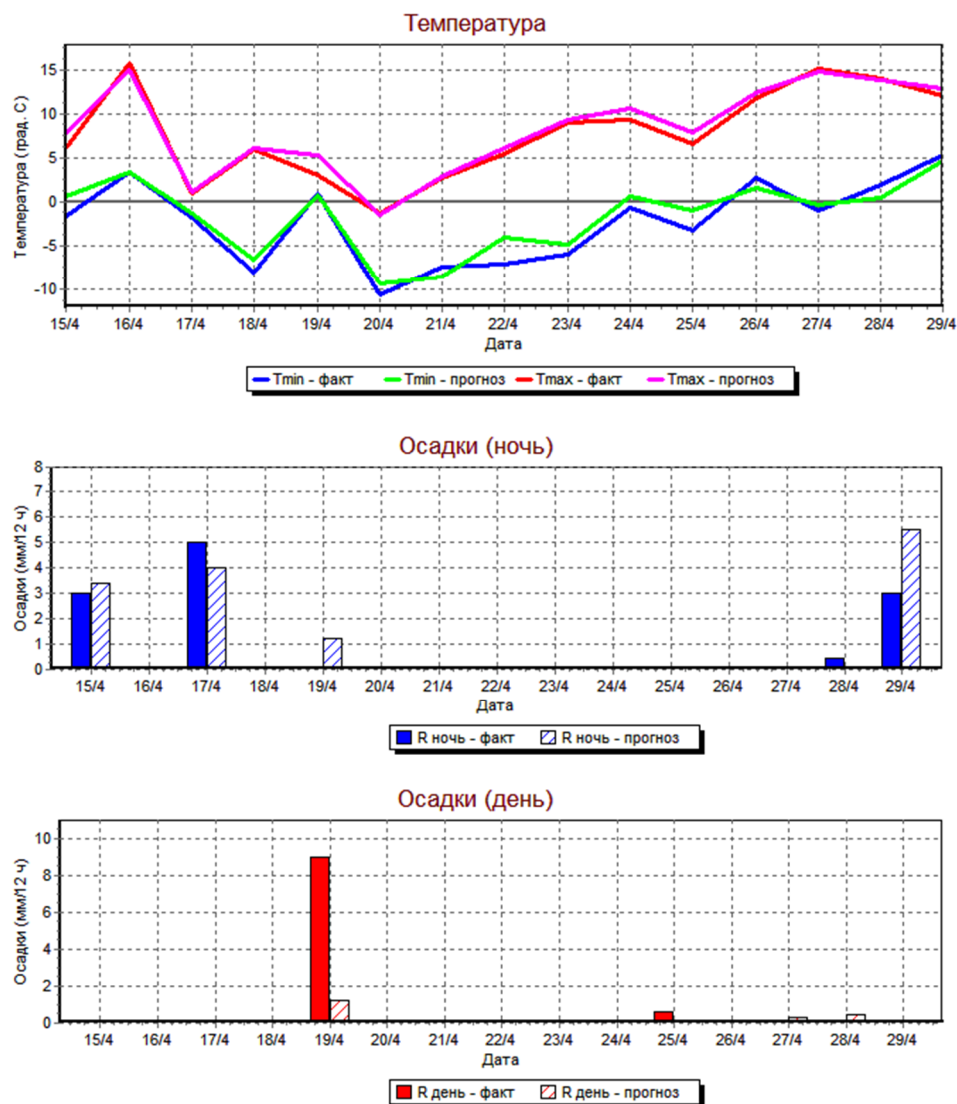


Рисунок Б.7 – Сравнение прогностических (по предлагаемому комплексному прогнозу) и фактических значений максимальной дневной и минимальной ночной температур воздуха, количества осадков, ежедневно публикуемые на сайте комплексного прогноза [9] в разделе «прошедшая погода»

Мониторинг качества прогноза



Рисунок Б.8 – Пример сравнения средних абсолютных погрешностей прогнозов, в том числе по предлагаемому комплексному прогнозу (обозначен как COMPL_V, красные кривые) на сайте «Погода для специалистов» <https://special.meteoinfo.ru/> в разделе «Мониторинг качества прогнозов» (администратор сайта А.В.Смирнов)

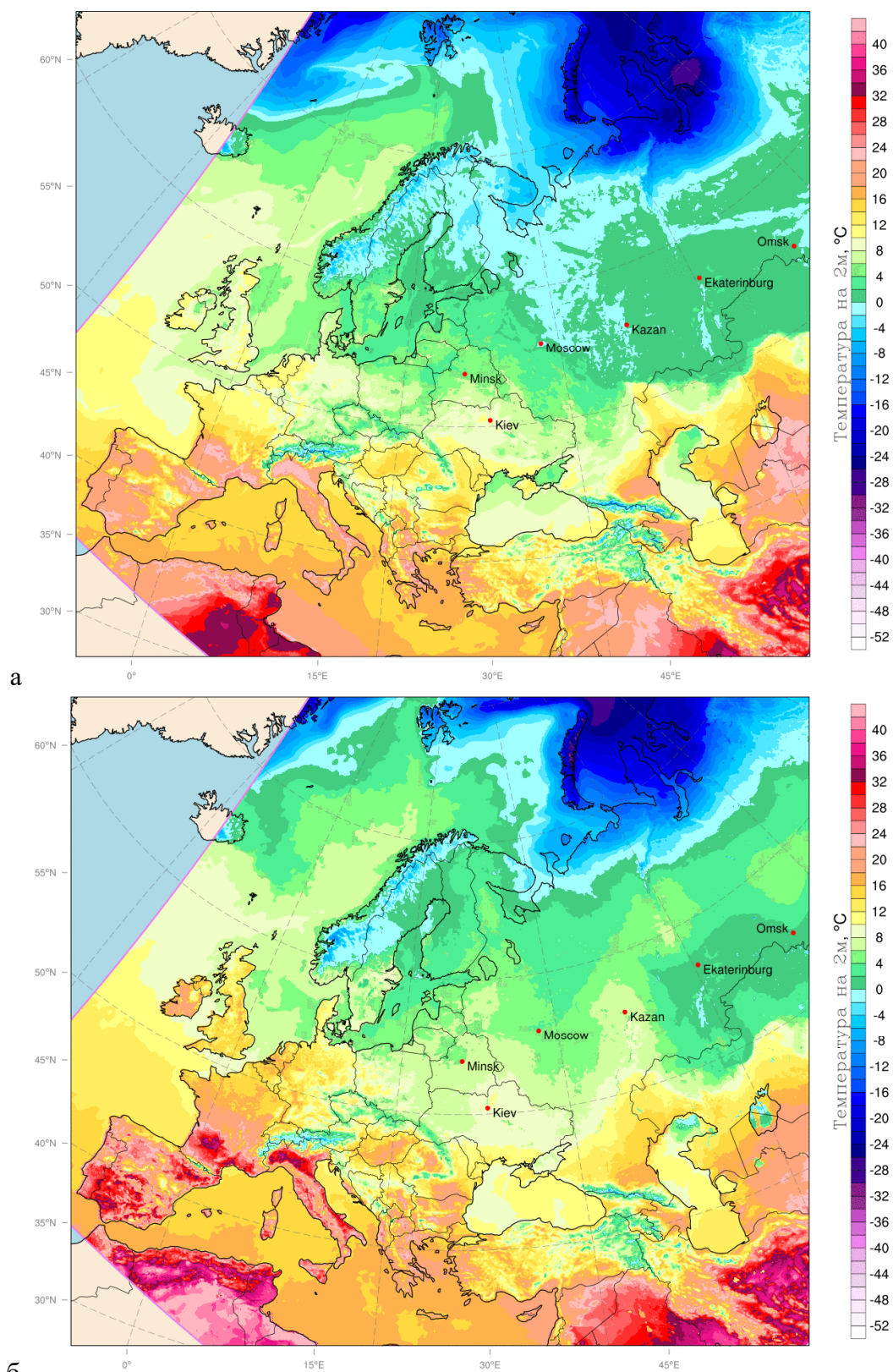


Рисунок Б.9 – Пример карт с прогнозом температуры воздуха на высоте 2 м от 00:00 UTC 30 марта 2021 года с заблаговременностью 108 часов (нарисованы Д.В.Блиновым):
 а. по модели COSMO-Ru в конфигурации COSMO-Ru6ENA;
 б. комплексным прогнозом с первым приближением по модели COSMO-Ru в конфигурации COSMO-Ru6ENA

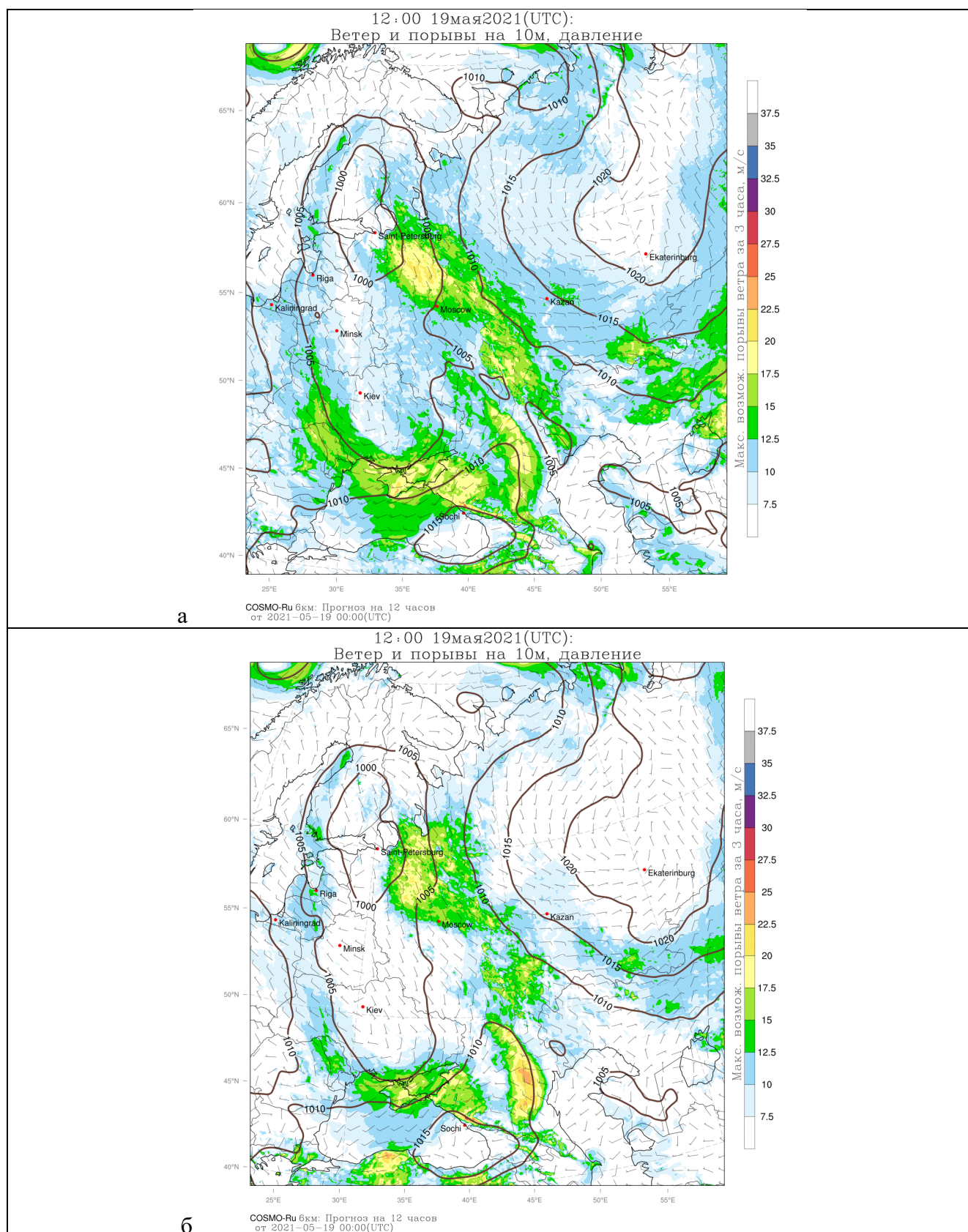


Рисунок Б.10 – Пример карт с прогнозом скорости ветра, порывов ветра на высоте 10 м и давления, приведенного на уровень моря, от 00:00BCB 19 мая 2021 года с заблаговременностью 12 часов (нарисованы Д.В.Блиновым):

- а. по модели COSMO-Ru в конфигурации COSMO-Ru6ENA;
- б. комплексным прогнозом с первым приближением по модели COSMO-Ru в конфигурации COSMO-Ru6ENA

**Приложение В. Письма руководителей и специалистов региональных управлений
Росгидромета**

От: Кожевникова Т. П. <t.kozhevnikova@pogoda-sv.ru>

Кому: shakotko@mecon.ru

Написано: 27 февраля 2020 г., 10:21:05

Тема: Использование карт комплексного прогноза по Приволжскому ФО Самарская область

Файлы: <none>

Уважаемая Елена Николаевна!

Прогнозисты отдела довольны картами комплексного прогноза, отмечают хорошую оправдываемость расчетных метеопараметров. Преимуществом является то, что прогностические данные можно визуально обозреть сразу по всем районам области.

Планируется ли в дальнейшем размещение на картах КП прогноза конвективных явлений (гроза, град, шквал) ?

--

С уважением,

Кожевникова Татьяна Петровна

Начальник отдела ОМП ФГБУ "Приволжское УГМС"

Телефон: 8(846) 952-99-09 внутр. номер: 141

Елена Николаевна, здравствуйте! В течение февраля 2020 года синоптики Саратовского ЦГМС использовали информацию с карт комплексного прогноза по Приволжскому федеральному округу. В течение месяца все прогнозируемые элементы оправдывались неплохо. Информация является нужной и полезной!!! Просим дополнительно нанести прогностическую информацию для севера области: Балаково (индекс 34086), Б. Карабулак (34072), Перелюб (35007). И для себя хотелось бы знать какие прогностические модели используются в этом комплексном прогнозе?

С уважением, Васильева Л.А.

From: Светлана Хрусталева <sslapt@mail.ru>
Sent: Tuesday, September 1, 2020 11:37 AM
To: bagrov@mecom.ru
Cc: shakotko55@mail.ru
Subject: письмо

Добрый день! Проанализировав комплексный прогноз (карты комплексного прогноза Методического кабинета ГМЦ России) в летние месяцы по станциям Забайкальского края (ФГБУ «Забайкальское УГМС»), хочется отметить высокую оправдываемость минимальной и максимальной температуры, в том числе и по г.Чита. В прогнозе температур чаще прослеживается занижение как ночных, так и дневных температур. Хуже оправдываются осадки по количеству, но, если оценивать по наличию, независимо от количества, то оправдываемость составляет более 85%, сильный ветер (более 15 м/с) в летний период 2020 году не часто наблюдался по территории Забайкальского края, поэтому оценить эффективность данного явления сложно. В оперативной работе синоптики Читинского ГМЦ используют данный метод, как консультативный, который имеет хорошую оправдываемость. Если есть такая возможность, хотелось бы расширить список наблюдательных подразделений по территории Забайкальского края, включить больше восточных и южных районов.

С уважением начальник отдела метпрогнозов Читинского ГМЦ,
Светлана Хрусталева

To: bagrov@mecom.ru
Subject: Об оценке

Добрый день!

Синоптикам Ульяновского ЦГМС - филиала ФГБУ "Приволжское УГМС" для работы очень полезны комплексные прогнозы. Спасибо Вам за дополнительную информацию. А вот прогностические изотермы АТ-850 нам не нужны на данных прогнозах, они итак имеются на прогностических картах.

--

С уважением, Шарафетдинова Регина Менсуровна Начальник отдела прогнозирования
Ульяновского ЦГМС- филиала ФГБУ "Приволжское УГМС"
Телефон: 8(8422)41-08-23

От: **pogoda(mecom)** <pogoda@hbrv.mecom.ru>
Date: пт, 6 нояб. 2020 г. в 11:46
Subject: О Комплексном прогнозе
To: <a.bagrov13@gmail.com>

Добрый день!

Представленный Вами прогноз по территории на картах очень удобен.

Мы всегда пользовались графическими прогнозами по пунктам, где заблаговременность до 144час. Обзор на картах безусловно более удобен в работе при анализе прогностической обстановки, чем графически.

Наши пожелания. Добавить на карту пункты по Хабаровскому краю -Чумикан, Чегдомын и если получится. то П.Осипенко.

Но одно другому не мешает. Необходимо добавить в список графических ансамблевых прогнозов эти же пункты по Хабаровскому краю и четыре пункта по Еврейской АО: Е.Никольское, Облучье, Смидовичи, Ленинское, которые есть на карте.

--

С уважением Агеева С.В.
ФГБУ "Дальневосточное УГМС"

РОСГИДРОМЕТ

Федеральное государственное
бюджетное учреждение
«МУРМАНСКОЕ УПРАВЛЕНИЕ
ПО ГИДРОМЕТЕОРОЛОГИИ И
МОНИТОРИНГУ ОКРУЖАЮЩЕЙ СРЕДЫ»
(ФГБУ «Мурманское УГМС»)

Шмидта ул., д. 23, г. Мурманск, 183038
Телефон: (815-2) 47-25-49; факс: (815-2) 47-24-06
e-mail: leader@kolgimet.ru; http://www.kolgimet.ru
ОКПО 02572737, ОГРН 1025100851522
ИНН/КПП 5191501269/519001001

Директору
ФГБУ «Гидрометцентр России»
Борщу С.В.

Эл. почта: shakotko@mecom.ru

06.08.2020 № 60/61-577

На №К-1090 от 26.12.2019

Уважаемый Сергей Васильевич!

Синоптики ФГБУ «Мурманское УГМС» успешно используют в своей работе Комплексные прогнозы «Методического кабинета» Гидрометцентра России по отдельным пунктам области (г.Мурманск, г.Мончегорск, г.Апатиты). Карты Комплексного прогноза по Мурманской области, после начала их публикации, также стали использоваться в работе специалистов. В результате был сделан вывод, что использование карт Комплексного прогноза возможно в качестве вспомогательного материала для обзора прогноза метеозадач по территории нашего региона. Для специалистов нагляднее и удобнее для использования прогноз, представленный по отдельным пунктам в графическом виде.

К сожалению, в последнее время в Комплексном прогнозе для г.Мончегорска и г.Апатиты перестали даваться экстремальные значения температуры воздуха. Просим, при возможности, восстановить предоставление прогностических значений экстремальных температур воздуха для указанных городов, так как данный вид информации имеет хорошую оправдываемость и ежедневно используется при составлении прогнозов. Для синоптиков Мурманского УГМС большой помощью в работе был бы, дополнительно к имеющимся, Комплексный прогноз (в графическом виде) для н.п.Териберка.

Начальник



О.М.Чаяс

Алдарова Наталия Владимировна,
гл.синоптик ОМП ГМЦ
(8152)47-23-69,
aldarova@kolgimet.ru



Росгидромет
ФГБУ «Центральное УГМС»
Владимирский центр по
гидрометеорологии и мониторингу
окружающей среды - филиал
Федерального государственного
бюджетного учреждения "Центральное
управление по гидрометеорологии и
мониторингу окружающей среды"
(Владимирский ЦГМС - филиал ФГБУ
«Центральное УГМС»)
Юридический адрес: Нововаганьковский пер., д.8,
Москва, ГСП-3, 123242
Почтовый адрес: ул. Стрелецкая д.20, Владимир,
600021
E-mail: cgms@vladimir.mecom.ru
т/ф. 8 (4922) 32-63-64
ОКПО 32940928, ОГРН 1127747295170
ИНН/КПП 7703782266/332743001

30.10.2019 г. №02/05/1074

Главному специалисту ОГКПП
Гидрометцентра России
Шакотько Н.Н.

Уважаемая Елена Николаевна!

Владимирский центр по гидрометеорологии и мониторингу окружающей среды – филиал Федерального государственного бюджетного учреждения «Центральное управление по гидрометеорологии и мониторингу окружающей среды» на Ваш запрос сообщает, что карты комплексного прогноза метеозлементов по областям будут использоваться синоптиками филиала при прогнозировании, т.к. имеют ряд положительных качеств:

- ✓ Удобство в применении;
- ✓ Возможность просмотра карт прогноза метеозлементов соседних областей;
- ✓ Большая заблаговременность;
- ✓ Добавлены пункты Владимирской области, для которых ранее не составлялся прогноз метеозлементов (г. Александров, г. Петушки, г. Муром, г. Вязники).

Поскольку карты прогнозов метеозлементов синоптики филиала только начали использовать в своей практике, высказать свои пожелания (замечания) будет возможно немного позднее. Но нужно отметить, что сильный ветер с порывами более 15 м/сек., который наблюдался по всей территории Владимирской области 28.10.2019 года, был отмечен на картах прогноза метеозлементов заблаговременно.

С уважением,
Начальник Владимирского ЦГМС –
филиала ФГБУ «Центральное УГМС»

И.В. Новикова

Исполнитель: синоптик 2 категории Лебабина Елена Николаевна,
Тел. (4922)32-63-64.