

ФЕДЕРАЛЬНОЕ ГОСУДАРСТВЕННОЕ БЮДЖЕТНОЕ УЧРЕЖДЕНИЕ  
«ГИДРОМЕТЕОРОЛОГИЧЕСКИЙ НАУЧНО-ИССЛЕДОВАТЕЛЬСКИЙ  
ЦЕНТР РОССИЙСКОЙ ФЕДЕРАЦИИ»

На правах рукописи

УДК 551.501.75

Мизяк Василий Геннадьевич

**АНСАМБЛЕВАЯ СИСТЕМА УСВОЕНИЯ ДАННЫХ С  
ИСПОЛЬЗОВАНИЕМ СПУТНИКОВЫХ  
НАБЛЮДЕНИЙ ВЕТРА**

Специальность 1.6.18 —

«Науки об атмосфере и климате»

Диссертация на соискание учёной степени  
кандидата физико-математических наук

Научный руководитель:

д.ф.-м.н.

Толстых Михаил Андреевич

Москва — 2024

## Оглавление

	Стр.
<b>Введение</b> . . . . .	5
 <b>Глава 1. Применяемые методы и подходы к усвоению</b>	
<b>метеорологических данных</b> . . . . .	16
1.1 Постановка задачи усвоения и основные определения . . . . .	16
1.2 Методы решения задачи усвоения . . . . .	19
1.2.1 Оптимальная интерполяция . . . . .	20
1.2.2 Вариационные методы усвоения данных . . . . .	22
1.2.3 Детерминированные и стохастические ансамблевые фильтры Калмана . . . . .	25
1.2.4 Локальный ансамблевый фильтр Калмана с преобразованием ансамбля . . . . .	30
1.2.5 Гибридные схемы усвоения . . . . .	35
1.3 Использование данных наблюдений в системах усвоения . . . . .	37
1.3.1 Виды наблюдений . . . . .	38
1.3.2 Оператор наблюдений . . . . .	41
1.3.3 Ошибки наблюдений . . . . .	42
1.3.4 Контроль качества наблюдений . . . . .	44
1.3.5 Локализация наблюдений . . . . .	45
1.3.6 Инфляция матрицы ковариаций ошибок первого приближения . . . . .	47
1.4 Спутниковые наблюдения ветра AMV . . . . .	49
1.5 Выводы к главе 1 . . . . .	52
 <b>Глава 2. Ансамблевая система усвоения Гидрометцентра России</b>	
2.1 Состав системы усвоения . . . . .	54
2.2 Глобальная полулагранжева модель атмосферы . . . . .	55
2.3 Настройка параметров, определяющих значения матрицы ковариаций ошибок наблюдений . . . . .	57
2.4 Применяемые способы контроля качества наблюдений . . . . .	58

2.5	Настройка параметров локализации наблюдений . . . . .	59
2.6	Настройка параметров инфляции матрицы ковариаций ошибок первого приближения . . . . .	62
2.7	Параллельная реализация алгоритма LETKF . . . . .	65
2.8	Задание начальных значений почвенных переменных . . . . .	70
2.9	Методика циклического запуска усвоения . . . . .	72
2.10	Выводы к главе 2 . . . . .	74
<b>Глава 3. Усвоение спутниковых наблюдений ветра AMV . . . . .</b>		<b>75</b>
3.1	Конфигурация системы усвоения . . . . .	75
3.2	Верификация результатов усвоения . . . . .	76
3.3	Использование данных наблюдений AMV . . . . .	78
3.4	Переопределение высоты наблюдений . . . . .	81
3.5	Автокорреляция ошибок наблюдений AMV и методы её учёта . .	85
3.5.1	Влияние прореживания наблюдений . . . . .	89
3.5.2	Влияние учёта корреляций ошибок наблюдений . . . . .	89
3.6	Влияние включения данных AMV на результаты среднесрочных прогнозов . . . . .	97
3.7	Выводы к главе 3 . . . . .	99
<b>Глава 4. Применение системы ансамблевого усвоения в глобальной системе ансамблевого среднесрочного прогноза . . . . .</b>		<b>104</b>
4.1	Технология глобального ансамблевого среднесрочного прогноза на основе ансамблевого усвоения данных . . . . .	104
4.1.1	Генерация начальных данных и центрирование . . . . .	105
4.1.2	Верификация ансамблевых прогнозов . . . . .	107
4.1.3	Результаты подключения данных ветра AMV . . . . .	110
4.2	Результаты оперативных испытаний системы ансамблевого прогноза . . . . .	115
4.3	Выводы к главе 4 . . . . .	125
<b>Заключение . . . . .</b>		<b>126</b>
<b>Список литературы . . . . .</b>		<b>128</b>

Приложение А. Акт внедрения технологии ансамблевого прогноза . . . . .	144
--	-----

## Введение

Жизнь человека во многом зависит от состояния атмосферы, которая является нашей естественной средой обитания. Погода оказывает существенное влияние на многие аспекты хозяйственной деятельности. Опасные погодные явления нередко становятся причиной масштабных бедствий, которые поражают значительные территории различных государств по всему миру. Примером таких явлений являются ураганы «Катрина», «Харви» в США в 2005 и 2017 годах, «Айк» на Гаити в 2008 году, аномальная жара с лесными пожарами в России в 2010 г. и странах Европы в 2022 г., разрушительные наводнения в Краснодарском крае в 2012 и 2023 годах, на Дальнем Востоке в 2013 году, в Сибири в 2019 году и другие. Технический прогресс заметно повысил способность человека противостоять таким погодным явлениям, в т. ч. путём прогноза их появления и развития.

Сегодняшний мир уже невозможно представить без своевременного и точного прогноза погоды, который крайне важен для таких отраслей, как сельское хозяйство, рыбная ловля, авиаперевозки, морской и речной транспорт, энергетика и многие другие. Ошибки в предсказании грядущих погодных условий могут помешать подготовке к возникающим опасным метеорологическим явлениям и вести к принятию ошибочных управленческих решений. При этом вовремя сделанное предупреждение может существенно уменьшить, а в некоторых случаях и совершенно исключить, материальный ущерб и людские потери. Этот факт является одной из причин развития методов прогноза погоды, являющегося исключительно важной задачей, для решения которой работают национальные метеослужбы и научно-исследовательские центры.

Развитие компьютерных технологий, берущее своё начало во второй половине XX века и продолжающееся в настоящее время, стало одной из причин развития **численного прогноза погоды** как важного направления научной и практической деятельности мировых прогностических центров.

Любая система численного прогноза погоды представляет собой сложный программный комплекс, который содержит помимо численной математической модели, воспроизводящей динамику атмосферной циркуляции, также блок, ответственный за подготовку начальных данных. Подготовка начальных данных,

необходимых для старта численной модели прогноза погоды, в метеорологии происходит с помощью т. н. **усвоения данных**, которое занимает центральное место в данной работе.

В метеорологии усвоением данных называют циклический процесс, в ходе которого определяются численные характеристики состояния атмосферы в определённые моменты времени с использованием имеющихся данных метеорологических наблюдений и некоторой априорной информации об исследуемом состоянии атмосферы. В результате получают трёхмерное распределение модельных переменных (две составляющие скорости ветра, температура и влажность, давление на поверхности Земли). Такое распределение принято называть **объективным анализом**.

Объективный анализ является максимально точной (в некотором смысле) оценкой состояния атмосферы. Он используется в качестве начальных данных при интегрировании численных моделей прогноза погоды. Поэтому значения составляющих его модельных переменных определены в точках модельной сетки. Размерности массивов данных, из которых состоит объективный анализ, велики и продолжают постоянно расти вследствие непрекращающегося роста разрешающих способностей сеток современных моделей.

Так, используемая в данной работе версия глобальной модели атмосферы ПЛАВ [21], имеет постоянное разрешение по пространству  $0,9^\circ \times 0,72^\circ$ , что составляет 400 узлов сетки по долготе и 251 узел по широте. Все переменные должны быть определены на 96 вертикальных уровнях свободной атмосферы, а также на поверхности Земли. Таким образом, общая размерность массивов данных, из которых состоит объективный анализ, составляет величину порядка  $10^9 - 10^{10}$ . Кроме того, в ходе решения задачи усвоения необходимо обработать данные метеонаблюдений, число которых с учётом наблюдений со спутников, в современных прогностических системах имеет порядок  $10^{12} - 10^{14}$ .

Такие внушительные объёмы обрабатываемых данных вызывают потребность в разработке и развитии методов, которые должны быть хорошо обоснованы математически и при этом эффективно выполняться на современных массивно-параллельных вычислительных системах. Особенно актуальным является получение хороших результатов при использовании недостаточного количества метеорологических наблюдений над некоторыми обширными участками поверхности Земли. Так, сеть аэрологических станций для территории Рос-

сии можно охарактеризовать как крайне неоднородную и при этом достаточно скудную. По состоянию на 1 января 2019 г. вся государственная наблюдательная сеть включает 1921 метеорологическую станцию и 3079 гидрометеорологических постов [7]. Расстояния между соседними станциями зачастую превышают 500 км и только в некоторых густонаселённых районах составляют менее 300 км. Такое количество несомненно не соответствует требованиям объективного анализа метеорологических полей, горизонтальное разрешение которых не превышает сотню километров. Аналогичная ситуация с наблюдениями существует и для других обширных территорий Земли (Африка, Центральная и Южная Азия, акватории океанов, приполярные регионы). Очевидно, что для преодоления указанных трудностей необходимо привлекать дополнительные источники информации, такие как многолетние наблюдения за температурным режимом исследуемой области (климатические поля), результаты моделирования и др.

По мнению Л.С. Гандина, который наряду с Ж. Матероном является одним из основоположников усвоения данных и автором одного из наиболее долго и успешно применяемого метода, называемого «Оптимальная интерполяция», термин «объективный анализ» не является удачным, т.к. в ходе его получения происходит не исследование свойств полей метеорологических элементов (собственно, анализ), а, скорее, восстановление значений этих полей по данным метеорологических наблюдений [5, с. 5]. Тем не менее, этот термин получил широкое применение уже тогда и в дальнейшем ещё более укрепился в своём использовании.

Объективный анализ, чаще называемый просто **«анализ»** или **«реанализ»**, используют не только для инициализации численных моделей прогноза погоды. Часто поля анализа применяются также для получения псевдонаблюдений. На основе конкретных значений анализов формируют выборки тестовых наблюдений, обладающих заданными известными свойствами (в первую очередь с известной функцией распределения и её параметрами). Полученные таким образом «наблюдения» полезны при различных исследованиях в области методов усвоения данных, для изучения свойств и характеристик реальных наблюдений и в других областях.

Поля анализа применяются также для оценки успешности численных прогнозов. При этом анализ выступает в качестве наилучшего приближения к ис-

тинному состоянию атмосферы. Всемирной метеорологической организацией определён перечень стандартных процедур [13], с помощью которых производится расчёт оценок ошибок прогнозов. Эти оценки рассчитываются всеми прогностическими центрами. Ежемесячно происходит обмен информацией об ошибках прогнозов. Эта информация доступна на сайте Ведущего центра по верификации детерминированных численных прогнозов (Lead Centre for Deterministic Forecast Verification LCDNV). В данной работе указанные процедуры также применяются для оценки качества прогнозов. Более подробное описание используемых оценок приведено в разделе 3.2.

Аналогично получают и используют поля анализов в области моделирования и исследования мирового океана [45; 81; 82; 128], а также в получивших развитие последнее время совместных моделях атмосферы, океана и морского льда [6; 12].

Первые попытки оценить состояние атмосферы с использованием анализа данных наблюдений были проделаны ещё в середине XIX века. Математически эти попытки основывались на простейшей интерполяции имеющихся тогда данных измерений основных параметров погоды, таких как температура, скорость ветра, давление, количество выпавших осадков. Тогда, в виду отсутствия вычислительных возможностей и недостаточности знаний, результат анализа очень сильно отклонялся от фактического состояния параметров атмосферы.

Следующей попыткой получения начальных данных для моделирования атмосферы стала также интерполяция первых измерений в свободной тропосфере над континентальной Европой для численного эксперимента Ричардсона [109]. Далее этот эксперимент с использованием первой вычислительной машины повторил Чарни [48]. В его эксперименте формирование начальных условий занимало довольно много времени.

В дальнейших исследованиях по подготовке начальных данных для численного прогноза погоды была реализована идея применения некоторой априорной информации о текущем состоянии атмосферы в сочетании с данными метеорологических наблюдений. Такой информацией является т. н. «**первое приближение**» (в англоязычной литературе *first guess* или *background*). В качестве первого приближения чаще всего используют модельный краткосрочный прогноз (чаще всего 6-ти часовой) на интересующий нас момент времени. Реже вместо краткосрочного прогноза используют климатические поля.

Основы применения первого приближения, в качестве которого был использован краткосрочный модельный прогноз, были заложены в работе Бергторссона и Дуза [36]. Спустя несколько лет этот подход был развит Крессманом [52] и состоял в нескольких итерациях анализа в так называемом методе последовательных поправок (SCM или Successive Correction Method). Этот метод стоял на вооружении во многих бюро прогнозов погоды довольно долгое время.

Далее техника статистической интерполяции была модифицирована Л. С. Гандиным в 1963 году [5]. Он получал стартовые поля численных моделей прогноза погоды путём некоторой оптимальной оценки состояния атмосферы, выполненной на основе взвешенной суммы первого приближения и метеорологических наблюдений.

Современные же методы усвоения данных в целом можно разделить на две группы. Одна группа методов основана на вариационных принципах. Для решения задачи проводится поиск минимума некоторого функционала, описывающего «близость» модельного решения и данных измерений в некоторой заданной метрике. Такой подход в литературе получил название «вариационный метод» решения задач усвоения. Эта группа методов сравнительно хорошо разработана и успешно применяется на практике, хотя и требует для своей реализации значительных вычислительных ресурсов, а также создания и поддержки сопряжённой прогностической модели.

Вторая группа основывается на последовательном подходе, в основе которого лежит использование алгоритма фильтра Калмана [77]. Физический смысл этого метода состоит в определении связи между известным (наблюдаемым) и неизвестным состоянием атмосферы во времени и/или пространстве, которые в линейном приближении задаются через математическое ожидание и ковариационные функции. Развитие методов усвоения показало, что этот подход несколько проще реализуем, чем вариационный. Корректное применение данного метода в чистом виде возможно только для линейных систем [139]. В случае сильной нелинейности модели и/или нарушении других ограничений, накладываемых на систему, возможно применение некоторых специальных методов получения субоптимальной оценки состояния системы. Описание методов и схем, основанных на алгоритме фильтра Калмана, приводится в главе 1 настоящей работы.

В настоящее время активно развиваются гибридные схемы, объединяющие оба указанных выше подхода к усвоению данных. Чаще всего эти схемы

основаны на получении с помощью фильтра Калмана матрицы ковариаций ошибок прогноза и дальнейшем её использовании в вариационном алгоритме (например, [131—134] и др.). Подобные схемы, несмотря на отсутствие строгого доказательства своей оптимальности и сходимости, обычно вполне работоспособны. Более подробное описание некоторых подходов приводится в разделе 1.2.5.

Важнейшую роль при реализации как вариационных методов, так и алгоритмов, основанных на фильтре Калмана, играет первое приближение. Ключевой особенностью этого источника информации является его регулярность и однородность — поля доступны на трёхмерных модельных сетках в любой исследуемый момент времени. Эти поля гидростатически сбалансированы и являются предварительной оценкой исследуемого состояния атмосферы. Второй источник — данные метеорологических наблюдений — является разнородной, нерегулярной и неполной информацией, получаемой от различных систем мониторинга. Несмотря на нерегулярность, этот источник информации является более объективным, т. к. представляет собой измерения величин в моменты времени, максимально приближенные к исследуемому. Число метеорологических наблюдений хотя и постоянно растёт, но всё же по-прежнему меньше количества узлов модельной сетки, которое также постоянно увеличивается. И первое приближение, и данные метеонаблюдений содержат неизвестную априори ошибку. При этом, как будет показано в последующих главах, ошибка получаемого в ходе усвоения данных анализа уменьшается по сравнению с ошибками, содержащимися в обоих источниках первоначальной информации по отдельности.

В настоящее время актуальной становится необходимость использования для подготовки анализа постоянно увеличивающихся массивов данных спутниковых наблюдений. Сложность усвоения таких данных стоит в том, что в ходе спутниковых измерений (в отличие от традиционных контактных наблюдений) происходит определение величин, которые не являются прямой информацией о состоянии переменных, входящих в модельные уравнения. Для получения значений из спутниковых данных температуры, влажности, скорости и направления ветра и других величин, необходимых для численных моделей, приходится применять сложные алгоритмы и процедуры, которые способны вносить свои ошибки в измеренные значения. Учёт этих ошибок и их свойств сам по себе является сложной задачей, решение которой требует значительных исследова-

ний и вычислительных ресурсов. Тем не менее, эта задача требует решения, т. к. данных традиционной наблюдательной сети недостаточно для подготовки качественных анализов.

На сегодняшний день самой совершенной системой усвоения большинством исследователей считается система, разработанная и применяемая в Европейском центре среднесрочных прогнозов погоды (ЕЦСПП, англ. The European Centre for Medium-Range Weather Forecasts — ECMWF). Эта система использует в своей работе гибридный алгоритм, в котором объединяются достоинства обоих упомянутых выше подходов к решению задачи усвоения данных [76].

В Гидрометцентре России оперативной является система усвоения данных, основанная на алгоритме трёхмерного вариационного усвоения [27]. Указанная система использует как традиционные контактные наблюдения, так и некоторые виды спутниковых наблюдений. В качестве первого приближения используются прогностические поля 6-часового прогноза NCEP (Вашингтон, США).

В конце 2011 года в Лаборатории перспективных численных методов Гидрометцентра России началась разработка системы усвоения, которая должна производить начальные данные для численной модели прогноза погоды ПЛАВ. В качестве схемы усвоения была выбрана схема Локального ансамблевого фильтра Калмана с преобразованием ансамбля (англ. Local Ensemble Transform Kalman Filter, LETKF), предложенного в [75]. Анализ тематической литературы, имеющейся к тому времени, указывал на перспективность данной схемы с точки зрения относительной простоты реализации, высокого качества показываемых результатов и эффективного выполнения на массивно-параллельных вычислительных машинах. Последний пункт был весьма важен в свете планов по повышению разрешения модели ПЛАВ и увеличивающегося объёма данных наблюдений, которые необходимо усваивать для повышения качества получаемых анализов. Кроме того, ансамблевая природа схемы LETKF позволяла использовать её в дальнейшем для получения ансамбля анализов в системе ансамблевого прогноза, разработка которой также стояла на повестке дня.

Первые результаты по разработке системы усвоения были опубликованы в [10; 29; 116]. Дальнейшая эволюция предполагала использование спутниковых данных наблюдений. Для этого были выбраны наблюдения ASCAT [19] и AMV [9]. Именно усвоению спутниковых данных AMV (Atmospheric Motion

Vector [42]), которые являются единственным источником о ветре в свободной атмосфере над некоторыми регионами Земли, и посвящена данная работа.

**Актуальность выполненного исследования.** Необходимость ансамблевой системы среднесрочного прогноза, воспроизводящей неопределённость прогностической модели атмосферы и начальных данных для её работы. Для этого размер ансамбля должен составлять не менее нескольких десятков. Система усвоения данных наблюдений на основе LETKF позволяет описать неопределённость начальных данных. Важным видом данных наблюдений являются спутниковые наблюдения ветра.

**Целью** диссертационной работы является разработка системы ансамблевого усвоения данных, использующей контактные наблюдения за состоянием атмосферы и спутниковые наблюдения ветра, и применение этой системы для ансамблевого среднесрочного прогноза погоды.

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

1. Разработать и применить методику настройки ансамблевой системы усвоения, позволяющую уменьшить ошибку вектора состояния системы.
2. Исследовать свойства ошибок спутниковых наблюдений ветра, разработать методику их учёта для использования в ансамблевой системе усвоения данных.
3. Разработать и испытать систему ансамблевого среднесрочного прогноза погоды, основанную на ансамблевой системе усвоения.

**Основные положения, выносимые на защиту:**

1. Разработанная методика настройки ансамблевой системы усвоения позволяет уменьшить ошибку и управлять разбросом ансамбля начальных данных, уменьшая ошибку ансамблевого прогноза погоды.
2. Использование схемы переопределения высоты наблюдений и учёт корреляций ошибок наблюдений спутниковых наблюдений ветра AMV повышает точность начальных данных для численных прогнозов погоды.
3. Разработанная при участии автора система усвоения позволила реализовать и внедрить новую систему ансамблевого среднесрочного прогноза погоды в Гидрометцентре России.

**Научная новизна:**

1. Впервые в России спутниковые данные ветра AMV использованы в ансамблевой системе усвоения данных.

2. Впервые в России применено моделирование ковариаций ошибок наблюдений в усвоении данных в метеорологии.

3. Выполнено новое исследование влияния свойств ошибок спутниковых наблюдений ветра на точность среднесрочного прогноза погоды.

**Теоретическая значимость.** Предложена методика настройки, позволяющая повысить точность результатов ансамблевой системы усвоения. Разработана схема учёта корреляции ошибок спутниковых наблюдений ветра, учитывающая различные типы спутников, их географическое расположение и каналы получения данных.

**Практическая значимость.** Разработана и реализована ансамблевая система усвоения данных, в которой использованы в т.ч. спутниковые наблюдения ветра. Усвоение спутниковых данных ветра повышает точность численных прогнозов, в которых полученная оценка состояния атмосферы использовалась в качестве начальных данных. Эта система усвоения использована в системе ансамблевого среднесрочного прогноза погоды, которая прошла испытания и внедрена в оперативную практику в Гидрометцентре России.

**Степень достоверности** полученных результатов обеспечивается результатами проведённых численных экспериментов и их сравнением с данными метеорологических наблюдений и оперативного анализа Гидрометцентра России. Приведённые в диссертационной работе результаты находятся в соответствии с результатами, полученными другими авторами.

**Апробация работы.** Основные результаты работы докладывались и обсуждались автором на следующих конференциях:

1. Международная суперкомпьютерная конференция «Научный сервис в сети Интернет: поиск новых решений», п. Абрау-Дюрсо, Краснодарский край, 17 – 22 сентября 2012 г.;

2. Международная молодёжная школа и конференция по вычислительно-информационным технологиям для наук об окружающей среде: «CITES-2015», г. Томск, 20 – 30 июня 2015 г.;

3. Школа-семинар «Современные методы дистанционных исследований и прогноза параметров среды в Арктике», г. Таруса, Калужская обл., 12 – 14 октября 2015 г.;

4. Международная научная конференция «Актуальные проблемы вычислительной и прикладной математики 2015» (АПВПМ-2015), г. Новосибирск, Академгородок, 19 – 23 октября 2015 г.;

5. Международная молодёжная школа и конференция по вычислительно-информационным технологиям для наук об окружающей среде: «CITES-2017», г. Таруса, г. Звенигород, 28 августа – 7 сентября 2017 г.;

6. Международная молодёжная школа и конференция по вычислительно-информационным технологиям для наук об окружающей среде: «CITES-2019», г. Москва, 27 мая – 6 июня 2019 г.;

7. International Conference on Computer Simulation in Physics and beyond, October 12-16, 2020, Moscow, Russia;

8. Международная молодёжная школа и конференция по вычислительно-информационным технологиям для наук об окружающей среде: «CITES-2021», г. Москва, 22 – 27 ноября 2021 г.;

9. Международная конференция и школа молодых учёных по измерениям, моделированию и информационным системам для изучения окружающей среды: «ENVIROMIS-2022», г. Томск, 12 – 17 сентября 2022 г.;

10. IV Всероссийская конференция с международным участием «Турбулентность, динамика атмосферы и климата», посвящённая памяти академика Александра Михайловича Обухова, г. Москва, 22–24 ноября, 2022 г.;

11. Международная конференция «Марчуковские научные чтения 2024», г. Новосибирск, 7 – 11 октября 2024 г.

Кроме того, автором были сделаны доклады по теме диссертационной работы на семинарах в ИВМ РАН и в Гидрометцентре России.

**Личный вклад.** Автор принимал участие во многих совместных этапах разработки ансамблевой системы усвоения, среди которых наиболее значимый вклад был внесён на этапе программной реализации параллельного алгоритма усвоения с помощью гибридной технологии MPI+OpenMP. Лично автором были выполнены следующие работы:

– разработка и внедрение системы верификации прогнозов, сделанных на основе начальной оценки состояния атмосферы, полученных с помощью разработанной системы усвоения;

– исследование влияния свойств ошибок наблюдений спутниковых наблюдений ветра AMV на точность получаемых результатов;

- разработка схемы переопределения высоты наблюдений AMV;
- разработка схемы использования коррелированных ошибок в локальном наборе данных наблюдений AMV;
- разработка методики настройки параметров схемы аддитивной инфляции в ансамблевой системе усвоения;
- разработка общего алгоритма и практическая реализация системы среднесрочного ансамблевого прогноза, в т. ч. системы ансамблевой верификации.

Основные результаты по теме диссертации изложены в 6 статьях, 4 из которых опубликованы в журналах, входящих в список ВАК. Кроме того, соискатель является соавтором 18 прочих статей, 2 монографий, 9 тезисов докладов. Получено 3 свидетельства Роспатента о государственной регистрации программы для ЭВМ. Имеется акт № 1 ЦМКП/2022/ от 26.12.2022 г. о внедрении технологии ансамблевого прогноза метеорологических полей с заблаговременностью до 10 суток на основе результатов настоящей диссертационной работы (Приложение А).

**Объем и структура работы.** Диссертация состоит из введения, четырёх глав, заключения и одного приложения. Полный объём диссертации составляет 144 страницы с 40 рисунками и 4 таблицами. Список литературы содержит 144 наименования.

## Глава 1. Применяемые методы и подходы к усвоению метеорологических данных

Ключевой идеей усвоения данных в современной метеорологии является получение наилучшей возможной оценки состояния атмосферы, используя для этого краткосрочные прогнозы погоды и данные метеорологических наблюдений в такой комбинации, которая учитывает статистические характеристики обоих входящих в неё источников информации. Таким образом, основными составляющими любой системы усвоения являются:

- прогностическая модель, используемая для получения краткосрочных прогнозов;
- метеорологические наблюдения;
- алгоритм, который используется для получения оценки состояния атмосферы в интересующий нас момент времени.

Постановка задачи усвоения данных, обозначения используемых величин и наиболее часто применяемые методы её решения рассмотрены в разделах 1.1 и 1.2 соответственно. В разделе 1.3 описываются особенности применения наблюдений при усвоении данных. Раздел 1.4 посвящён спутниковым данным ветра AMV. Выводы к главе 1 даются в разделе 1.5.

### 1.1 Постановка задачи усвоения и основные определения

В общей постановке задача усвоения данных состоит в оценке состояния управляющих параметров некоторой системы в определённый момент времени, используя информацию о связях между этими параметрами, об их эволюции во времени и о наблюдениях за ними. Применительно к моделированию атмосферы можно сказать, что с помощью имеющихся наблюдений и первого приближения необходимо оценить состояние атмосферы в точках модельного пространства  $X$ .

Рассматривается математическая модель, которая описывает эволюцию параметров состояния атмосферы:

$$\frac{d\mathbf{x}}{dt} = M(\mathbf{x}, t). \quad (1.1)$$

Здесь  $\mathbf{x} \in \mathbb{X}$  – вектор состояния, который имеет размерность  $n$ , равную произведению количества модельных переменных на количество точек модельного пространства (порядка  $10^9$ – $10^{10}$  в реальных задачах). Этот вектор является приближением (с некоторой ошибкой  $\varepsilon^M$ ) вектора *реального* состояния атмосферы  $\mathbf{x}^t$ , которое нам не известно.  $M$  – динамический оператор модели, осуществляющий переход системы из состояния, описываемого вектором  $\mathbf{x}_j = \mathbf{x} \Big|_{t_j}$ , в состояние  $\mathbf{x}_{j+1} = \mathbf{x} \Big|_{t_{j+1}}$ ,  $t_{j+1} > t_j > 0$ . В подавляющем большинстве случаев оператор  $M$  нелинейный и описывает реальное состояние атмосферы с ошибкой  $\varepsilon^M$ , называемой ошибкой модели. Кроме того, в реальных задачах моделирования атмосферы этот оператор является дискретным, что вносит дополнительную составляющую в ошибку модели.

Дискретная модель, описывающая переход от момента времени  $t_j$  в момент времени  $t_{j+1}$ , имеет следующий вид:

$$\begin{cases} \mathbf{x}_{j+1} = M_j(\mathbf{x}_j), \\ \mathbf{x}_{j+1}^t = \mathbf{x}_{j+1} + \varepsilon_j^M. \end{cases} \quad (1.2)$$

Кроме модели также имеются данные наблюдений в виде вектора  $\mathbf{y}_j^o \in \mathbb{Y}$  размерности  $m$  (порядка  $10^{12}$ – $10^{14}$ ), который имеет следующий вид:

$$\mathbf{y}_j^o = H_j(\mathbf{x}_j^t) + \varepsilon_j^o, \quad (1.3)$$

где  $\mathbb{Y}$  – пространство наблюдений,  $\varepsilon_j^o$  – ошибка наблюдений,  $H$  – оператор наблюдений.

Вектор  $\mathbf{y}_j^o$  состоит из наблюдений, которые расположены в точках, чаще всего отличных от узлов модельной сетки. Оператор  $H$  является в общем случае нелинейным, а также может зависеть от  $\mathbf{x}_j$ . Он задаёт отображение вектора состояния модели из модельного пространства  $\mathbb{X}$  в пространство наблюдений  $\mathbb{Y}$ .

Для интегрирования модели необходимо иметь начальные условия. Другими словами, чтобы получить вектор  $\mathbf{x}_{j+1}$ , необходимо иметь вектор  $\mathbf{x}_j$ , максимально точно отражающий реальное состояние атмосферы в момент времени  $t_j$ . Уточнённый вектор состояния в момент очередного старта модели  $t_j$  можно получить с помощью доступных на этот момент времени данных наблюдений и результатов моделирования состояния атмосферы на прошлом шаге по времени, комбинируя оба этих источника информации таким образом, чтобы получить оптимальный в каком-либо смысле вектор состояния в интересующий нас момент времени  $t_j$ .

Такой уточнённый вектор называют *анализом*. Задача усвоения данных состоит в определении этого вектора. Ниже приведём наиболее часто используемые обозначения параметров, с которыми необходимо оперировать для решения этой задачи, опуская где можно индекс по времени  $j$ :

$$\mathbf{x}^t = (x_1^t, \dots, x_n^t)^T \quad (1.4)$$

– подлежащий оценке неизвестный вектор истинного состояния атмосферы размерности  $n$ ;

$$\mathbf{x}^a = (x_1^a, \dots, x_n^a)^T \quad (1.5)$$

– искомый вектор объективного анализа размерности  $n$ , является решением задачи нахождения оптимальной оценки неизвестного вектора  $\mathbf{x}^t$ ;

$$\mathbf{x}^b = (x_1^b, \dots, x_n^b)^T \quad (1.6)$$

– вектор первого приближения, имеет размерность  $n$ , является, как правило, краткосрочным прогнозом атмосферной модели на той же сетке, в узлах которой требуется получить анализ, т.е.  $\mathbf{x}_j^b = M_{j-1}(\mathbf{x}_{j-1}^a)$ ;

$$\mathbf{y}^o = (y_1^o, \dots, y_m^o)^T \quad (1.7)$$

– вектор наблюдений размерности  $m$ ;

Также вводятся следующие обозначения для ошибок:

$\varepsilon^a = \mathbf{x}^a - \mathbf{x}^t$  – ошибка анализа;

$\varepsilon^o = \mathbf{y}^o - H(\mathbf{x}^t)$  – ошибка наблюдений;

$\varepsilon^b = \mathbf{x}^b - \mathbf{x}^t$  – ошибка первого приближения.

Важными являются следующие ограничения, накладываемые на описанные выше ошибки:

1. Все ошибки несмещённые и имеют Гауссову плотность распределения:

$$E[\varepsilon^a] = E[\varepsilon^o] = E[\varepsilon^b] = 0. \quad (1.8)$$

2. Ошибки первого приближения и ошибки наблюдений некоррелированы:

$$\text{cov}(\varepsilon^b, \varepsilon^o) = E[\varepsilon^b (\varepsilon^o)^T] = E[\varepsilon^o (\varepsilon^b)^T] = 0. \quad (1.9)$$

Таким образом, ошибки наблюдений и ошибки первого приближения при сделанных допущениях могут быть полностью описаны своими ковариационными матрицами:

$\mathbf{R} = E[\varepsilon^o (\varepsilon^o)^T]$  – матрица ковариаций ошибок наблюдений;

$\mathbf{P}^b = E[\varepsilon^b (\varepsilon^b)^T]$  – матрица ковариаций ошибок первого приближения.

3. Оператор наблюдений  $H$  полагается линейным, либо линеаризуем:

$$H(\mathbf{x} + \delta\mathbf{x}) - H(\mathbf{x}) \approx \mathbf{H}(\delta\mathbf{x}), \quad (1.10)$$

где  $\delta\mathbf{x}$  – малое отклонение от  $\mathbf{x}$ ,  $\mathbf{H}$  – матрица  $m \times n$ .

## 1.2 Методы решения задачи усвоения

Точное решение описанной выше задачи возможно лишь в ограниченном числе случаев, которые не могут быть применены при определении начальных условий для численных моделей атмосферы. Одним из примером таких случаев является линейный оператор наблюдений (1.10). В реальных задачах наблюдения часто представляют собой величины, которые существенно отличаются от модельных переменных, и являются элементами других функциональных пространств. Поэтому условие линейности оператора наблюдений может удовлетворяться в основном в тестовых задачах и/или для некоторых традиционных видов наблюдений. Точное вычисление ковариационных матриц  $\mathbf{P}^b$  и  $\mathbf{R}$  для реальных величин также невозможно, т. к. размеры матриц составляют величины порядков  $10^{10} \times 10^{10}$  и  $10^{14} \times 10^{14}$  соответственно.

Кроме того, в реальных задачах геофизики не выполняются условия несмещённости и гауссовости распределения ошибок первого приближения и ошибок наблюдений ввиду нелинейности модели и оператора наблюдений, технологии получения наблюдений для некоторых видов и других причин [85]. Возможные методы преодоления указанных трудностей представлены в работах [33; 57; 85] и др.

Эти и другие факты обуславливают необходимость применения статистического аппарата, который сам по себе исключает поиск точного решения. Применяемые допущения и/или предположения характеризуют методы усвоения данных. Далее будут рассмотрены некоторые из них, которые наиболее часто применяются в настоящее время в задачах моделирования атмосферы. Более подробное описание приведённых ниже методов можно найти в [78].

### 1.2.1 Оптимальная интерполяция

Одним из первых *обоснованных* методов получения объективного анализа, применяемым для усвоения данных в моделях численного прогноза погоды, стал **метод оптимальной интерполяции**. Уравнения многоэлементной оптимальной интерполяции были получены и применены Л. С. Гандиным в 1963 году [5]. Активно этот метод использовался в моделях ведущих прогностических центров с 80-х годов прошлого столетия. В настоящее время модификации этого метода используются в исследованиях в области моделирования океана [81]. Особенностью метода является присваивание наблюдениям весов, которые связывают с ошибками наблюдений. Также в уравнениях оптимальной интерполяции полагается, что оператор наблюдений линейный или может быть линеаризован (1.10).

Решение задачи вектор  $x^a$  ищется в виде

$$\mathbf{x}^a = \mathbf{x}^b + \mathbf{W} [\mathbf{y}^o - H(\mathbf{x}^b)], \quad (1.11)$$

где  $\mathbf{W}$  – весовая матрица размерности  $n \times m$ , коэффициенты которой подлежат определению.

Матрица  $\mathbf{W}$  определяется как оптимальная матрица многомерной регрессионной линейной модели. Если существуют два несмещённых случайных вектора  $\xi = (\xi_1, \dots, \xi_n)^T$  и  $\eta = (\eta_1, \dots, \eta_l)^T$  связанные линейным регрессионным соотношением  $\xi = \mathbf{W}\eta - \varepsilon_\xi$ , то оценка по методу наименьших квадратов (МНК-оценка)  $\mathbf{W}$

$$\mathbf{W} = E [\xi \eta^T] (E [\eta \eta^T])^{-1}. \quad (1.12)$$

Из (1.11) имеем

$$\mathbf{x}^t = \mathbf{x}^b + \mathbf{W} (\mathbf{y}^o - H(\mathbf{x}^b)) - \varepsilon^a. \quad (1.13)$$

Тогда, полагая  $\xi = \mathbf{x}^t - \mathbf{x}^b$ ,  $\eta = \mathbf{y}^o - H(\mathbf{x}^t) = \varepsilon^o - \mathbf{H}\varepsilon^b$ , получим оценку для весовой матрицы

$$\mathbf{W} = E \left[ -\varepsilon^b (\varepsilon^o - \mathbf{H}\varepsilon^b)^T \right] \left( E \left[ (\varepsilon^o - \mathbf{H}\varepsilon^b) (\varepsilon^o - \mathbf{H}\varepsilon^b)^T \right] \right)^{-1},$$

откуда, раскрывая скобки и учитывая предположения об ошибках (1.8) и (1.9), окончательно получим следующее выражение для весовой матрицы, которая применяется в методе оптимальной интерполяции

$$\mathbf{W} = \mathbf{P}^b \mathbf{H}^T (\mathbf{R} + \mathbf{H} \mathbf{P}^b \mathbf{H}^T)^{-1}. \quad (1.14)$$

Для матрицы ковариаций ошибок анализа получим следующую оценку

$$\mathbf{P}^a = E \left[ \varepsilon^a (\varepsilon^a)^T \right] = (\mathbf{I} - \mathbf{W} \mathbf{H}) \mathbf{P}^b, \quad (1.15)$$

где  $\mathbf{I}$  – единичная матрица размерности  $n \times n$ .

Из (1.14) и (1.15) вытекают следующие свойства.

1. Анализ есть сумма первого приближения (прогноза) и взвешенного оптимальной весовой матрицей инкремента наблюдений  $\mathbf{d} = \mathbf{y}^o - H(\mathbf{x}^b)$ .

2. Чем больше ошибка первого приближения, тем сильнее необходимо для получения анализа корректировать первое приближение с помощью наблюдений. В идеальной модели (когда отсутствует ошибка моделирования, а значит и ошибка первого приближения) необходимость в наблюдениях отсутствует.

3. Используя (1.14) и (1.15), можно показать, что

$$(\mathbf{P}^a)^{-1} = (\mathbf{P}^b)^{-1} + \mathbf{H}^T \mathbf{R}^{-1} \mathbf{H}, \quad (1.16)$$

т.е. точность анализа есть сумма точностей первого приближения и спроецированных в модельное пространство наблюдений.

4. При возрастании ошибки первого приближения анализ сильнее смещается в сторону интерполированных на модельную сетку наблюдений. И наоборот, чем меньше ошибка первого приближения, тем сильнее анализ «доверяет» смоделированному полю. Так, в случае  $\mathbf{P}^b = 0$  вообще отпадает всякая необходимость в использовании наблюдений.

Существенными недостатками метода являются:

1. Необходимость обращения матрицы большой размерности, что обуславливает введение ограничений на используемое количество наблюдений для каждой точки сетки. В свою очередь, это вызывает необходимость разработки дополнительных алгоритмов для отбора наблюдений. Кроме того, использование разных наборов данных наблюдений для разных точек сеток является причиной появления дополнительного шума в полях анализа.

2. Невозможность усвоения информации данных некоторых видов наблюдений спутников и радаров, т. к. эту информацию трудно представить в виде метеорологических параметров без значительной потери точности.

3. Существенное ограничение в виде необходимости линеаризации оператора наблюдений. В реальных задачах этот оператор нелинейный, т. к. наблюдения не располагаются в точках сетки и чаще всего измеряют не модельные переменные, а величины, связанные нелинейными соотношениями с ними.

### 1.2.2 Вариационные методы усвоения данных

В ходе дальнейших исследований в области методов решения задач усвоения данных в [15] была предложена идея минимизации стоимостного функционала, который связан с данными наблюдений и минимизируется на траекториях (решениях) рассматриваемой модели. Фактически этот функционал определяет меру отклонения данных наблюдений от полей первого приближения. Таким

образом, задача усвоения данных была сформулирована как задача оптимального управления. Эта идея оказалась очень продуктивной и легла в основу разработки т.н. **методов вариационного усвоения**. Теоретические основы этих методов могут быть найдены, например, в работах [16; 55; 86; 91].

Для **трёхмерного вариационного усвоения** минимизируемый функционал имеет следующий вид:

$$J(\mathbf{x}) = \frac{1}{2} (\mathbf{y}^o - H(\mathbf{x}))^T \mathbf{R}^{-1} (\mathbf{y}^o - H(\mathbf{x})) + \frac{1}{2} (\mathbf{x}^b - \mathbf{x})^T (\mathbf{P}^b)^{-1} (\mathbf{x}^b - \mathbf{x}). \quad (1.17)$$

К выражению (1.17) можно прийти различными путями. Один из способов — использование теоремы Байеса для получения апостериорной плотности вероятности вектора состояния системы при имеющихся данных наблюдений  $\mathbf{y}^o$  и предварительной (априорной) оценке, в качестве которой используется вектор первого приближения  $\mathbf{x}^b$ . Решением задачи является значение вектора  $\mathbf{x}^a$ , при котором достигается максимум апостериорной плотности вероятности.

По теореме Байеса

$$p(\mathbf{x}|\mathbf{y}^o) = \frac{p(\mathbf{y}^o|\mathbf{x})p(\mathbf{x})}{p(\mathbf{y}^o)}. \quad (1.18)$$

Априорной оценкой вектора состояния  $\mathbf{x}$  полагается вектор  $\mathbf{x}^b$ . Отсюда, из предположения о распределении ошибки первого приближения  $\varepsilon^b$ , имеем  $p(\mathbf{x}) \sim N(\mathbf{x}^b, \mathbf{P}^b)$ . Условная плотность вероятности наблюдений  $\mathbf{y}^o$  при реализации состояния  $\mathbf{x}$  имеет вид  $p(\mathbf{y}^o|\mathbf{x}) \sim N(H(\mathbf{x}), \mathbf{P}^b)$ . Входящая в знаменатель плотность  $p(\mathbf{y}^o)$  не зависит от переменной  $\mathbf{x}$ , по которой проводится максимизация  $p(\mathbf{x}|\mathbf{y}^o)$ , поэтому может быть включена в константу  $C$ . Тогда, после подстановки выражений для гауссовых плотностей вероятностей, указанных выше, получим следующее выражение для оценки плотности вероятности  $\mathbf{x}$  по данным наблюдений  $\mathbf{y}^o$ :

$$p(\mathbf{x}|\mathbf{y}^o) \sim C \exp \left[ -\frac{1}{2} (\mathbf{y}^o - H(\mathbf{x}))^T \mathbf{R}^{-1} (\mathbf{y}^o - H(\mathbf{x})) \right] \times \\ \times \exp \left[ -\frac{1}{2} (\mathbf{x}^b - \mathbf{x})^T (\mathbf{P}^b)^{-1} (\mathbf{x}^b - \mathbf{x}) \right] = C \exp(-J(\mathbf{x})). \quad (1.19)$$

Максимум функции (1.19) будет доставлять минимум (1.17), который достигается в точке  $\mathbf{x}^a$ , где градиент  $\nabla J$  равен нулю. Вторая производная (1.17) по

$\mathbf{x}$  приводит к квадратной матрице Гессе. Из функционального анализа известно, что экстремум скалярной функции будет минимумом, если матрица Гессе положительно определена. Постановка задачи предполагает выполнение этого условия, т. к. ковариационные матрицы ошибок наблюдений  $\mathbf{R}$  и ошибок первого приближения  $\mathbf{P}^b$  являются положительно определёнными.

Если выполняется условие (1.10), задача минимизации функционала (1.17) имеет единственное решение, которое совпадает с решением задачи при применении метода оптимальной интерполяции. Это решение легко получить аналитически [78, с. 171]:

$$\mathbf{x}^a = \mathbf{x}^b + \left( (\mathbf{P}^b)^{-1} + \mathbf{H}^T \mathbf{R}^{-1} \mathbf{H} \right)^{-1} \mathbf{H}^T \mathbf{R}^{-1} (\mathbf{y}^o - H(\mathbf{x}^b)) = \mathbf{x}^b + \mathbf{W} [\mathbf{y}^o - H(\mathbf{x}^b)]. \quad (1.20)$$

Следует отметить, что и в методе оптимальной интерполяции, и в вариационном усвоении применяется постоянная, т. е. не зависящая от шага по времени, матрица  $\mathbf{P}^b$ .

Определение минимума (1.17) является сложной задачей ввиду многих причин, в т. ч. из-за большой размерности и использования более близкого к реальности нелинейного оператора наблюдений  $H$ . В книге [53] изложены различные подходы к получению решения этой задачи.

Впервые в оперативной практике алгоритм вариационного усвоения 3D-VAR был реализован в Национальном центре прогнозирования окружающей среды (NCEP) в 1992 году [103]. Далее последовательно этот метод был внедрён в оперативные схемы усвоения в Европейском центре среднесрочных прогнозов погоды (ECMWF), Канадском метеорологическом центре, Английской метеослужбе (MetOffice) и в других центрах. В Гидрометцентре России система усвоения на основе 3D-VAR применяется по сей день [27].

**Четырёхмерное вариационное усвоение** данных наблюдений (**4D-VAR**) стало логическим продолжением развития вариационного подхода. При использовании этого метода усваиваются наблюдения, попавшие в некоторый временной интервал, называемый окном усвоения. Функционал минимизирует отклонение модельных прогнозов, стартовавших с искомого анализа, от наблюдений в моменты времени, принадлежащие окну усвоения:

$$\begin{aligned}
J(\mathbf{x}_0) = & \frac{1}{2} (\mathbf{x}_0^b - \mathbf{x}_0)^T (\mathbf{P}_0^b)^{-1} (\mathbf{x}_0^b - \mathbf{x}_0) + \\
& + \frac{1}{2} \sum_{j=0}^N \left[ (\mathbf{y}_j^o - H_j(\mathbf{x}_j))^T \mathbf{R}_j^{-1} (\mathbf{y}_j^o - H_j(\mathbf{x}_j)) \right].
\end{aligned} \tag{1.21}$$

Здесь  $j$  – индекс по времени,  $N$  – количество временных моментов, попавших в окно усвоения,  $\mathbf{x}_j = \mathbf{x}|_{t_j} = M_{t_0 \rightarrow t_j}(\mathbf{x}_0)$  – спрогнозированный моделью вектор состояния на момент времени  $t_j$  при начальном значении  $\mathbf{x}_0$  в момент времени  $t_0$ ,  $M_{t_0 \rightarrow t_j}$  – оператор модели, с помощью которого получают прогноз  $\mathbf{x}_j$  из начального состояния  $\mathbf{x}_0$ . Минимизация производится по переменной  $\mathbf{x}_0$ , от которой зависят в том числе и величины  $\mathbf{x}_j$  в различные моменты времени внутри окна усвоения.

Численные эксперименты по сравнению алгоритмов 4D-VAR с несколькими окнами усвоения данных (6, 12 и 24 часа) и 3D-VAR показали [107] заметное преимущество метода 4D-VAR с окнами усвоения 6 и 12 часов. Одно из первых описаний практического применения схемы 4D-VAR приводится в [51] (ЕЦСПП в 1997 году).

### 1.2.3 Детерминированные и стохастические ансамблевые фильтры Калмана

Методы решения задачи усвоения, основанные на алгоритме фильтра Калмана, реализуют последовательную схему, в ходе которой на каждом шаге усвоения матрица ковариаций ошибок прогнозов обновляется с помощью прогностической модели. В числе первых данный подход был применён в [77] для оценки вектора состояния навигационной системы, в которой использовались неполные и неточные измерения, подверженные влиянию шума. В настоящее время различные модификации фильтра Калмана активно и успешно применяются во многих областях науки и техники.

Применительно к усвоению метеорологических данных, решение задачи ищется с учётом наблюдений в момент анализа, прогноза на интервале между

предыдущим и текущим сроком анализа и ковариаций ошибок полей первого приближения. Уравнения прогностической модели используются не только для получения прогноза среднего состояния (первого приближения), но и для прогноза ковариаций ошибок анализа, который учитывает ковариации ошибок самой прогностической модели. По сравнению с 4D-VAR, в этом случае появляется явная возможность учесть ошибки прогностической модели. Кроме того, применение фильтров Калмана не накладывает ограничений на интервал времени окна усвоения.

Классический алгоритм формулируется для линейных динамических систем, в которых эволюция вектора состояния описывается следующим образом ( $\mathbf{M}_{t_{j-1} \rightarrow t_j}$  – линейный оператор):

$$\mathbf{x}_j = \mathbf{M}_{t_{j-1} \rightarrow t_j} \mathbf{x}_{j-1} \quad (1.22)$$

Оператор наблюдений также полагается линейным, т.е.  $H_j(\mathbf{x}) \equiv \mathbf{H}_j \mathbf{x}$ . Ошибка прогностической модели  $\varepsilon_j^M$  полагается несмещённой, имеет известную матрицу ковариаций  $\mathbf{Q}_j$  и не зависит от ошибки анализа  $\varepsilon^a$ :

$$\varepsilon_j^M = \mathbf{M}_{t_{j-1} \rightarrow t_j} \mathbf{x}_{j-1}^t - \mathbf{x}_j^t, \quad E[\varepsilon^M] = 0, \quad E[\varepsilon_j^M (\varepsilon_j^M)^T] = \mathbf{Q}_j, \quad E[\varepsilon_j^M (\varepsilon_j^a)^T] = 0. \quad (1.23)$$

Далее проводят те же рассуждения, что и в методе оптимальной интерполяции (см. п. 1.2.1), полагая  $\mathbf{x}_j = \mathbf{x}_j^b$  и  $\mathbf{x}_{j-1} = \mathbf{x}_{j-1}^a$  в формуле (1.22). Ключевым отличием данного алгоритма от метода оптимальной интерполяции является обновление матрицы  $\mathbf{P}_j^b$  на каждом временном шаге, в то время как матрица ковариаций ошибок первого приближения при оптимальной интерполяции (как и в вариационных методах) полагается постоянной. Остальные входящие в уравнения данного метода матрицы и вектора также меняются на каждом шаге.

Таким образом, рекуррентный алгоритм нахождения оптимальной МНК-оценки неизвестного в момент времени  $t_j$  вектора состояния системы (1.2) с помощью Калмановской фильтрации при известных величинах  $\mathbf{x}_{j-1}^a, \mathbf{P}_{j-1}^a, \mathbf{y}_j^o, \mathbf{H}_j, \mathbf{R}_j, \mathbf{M}_{t_{j-1} \rightarrow t_j}, \mathbf{Q}_j$  выглядит следующим образом:

1.  $\mathbf{x}_j^b = \mathbf{M}_{t_{j-1} \rightarrow t_j} \mathbf{x}_{j-1}^a$  – вычисление вектора первого приближения, т.е. прогноз состояния атмосферы на текущий момент, используя анализ на предыдущем шаге.

2.  $\mathbf{P}_j^b = \mathbf{M}_{t_{j-1} \rightarrow t_j} \mathbf{P}_{j-1}^a \mathbf{M}_{t_{j-1} \rightarrow t_j}^T + \mathbf{Q}_j$  – вычисление матрицы ковариаций ошибок первого приближения, используя матрицу линейной модели и матрицу ковариаций её ошибок.

3.  $\mathbf{K}_j = \mathbf{P}_j^b \mathbf{H}_j^T (\mathbf{R}_j + \mathbf{H}_j \mathbf{P}_j^b \mathbf{H}_j^T)^{-1}$  – вычисление весовой матрицы.

4.  $\mathbf{x}_j^a = \mathbf{x}_j^b + \mathbf{K}_j [\mathbf{y}_j^o - \mathbf{H}_j \mathbf{x}_j^b]$  – вычисление вектора анализа.

5.  $\mathbf{P}_j^a = (\mathbf{I} - \mathbf{K}_j \mathbf{H}_j) \mathbf{P}_j^b$  – вычисление матрицы ковариаций ошибок анализа, необходимой на этапе 2 во время следующего шага усвоения.

В реальных задачах этот алгоритм в приведённом выше виде не применим, т. к. условие линейности операторов модели  $M_{t_{j-1} \rightarrow t_j}$  и наблюдений  $H_j$  не выполняются. Для обхода указанного ограничения в некоторых случаях могут быть использованы линеаризации обоих операторов:

$$H_j(\mathbf{x}) \approx \mathbf{H}_j \mathbf{x} + \delta \mathbf{x}, \quad M_{t_{j-1} \rightarrow t_j}(\mathbf{x}) \approx \mathbf{M}_{t_{j-1} \rightarrow t_j} \mathbf{x} + \delta \mathbf{x}$$

Тогда в ходе усвоения с использованием алгоритма фильтра Калмана при вычислении матриц на шагах 2, 3 и 5 приведённого выше алгоритма вместо линейных операторов используются их линеаризации, а в шагах 1 и 4 используются сами нелинейные операторы. Такой переход является основой **расширенного фильтра Калмана** [57]. Применение данного подхода в реальных задачах затруднено из-за больших размерностей линеаризованных операторов модели и наблюдений и необходимости вследствие этого использования больших вычислительных ресурсов.

Менее вычислительно затратным стал подход, применяемый в группе методов, которые получили название **ансамблевые фильтры Калмана** [37; 58; 66; 70–72]. В схемах усвоения данных, основанных на таком подходе, ансамбль полей первого приближения  $\{\mathbf{x}^{b(i)}, i = \overline{1, k}\}$  (где  $k$  – размерность ансамбля, на практике используются ансамбли с количеством участников порядка нескольких десятков или сотен) генерируется из ансамбля начальных условий  $\{\mathbf{x}^{a(i)}, i = \overline{1, k}\}$ , распределенных согласно результатам предыдущего анализа.

Матрица ковариаций ошибок первого приближения  $\mathbf{P}_j^b$  при этом вычисляется, как статистика ансамбля (здесь и далее индекс  $j$  опущен, если все величины в выражении относятся к одному и тому же моменту времени  $t_j$ ):

$$\mathbf{P}^b = (k - 1)^{-1} \mathbf{X}^b (\mathbf{X}^b)^T, \quad (1.24)$$

где  $\mathbf{X}^b = \{\mathbf{x}^{b(i)} - \bar{\mathbf{x}}^b, i = \overline{1, k}\}$  – матрица (размерностью  $n \times k$ ) отклонений от среднего,  $\bar{\mathbf{x}}^b$  – среднее по ансамблю.

Преимущество данного подхода заключается в отсутствии необходимости трудоёмкой разработки и поддержки линейно-касательной прогностической модели. Кроме того, использование ансамбля для аппроксимации матрицы  $\mathbf{P}^b$  позволяет избегать необходимости явного включения в расчёты ошибки модели  $\mathbf{Q}$ , а также проводить распараллеливание вычислений на многопроцессорной вычислительной технике.

Основное различие между существующими схемами, основанными на ансамблевом подходе, состоит в генерации ансамбля анализов. Первое семейство схем основано на возмущении наблюдений методом Монте-Карло [58; 71]. Этот подход называют **стохастическим**. В стохастических разновидностях фильтра ансамбль анализов получается путём усвоения различных наборов данных наблюдений для каждого отдельного члена ансамбля первых приближений. Ансамбль наблюдений получают, добавляя к реальным данным наблюдений  $\mathbf{y}^o$  случайный шум  $\varepsilon^{o(i)}$ , который генерируются согласно ковариационной матрице ошибок наблюдений:

$$\{\mathbf{y}^{o(i)} = \mathbf{y}^o + \varepsilon^{o(i)}, i = \overline{1, k}\}, \quad \varepsilon^{o(i)} \sim N(0, \mathbf{R}). \quad (1.25)$$

Другой подход применяется в **детерминированных** ансамблевых фильтрах Калмана. Имеющиеся наблюдения используются в формуле п. 4 приведённого выше алгоритма для коррекции среднего по ансамблю первого приближения  $\bar{\mathbf{x}}^b$ . В этом случае получают единственный вектор анализа  $\mathbf{x}^a = \bar{\mathbf{x}}^b + \mathbf{K} [\mathbf{y}^o - \mathbf{H}\bar{\mathbf{x}}^b]$ , из которого в последствии необходимо сформировать ансамбль анализов для вычисления ансамбля первых приближений в следующем цикле усвоения.

Для этих целей наиболее часто применяется стратегия, принятая в ансамблевых фильтрах **квадратного корня**. Ансамбль анализов генерируется таким образом, чтобы из него получалась матрица ковариаций первого приближения в виде  $\mathbf{P}^a = (\mathbf{I} - \mathbf{K}\mathbf{H}) \mathbf{P}^b$  как и в первоначальном алгоритме фильтра Калмана. Такую матрицу можно разложить аналогично (1.24):

$$\mathbf{P}^a = (k - 1)^{-1} \mathbf{X}^a (\mathbf{X}^a)^T, \quad (1.26)$$

где  $\mathbf{X}^a$  – матрица размерностью  $n \times k$ , каждый  $i$ -ый столбец которой  $\mathbf{X}^{a(i)}$ , аналогично (1.24), может трактоваться, как вектор отклонения  $i$ -го анализа  $\mathbf{x}^{a(i)}$  от среднего по ансамблю  $\mathbf{x}^a$ :

$$\mathbf{X}^{a(i)} = \mathbf{x}^{a(i)} - \mathbf{x}^a, i = \overline{1, k}. \quad (1.27)$$

Тогда, зная  $\mathbf{P}^a$ , можно вычислить  $\mathbf{X}^a$ , используя формулу:

$$\mathbf{X}^a = \sqrt{k-1} (\mathbf{P}^a)^{\frac{1}{2}} = \sqrt{k-1} [(\mathbf{I} - \mathbf{K}\mathbf{H}) \mathbf{P}^b]^{\frac{1}{2}}. \quad (1.28)$$

В результате каждый  $i$ -ый член ансамбля анализов вычисляется как

$$\mathbf{x}^{a(i)} = \mathbf{x}^a + \mathbf{X}^{a(i)}, i = \overline{1, k}. \quad (1.29)$$

Теоретические обоснования описанной стратегии приведены в [125]. Применение алгоритма, использующего ансамблевую калмановскую фильтрацию квадратного корня, реализовано в [32; 37; 74; 75; 100; 137; 144] и во многих других работах.

Сравнение реализованных в квазиоперативном режиме схем усвоения реальных наблюдений показывают, что ансамблевые фильтры дают лучшие результаты, чем схемы трёхмерного вариационного усвоения, при этом хуже, чем схемы четырёхмерного вариационного усвоения [72]. В [138] установлено, что применение ансамблевого фильтра квадратного корня для глобальной модели среднесрочных прогнозов модели NCEP для усвоения всех доступных оперативных наблюдений в атмосфере (исключая спутниковые данные) позволяет получить лучшие прогнозы, чем использование трёхмерной вариационной оперативной системы на том же наборе данных наблюдений.

Основные преимущества ансамблевых фильтров видятся в следующем:

1. Более простая реализация (ввиду отсутствия необходимости в линейно-касательной модели и сложной процедуры минимизации функционала).
2. Независимость от прогностической модели.
3. Генерация ансамбля начальных условий для ансамблевых систем прогноза погоды.

В качестве основного недостатка приводится существенно меньшая размерность ансамбля в сравнении с размерностью вектора состояния, что ска-

зывается на недооценке матрицы ковариаций ошибок первого приближения. Преодолеть этот недостаток помогает применяемая в [75; 100] локализация наблюдений вокруг точек модельной сетки.

#### 1.2.4 Локальный ансамблевый фильтр Калмана с преобразованием ансамбля

Применяемая в данной работе система усвоения [116] основана на алгоритме, который носит название *Локальный ансамблевый фильтр Калмана с преобразованием ансамбля* (**Local Ensemble Transform Kalman Filter, LETKF**).

Алгоритм LETKF впервые был описан в [75]. Он совместил в себе идеи, которые стали основой схем [37] (ETKF) и [100] (LEKF). Использование схемы LETKF для усвоения реальных данных наблюдений (в т. ч. спутниковых) приводится в [9; 19; 34; 40; 93; 94; 116; 121; 138]. Результаты сравнения алгоритма LETKF с вариационными схемами опубликованы в [79; 95]. Проведённые сравнения на различных моделях при усвоении искусственно смоделированных наблюдений свидетельствуют о том, что схема LETKF позволяет получить более точные анализы, чем при использовании 3D-VAR, и немного менее точные, чем при использовании схемы 4D-VAR.

С февраля 2012 года основанная на алгоритме LETKF система усвоения данных в оперативном режиме производит начальные данные для региональной модели COSMO-ME в Итальянском Национальном центре метеорологии и климатологи (CNMCA, [39]). В модифицированном виде схема LETKF применяется в блоках усвоения систем ансамблевого прогноза в Японской (JMA) и Германской (DWD) метеослужбах. Активная разработка систем усвоения, основанных на алгоритме LETKF, проводится в Бразильской метеослужбе (CPTEC, [92]) и Гидрометцентре России ([9; 10; 19; 29; 116]).

Для вывода формул алгоритма будем предполагать случай, когда все наблюдения доступны в момент анализа  $t_n$ . Следуя логике ансамблевого подхода в Калмановской фильтрации, будем полагать, что матрица ковариаций ошибок

первого приближения аппроксимируется в виде (1.24), т.е.

$$\mathbf{P}^b = (k - 1)^{-1} \mathbf{X}^b (\mathbf{X}^b)^T,$$

где  $\mathbf{X}^b = \{\mathbf{x}^{b(i)} - \bar{\mathbf{x}}^b, i = \overline{1, k}\}$  – матрица (размерностью  $n \times k$ ) отклонений от среднего,  $\bar{\mathbf{x}}^b$  – среднее по ансамблю.

Далее предположим, что существует такой вектор  $\mathbf{x} \in \mathbb{X}$ , который может быть представлен в виде

$$\mathbf{x} = \bar{\mathbf{x}}^b + \mathbf{X}^b \mathbf{w}, \quad (1.30)$$

где  $\mathbf{w}$  – некоторый случайный вектор размерности  $k$ , имеющий Гауссово распределение  $\mathbf{w} \sim N(\mathbf{0}, (k - 1)^{-1} \mathbf{I}_k)$ , где  $\mathbf{I}_k$  – единичная матрица размерности  $k$ . Матрица  $\mathbf{X}^b$  является линейным оператором проецирования из пространства ансамбля в модельное пространство.

Используя (1.30) как ограничение в задаче минимизации (1.17), после преобразования получим модифицированный функционал, зависящий от вектора  $\mathbf{w}$ :

$$\tilde{J}(\mathbf{w}) = (k - 1) \mathbf{w}^T \mathbf{w} + [\mathbf{y}^o - H(\bar{\mathbf{x}}^b + \mathbf{X}^b \mathbf{w})]^T \mathbf{R}^{-1} [\mathbf{y}^o - H(\bar{\mathbf{x}}^b + \mathbf{X}^b \mathbf{w})]. \quad (1.31)$$

Вычислим далее ансамбль проекций первых приближений в пространстве наблюдений:

$$\{\mathbf{y}^{b(i)} = H(\mathbf{x}^{b(i)}), i = \overline{1, k}\}, \quad (1.32)$$

а также среднее по ансамблю и матрицу отклонений от среднего, аналогично  $\mathbf{X}^b$ :

$$\bar{\mathbf{y}}^b = k^{-1} \sum_{i=1}^k \mathbf{y}^{b(i)}, \mathbf{Y}^b = \{\mathbf{y}^{b(i)} - \bar{\mathbf{y}}^b, i = \overline{1, k}\}. \quad (1.33)$$

Если существует линеаризация оператора  $H$ , то выполняется приближённое равенство:

$$H(\bar{\mathbf{x}}^b + \mathbf{X}^b \mathbf{w}) \approx \bar{\mathbf{y}}^b + \mathbf{Y}^b \mathbf{w}, \quad (1.34)$$

используя которое для (1.31) окончательно получим функционал, который подлежит минимизации.

$$\tilde{J}^*(\mathbf{w}) = (k - 1) \mathbf{w}^T \mathbf{w} + [\mathbf{y}^o - \bar{\mathbf{y}}^b - \mathbf{Y}^b \mathbf{w}]^T \mathbf{R}^{-1} [\mathbf{y}^o - \bar{\mathbf{y}}^b - \mathbf{Y}^b \mathbf{w}]. \quad (1.35)$$

Выражение (1.35) возникает также при решении задачи усвоения, в которой необходимо найти оптимальное значение вектора  $\mathbf{w}$ , если его приближение  $\mathbf{w}^b = 0$ , матрица ковариаций первого приближения  $\tilde{\mathbf{P}}^b = (k-1)^{-1}\mathbf{I}_k$ , вектор наблюдений  $\mathbf{y}^o - \bar{\mathbf{y}}^b$ , а линейный оператор наблюдений  $\mathbf{Y}^b$ . Решение, полученное, например, с помощью метода классического фильтра Калмана, даёт следующие выражения для оптимальной оценки вектора  $\mathbf{w}$ :

$$\bar{\mathbf{w}}^a = \tilde{\mathbf{P}}^a (\mathbf{Y}^b)^T \mathbf{R}^{-1} (\mathbf{y}^o - \bar{\mathbf{y}}^b), \quad (1.36)$$

$$\tilde{\mathbf{P}}^a = \left[ (k-1)\mathbf{I}_k + (\mathbf{Y}^b)^T \mathbf{R}^{-1} \mathbf{Y}^b \right]^{-1}. \quad (1.37)$$

Тогда, переходя снова в модельное пространство  $\mathbb{X}$ , получим следующие оценки оптимального вектора  $\mathbf{x}$  и матрицы ковариаций его ошибок:

$$\bar{\mathbf{x}}^a = \bar{\mathbf{x}}^b + \mathbf{X}^b \bar{\mathbf{w}}^a, \quad (1.38)$$

$$\mathbf{P}^a = \mathbf{X}^b \tilde{\mathbf{P}}^a (\mathbf{X}^b)^T. \quad (1.39)$$

Для формирования ансамбля анализов, необходимого для получения ансамбля первых приближений на следующем шаге усвоения, воспользуемся методом, предложенным в ансамблевых фильтрах квадратного корня [125]. Полагая, что матрица  $\mathbf{P}^a$  имеет представление в виде (1.26), и используя выражение (1.39), получим:

$$\mathbf{P}^a = (k-1)^{-1} \mathbf{X}^a (\mathbf{X}^a)^T = \mathbf{X}^b \tilde{\mathbf{P}}^a (\mathbf{X}^b)^T = (k-1)^{-1} \mathbf{X}^b \mathbf{W}^a (\mathbf{X}^b \mathbf{W}^a)^T, \quad (1.40)$$

где  $\mathbf{W}^a = \left[ (k-1)\tilde{\mathbf{P}}^a \right]^{1/2}$  — матрица квадратного корня для симметричной матрицы  $(k-1)\tilde{\mathbf{P}}^a$ , которую можно найти методами линейной алгебры согласно [37]. Если  $\mathbf{D}$  — диагональная матрица, у которой на диагонали расположены собственные значения матрицы  $(k-1)\tilde{\mathbf{P}}^a$ , а  $\mathbf{Z}$  — ортогональная матрица, составленная из собственных векторов матрицы  $(k-1)\tilde{\mathbf{P}}^a$ , то

$$\mathbf{W}^a = \left[ (k-1)\tilde{\mathbf{P}}^a \right]^{1/2} = \mathbf{Z} \mathbf{D}^{1/2} \mathbf{Z}^T. \quad (1.41)$$

Тогда матрица отклонений членов ансамбля анализов от среднего может быть вычислена по формуле

$$\mathbf{X}^a = \mathbf{X}^b \mathbf{W}^a. \quad (1.42)$$

Далее определим весовую матрицу  $\mathbf{w}^a$  размерности  $k$ , каждый  $i$ -ый столбец которой состоит из полученного анализа в пространстве ансамбля, т. е. из суммы  $\bar{\mathbf{w}}^a$  и  $i$ -го столбца матрицы  $\mathbf{W}^a$ :

$$\mathbf{w}^{a\langle i \rangle} = \bar{\mathbf{w}}^a + \mathbf{W}^{a\langle i \rangle}, i = \overline{1, k}. \quad (1.43)$$

Тогда из уравнений (1.29), (1.38) и (1.42) получим следующее выражение для определения значения  $i$ -го вектора в ансамбле анализов:

$$\mathbf{x}^{a(i)} = \bar{\mathbf{x}}^b + \mathbf{X}^b \mathbf{w}^{a\langle i \rangle}, i = \overline{1, k}. \quad (1.44)$$

Заметим, однако, что ранги матриц  $\mathbf{P}^b$  (1.24) и  $\mathbf{P}^a$  (1.39) не превышают число  $k - 1$ , а значит в случае применения формул описанного алгоритма для получения вектора глобального анализа ими можно описать свойства распределений первого приближения и анализа только в подпространстве гораздо меньшей в сравнении с модельным размерности. Поэтому повышение точности априорной оценки состояния системы за счёт проведения процедуры усвоения данных наблюдений возможно только в этом подпространстве, в то время как «неучтённые» ошибки не корректируются и в дальнейшем растут на этапе прогноза (т. е. при получении первого приближения для следующего шага). Этот факт особенно важен при использовании полученного анализа в качестве начальных данных для среднесрочных прогнозов погоды, т. к. в этом случае ошибка в начальных данных возрастает многократно в ходе моделирования.

Тем не менее, если вместо глобального модельного пространства использовать некоторую достаточно маленькую локальную область, то размеры пространств ансамбля и используемой локальной области становятся сопоставимы. В работах [98; 104] было выявлено, что в случае глобальных моделей прогноза погоды, локальные регионы достаточно малого размера могут рассматриваться как нестабильные системы с низким разрешением, состояние которых обусловлено динамикой в соседних регионах. Поэтому, применяя ансамблевое усвоение для локальных областей, и «собирая» в дальнейшем полученные решения в вектор глобального анализа, можно многократно увеличить эффективный размер

пространства, в котором оценка вектора состояния системы (т. е. анализ), будет близка к оптимальной.

Кроме того, при использовании приближения (1.24) для вычисления матрицы  $\mathbf{P}^b$ , структура получившейся матрицы предполагает появление т. н. ложных корреляций между далеко отстоящими друг от друга точками модельной сетки. В результате возможна ситуация, когда значение анализа в разных точках случайным образом подвержено влиянию наблюдений, которые в реальности не должны быть учтены. В [66] показано, что относительная ошибка оценки ковариации значительно увеличивается при уменьшении настоящей ковариации или при уменьшении размера ансамбля. Так как в большинстве случаев ковариации между отдалёнными точками являются низкими, это может приводить к неверному влиянию отдалённых наблюдений. В [66] также показано, что использование оценок ковариаций с высокой относительной ошибкой может приводить к ухудшению качества фоновое состояние при использовании анализа в шаге распространения.

Однако существуют методы для борьбы с описанными выше проблемами: локализация и инфляция. Краткое описание применяемых методов локализации и инфляции приведено в разделах 1.3.5 и 1.3.6 соответственно.

Большинство реальных данных наблюдений поступает в моменты времени, отличные от времени анализа. Типичное время, которое проходит между двумя сроками усвоения, составляет 6 часов. Это время, как было замечено выше, называется окном усвоения. За прошедшее между двумя сроками усвоения время погодные условия могут значительно поменяться. Таким образом, значения наблюдений, поступающие из соседних точек в течение окна усвоения данных могут весьма различаться. Поэтому «привязка» наблюдений к конкретному моменту времени позволяет повысить точность анализа, в сравнении с подходами, в которых игнорируется или момент появления наблюдения, или оно само в случае, если далеко отстоит во времени от момента анализа.

Явный учёт момента времени внутри окна усвоения происходит в методе четырёхмерного вариационного усвоения, который считается наиболее точным алгоритмом. Максимальная точность достигается за счёт минимизации расстояния между модельной траекторией и наблюдениями в пространстве наблюдений в сочетании со статистикой ошибок модели (ковариационной матрицей ошибок первого приближения).

Несмотря на то, что представленный в этой главе алгоритм LETKF был сформулирован для наблюдений, доступных в момент времени анализа, он легко расширяется на случай наблюдений, поступающих в течение окна усвоения нерегулярно. Сутью подхода, осуществляемого в LETKF, является аппроксимация оптимального модельного состояния, которое считается анализом, линейной комбинацией априорных оценок, составляющих ансамбль первых приближений. Этот подход также можно расширить и на четырёхмерный случай, аппроксимируя оптимальную модельную траекторию линейной комбинацией априорных оценок таких траекторий, составляющих ансамбль траекторий первых приближений. Описанную схему, которая получила название **4D-LETKF**, можно достаточно легко применить в рамках представленного выше алгоритма. Более подробное описание и теоретическое обоснование приведено в [68]. В данной работе полагалось, что все наблюдения относятся к моменту анализа, поэтому описанное четырёхмерное расширение не применялось.

Основным преимуществом алгоритма LETKF является высокая степень распараллеливания по данным, которая возможна ввиду того, что данные наблюдений усваиваются локально для каждой точки модельной сетки. Это качество является определяющим при постоянно растущих объёмах данных, используемых в оперативных глобальных моделях, т. к. эффективное распараллеливание алгоритма позволяет компенсировать большой объём обрабатываемых данных большим количеством процессорных ядер и высокой производительностью вычислительной системы, которые также постоянно увеличиваются.

### 1.2.5 Гибридные схемы усвоения

Вариационные методы и фильтр Калмана дают одинаковый результат, если все ошибки гауссовы, не зависят друг от друга, имеют нулевые математические ожидания и операторы модели и наблюдений линейны (см. ограничения в п. 1.2). В реальных задачах, как уже было отмечено, эти условия не выполняются. Поэтому актуально стоит вопрос сравнения вариационного и применяющего фильтрацию подходов к решению задачи усвоения данных ([134; 142] и др.). Подробное исследование этого вопроса проведено в [80], где сделан

акцент на сравнение ансамблевых разновидностей фильтров Калмана и алгоритма 4D-VAR.

Преимущества схемы 4D-VAR авторы видят в возможности усваивать асинхронно поступающие наблюдения, а также в получении более гладких, сбалансированных полей модельных переменных. Недостатки же очевидны: необходимость в касательно-линейной модели и малая степень параллелизма алгоритма, что делает его трудно реализуемым.

Попытки объединить достоинства обоих подходов и нивелировать недостатки приводят к возникновению **гибридных схем усвоения**. Пожалуй, самый очевидный способ построения гибридной схемы – использование в вариационном усвоении аппроксимации матрицы  $\mathbf{P}^b$ , полученной из ансамбля первых приближений (1.24). В [143] гибридная схема усвоения, названная E4DVAR, строится на предположении, что входящая в уравнение (1.21) матрица ковариаций ошибок первого приближения вычисляется как

$$\mathbf{V} = \beta \mathbf{P}^b + (1 - \beta) \mathbf{V}^s, \quad (1.45)$$

где  $\mathbf{V}^s$  – классическая матрица вариационных методов, полученная путём статистической оценки,  $\mathbf{P}^b$  – вычислена из ансамбля первых приближений,  $\beta \in [0; 1]$  – вес, который подобран эмпирическим путём.

В сочетании с моделью Лоренца [90] подобный подход позволяет получать более точные анализы в сравнении с применением по отдельности четырёхмерного вариационного метода и ансамблевого фильтра. При этом минимизация функционала по-прежнему использует касательно-сопряжённую модель. Уйти от этого использования удаётся в гибридном методе, названном En4DVAR [87]. Результаты этой работы показывают, что гибридный алгоритм En4DVAR даёт результат, сравнимый по качеству с классическим 4DVAR и при этом выполняется с гораздо меньшими вычислительными затратами на минимизацию.

Существует возможность применять указанную матрицу и в схемах ансамблевой фильтрации, повышая таким образом эффективный размер матрицы ковариаций ошибок первого приближения, который в случае аппроксимации с помощью выражения (1.24) ограничен размерностью ансамбля  $k$ .

С некоторыми различными подходами к реализации гибридного усвоения можно ознакомиться в [2; 38; 67; 75; 83; 131; 133] и во многих других работах. В целом преобладающим мнением в сообществе исследователей, занимающихся

проблемами усвоения данных, является мнение о том, что гибридные схемы в будущем будут доминировать среди всех методов. Оперативная система усвоения данных в ЕЦСПП [38] уже основана на гибридном подходе, при котором перед каждым запуском вариационного усвоения 4D-VAR происходит расчёт матрицы ковариаций ошибок первого приближения на основе ошибок ансамблевого прогноза, вычисленных путём интегрирования модели, стартовавшей с начальных данных, которые в свою очередь были получены из 10 независимых запусков 4d-VAR с более грубым разрешением. В каждом из этих 10 запусков данные наблюдений возмущаются случайным образом.

Кроме комбинирования вариационного подхода и разновидностей фильтров Калмана, исследователи в последнее время стали активно изучать способы применения методов машинного обучения в усвоении данных. Пример такого применения приводится в статье [127], где для восстановления концентрации морского льда по микроволновым спутниковым измерениям применяется нейросеть, которая даёт результат, превосходящий несколько традиционных методов.

В целом же большинство из гибридных схем пока ещё недостаточно обоснованы математически и не очень глубоко исследованы на предмет сходимости к оптимальному решению задачи усвоения.

### 1.3 Использование данных наблюдений в системах усвоения

Данные метеорологических наблюдений являются очень важной частью всех без исключения систем усвоения, вокруг которой и построены все алгоритмы и методы получения оптимальной оценки состояния атмосферы в точках модельной сетки. Именно они являются наиболее объективным источником информации, на основании которого вносятся изменения в априорные, а значит изначально менее точные, оценки вектора состояния атмосферной системы.

Современные массивы данных наблюдений имеют очень большие размеры. При этом их количество продолжает расти, т. к. изобретаются новые инструменты, позволяющие получить актуальные данные о характеристиках состояния атмосферы. На сегодняшний день в каждый срок анализа с учётом данных

со спутников имеется от  $10^{12}$  до  $10^{14}$  наблюдений. Как уже было отмечено выше, большинство из них не обеспечивают измерения модельных переменных, и расположены при этом в точках, отличных от пространственных узлов модельных сеток. Некоторые виды наблюдений (например, спутниковые наблюдения AMSU) представляют собой спектральные характеристики целых атмосферных слоёв, поэтому для них не существует понятия «точка расположения». Таким образом, изучение свойств и характеристик различных видов наблюдений, с учётом технологии их получения, является сложной и важной составляющей усвоения данных.

### 1.3.1 Виды наблюдений

Данные метеорологических наблюдений в настоящее время получают, используя принципиально разные инструменты. Наблюдения за состоянием атмосферы принято разделять на контактные и дистанционные, к которым в т.ч. относятся и спутниковые.

К контактными видам наблюдений относятся:

- наблюдения с наземных станций (SYNOP) и аэродромов (METAR), а также судовые наблюдения (SHIP);
- радиозондовые наблюдения (TEMP) и шаропилотные наблюдения за ветром (PILOT);
- наблюдения с борта воздушных судов (AIREP);
- наблюдения с дрейфующих буёв (DRIBU).

Наиболее значимыми спутниковыми наблюдениями на сегодняшний день являются следующие:

- микроволновые наблюдения уходящей радиации (AMSU);
- данные зондирования в диапазоне инфракрасного излучения (IASI, AIRS);
- радиозатменные наблюдения задержки радиосигнала в атмосфере (GPS);
- скаттерометрические наблюдения рассеянной радиации (ASCAT);
- наблюдения ветра по вычислению облачных структур на спутниковых изображениях (AMV).

Далее приведено краткое описание упомянутых выше типов наблюдений.

**Наблюдения с наземных станций (SYNOP) и аэродромов (METAR), а также судовые наблюдения (SHIP)** обеспечивают измерения давления, температуры, ветра, влажности, осадков, снежного покрова, облачности и других метеорологических величин на метеорологических станциях, в метеослужбах аэропортов и аэродромов, на борту метеорологических судов. Наблюдения расположены в различных точках поверхности Земли. В Гидрометцентр России сводки с этими наблюдениями поступают 4 раза в сутки. За один сеанс доступно порядка 7600 сводок с наземных станций. Общее же количество станций (с учётом судовых наблюдений) достигает 10000.

**Радиозондовые наблюдения (TEMP)** представляют информацию о геопотенциале, температуре, горизонтальном ветре и влажности в свободной атмосфере. По состоянию на конец 2016 года вся мировая аэрологическая сеть насчитывала порядка 800 станций, с которых передаются данные. В России находится максимальное количество таких станций — 111. Для большинства станций данные доступны дважды в сутки — в 00 UTC и в 12 UTC. Некоторые российские станции передают сводки только один раз в сутки.

**Наблюдения с борта воздушных судов (AIREP)** обеспечивают измерения температуры, давления, ветра, а также влажности в свободной атмосфере. Измерения производятся на взлёте и посадке, а также во время движения по трассе полёта. У этих наблюдений нет определённых сеансов, т. е. данные поступают постоянно. За одни сутки в Гидрометцентре России доступно около 30-40 тысяч сводок, в зарубежных центрах — 250-300 тысяч.

**Наблюдения с дрейфующих буёв (DRIBU)** обеспечивают информацией о приводном давлении и ветре. Количество буёв — около 200. Сводки доступны 4 раза в сутки.

**Микроволновые наблюдения уходящей радиации (AMSU)** позволяют получить температурные профили в свободной атмосфере по уходящей микроволновой радиации. Данные имеют глобальное покрытие Земли и поступают постоянно. В течение суток доступно порядка 140 тысяч профилей. В ходе многочисленных исследований неоднократно было выявлено, что данный вид наблюдений вносит наибольший вклад в точность среди всех используемых в настоящее время видов наблюдений (например, [56]).

**Данные зондирования в диапазоне инфракрасного излучения (IASI, AIRS)** очень похожи на предыдущий тип наблюдений за исключением того, что зондирование происходит в инфракрасном диапазоне, что ограничивает наблюдаемые зоны участками атмосферы, свободными от облачности.

**Радиозатменные наблюдения задержки радиосигнала в атмосфере (GPS)** позволяют измерить коэффициент преломления воздуха (или т. н. угол рефракции) в атмосфере по задержке радиосигнала, который излучает высокоорбитальный радионавигационный спутник GPS. Приёмником сигнала, прошедшего через атмосферу, может быть как низкоорбитальный спутник, так и наземный измеритель. Угол рефракции зависит от температуры, давления и влажности, которые восстанавливаются с помощью специальной модели. Данные этого вида наблюдений имеют высокое разрешение по вертикали, но низкое по горизонтали. В сутки имеется от 2000 до 3000 вертикальных профилей коэффициента преломления.

**Скаттерометрические наблюдения рассеянной радиации (ASCAT)** позволяют получать данные о приводном ветре над обширными акваториями океанов. Скаттерометром называется активный зонд, излучающий микроволновый сигнал и принимающий его рассеянное отражение. По интенсивности отражённого сигнала определяется скорость и направление ветра над поверхностью воды. Наблюдения ASCAT обеспечивают глобальное покрытие данными поверхности океана, за исключением областей, которые покрыты льдом. В отсутствие льда для каждой точки имеется от 10 до 45 наблюдений в течение месяца.

**Наблюдения ветра по вычислению облачных структур на спутниковых изображениях (AMV)** позволяют получить данные о ветре в свободной атмосфере. Большим преимуществом этого вида наблюдений, обуславливающим необходимость его использования в системах усвоения данных, является глобальное покрытие земного шара данными AMV. В каждый срок усвоения доступно порядка 200000 наблюдений. Зачастую наблюдения AMV являются единственным источником данных о ветре в тропосфере над некоторыми акваториями океанов и в верхних широтах [60], в т. ч. арктических. Кроме того, этому виду наблюдений свойственно высокое временное разрешение, а характеристики измерений легко приводятся к модельным переменным. Наблюдениям AMV посвящён раздел 1.4 данной работы.

### 1.3.2 Оператор наблюдений

Использование оператора наблюдений  $H$  является необходимым для получения «образа» модельного состояния  $\mathbf{x} \in \mathbb{X}$  в пространстве наблюдений  $\mathbb{Y}$ . С его помощью в схемах усвоения производится оценка «близости» полученного образа к данным наблюдений  $\mathbf{y}^o$ .

При усвоении большинства контактных (точечных) наблюдений, оператор  $H$  производит интерполяцию модельных величин, совпадающих с наблюдаемыми, из точек сетки в точки расположения наблюдений. Для такого оператора довольно легко вводится линеаризация  $\mathbf{H}$ , которая представляет из себя матрицу  $m \times n$ . Во многих исследовательских задачах считается, что наблюдения расположены в точках модельной сетки, и тогда оператор наблюдений  $H$  совпадает с единичной матрицей  $\mathbf{I}$  размерности  $m \times m$ .

При усвоении спутниковых данных оператор наблюдений  $H$  представляет собой сложную модель. Как видно из описания спутниковых наблюдений в разделе 1.3.1, инструменты, установленные на спутниках, измеряют величины и характеристики, которые весьма сильно отличаются от переменных, используемых в численных моделях.

Для того, чтобы получить значения модельных переменных (например, температуры, влажности, давления и др.) из измеряемых спутниковым прибором характеристик, обычно используются т. н. температурно-влажностное зондирование атмосферы, которое базируется на интерпретации или «обращении» данных измерений интенсивности уходящего теплового излучения с помощью радиометров, «косвенных» по отношению к искомым профилям модельных величин.

Методы решения задачи температурно-влажностного зондирования атмосферы основаны на решении нелинейного переноса излучения в атмосфере. Обычно для получения профилей температуры и влажности из данных спутникового зондирования используются численные модели.

Оператор наблюдений  $H$  при этом необходим для проведения процедуры восстановления характеристик, измеряемых спутниковым прибором. Эти характеристики восстанавливаются с помощью обратной задачи температурно-влажностного зондирования атмосферы, которая также решается численно и

является существенно нелинейной. Этот факт обуславливает сложность применяемого при усвоении спутниковых данных оператора наблюдений  $H$ . Так, в оперативной системе усвоения Гидрометцентра России [27] оператор наблюдений, используемый для усвоения радиозатменных спутниковых данных GPS, представляет собой 100 тысяч строк кода, написанного на языке Фортран.

В настоящей работе используется оператор наблюдений, представляющий собой трилинейный интерполятор в трёхмерном пространстве (для переменных в свободной атмосфере) и билинейный в двумерном пространстве (для переменных на поверхности Земли). В качестве вертикальной координаты используется логарифм давления.

### 1.3.3 Ошибки наблюдений

Большинство применяемых в метеорологии методов усвоения данных построены на предположении о том, что ошибки наблюдений имеют гауссово распределение с нулевым математическим ожиданием и некоторой заданной матрицей ковариаций  $\mathbf{R}$ . Считается также, что ошибки наблюдений не зависят от ошибок первого приближения. Вместе все эти предположения приводят к оценке вектора состояния, при вычислении которой используется обратная матрица  $\mathbf{R}^{-1}$ .

В зависимости от типов усваиваемых данных и количества наблюдений для одного срока усвоения, размерность матрицы  $\mathbf{R}$  может составлять величину порядка от  $10^8 \dots 10^{10}$  (при использовании только контактных наблюдений) до  $10^{12} \dots 10^{14}$  (при использовании и контактных, и спутниковых данных). Поэтому обращение такой матрицы может быть очень дорогой операцией с точки зрения вычислительных затрат.

Для того, чтобы максимально упростить вычисление обратной матрицы, предполагают, что матрица  $\mathbf{R}$  имеет диагональный вид, т. е. не учитывают возможные автокорреляции между составляющими вектора наблюдений  $\mathbf{y}^o$ . В этом случае необходимо задать только дисперсии ошибок, которые будут располагаться на главной диагонали матрицы  $\mathbf{R}$ .

Ошибки наблюдений, в соответствии с [54], могут быть представлены в виде следующей разности:

$$\varepsilon^o = \mathbf{y}^o - H(\mathbf{x}^t) = \mathbf{y}^o - \mathbf{y}^t + \mathbf{y}^t - H(\mathbf{x}^t) = \varepsilon^i - \varepsilon^H, \quad (1.46)$$

где  $\mathbf{y}^t$  — истинное значение измеряемого параметра,  $\varepsilon^i$  — инструментальная ошибка,  $\varepsilon^H$  — составляющая ошибки, которая сложным образом зависит от

- способа измерений (контактное или интегральное измерение);
- репрезентативности измерений (значения модельных переменных в точке сетки по сути являются интегральной оценкой состояния в ячейке сетки, тогда как измеренные значения являются точечной оценкой в случае контактных наблюдений и интегральной оценкой по всему вертикальному профилю в случае многих спутниковых наблюдений);
- точности оператора наблюдений  $H$ .

Ошибки наблюдений и их свойства являются предметом многочисленных исследований. В [129] показано, что ошибка репрезентативности может влиять на возникновение корреляций в ошибках наблюдений, которые свойственны всем наиболее значимым видам спутниковых наблюдений [43; 44; 64; 120]. В исследовании [54] показано, что в случае отсутствия учёта корреляций при усвоении данных наблюдений, в ошибках которых существует автокорреляция, полученный анализ менее точен, чем в случае усвоения на таких же условиях существенно прореженного набора данных тех же наблюдений. Наиболее же точным является анализ, при получении которого использовался полный набор коррелированных наблюдений с предположением о том, что ошибки наблюдений автокоррелированы. Другими словами, если в схеме усвоения используется диагональная матрица  $\mathbf{R}$ , то для получения более точного анализа необходимо обоснованно прореживать имеющийся набор данных наблюдений с учётом радиуса корреляции в их ошибках [84]. Такое прореживание может проводиться путём присвоения некоторой пространственной ячейке т. н. «супернаблюдения», значение которого является взвешенной суммой всех входящих в рассматриваемую ячейку наблюдений [53]. При этом дисперсии (т. е. диагональные элементы матрицы  $\mathbf{R}$ ) искусственно завышают с тем, чтобы придать меньший вес наблюдениям, ковариации ошибок которых заданы некорректно [138].

Очевидное решение — использование недиагональной матрицы ковариаций ошибок наблюдений  $\mathbf{R}$ . При использовании недиагональной матрицы  $\mathbf{R}$

общего вида необходимо решать задачу обращения матрицы, что существенно увеличивает вычислительные затраты на вычисление анализа. Количество наблюдений даже в локальных областях (в случае использования локализации при вычислении анализа) при этом достаточно велико. Можно, тем не менее, попробовать подобрать вид матрицы ковариаций, которую будет относительно просто обратить. Например, в [119] исследовалось влияние коррелированных ошибок наблюдений на точность анализа. При этом рассматривались различные аппроксимации недиагональных матриц ковариаций ошибок наблюдений: диагональная аппроксимация (с увеличением дисперсий), ковариационные матрицы Маркова (процесса авторегрессии первого порядка, [140]) и разложение по собственным значениям. Ошибки усваиваемых наблюдений были сгенерированы как матрица ковариаций процесса авторегрессии первого порядка и матрица ковариаций процесса авторегрессии второго порядка. Наилучшие результаты показали ковариационные матрицы Маркова, хотя следует заметить, что этот результат зависит от того, как была выбрана «настоящая» матрица ковариаций ошибок наблюдений, по которой они были сгенерированы.

### 1.3.4 Контроль качества наблюдений

Данные метеорологических наблюдений могут быть подвержены грубым ошибкам, которые иногда возникают вследствие человеческого фактора при вводе наблюдаемых значений в отчётный бюллетень, обрывов каналов связи, сбоев в работе алгоритмов вычисления наблюдаемых значений и др. причин. Таким ошибкам подвержена очень малая часть всех имеющихся наблюдений, но они способны значительно уменьшать точность анализа (и, как следствие, прогноза) в окрестности той точки физического пространства, в которой была допущена грубая ошибка наблюдения. Для выявления грубых ошибок наблюдений используются т. н. процедуры контроля качества наблюдений.

Традиционно такие процедуры основаны на проверке правдоподобия измеренного значения. Например, если в наборе данных наблюдений появляется значение, которое никогда из физических свойств не может наблюдаться (температура выше  $0^{\circ}\text{C}$  на уровнях выше 100 гПа в пределах модельной атмо-

сферы), то такое наблюдение должно быть исключено из использующихся для усвоения данных. Для проверки правдоподобия используются также абсолютные значения отклонений наблюдаемого значения от спрогнозированного и/или климатического. Такое абсолютное значение не должно превышать некоторого установленного заранее числа, характеризующегося стандартным отклонением величины наблюдения в рассматриваемой точке. Также правдоподобие измеренного значения можно оценить с помощью проверки согласованности различных полученных величин наблюдений в одной и той же точке. Классическим примером наблюдения, не прошедшего проверку правдоподобия, является показывающее дождь на станции, над которой в момент наблюдения отсутствовали облака.

В последние годы были разработаны и внедрены сложные процедуры контроля качества наблюдений, в которых происходит не только принятие решения об использовании или отклонении наблюдений для усвоения данных, но и их коррекция на основании статистических и гидростатических проверок. В системах, комплексно реализующих такие процедуры, широко используются и вариационные принципы контроля качества наблюдений. Основы методов контроля качества наблюдений были заложены в [4; 61; 89]. Описание современных систем, используемых для контроля качества наблюдений, приводится в [31; 50; 123] и мн. др. источниках.

### 1.3.5 Локализация наблюдений

Локализация позволяет избавиться от ложных дальних корреляций и повысить «эффективный размер» ансамбля ([66; 99]). Существует два подхода для локализации ковариаций: локализация ковариаций фоновое состояние (неявная) и локализация ковариаций наблюдений (явная).

Неявный подход применяется в [66; 70; 71; 137]. Такая локализация подразумевает умножение соответствующих элементов матрицы  $\mathbf{P}^b$  на некоторые веса таким образом, чтобы наблюдения, находящиеся на некотором удалении, переставали оказывать влияние на модельное состояние. Этого можно добиться

путём поэлементного умножения матрицы ковариаций ошибок фонового состояния на матрицу локализации (корреляционную матрицу)  $\mathbf{L}^b$ :

$$\mathbf{P}^{/b} = \mathbf{P}^b \circ \mathbf{L}^b.$$

Здесь  $\circ$  обозначает поэлементное умножение (умножение Шура, умножение Адамара). Если матрица локализации является матрицей ковариаций, тогда матрица  $\mathbf{P}^b$  также является ковариационной матрицей. Коэффициенты матрицы локализации  $0 \leq L_{ij}^b \leq 1$  близки к нулю в тех точках, где требуется уменьшить ковариацию. Элементы  $\mathbf{L}^b$  могут зависеть от расстояния между наблюдениями и расчётной точкой сетки [66], или от расстояния между двумя точками сетки [99]; обычно используется корреляционная функция Гаспари-Кона с локальным носителем [62]

$$f(x) = \begin{cases} x^2 \left( -\frac{5}{3} + x \left( \frac{5}{8} + x(0,5 - 0,25) \right) \right) + 1, & 0 \leq |x| \leq 1, \\ x \left( -5 + x \left( -\frac{5}{3} + x \left( \frac{5}{8} + x(-0,5 + x/12) \right) \right) \right) + 4 - \frac{2}{3x}, & 1 < |x| \leq 2, \\ 0, & |x| > 2, \end{cases} \quad (1.47)$$

где  $x = r/c$  — аргумент функции,  $r$  — расстояние между узлом анализа и точкой наблюдения,  $c$  — радиус локализации наблюдений вокруг узла анализа.

Локализация ковариаций фонового состояния (неявная) не может быть применена к некоторым ансамблевым схемам [110; 111].

Явная локализация состоит в начальном выборе и дальнейшем использовании для получения анализа только тех наблюдений, которые попадают в область вокруг рассматриваемой точки, ограниченной заданным радиусом локализации. Этот подход применяется в данной работе, а также, например, в [32; 70; 74; 75; 100; 101] и др. Такой подход в некотором смысле похож на метод оптимальной интерполяции: анализ в каждой точке расчётной сетки получается отдельно от остальных точек, используя наблюдения, лежащие внутри некоторой области вокруг расчётной точки. Это можно представить как поэлементное умножение обратной матрицы ковариаций ошибок наблюдений на матрицу локализации  $\mathbf{L}^o$  ( $0 \leq L_{ij}^o \leq 1$ , при этом элемент матрицы локализации равен нулю, если наблюдение не является локальным:

$$(\mathbf{R}')^{-1} = (\mathbf{R}_{[g]})^{-1} \circ \mathbf{L}, \quad (1.48)$$

где  $\mathbf{R}'$  — локальная матрица ковариаций ошибок наблюдений,  $\mathbf{R}_{[g]}$  — глобальная матрица ковариаций ошибок наблюдений,  $\mathbf{L} : 0 \leq L_{ij} \leq 1$  — матрица локализации, элементы которой зависят от расстояния между точкой наблюдения и рассматриваемой точкой модельной сетки ( $L_{ij} = 0$  если расстояние превышает радиус локализации);  $\circ$  — операция поэлементного умножения. Такая «гладкая локализация» соответствует постепенному увеличению присвоенной значению наблюдения неопределённости. Не попадающие в область локализации наблюдения имеют «бесконечную» неопределённость и, следовательно, никакого влияния на анализ не оказывают. Кроме того, «гладкая локализация» способствует получению более гладких полей глобального анализа.

Этот подход применим к любой ансамблевой схеме усвоения данных.

В [110] было показано, что обе схемы локализации дают похожие результаты. Оба метода уменьшают ложные корреляции. Локализация ковариаций увеличивает ранг ковариационной матрицы ошибок фонового состояния. Локализация наблюдений использует тот факт, что в локальных областях, в которых наблюдается быстрый рост ошибки, атмосферу можно рассматривать как низкоразмерную систему [105]; также для каждого локального анализа выбирается различная комбинация членов ансамбля. В [65] также делается вывод о сравнимости двух подходов.

Стоит также отметить, что для спутниковых данных, которые представляют собой интегральные наблюдения, вертикальная локализация является сложной задачей, обуславливающей практическую невозможность использования таких наблюдений для усвоения с помощью алгоритма LETKF (см. 1.2.4).

### 1.3.6 Инфляция матрицы ковариаций ошибок первого приближения

Типичной проблемой ансамблевых фильтров Калмана является недооценка ковариаций ошибок анализа. Это связано с особенностями ансамблевого подхода: оценка матрицы ковариации ошибок проводится в пространстве суще-

ственно меньшей размерности, чем пространство модели  $\mathbb{X}$ . Недооценка неопределённости априорной оценки (т. е. фактически переоценка точности первого приближения) приводит к тому, что в схеме первое приближение «перевешивает» данные наблюдений [137]. Если несоответствие становится слишком большим в течение долгого времени, то наблюдения практически игнорируются, и апостериорная оценка (вектор анализа) с каждым шагом усвоения всё дальше отдалается от истинного значения, больше доверяя первому приближению.

Используемые подходы можно условно разделить на две группы: физические, применяемые на шаге прогноза, и статистические, применяемые на шаге анализа. На шаге прогноза возможно применение различных параметризаций неадиабатических процессов подсеточного масштаба для разных участников ансамбля [73] или различных параметров этих параметризаций. Также в некоторых реализациях применяют схему стохастического обратного рассеяния кинетической энергии (Stochastic Kinetic Energy Backscatter, SKEB) [117], стохастические параметризации процессов подсеточного масштаба [46] (т.н. «стохастическую физику», Stochastically Perturbed Parametrization Tendencies, SPPT) и другие подходы, которые изложены в [102].

Подходы к решению проблемы недооценки ковариаций ошибок анализа, применяемые на шаге анализа, основаны на искусственном увеличении ковариации или в ансамбле первых приближений (т. н. мультипликативная инфляция, искусственное увеличение элементов матрицы первого приближения  $\mathbf{P}^b$ ), или в ансамбле анализов (т. н. аддитивная инфляция, искусственное внесение случайной части к каждому члену ансамбля анализов). В каком-то смысле применение таких увеличенных ковариаций аналогично добавлению слагаемого, моделирующего ковариацию ошибок модели в уравнениях классического фильтра Калмана.

Мультипликативная инфляция может трактоваться как применение понижающего коэффициента к влиянию прошлых наблюдений на текущем шаге анализе. Фактор инфляции определяет масштаб времени, в течение которого наблюдения оказывают значительное влияние на анализ, избегая таким образом переоценки точности первого приближения. Таким образом, инфляция матрицы ковариаций ошибок первого приближения локализует анализ во времени. Этот эффект особенно желателен при наличии систематической ошибки моде-

ли, потому что в таком случае модель ограничивает распространение ошибок, накопившихся в текущем анализе от прошлых наблюдений.

При использовании мультипликативной инфляции компоненты матрицы ковариации ошибок первого приближения в пространстве ансамбля (которой, согласно (1.35), является  $(k-1)^{-1}\mathbf{I}_k$ ) умножаются на эмпирически подобранный коэффициент инфляции  $\rho > 1$  [75]. Тогда выражение (1.37) для вычисления матрицы ковариации ошибок анализа в пространстве ансамбля принимает следующий вид:

$$\tilde{\mathbf{P}}^a = \left[ \frac{(k-1)}{\rho} \mathbf{I}_k + (\mathbf{Y}^b)^T \mathbf{R}^{-1} \mathbf{Y}^b \right]^{-1}. \quad (1.49)$$

Использование аддитивной инфляции подразумевает добавление к полученным членам ансамбля анализов на каждом шаге случайного шума  $\varepsilon^{In(i)}$  с заданным гауссовым распределением. Тогда на шаге прогноза получим

$$\mathbf{x}^{b(i)} = M \left( \mathbf{x}^{a(i)} + \varepsilon^{In(i)} \right), i = \overline{1, k}. \quad (1.50)$$

## 1.4 Спутниковые наблюдения ветра AMV

В данной работе рассматривается усвоение данных спутниковых наблюдений ветра **AMV** (Atmospheric Motion Vectors). Эти наблюдения обеспечивают глобальное покрытие земного шара данными о ветре, в т.ч. и в арктических широтах, что крайне важно ввиду практически полного отсутствия других источников информации об атмосферной циркуляции в этих регионах.

В ходе различных исследований было установлено, что наблюдения AMV вносят заметный, хотя и не решающий, вклад в повышение точности анализов и численных прогнозов. Так, в статье [56] показано, что если в ходе усвоения в ЕЦСПП используются наблюдения AMV с геостационарных спутников, то точность прогноза на сутки увеличивается примерно на 7%. Наблюдения AMV с полярно-орбитальных спутников дают положительный вклад ещё в 1%. Результаты похожего исследования приводятся в статье [59] для системы усвоения

MetOffice. Общий вклад наблюдений AMV в точность составляет порядка 12% (см. рисунок 1.1).

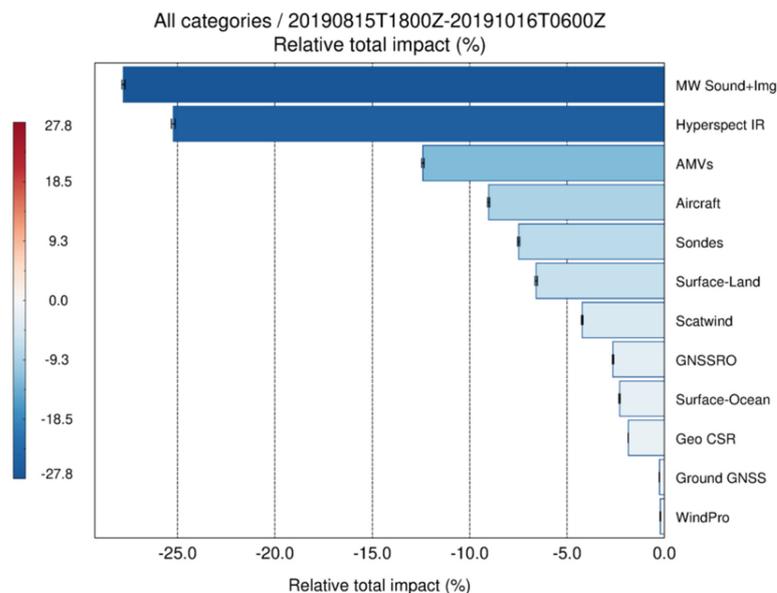


Рисунок 1.1 — Относительный общий эффект всех наблюдений на точность прогноза в Met Office на 24 часа [59]

Наблюдения AMV получают, отслеживая передвижения облаков или скоплений водяного пара по сериям спутниковых изображений, сделанных в различных диапазонах частот (видимый, инфракрасный, водяной пар). По положению одинаковых структур на изображениях, сделанных спутниками через равные промежутки времени, делают вывод о скорости и направлении ветра в точке наблюдения. Такое представление данных существенно упрощает применяемый при их усвоении оператор наблюдений  $H$ .

Наблюдения AMV производятся в таких центрах, как EUMETSAT, NOAA-NESDIS, JMA, CMA, KMA, IMD. В настоящее время доступны данные с геостационарных спутников Meteosat 8-11, Himawari 8-9, GOES 12-17, FY 2A-2H, FY 4A, INSAT 3D, 3DR, 3DS, COMS, GEO-KOMPSAT 2A, а также полярно-орбитальных спутников Metop A-C, Terra, Aqua/MODIS, NOAA 15-20 [59].

Наблюдения AMV распространяются в коде BUFR [14] и содержат следующую информацию:

- географические координаты (широта и долгота) точки наблюдения;
- скорость и направление ветра в этой точке;
- давление, на котором наблюдается ветер;
- спутник и диапазон, изображения с которых обрабатывались при получении наблюдения;

- индикатор качества наблюдения, рассчитанный производителем.

Технология получения наблюдений состоит в выборе и отслеживании некоторой цели на последовательности из трёх или, реже, двух спутниковых изображений. Схематически процесс отслеживания цели на спутниковом изображении представлен на рисунке 1.2.

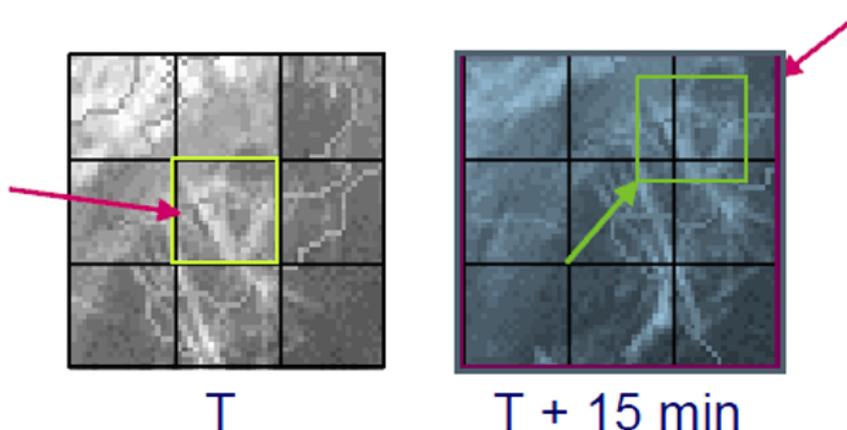


Рисунок 1.2 — Схема наблюдения скорости и направления ветра по отслеживанию положения одинаковых структур на последовательных спутниковых снимках, рисунок из [60]

Принимаются следующие предположения (и ограничения) [42]:

- выбранная цель движется с той же скоростью и в том же направлении, что локальный ветер;
- отслеженное движение представляет движение вершины облака (на этом основан выбор высоты, на которой наблюдается ветер).

Далее проводят следующие действия:

- выбор цели на начальной картинке;
- определение положения цели на последующих изображениях (ищется лучшее совпадение всех пикселей из всех возможных вариантов, используя кросс-корреляции);
- определение скорости и направления ветра;
- сопоставление высоты наблюдению, т.е. определение давления, на котором наблюдается ветер (чаще всего в давление преобразовывается температура, определённая по насыщенности цвета цели, с использованием прогностических профилей температуры);
- автоматический контроль качества (вычисление индикатора качества).

Недостатками данного вида наблюдений являются в первую очередь неопределённость вертикальной привязки в пространстве, низкое вертикальное разрешение и возможность смещений из-за изменения формы облаков при их движении. Реализованные автором методы преодоления трудностей, возникающих из-за указанных недостатков, и их теоретические обоснования приводятся в главе 3.

## 1.5 Выводы к главе 1

Приведённый в главе 1 обзор показывает, что усвоение данных является важным направлением в области численного моделирования прогноза погоды, обеспечивая модели начальными данными на регулярной сетке. Результатом работы алгоритмов усвоения данных является наилучшая возможная оценка состояния атмосферы. Источниками получения такой оценки являются краткосрочный прогноз на регулярной сетке и данные метеонаблюдений. Кроме того, необходимо знать или оценивать статистические характеристики этих источников, в первую очередь их ковариационные матрицы.

В настоящее время сформировалось два основных подхода к решению задачи усвоения данных. Вариационный подход даёт наилучшие результаты в своей четырёхмерной разновидности, но требует разработки очень сложной линейно-касательной прогностической модели. Также вариационные подходы хуже используют современные возможности массивно-параллельных вычислительных комплексов.

Различные схемы ансамблевых фильтров Калмана значительно легче реализуемы и при этом несколько превосходят результаты использования трёхмерных вариационных схем. Одним из достоинств всех схем усвоения, основанных на фильтре Калмана, является обновляемая на каждом шаге усвоения матрица ковариаций ошибок первого приближения  $\mathbf{P}^b$ , учитывающая текущее состояние атмосферы. Существенное влияние на результат усвоения оказывает локализация наблюдений. При этом для увеличения точности результатов необходимо применять инфляцию матриц ковариаций ошибок первого приближения и наблюдений.

Большую сложность представляет усвоение данных спутникового зондирования ввиду нетривиального перевода измеряемых величин в модельное пространство и обратно, а также из-за наличия автокорреляций ошибок наблюдений. При этом спутниковые наблюдения ветра AMV, которые используются в данной работе для включения в число усваиваемых в разработанной системе усвоения данных, локализованы в физическом пространстве, поэтому легко могут быть использованы в алгоритме LETKF.

AMV являются одним из видов спутниковых наблюдений, позволяющим восполнить недостаток информации от сетей традиционных контактных метеорологических наблюдений. Эти наблюдения содержат значения ветра, основанные на отслеживании облаков жидкой и кристаллической фазы, а также водяного пара на спутниковых изображениях. Они имеют высокое временное разрешение, обеспечивая при этом глобальное покрытие данными о ветре, в т.ч. и в арктических широтах. Это крайне важно ввиду малого количества данных контактных и спутниковых наблюдений об атмосферной циркуляции в высоких широтах. В проводимых ранее исследованиях влияния различных видов наблюдений на точность численных прогнозов погоды [47; 56; 59] было выявлено, что наблюдения AMV вносят заметный, хотя и не решающий, вклад в повышение точности анализов и прогнозов.

## Глава 2. Ансамблевая система усвоения Гидрометцентра России

Данная глава посвящена разработанной ансамблевой системе усвоения, состав которой приводится в разделе 2.1. Система была разработана при участии автора [116]. Дальнейшие шаги по её развитию, в которых автор активно участвовал, были опубликованы в статьях [9—11; 20; 24; 26]. В настоящее время описываемая система усвоения оперативно применяется в составе глобальной системы ансамблевого среднесрочного прогноза погоды на основе модели ПЛАВ в Гидрометцентре России [8], описание которой приводится в главе 4.

Описание модели ПЛАВ, используемой в описываемой ансамблевой системе усвоения для генерации первого приближения, приводится в разделе 2.2. Раздел 2.3 посвящён описанию настройки параметров матрицы ковариаций ошибок наблюдений **R**. Применяемые способы контроля качества наблюдений описываются в разделе 2.4. Разделы 2.5 и 2.6 посвящены таким важным вопросам, возникающим при реализации LETKF, как локализация наблюдений вокруг точек модельной сетки и инфляция матрицы ковариаций ошибок первого приближения, позволяющей избежать переоценки точности первого приближения при вычислении анализа. В разделе 2.7 описана практическая параллельная реализация фильтра для запуска на массивно-параллельных вычислительных кластерах. Описание задания почвенных переменных и запуска полного комплекса программ системы усвоения приведено в разделах 2.8 и 2.9. Результаты работ, проведённых в рамках данной главы, опубликованы в статьях [9; 116].

### 2.1 Состав системы усвоения

Применяемая в данной работе система усвоения основана на алгоритме LETKF и использует краткосрочные прогнозы глобальной модели атмосферы ПЛАВ для получения полей первого приближения на модельной сетке. Поля имеют постоянное горизонтальное разрешение по пространству  $0,9^\circ \times 0,72^\circ$  с 96 вертикальными гибридными уровнями или 28 вертикальными сигма-уровнями. В состав ансамбля входят 40 участников. Код основной программы написан на

языке Фортран с использованием библиотек параллельного программирования MPI и OpenMP. Для предварительной обработки данных спутниковых наблюдений AMV написаны скрипты на языке Python. Усвоение почвенных переменных происходит в отдельном блоке (2.8). Для оценки результатов работы системы используется блок верификации, описание которого приводится в разделе 3.2.

## 2.2 Глобальная полулагранжева модель атмосферы

Ядром описываемой системы является вычислительно эффективная полулагранжева глобальная конечно-разностная модель общей циркуляции атмосферы, получившая имя ПЛАВ (от ПолуЛагранжева, основанная на уравнении Абсолютной завихренности, в английском варианте – SL-AV) [21; 23; 26]. Модель была разработана в Институте вычислительной математики РАН им. Г.И. Марчука и в Гидрометцентре России, где используется в качестве оперативной модели среднесрочных и долгосрочных численных прогнозов погоды.

Модель ПЛАВ состоит из блока решения уравнений динамики атмосферы (далее – динамического ядра) собственной разработки и набора параметризаций процессов подсеточного масштаба, в основном заимствованных.

Особенностями динамического блока являются использование вертикальной компоненты абсолютной завихренности и дивергенции в качестве прогностических переменных, а также применение конечных разностей четвёртого порядка на несмещенной сетке для аппроксимации неадвективных слагаемых уравнений. С 2017 года во всех версиях модели применяется гибридная вертикальная координата [28].

В модели используется полуявная схема интегрирования по времени [18] и полулагранжев подход к дискретизации адвективных членов уравнений [118]. Данное сочетание позволяет использовать шаги по времени значительно больше ограничения, определяемого из условия устойчивости Куранта. Практически можно использовать шаг по времени в 3-5 раз больше, чем максимально возможный при использовании эйлеровой схемы расчёта адвективных членов и/или явной схемы интегрирования по времени. Подробное описание динамического блока модели приведено в [23; 126].

Эффект процессов, происходящих на масштабе более мелком, чем масштаб сетки (например, глубокой конвекции), во всех гидродинамических моделях описывается с помощью известных в моделях значений термодинамических параметров атмосферы на масштабе сетки и ряда введённых из физических соображений зависимостей и параметров. Используемое в моделях описание этих процессов принято называть параметризациями.

В настоящее время блок параметризаций модели ПЛАВ можно условно разделить на три основных части:

- радиационные процессы в атмосфере: применяются свободно распространяемые пакеты CLIRAD SW [49; 122] и RRTMG LW [96] для коротко- и длинноволновой радиации соответственно;

- расчет тепло- и влагообмена в почве с учётом фазовых состояний воды;

- описание всех остальных процессов подсеточного масштаба в атмосфере (глубокая и мелкая конвекция, планетарный пограничный слой, гравитационно-волновое сопротивление, образование и выпадение осадков, облачность и пр.) и на поверхности Земли (параметризация тепло- и влагообмена с подстилающей поверхностью с учётом типа растительности, наличия снега и пр.) В этих параметризациях применяются алгоритмы, которые были разработаны под руководством Ж.-Ф. Желена консорциумом по мезомасштабному прогнозу погоды ACCORD /LACE, также называемые ALARO [124]. Эти алгоритмы были адаптированы и уточнены авторами модели ПЛАВ.

Для повышения точности описания радиационных потоков авторами модели ПЛАВ была разработана новая параметризация морских слоисто-кучевых облаков, представленная в [25].

В данной работе, если не указано специально, используется версия модели ПЛАВ2018, которая имеет горизонтальное разрешение  $0,9^\circ \times 0,72^\circ$  по долготе и широте соответственно и 96 уровней по вертикали с использованием гибридной вертикальной координаты.

### 2.3 Настройка параметров, определяющих значения матрицы ковариаций ошибок наблюдений

В ходе численных экспериментов проводилось квазиоперативное усвоение данных в непрерывном цикле с шагом усвоения 6 часов. В экспериментах использовались реальные данные следующих видов наблюдений:

- наблюдения с наземных станций (SYNOP), а также судовые наблюдения (SHIP);
- радиозондовые наблюдения (TEMP);
- наблюдения с борта воздушных судов (AIREP);
- скаттерометрические наблюдения рассеянной радиации (ASCAT);
- наблюдения ветра по вычислению облачных структур на спутниковых изображениях (AMV).

Для всех участвующих в усвоении видов наблюдений при разработке системы усвоения и её первоначальной настройки использовался диагональный вид матрицы ковариаций ошибок наблюдений  $\mathbf{R}$ , т. е. возможные автокорреляции между составляющими вектора наблюдений  $\mathbf{y}^o$  не учитывались на начальном этапе.

Подбор диагональных элементов производился в ходе экспериментов на начальных стадиях разработки описываемой ансамблевой системы усвоения А.В. Шляевой (для наблюдений радиозондов, приземных станций и наблюдений с самолётов), В.С. Рогутковым (для спутниковых наблюдений ASCAT) и автором представленной диссертационной работы (для спутниковых наблюдений ветра AMV). Каждый эксперимент состоял из 120 прогнозов в течение 30 дней. Первоначальные значения ковариаций были взяты из работы [69].

Используемые значения приведены в таблице 1. Значения этих параметров были взяты из работы [44]. Их подбор и оценка чувствительности точности анализов и прогнозов к их значениям в рамках данной работы не проводились. Таблица 1 — Таблица значений диагональных элементов матрицы ковариаций ошибок наблюдений для различных видов наблюдений

Вид наблюдений	Диапазон значений высоты	T, К	Q	U, m/s	U, m/s	Ps, hPa
Радиозонды	Выше 400 гПа	1,4	0,18	2,0	2,0	—
	От 400 до 700 гПа	1,2	0,04	2,6	2,6	—
<i>продолжение следует</i>						

<i>(продолжение)</i>						
Вид наблюдений	Диапазон значений высоты	T, К	Q	U, m/s	U, m/s	Ps, hPa
	Ниже 700 гПа	1,0	0,05	1,8	1,8	—
Наземные станции	Уровень поверхности	1,0	0,01	2,0	2,0	3,0
Корабельные наблюдения	Уровень океана	1,0	0,01	2,2	2,2	3,5
Самолётные наблюдения	Выше 400 гПа	1,4	0,18	2,0	2,0	—
	От 400 до 700 гПа	1,2	0,04	2,6	2,6	—
	Ниже 700 гПа	1,4	—	2,8	2,8	—
ASCAT	Уровень океана	—	—	2,8	2,8	—
GOES-13/15	Выше 400 гПа	—	—	9,5	9,5	—
	От 400 до 700 гПа	—	—	5,7	5,7	—
	Ниже 700 гПа	—	—	6,5	6,5	—
METEOSAT-7	Выше 400 гПа	—	—	8,9	8,9	—
	От 400 до 700 гПа	—	—	5,7	5,7	—
	Ниже 700 гПа	—	—	6,2	6,2	—
METEOSAT-9	Выше 400 гПа	—	—	8,5	8,5	—
	От 400 до 700 гПа	—	—	5,1	5,1	—
	Ниже 700 гПа	—	—	6,0	6,0	—
HIMAWARI-7	Выше 400 гПа	—	—	9,3	9,3	—
	От 400 до 700 гПа	—	—	5,2	5,2	—
	Ниже 700 гПа	—	—	6,3	6,3	—
LeoGeo	Выше 400 гПа	—	—	9,5	9,5	—
	От 400 до 700 гПа	—	—	6,5	6,5	—
	Ниже 700 гПа	—	—	6,5	6,5	—

## 2.4 Применяемые способы контроля качества наблюдений

В описываемой ансамблевой системе наблюдения не применяются измерения, которые заведомо не удовлетворяют критериям качества, описанным ниже. Так, в усвоении не участвуют наблюдения с отрицательной абсолютной температурой, заведомо низким давлением (ниже 800 гПа для корабельных или наземных наблюдений на уровне моря или немного выше) и т.д. Такие наблюдения исключаются на этапе чтения из отчётов, содержащих измерения.

Далее измеренное значение сравнивается со значением первого приближения в точке наблюдения. В случае, если отклонение наблюдаемого значения от значения первого приближения превышает заданное количество ансамблевых стандартных отклонений, наблюдение игнорируется и в дальнейшем в усвоении не участвует. В данной работе в качестве такого порогового значения используется 5 стандартных отклонений.

Приземное давление аналогично [136] корректируется на разницу между билинейно интерполированной в точку станции модельной орографией и высотой станции. Для коррекции используется уравнение гидростатики, коррекция выполняется в случае, если разница между модельной орографией и высотой станции меньше 600 метров, в противном случае наблюдение исключается.

Для контроля качества наблюдений ветра AMV мы используем сравнение высоты наблюдения с уровнем тропопаузы. Наблюдения, лежащие выше этого уровня, игнорируются. Кроме того, на этапе переопределения высоты наблюдений AMV, подробно описанном в разделе 3.4, фактически проводится проверка на непрерывность измеряемой величины в физическом пространстве. Так, подозрительным является значение ветра, которое более чем в два раза отличается по величине и более, чем на 60 градусов по направлению отличается от наблюдаемого в соседних точках наблюдений.

## 2.5 Настройка параметров локализации наблюдений

В данной работе локализация наблюдений около рассматриваемой точки модельной сетки проводится в цилиндрической области, при нахождении в которой наблюдение оказывает влияние на значение получаемого анализа. Высота и радиус этого цилиндра подбираются эмпирическим путём. Влияние наблюдения должно уменьшаться с увеличением расстояния от рассматриваемой точки анализа. Высота цилиндра (т. е. величина радиуса локализации по вертикали) зависит от предполагаемого влияния наблюдений на анализ в точках, расположенных на соседних вертикальных уровнях.

Оператор наблюдений  $H$  для данных с приземных станций представляет собой билинейный интерполятор спрогнозированного значения величины из точки модельной сетки в точку наблюдения. Для наблюдений в свободной атмосфере применяется трилинейная интерполяция, где в качестве вертикальной координаты используется логарифм давления.

Большая часть наблюдений, которые оказывают влияние на значение анализа в некоторой точке модельной сетки, также учитываются и при вычислении анализа в соседней точке. Поэтому в полученном пространственном распреде-

лении метеоэлемента наблюдается отсутствие скачков значений при переходе от одной точки к соседней.

Использование полученного глобального анализа в качестве начальных данных для численного прогноза погоды подразумевает, помимо гладкости полей, выполнение физических законов сохранения. Поэтому перед тем, как запустить численную модель, полученный в ходе усвоения данных анализ проходит через инициализацию цифровым фильтром, суть которой состоит в т. н. балансировке пространственных распределений модельных величин.

Вертикальная локализация осуществляется в терминах логарифма давления. Величина горизонтальных и вертикальных радиусов локализаций варьируется в зависимости от составляющей вектора анализа, вида наблюдений, вертикального уровня и географических координат рассматриваемой точки. Радиусы локализации задаются в заранее определённых референсных точках, значения радиуса локализации в остальных точках получаются при помощи линейной интерполяции. В данной реализации эти точки расположены на линиях разграничения глобальных регионов по широте ( $-70^\circ$ ,  $-20^\circ$ ,  $20^\circ$ ,  $70^\circ$ ) для горизонтальных радиусов локализации и на поверхностях 100 гПа, 250 гПа, 500 гПа, 850 гПа и 925 гПа. Такой подход позволяет гибко использовать радиусы локализации в качестве настроечных параметров, подбирая значения, обеспечивающие как можно более точный анализ.

Конкретные значения радиусов локализации могут определяться из различных соображений. Один из подходов был предложен в [106], где радиус локализации зависит от количества измерений. Аналогичным образом задаются радиусы локализации и в данной работе. Контактные наблюдения распределены неравномерно (см. рисунок 2.1). Так, наименьшую плотность наблюдений можно увидеть в приполярных областях, над акваториями океанов и малонаселёнными регионами Африки, Евразии и обеих Америк.

В ходе экспериментов по подбору радиусов локализации, проведённых на начальном этапе развития системы, было установлено, что система усвоения даёт наиболее точные результаты, когда в регионах с малым количеством наблюдений радиусы локализации наблюдений имеют величину 1500 км для наземных наблюдений. Для наблюдений радиозондов радиусы локализаций в таких регионах имеют значения от 1500 км до 2500 км на больших высотах. В регионах с большим количеством наблюдений радиусы локализации варьируются в диа-

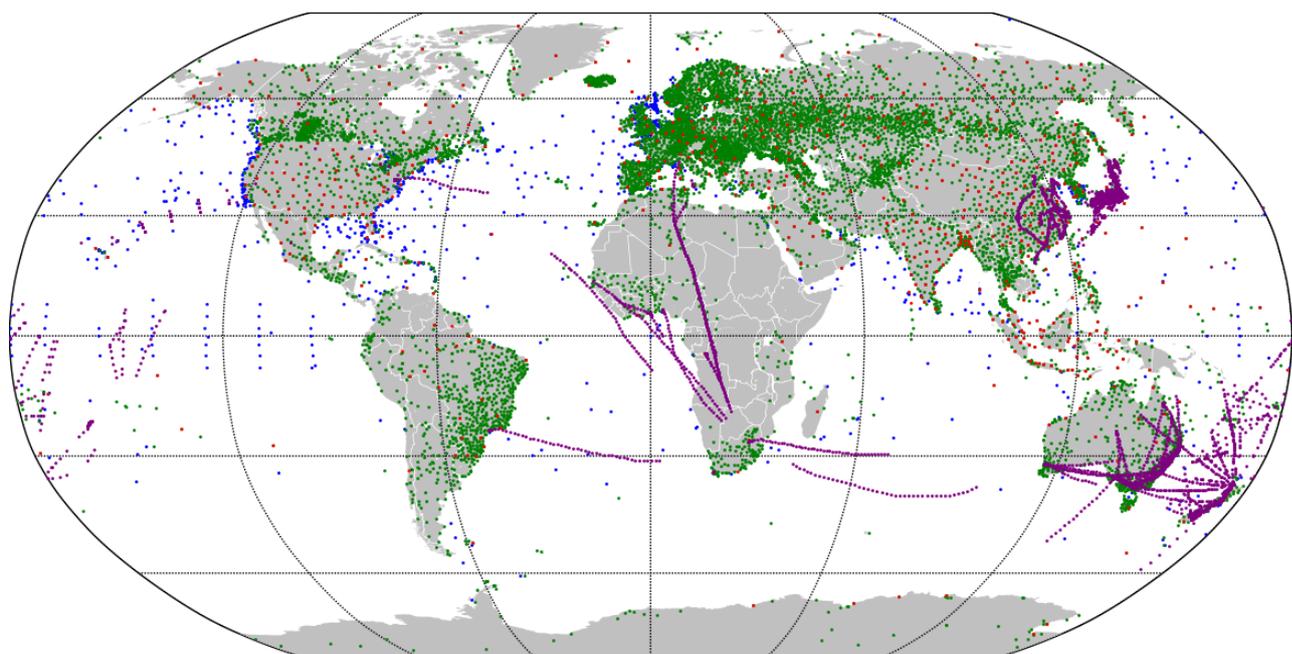


Рисунок 2.1 — Распределение доступных для использования наблюдений в Гидрометцентре России 24 ноября 2019 года (зелёным цветом показаны наблюдения с наземных станций, красным – радиозондовые наблюдения, синим – корабельные наблюдения, фиолетовым – наблюдения с самолётов)

пазоне от 500 км для наземных станций и радиозондов на нижних уровнях до 1500 км для радиозондов на высоких уровнях. Для всех широт в верхних слоях атмосферы используются на 15 – 20% бóльшие значения радиуса локализации, чем в пограничном слое.

Кроме того, были проведены многочисленные эксперименты, направленные на исследование зависимости точности анализов от радиусов локализации для спутниковых наблюдений AMV в различных географических областях на различных высотах. Наименьшие радиусы в настоящее время в системе имеют наблюдения AMV с геостационарных спутников на высотах от 750 до 490 гПа (300-400 км). Наибольшие радиусы имеют данные с полярно-орбитальных спутников на уровнях выше 400 гПа (1250 км).

## 2.6 Настройка параметров инфляции матрицы ковариаций ошибок первого приближения

Как было отмечено в 1.3.6, инфляция реализуется или путём воздействия на поля во время работы модели (физическая), или путём воздействия на поля во время работы усвоения (статистическая). Физическая инфляция реализована К. А. Алиповой (при активном участии автора данной диссертационной работы) с помощью возмущений параметров и тенденций параметризаций неадиабатических процессов подсеточного масштаба для разных участников ансамбля в ходе разработки системы ансамблевого прогноза, в составе которой работает описываемая система усвоения. Подбор параметров производился во время экспериментов по внедрению в модель ПЛАВ блока стохастических параметризаций. Более подробное описание можно найти в [30].

Статистическая инфляция реализована путём внедрения в систему усвоения мультипликативной и аддитивной инфляций.

Параметр мультипликативной инфляции  $\rho$  (см. раздел 1.3.6) в данной реализации может быть задан как функция вертикального уровня и широты. Исследования по влиянию его значения на результаты работы системы усвоения проводились на раннем этапе развития системы в соответствии с рекомендациями из статьи [75]. Первоначально для всех регионов его значение было принято равным 1,01 и постепенно увеличивалось. В ходе численных экспериментов (серии из 120 прогнозов в течение 30 дней) было выявлено, что наименьшие ошибки анализов относительно радиозондовых наблюдений полученные анализы имеют при значениях параметра  $\rho$ , представленных в таблице 2.

Таблица 2 — Таблица значений параметра мультипликативной инфляции  $\rho$

Величины	Диапазон значений высоты	Внетропическая часть северного полушария	Тропики	Внетропическая часть южного полушария
В свободной атмосфере	Выше 400 гПа	$\rho = 1,15$	$\rho = 1,17$	$\rho = 1,15$
	От 400 до 700 гПа	$\rho = 1,20$	$\rho = 1,22$	$\rho = 1,20$
	Ниже 700 гПа	$\rho = 1,24$	$\rho = 1,26$	$\rho = 1,25$
Приземные	—	$\rho = 1,29$	$\rho = 1,35$	$\rho = 1,31$

Аддитивная инфляция реализована с помощью вычисления анализа по формуле 2.1:

$$\tilde{\mathbf{x}}^{a(i)} = \mathbf{x}^{a(i)} + \varepsilon^{In(i)}. \quad (2.1)$$

Случайный шум  $\varepsilon^{In(i)}$  моделируется в соответствии с уравнением 2.2 из [102]:

$$\varepsilon^{In(i)} = r \mathbf{x}^{a(i)}. \quad (2.2)$$

Здесь  $r$  является случайным множителем, который вычисляют с помощью сферических функций:

$$r = \sum_{mn} \hat{r}_{mn} Y_{mn}, \quad (2.3)$$

где  $Y_{mn}$  — сферические гармоники,  $m$  и  $n$  — зональное и полное волновые числа соответственно. Спектральные коэффициенты  $\hat{r}_{mn}$  изменяются соответствии с авторегрессионной моделью первого порядка:

$$\hat{r}_{mn}(t + \Delta t) = \hat{r}_{mn}(t) + \sigma_n \eta_{mn}(t), \quad (2.4)$$

где  $\eta_{mn} \in \mathbb{C}$ , их действительная и мнимая части являются независимыми гауссовскими случайными числами с единичной дисперсией и нулевым средним. Кроме того,  $\eta_{mn}$  также независимы от различных сферических гармоник.

Коэффициенты  $\hat{r}_{mn}$  являются случайными числами, которые имеют гауссово распределение, некоррелированы по высоте и независимы во времени для разных модельных переменных и для разных  $m$  и  $n$ . Дисперсии величин  $\hat{r}_{mn}$  задаются константами так, чтобы полученный шум имел заданный радиус корреляции по горизонтали и вертикали и заданную дисперсию. Радиус корреляции и дисперсия сгенерированного шума настраиваются в зависимости от модельного уровня, широты и переменной, как и радиусы локализаций наблюдений. Такой подход помогает смоделировать ошибку модели, которая представляет собой стохастический шум с нулевым математическим ожиданием и неизвестной матрицей ковариаций  $\mathbf{Q}$ .

Для подбора величин параметров, оказывающих влияние на аддитивную инфляцию, автором данной работы были проведены многочисленные эксперименты, в ходе которых рассчитывались и обрабатывались прогнозы в непрерывный период времени от месяца до сезона. В ходе этих экспериментов установлено, что наиболее заметный отклик наблюдается при изменении параметров аддитивной инфляции приземного давления. Типичное поле случайного шума, генерируемого описываемой системой усвоения для приземного давления, показано на рисунке 2.2.

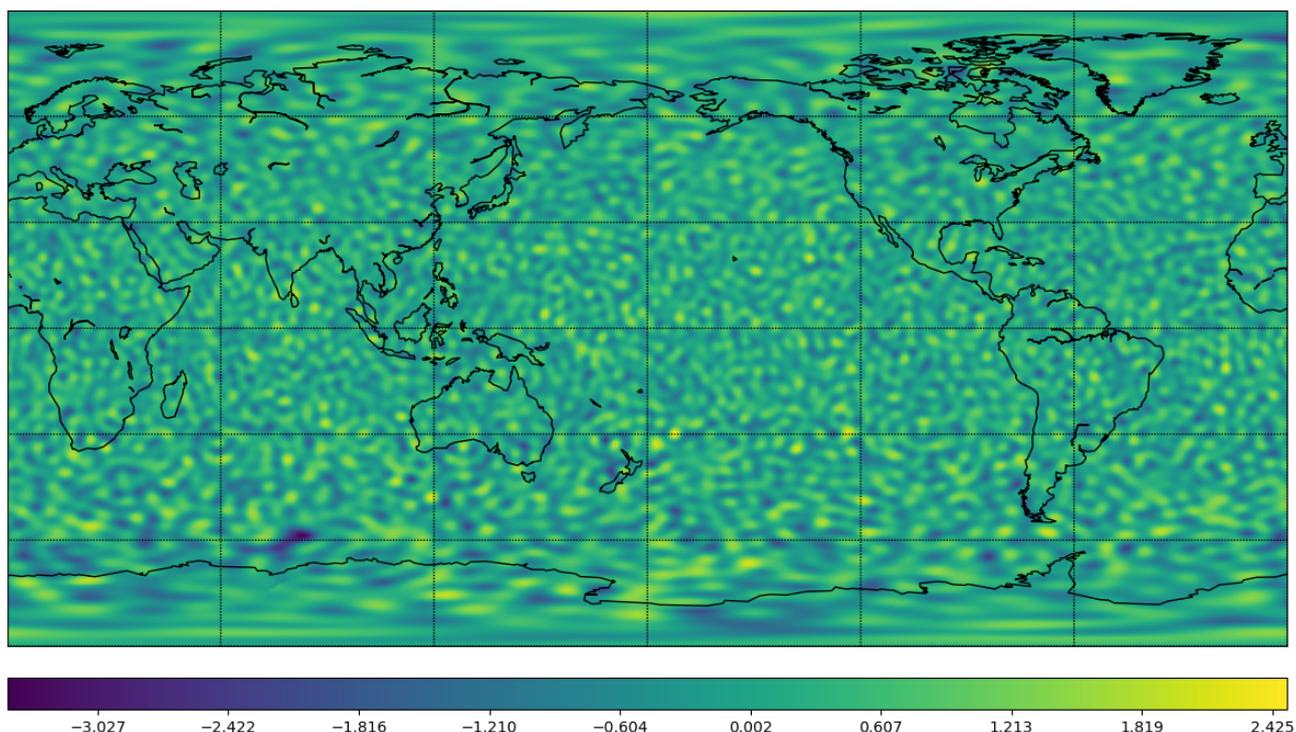


Рисунок 2.2 — Типичное поле случайного шума при использовании аддитивной инфляции в системе усвоения для приземного давления, единицы измерения – гПа

Подбор параметров, отвечающих за амплитуду случайного шума, производился преимущественно в ходе разработки системы ансамблевого прогноза для контроля разброса ансамбля прогнозов. Наиболее простым способом проверки того, насколько адекватно генерируемый ансамбль воспроизводит неопределённость смоделированного состояния, является сопоставление разброса ансамбля со среднеквадратической ошибкой среднего по ансамблю прогноза. Параметры амплитуды случайного шума, реализующего механизм аддитивной инфляции в ансамблевой системе усвоения, подбирались таким образом, чтобы разброс ансамбля максимально совпадал со среднеквадратической ошибкой прогноза на заблаговременности 24 и 48 часов. На графиках рисунков 2.3 и 2.4 для примера представлены осреднённые за июнь 2023 года зависимости среднеквадратических ошибок и разбросов для давления на уровне моря и меридиональной составляющей скорости ветра на уровне 850 Гпа во внетропической части северного полушария до и после применения оптимальных значений параметров амплитуды аддитивной инфляции.

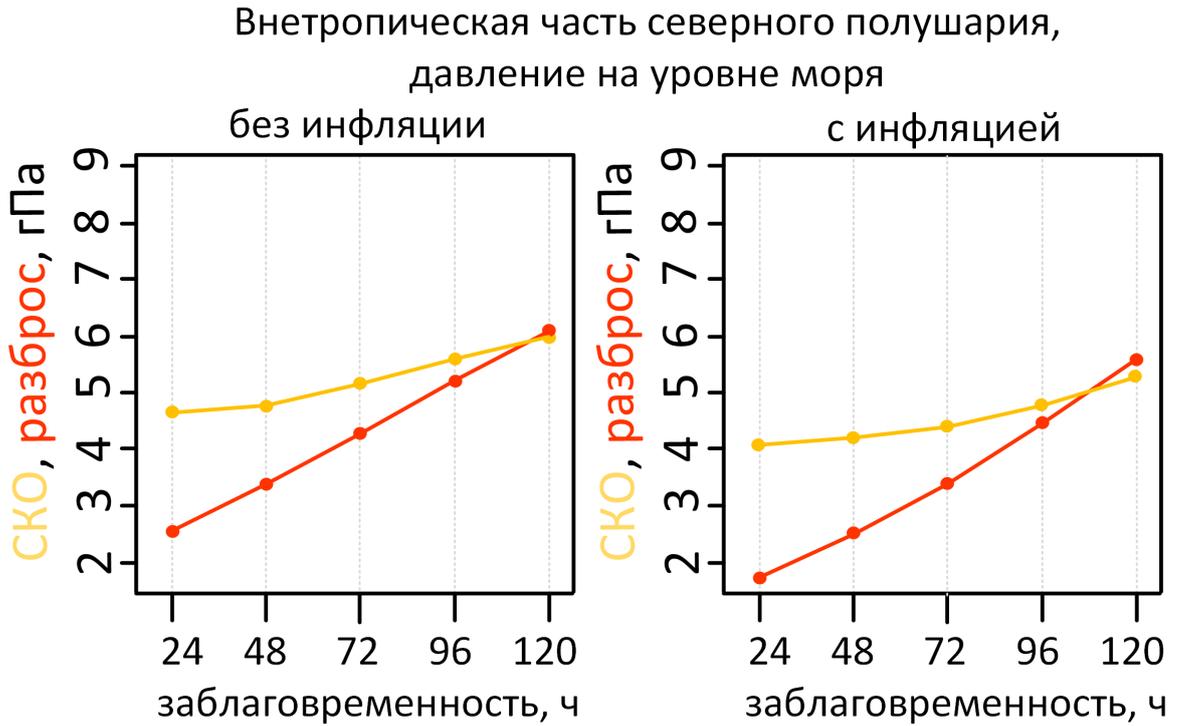


Рисунок 2.3 — Графики зависимости осреднённых за июнь 2023 года среднеквадратических ошибок (СКО, красный цвет) и разбросов по ансамблю (разброс, оранжевый цвет) давления на уровне моря от заблаговременности прогноза до настройки инфляции (слева) и после применения инфляции (справа). Регион – внетропическая часть северного полушария

## 2.7 Параллельная реализация алгоритма LETKF

Оригинальный алгоритм LETKF [75] предполагает выполнение следующих шагов для получения вектора анализа:

1. Вычисление ансамбля (глобальных) первых приближений в точках наблюдений  $\mathbf{y}_g^b = H(\mathbf{x}_g^b)$ , среднее по ансамблю  $\bar{\mathbf{y}}_g^b = k^{-1} \sum_{i=1}^k \mathbf{y}_g^{b(i)}$ , матрицу отклонений от среднего  $\mathbf{Y}_g^b = \{\mathbf{y}_g^{b(i)} - \bar{\mathbf{y}}_g^b, i = \overline{1, k}\}$ . Здесь нижний индекс  $g$  обозначает, что соответствующие величины определены в глобальном модельном пространстве  $\mathcal{X}$ .

2. Вычисление среднего по локальному ансамблю первых приближений  $\bar{\mathbf{x}}^b = k^{-1} \sum_{i=1}^k \mathbf{x}^{b(i)}$  и матрицу отклонений от среднего  $\mathbf{X}^b = \{\mathbf{x}^{b(i)} - \bar{\mathbf{x}}^b, i = \overline{1, k}\}$ .

3. Выбор из глобального вектора  $\mathbf{y}_g^o$  локальных наблюдений  $\mathbf{y}^o$ , находящихся на не превышающем радиус локализации расстоянии от узла сетки, соответствующего

Внетропическая часть северного полушария,  
меридиональная составляющая ветра на уровне 850 гПа  
без инфляции

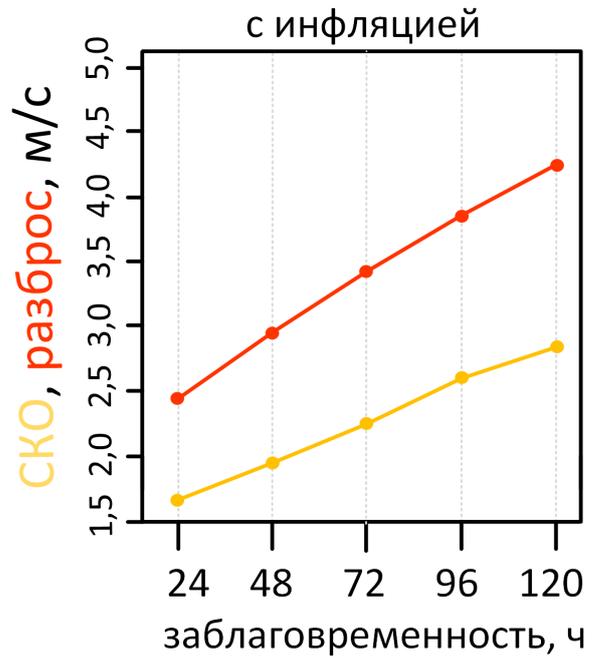
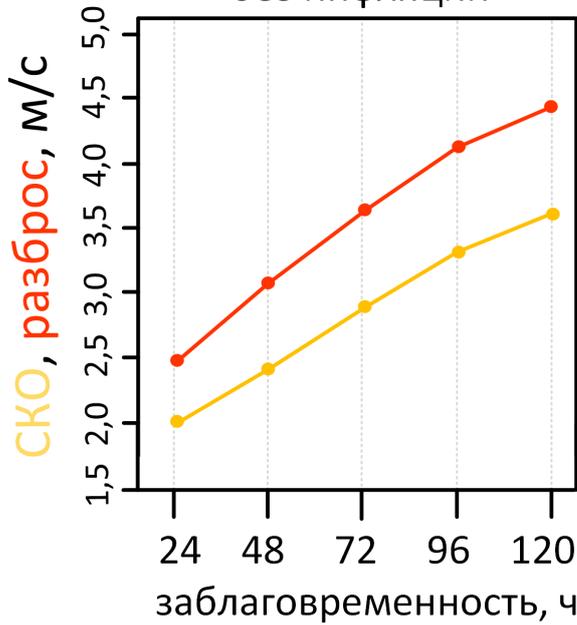


Рисунок 2.4 — Графики зависимости осреднённых за июнь 2023 года среднеквадратических ошибок (СКО, красный цвет) и разбросов по ансамблю (разброс, оранжевый цвет) меридиональной составляющей ветра на уровне 850 гПа от заблаговременности прогноза до настройки инфляции (слева) и после применения инфляции (справа). Регион – внетропическая часть северного полушария

ющих векторов первого приближения и отклонений в пространстве наблюдений  $\mathbf{y}^b$  и  $\mathbf{Y}^b$ .

4. Вычисление матрицы  $\mathbf{C} = (\mathbf{Y}^b)^T [(\mathbf{R}_{[g]})^{-1} \circ \mathbf{L}]$ .

5. Вычисление матрицы  $\tilde{\mathbf{P}}^a = \left[ \frac{(k-1)}{\rho} \mathbf{I}_k + (\mathbf{Y}^b)^T \mathbf{R}^{-1} \mathbf{Y}^b \right]^{-1}$ .

6. Вычисление матрицы  $\mathbf{W}^a = \left[ (k-1) \tilde{\mathbf{P}}^a \right]^{1/2}$  согласно (1.41).

7. Вычисление вектора среднего анализа в пространстве ансамбля  $\bar{\mathbf{w}}^a = \tilde{\mathbf{P}}^a \mathbf{C} (\mathbf{y}^o - \bar{\mathbf{y}}^b)$  и матрицы  $\mathbf{w}^a$ , в которой каждый столбец  $\mathbf{w}^{a(i)} = \bar{\mathbf{w}}^a + \mathbf{W}^{a(i)}$ ,  $i = \overline{1, k}$ .

8. Вычислить вектор среднего анализа  $\bar{\mathbf{x}}^a = \bar{\mathbf{x}}^b + \mathbf{X}^b \bar{\mathbf{w}}^a$  и векторы ансамбля анализов  $\mathbf{x}^{a(i)} = \bar{\mathbf{x}}^b + \mathbf{X}^b \mathbf{w}^{a(i)}$ ,  $i = \overline{1, k}$ .

Схема вычислений представлена на рисунке 2.5.

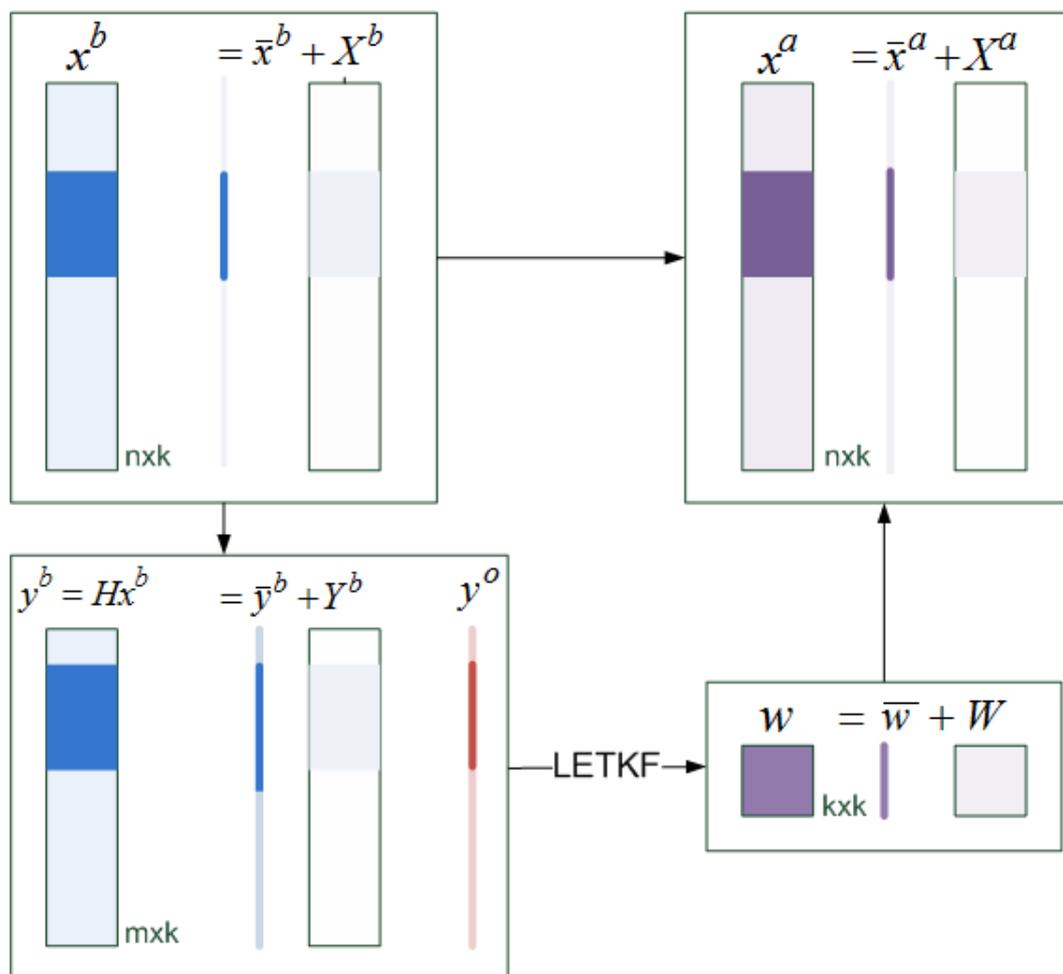


Рисунок 2.5 — Схема вычислений в алгоритме LETKF

Практическая реализация системы усвоения на основе этого алгоритма, применяемая в данной работе, представлена в [9; 10; 19; 116]. В настоящее время эта система усвоения использует информацию следующих видов метеонаблюдений:

- радиозонды TEMP (температура, относительная влажность и составляющие скорости ветра на 16 стандартных уровнях от 1000 гПа до 10 гПа, около 650 наблюдений на каждом уровне в 00ч ВСВ и в 12ч ВСВ);
- наземные и корабельные наблюдения SYNOP, SHIP (приземное давление, температура и относительная влажность на уровне 2 м, порядка 10000 наблюдений в 00ч ВСВ и в 12ч ВСВ);
- самолетные наблюдения AIREP (температура и составляющие скорости ветра, порядка 10000 наблюдений в 0000ч ВСВ и в 12ч ВСВ);
- спутниковые наблюдения ветра SATOB и AMV (составляющие скорости ветра, до 800000 наблюдений в день);

– скаттерометрические наблюдения рассеянной радиации ASCAT (составляющие скорости ветра над водными поверхностями, до 200000 наблюдений в день).

Вычисления в фильтре выполняются в режиме гибридного распараллеливания с помощью технологий MPI и OpenMP. Широтные полосы обрабатываются MPI—процессами, что ограничивает возможное количество используемых MPI—процессов разрешением сетки по широте. Внутри каждой широтной полосы используется равномерное разбиение расчётной области на долготные части, количество которых определяется количеством заданных OpenMP нитей.

Приведём описание этой схемы. Пусть имеется ансамбль из  $N_f$  полей, которые необходимо обработать на  $N_p$  MPI—процессах. Каждое поле имеет  $N_{lev}$  уровней с  $N_{lon}$  узлами по долготе и  $N_{lat}$  узлами по широте.

После инициализации на каждом MPI—процессе происходит расчёт количества обрабатываемых широт, чтение всех имеющихся на момент анализа наблюдений. Потом каждый из первых  $N_f$  MPI—процессов читает своё поле первого приближения и применяет прямой оператор наблюдений к прочитанному полю первого приближения. Таким образом, по окончании этой части первые  $N_f$  MPI—процессов имеют по одному глобальному полю первого приближения  $\mathbf{x}_g^b(i)$ , глобальное вектор наблюдений  $\mathbf{y}_g^o$  и значений первого приближения в точках наблюдений  $\mathbf{y}_g^b(i)$ . Все остальные MPI—процессы имеют свою копию  $\mathbf{y}_g^o$  и ждут получения своих порций обрабатываемых данных.

Далее происходит пересылка широтных полос полей первого приближения и значений первого приближения в точках наблюдений между всеми MPI—процессами. Схема пересылки полей первого приближения представлена на рисунке 2.6. По окончании процесса обмена данными каждый из MPI—процессов получает  $N_f$  широтных полос и  $N_f$  векторов значений первого приближения в точках наблюдений. После этого происходит расчёт средних величин  $\bar{\mathbf{x}}_{ip}^b, \bar{\mathbf{y}}^b$  и матриц  $\mathbf{X}_{ip}^b, \mathbf{Y}^b$ , где  $ip$  — номер MPI—процесса.

Все MPI—процессы инициализируют в себе заданное количество OpenMP нитей, каждая из которых получает свою часть обрабатываемых долгот и выполняет расчёт локального вектора анализа для каждой точки из своей широтно-долготной области. После завершения вычисления анализа OpenMP нити аннулируются, а каждый MPI—процесс имеет ансамбль из  $N_f$  широтных полос полей анализа. Происходит обратная пересылка данных, в ходе которой первые

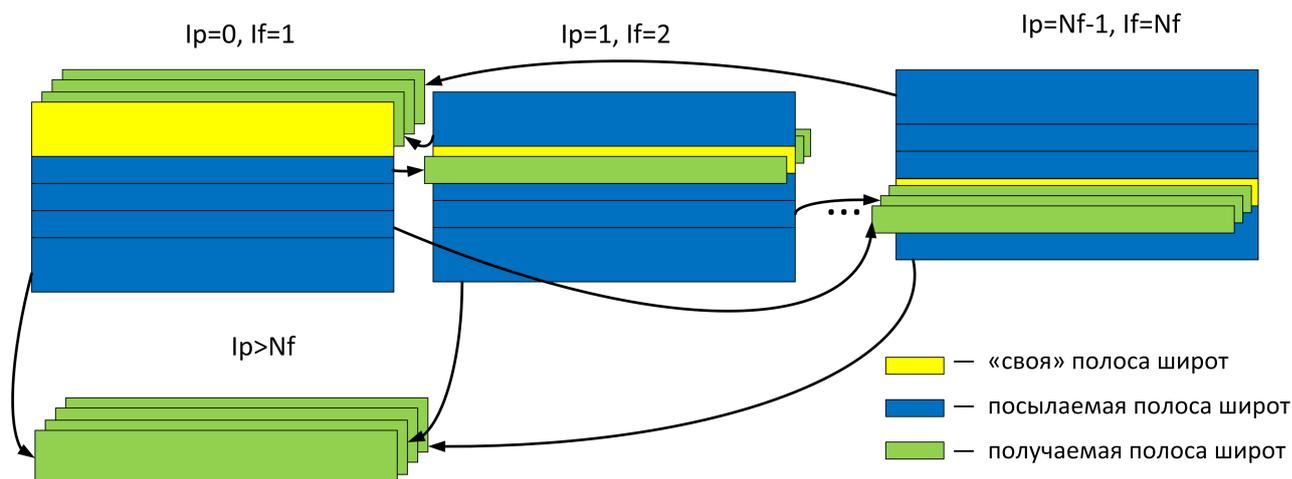


Рисунок 2.6 — Схема пересылок данных между процессами. Жёлтым отмечена порция данных для обработки, которую процесс получает сразу после чтения, синим — данные для пересылки другим процессам, зелёным — данные, получаемые от других процессов.  $I_p$  — номер процесса,  $I_f$  — номер члена ансамбля,  $N_f$  — размер ансамбля (количество его членов)

$N_f$  MPI-процессов получают по одному глобальному полю анализа. После получения всех широтных полос от соответствующих MPI-процессов происходит добавление аддитивной инфляции к полученному полю, и результат записывается в файл для использования в следующем цикле усвоения для генерации ансамбля первых приближений. Средний анализ, используемый в качестве решения задачи усвоения, записывается в результирующий файл MPI-процессом с номером  $i_p = 0$ . Этот анализ может быть в дальнейшем использован для получения среднесрочного детерминированного прогноза. Все поля имеют формат, пригодный для использования в модели ПЛАВ.

Разбиение расчётной области на равномерные широтные полосы приводит к существенной разбалансировке параллельного алгоритма, т. к. в этом случае далеко отстоящие от экватора полосы имеют для обработки существенно меньшее количество наблюдений. Для нейтрализации этого эффекта было применено разбиение на широтные полосы, ширина которых зависит от географического расположения. Чем ближе широта к полюсу, тем шире полоса обрабатываемых широт. Таким образом, большее количество модельных точек получают процессоры, которые обрабатывают регионы с меньшим количеством наблюдений. Такой подход позволил несколько сбалансировать загрузку процессоров и уменьшить (на 20%) общее время работы параллельной программы, которая

при горизонтальном разрешении полей  $400 \times 250$  показывает хорошее ускорение. В качестве примера на рисунке 2.7 показан график из работы [10], на котором приводится параллельное ускорение при запуске программы на компьютере РСК «Торнадо», применяемого в Главном вычислительном центре Росгидромета в период времени с 2011 по 2016 годы. Запуски проводились для 40 членов ансамбля с использованием функций одностороннего доступа из стандарта MPI-2. Параллельный алгоритм вычислений в программе с тех пор не менялся.

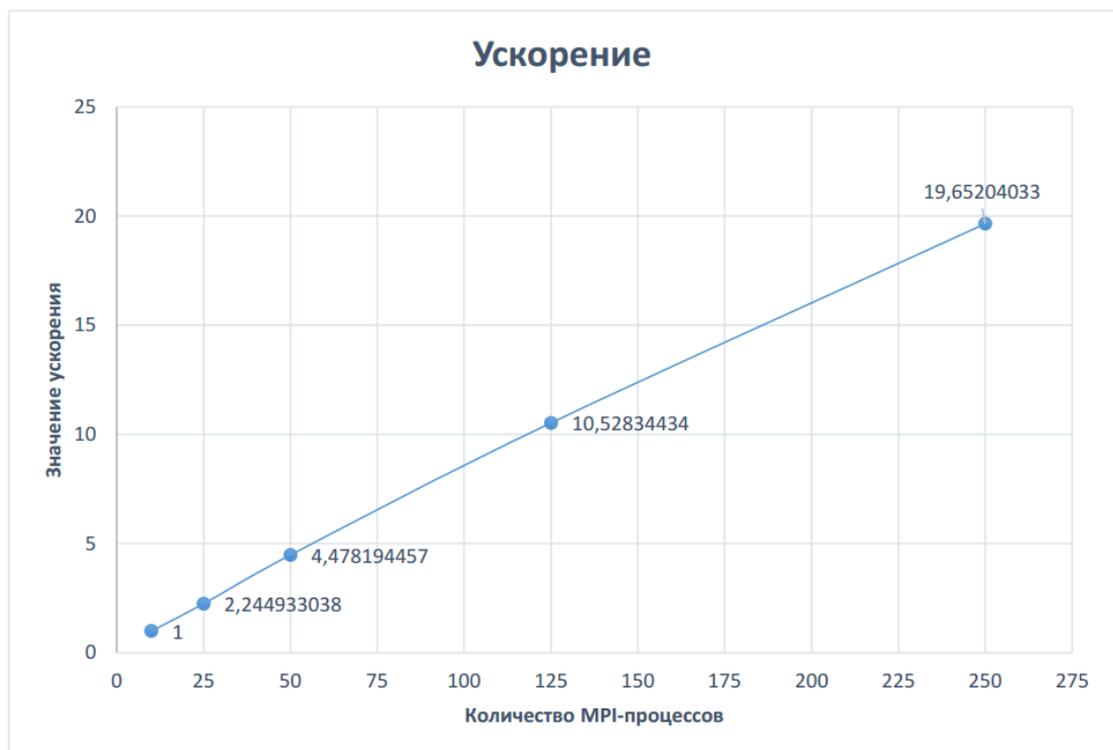


Рисунок 2.7 — Ускорение параллельной программы, реализующей ансамблевое усвоение с помощью алгоритма LETKF, при 40 членах ансамбля

## 2.8 Задание начальных значений почвенных переменных

Блок подготовки почвенных переменных выполняет усвоение приземных наблюдений температуры и влажности на сети метеостанций, вычисляя поля

приземной температуры и относительной влажности. Алгоритм усвоения основан на схеме оптимальной интерполяции.

Для усвоения температуры и влажности на уровне 2 метров (необходимых для работы блока почвенного анализа) используются данные, полученные с наземных станций: температура и относительная влажность. Для усвоения приземного давления, а также температуры и составляющих скорости ветра в свободной атмосфере используются данные наземных станций (приземное давление), спутниковых наблюдений ветра (составляющие скорости ветра), и наблюдения радиозондов (температура и составляющие скорости ветра).

Задаваемая для расчёта анализа ковариационная матрица ошибок содержит зависимость от расстояния по горизонтали и от разности высот. Функция корреляции имеет следующий вид:

$$\mu_{ij} = \exp\left(\frac{-1}{2} \left[ \frac{r_{ij}^2}{R} + \frac{z_{ij}^2}{Z} \right]\right), \quad (2.5)$$

где  $r_{ij}$  — расстояние между точками  $i$  и  $j$ ,  $z_{ij}$  — разность высот между точками  $i$  и  $j$ ,  $R$  — коэффициент, задающий радиус корреляции,  $Z$  — максимальная разность высоты коррелирующих между собой точек.

В реализованном анализе возможно задание различных радиусов корреляции для разных широт, а также различных радиусов корреляции для океана и суши. Радиусы корреляции задаются отдельно для «широтных полос» на суше и для океана. Затем инкременты анализов используются для инициализации температуры и влагосодержания почвы [22]. Этот блок позволяет перед запуском модели ПЛАВ задать начальные значения приземной температуры, влажности и почвенных переменных (поверхностную температуру почвы, температуру глубинного слоя, влагосодержание приповерхностного и глубинного слоя). Блок усвоения почвенных характеристик согласован с применяемой в ПЛАВ параметризацией процессов на поверхности суши с учётом растительности [35; 41; 97] (ISBA, Interaction Soil Biosphere Atmosphere).

В блоке усвоения почвенных переменных в виду отсутствия оперативных наблюдений для параметров почвы для коррекции значений почвенных переменных в каждой точке горизонтальной сетки используются инкременты анализа температуры и влажности на высоте 2 метра. Реализована схема коррекции почвенных переменных, предложенная в [63].

Метод коррекции температуры почвы основан на вычислении инкремента анализа температуры на высоте 2 метров в каждой точке сетки:

$$\Delta T_s = \Delta T_{2m},$$

$$\Delta T_p = \frac{\Delta T_{2m}}{2\pi},$$

где  $\Delta$  — инкремент, т. е. разность значений анализа и значений первого приближения (6-часовой прогноз, стартовавший 6 часов назад).

Коррекция влагосодержания поверхностного слоя почвы  $\omega_s$  и основного слоя почвы  $\omega_d$  линейно вычисляется из разности поля первого приближения и поля анализа для приземных переменных посредством оптимальной интерполяции:

$$\omega T_s = a_s^T \Delta T_{2m} + a_s^H \Delta RH_{2m},$$

$$\omega T_p = a_p^T \Delta T_{2m} + a_p^H \Delta RH_{2m},$$

где  $a_s^T, a_s^H, a_p^T, a_p^H$  — оптимальные коэффициенты, минимизирующие среднеквадратичные ошибки температуры и относительной влажности на уровне 2 м, которые зависят от структуры почвы, локального солнечного времени и характеристик подстилающей поверхности ([22]).

Главным отличием используемого в ансамблевой системе усвоения компонента подготовки почвенных переменных от исходной версии, которая применяется в модели ПЛАВ детерминированного прогноза [21], является подготовка ансамбля из 40 разных анализов стартовых полей на 2 м по разным первым приближениям из ансамбля первых приближением (выполнено А.В. Шляевой на начальном этапе разработки системы).

## 2.9 Методика циклического запуска усвоения

Методика запуска усвоения в цикле основана на скриптах, написанных на языке сценариев `bash`, которые запускают каждый следующий компонент системы усвоения в зависимости от выполнения предыдущего компонента. Цикл

усвоения и зависимости различных компонентов цикла показаны на рисунке 2.8.

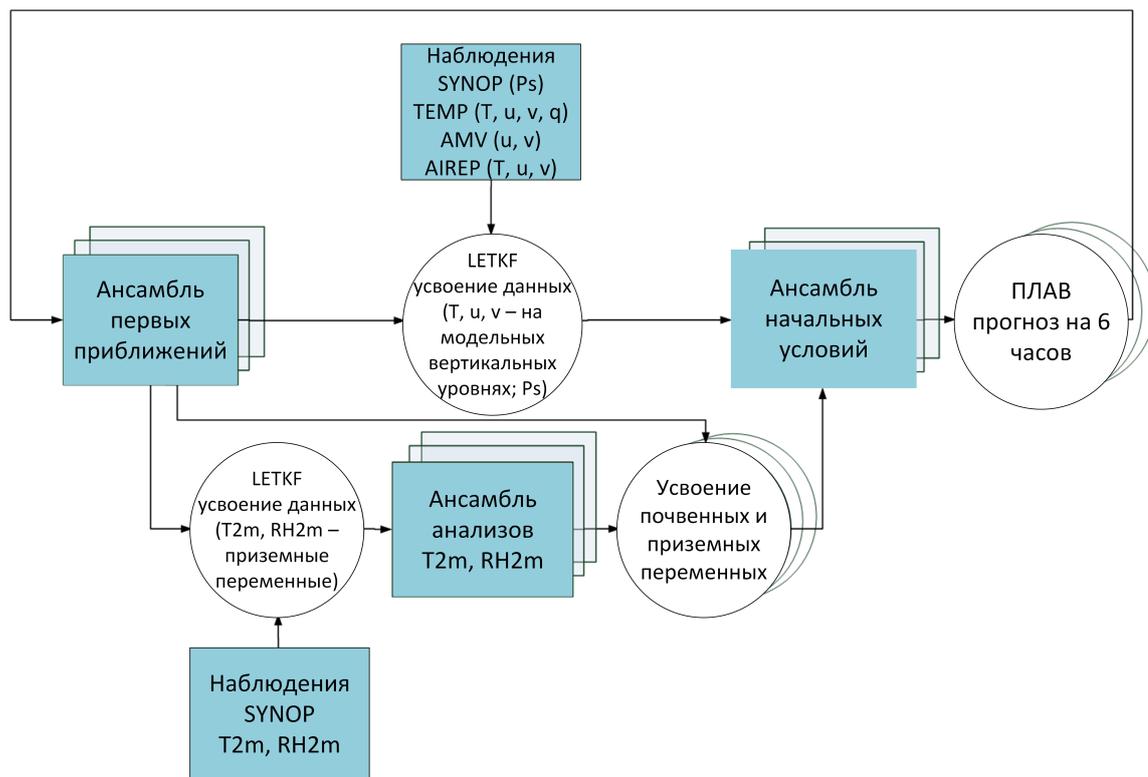


Рисунок 2.8 — Схема циклического усвоения (из [116])

На рисунке компоненты системы усвоения имеют следующие обозначения:

- SLAV – запуск ансамблевого прогноза модели ПЛАВ для получения первого приближения;
- LETKF – запуск усвоения данных на основе LETKF;
- SURFACE and SOIL – запуск ансамблевого почвенного анализа по инкрементам, полученным анализом LETKF.

Верификация результатов анализа LETKF по данным наблюдений и оперативных анализов, а также вычисление оценок, производятся отдельно (на рисунке не показано).

Моменты запуска соответствуют 00, 06, 12 и 18 часов всемирного скоординированного времени. В 00 и 12 часов со среднего по ансамблю анализа стартует расчёт модели на 5 и 10 суток соответственно. В остальные моменты времени рассчитываются только первые приближения (прогнозы на ближайшие 6 часов).

## 2.10 Выводы к главе 2

В результате представленных в главе 2 работ была создана методика настройки ансамблевой системы усвоения. Она основана на подборе следующих настроечных параметров:

- радиусы локализации наблюдений;
- значения мультипликативной инфляции;
- радиусы влияния и амплитуды шума аддитивной инфляции.

Эффективность применения указанных параметров оценивалась путём расчёта показателей успешности прогнозов и среднеквадратических ошибок полученных начальных данных по данным оперативного анализа Гидрометцентра и наблюдений радиозондов.

Методика применена при переходе ансамблевой системы усвоения с 28 сигма-уровней на 96 гибридных уровней, а далее (вне рамок диссертации) для ансамблевой системы усвоения на основе модели ПЛАВ20.

### Глава 3. Усвоение спутниковых наблюдений ветра AMV

В данной главе приводится описание усвоения спутниковых наблюдений ветра AMV в рамках описываемой системы усвоения. Эксперименты по включению этих наблюдений в число усваиваемых помогли внести ряд улучшений, которые позволяют использовать описываемую систему усвоения для подготовки начальных данных в оперативной системе ансамблевого прогноза Гидрометцентра России. Все приведённые здесь эксперименты проводились автором настоящей работы.

В разделе 3.1 приводится описание системы усвоения, которая использовалась в приведённых численных экспериментах. Способы уменьшения ошибки при усвоении данных наблюдений AMV, такие как прореживание наблюдений, переопределение высоты и учёт корреляции ошибок наблюдений, приведены в разделах 3.4, 3.5.1 и 3.5.2. В 3.7 представлены выводы к текущей главе.

Результаты данной главы опубликованы в работах [9; 11].

#### 3.1 Конфигурация системы усвоения

Как было отмечено выше, все численные эксперименты по усвоению данных проводились с помощью основанной на алгоритме LETKF ансамблевой системы [9; 10; 19; 29; 116]. Для генерации ансамбля первых приближений в экспериментах по прореживанию наблюдений и внедрению корреляций их ошибок использовалась глобальная модель атмосферы ПЛАВ [21]. В экспериментах по переопределению высоты наблюдений использовалась модель в версии с постоянным горизонтальным разрешением по пространству  $0,9^\circ \times 0,72^\circ$  и 28 вертикальными сигма-уровнями.

Проводились эксперименты по квазиоперативному усвоению в непрерывном цикле с шагом усвоения 6 часов. В экспериментах использовались данные указанных в разделе 2.3 наблюдений.

В проведённых экспериментах отдельно усваивались температура и относительная влажность на уровне 2 метров, а также приземное давление, темпера-

тура и составляющие скорости ветра на вертикальных гибридных уровнях модели. Инкременты анализа температуры и относительной влажности на уровне 2 метров использовались для анализа почвенных переменных аналогично [3].

Для оценки влияния, которое оказывало использование спутниковых данных AMV на точность получаемого анализа, первые численные эксперименты проводились без усвоения этого вида данных.

Оценка эффективности применяемых изменений проводилась с использованием оперативной технологии расчёта ошибок детерминированных среднесрочных прогнозов. Описание системы верификации приводится в разделе 3.2. В качестве стартовых полей для моделирования использовались анализы  $\mathbf{x}^a$ , полученные с помощью описываемой в данной работе системы усвоения.

Генерация участников ансамбля на первом шаге проводилась путём добавления к оперативному анализу Гидрометцентра России шума, аналогичного используемому для реализации аддитивной инфляции, описанной в разделе 1.3.6, с настройками из раздела 2.6. Для выхода фильтра в рабочий режим, в котором разброс ансамбля первых приближений примерно соответствует среднеквадратической ошибке детерминированного прогноза, стартовавшего со среднего анализа, использовалось циклическое усвоение в течение одного календарного месяца.

Вычисления в программе, реализующей описываемую систему усвоения, распараллелены с помощью систем параллельных вычислений MPI и OpenMP (см. раздел 2.7). Параллельная программа может эффективно использовать количество процессорных ядер, сопоставимое с числом широт модельной сетки [10].

### 3.2 Верификация результатов усвоения

Применяемая система усвоения имеет в своём составе компонент верификации результатов. Этот компонент имеет свидетельство Роспатента [17].

Оценки ошибок рассчитываются на основе сравнения прогнозов с данными оперативного анализа, который применялся в качестве начальных данных при моделировании, и с данными наблюдений радиозондов из списка, утвер-

ждённого Ведущим центром по верификации среднесрочных прогнозов погоды. По данным оперативного анализа Гидрометцентра России верифицируются следующие поля:

- давление на уровне моря (PMSL или P0 на рисунках далее);
- геопотенциальные высоты изобарических поверхностей 850, 500, 250 гПа (соответственно H850, H500, H250);
- температура на поверхностях 850, 500, 250 гПа (соответственно T850, T500, T250);
- скорость ветра (зональная и меридиональная составляющие) на поверхностях 850, 500, 250 гПа (соответственно U850, U500, U250 и V850, V500, V250);
- модуль скорости ветра на поверхностях 850, 500, 250 гПа (соответственно W850, W500, W250).

При верификации прогностические поля и поля оперативного анализа переинтерполируются на сетку  $1,5^\circ \times 1,5^\circ$ . Перевод прогностических полей на сетку более грубого разрешения производится с помощью метода интерполяции с весами, пропорциональными площади ячейки сетки.

Верификация по данным наблюдений проводится для указанных выше полей, за исключением давления на уровне моря.

Все рассчитанные оценки осредняются по следующим регионам:

- внетропическая часть Северного полушария ( $20^\circ - 90^\circ$  с.ш.);
- внетропическая часть Южного полушария ( $90^\circ - 20^\circ$  ю.ш.);
- тропики ( $20^\circ$  ю.ш. –  $20^\circ$  с.ш.);
- Азия ( $25^\circ - 65^\circ$  с.ш.,  $60^\circ - 145^\circ$  в.д.);
- приполярная часть Северного полушария ( $60^\circ - 90^\circ$  с.ш.);
- приполярная часть Южного полушария ( $90^\circ - 60^\circ$  ю.ш.).

Для оценки успешности рассчитываются следующие показатели:

- среднеквадратическая векторная оценка для ветра (как для верификации по оперативному анализу Гидрометцентра России, так и для верификации по данным наблюдений);
- средняя абсолютная ошибка для всех полей (как для верификации по оперативному анализу Гидрометцентра России, так и для верификации по данным наблюдений);

- средняя ошибка для всех полей (как для верификации по оперативному анализу Гидрометцентра России, так и для верификации по данным наблюдений);
- среднеквадратическая ошибка для всех полей кроме ветра (как для верификации по оперативному анализу Гидрометцентра России, так и для верификации по данным наблюдений);
- коэффициент корреляции аномалий прогнозов и анализов (только для верификации по данным оперативного анализа);
- градиентная ошибка для температуры, давления на уровне моря и высоты стандартных уровней (только для верификации по данным оперативного анализа).

Формулы для расчёта оценок приведены в [13]. Для оценки статистической значимости применяются оценки доверительных интервалов с использованием  $t$ -критерия Стьюдента и метод бутстрепа.

### 3.3 Использование данных наблюдений AMV

В настоящее время в Гидрометцентр России поступают данные наблюдения AMV со следующих геостационарных спутников:

- GOES-18 (NOAA, США, далее – GOES18) – расположен над 75-м меридианом западной долготы, наблюдается акватория Тихого океана, континентальные части Северной и Южной Америк;
- GOES-16 (NOAA, США, далее – GOES16) – расположен над 135-м меридианом западной долготы, наблюдается акватория Тихого океана, западные побережья Северной и Южной Америк;
- HIMAWARI-9 (JMA, Япония, далее – HMWR9) – расположен над 145-м меридианом восточной долготы, наблюдается акватория Тихого океана, Юго-Восточная Азия, Австралия;
- METEOSAT-9 (EUMETSAT, Европа, далее – MET9) – расположен над 57-м меридианом восточной долготы, наблюдается акватория Индийского океана, Юго-Восточная и Средняя Азия, Ближний Восток, Восточная Африка;

– METEOSAT-11 (EUMETSAT, Европа, далее – MET11) — расположен над 9-м меридианом восточной долготы, наблюдается акватория Атлантического океана, Европа и Ближний Восток, Африка, Восточная часть Южной Америки.

Спутниковые наблюдения AMV производятся в следующих диапазонах:

- IR — инфракрасный диапазон;
- SWIR — коротковолновый инфракрасный диапазон;
- VIS — видимый диапазон;
- WV — диапазон водяного пара.

Кроме того, доступны наблюдения, полученные путём обработки снимков с полярных низко-орбитальных спутников, имеющих общие коды LeoGeo (LG) и AVHRR. Эти наблюдения получены обработкой изображений, сделанных в IR диапазоне.

Таким образом, первоначальная обработка подразумевает разделение всего набора доступных на срок анализа наблюдений на 22 поднабора:

- LG.IR;
- AVHRR.IR;
- GOES18.IR;
- GOES18.SWIR;
- GOES18.VIS;
- GOES18.WV;
- GOES16.IR;
- GOES16.SWIR;
- GOES16.VIS;
- GOES16.WV;
- HMWR9.IR;
- HMWR9.SWIR;
- HMWR9.VIS;
- HMWR9.WV;
- MET9.IR;
- MET9.SWIR;
- MET9.VIS;
- MET9.WV;
- MET11.IR;
- MET11.SWIR;

- MET11.VIS;
- MET11.WV.

Для лучшего соответствия наблюдаемых величин срокам подготавливаемых анализов, из всех имеющихся наблюдений для усвоения выбираются только те, которые попадают в окно  $\pm 1$  час от срока анализа. Даже с учётом такого прореживания по времени, количество участвующих в усвоении наблюдений составляет порядка 200 000.

Распределение наблюдений AMV с различных спутников, доступных в Гидрометцентре России для использования в описываемой системе усвоения, приведено на рисунке 3.1.

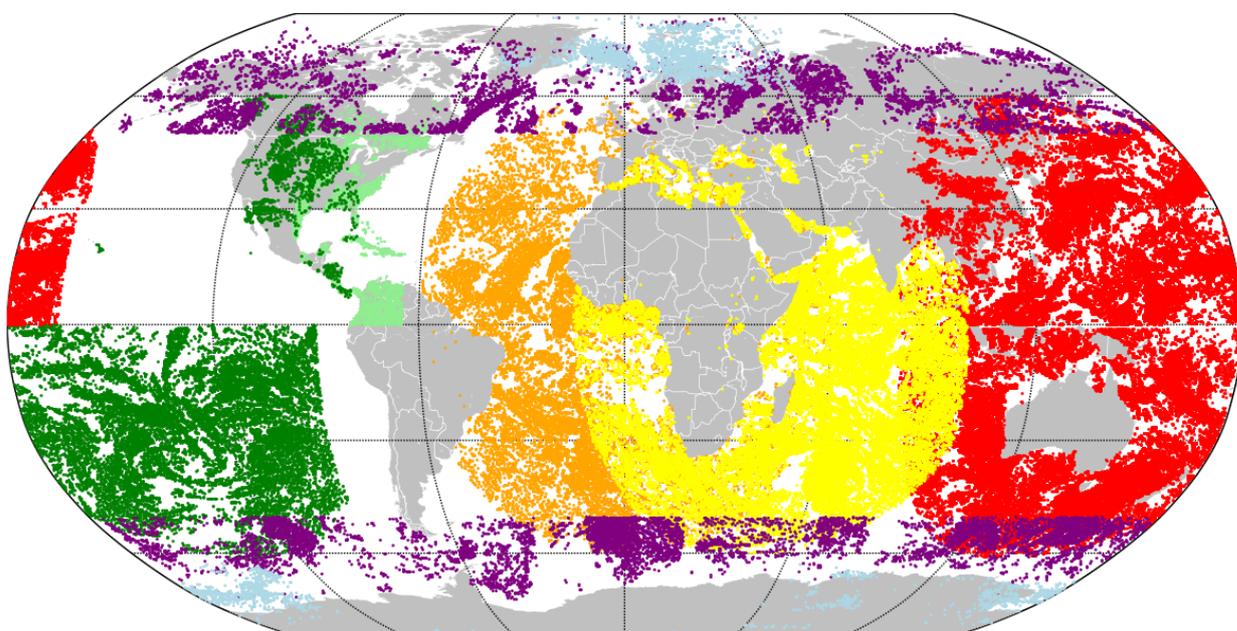


Рисунок 3.1 — Пример распределения используемых спутниковых наблюдений ветра AMV с различных спутников по данным на 01 июня 2023 г. (зелёным цветом показаны наблюдения со спутника GOES-18, светло-зелёным – со спутника GOES-16, красным – со спутника HIMAWARI-9, жёлтым – со спутника METEOSAT-9, оранжевым – со спутника METEOSAT-11, голубым – со спутника AVHRR, фиолетовым – со спутников LeoGeo)

### 3.4 Переопределение высоты наблюдений

Кроме наличия в ошибках наблюдений AMV автокорреляций, нарушающих одно из основных предположений большинства современных методов усвоения данных, серьёзной проблемой наблюдений AMV является определение высоты наблюдения. Ошибка, вносимая при определении высоты, составляет большую часть общей ошибки наблюдения (до 70% в соответствии с [60]). Источники её следующие [60]:

- выбор пикселей на спутниковом изображении, используемых для определения высоты;
- выбор уровня облака (его вершины или середины) для оценки давления в облаке;
- ограничения в методах определения давления в облаках.

Составляющая ошибки, вносимая при определении высоты наблюдения AMV, является самой существенной частью общей ошибки, которая свойственна наблюдению. На рисунке 3.2 показана иллюстрация влияния одинаковой ошибки ( $\pm 50$  гПа), вносимой при определении высоты, на общую ошибку наблюдения скорости ветра [113].

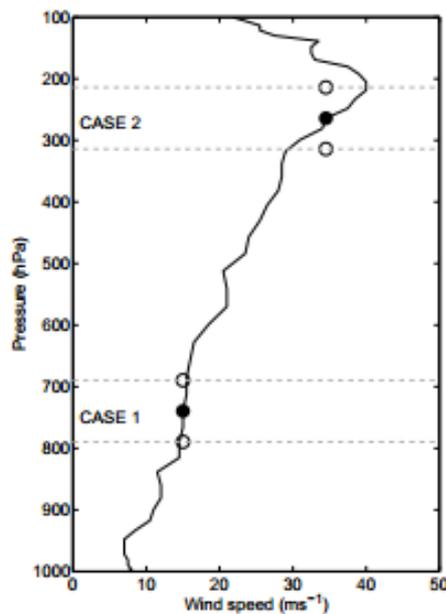


Рисунок 3.2 — Типичный вертикальный профиль ветра и возможная ошибка при одинаковой погрешности в определении давления на разных вертикальных уровнях [113, с. 2]

Исследования ошибок в наблюдениях АМВ показывают, что ошибка присвоенной наблюдению высоты может достигать 150—160 гПа в зависимости от применяемого метода определения, спутника и канала [112].

Естественным решением данной проблемы является переопределение высоты наблюдения АМВ в том случае, если присвоенная высота используемого в усвоении наблюдения не удовлетворяет некоторым установленным критериям. Такое переопределение возможно проводить с учётом спрогнозированного значения скорости и направления ветра в окрестностях точек физического пространства, в которых наблюдаемый с помощью данных АМВ ветер может содержать ошибку определения высоты.

Одним из способов переопределения высоты наблюдения АМВ является вычисление т.н. наилучшего давления, на котором разница между наблюдаемым вектором и вектором первого приближения минимальна. Такой подход применяется в системах усвоения ЕЦСПП и MetOffice [114]. В обоих центрах отмечено положительное влияние наблюдений АМВ на точность модельных прогнозов. В данной работе применяется алгоритм, предложенный в статье [141]. Его описание и детали применения описаны ниже. Кроме того, возможна коррекция высоты данных АМВ с помощью данных наблюдений, полученных с лидаров [135].

Для определения новой высоты или принятия решения об использовании присвоенного производителем значения скорости и направление наблюдаемого ветра АМВ сравниваются с модельными прогнозами.

Для каждого  $i$ -го наблюдения вектора скорости  $\mathbf{V}_i$  вычисляется коэффициент согласованности с  $m$ -м соседним вектором скорости  $\mathbf{V}_m$ :

$$\rho_{im} = D_{im}L_{im}, \quad (3.1)$$

где  $L_{im} = 1 - \frac{|\mathbf{V}_i| - |\mathbf{V}_m|}{\max(|\mathbf{V}_i|, |\mathbf{V}_m|)}$  — коэффициент согласованности скорости,

$D_{im} = \cos \theta_{im}$  — коэффициент согласованности направления,

$\theta_{im}$  — разность направления векторов.

Общая согласованность ветра АМВ в  $i$ -ой точке с окружающим полем ветра внутри заданного радиуса  $r$  на заданной в наблюдении высоте вычисляется

по формуле:

$$\bar{\rho}_i = \frac{\sum_m \rho_{im} G_{im}}{\sum_m G_{im}}, \quad (3.2)$$

где весовая функция  $G_{im}$  зависит от расстояния  $r_{im}$  между точкой  $i$ -го наблюдения и  $m$ -й точкой модельной сетки и вычисляется как

$$G_{im} = \begin{cases} \exp\left(1 - \frac{2r_{im}}{r}\right), & r_{im} \leq r, \\ 0 & r_{im} > r. \end{cases} \quad (3.3)$$

В зависимости от значения  $\bar{\rho}_i$  принимается решение о необходимости переопределения высоты. Если значение  $\bar{\rho}_i$  превышает некоторое заданное критическое значение  $\rho_0$  (в нашем случае 0,7), то значение высоты остаётся прежним. В случае, если этого не происходит, берутся несколько модельных вертикальных уровней в пределах, зависящих от первоначальной высоты наблюдения (максимум 250 гПа), и процедура повторяется для всех попадающих в этот промежуток уровней. Новая высота соответствует модельному уровню, на котором коэффициент согласованности  $\bar{\rho}_i$  максимален. Если на всех использованных для переопределения высоты уровнях  $\bar{\rho}_i < \rho_0$ , то наблюдение исключается из числа подлежащих усвоению.

В качестве радиуса, внутри которого поле спрогнозированного ветра полагалось подходящим для сравнения, был выбран радиус  $R = 200$  км. Как было отмечено выше, процедура переопределения исключает из числа используемых наблюдений те, для которых не удаётся найти высоту с хорошим согласованием наблюдаемого ветра со спрогнозированным, т.е. при  $\bar{\rho}_i < 0,7$ . В среднем от 5 до 10 % имеющихся наблюдений отфильтровываются на этом этапе.

Процедура переопределения высоты наблюдений AMV позволила уменьшить ошибки модельных прогнозов на заблаговременностях от 1 до 5 суток. На рисунках 3.3, 3.4 и 3.5 приводится сравнение осреднённых за январь 2015 года среднеквадратических ошибок прогностических переменных по внетропической части Северного полушария ( $20^\circ - 90^\circ$  с.ш.) и Южного полушария ( $20^\circ - 90^\circ$  ю.ш.) в зависимости от заблаговременности прогноза. Красный цвет — прогнозы, рассчитанные по анализам, включающим в себя усвоение AMV без переопределения высоты; синий цвет — прогнозы, стартовавшие с анализов, при получении которых высота AMV переопределялась. Как было указано выше,

для этих экспериментов применялась модель ПЛАВ в версии с постоянным горизонтальным разрешением по пространству  $0,9^\circ \times 0,72^\circ$  и 28 вертикальными сигма-уровнями. Верификации подвергались прогнозы, стартовавшие в срок 00 ВСВ. Усваивались наблюдения со спутника Meteosat-7, изображения в инфракрасном диапазоне. Приводятся средние за месяц значения и доверительные интервалы с уровнем доверия 0,95.

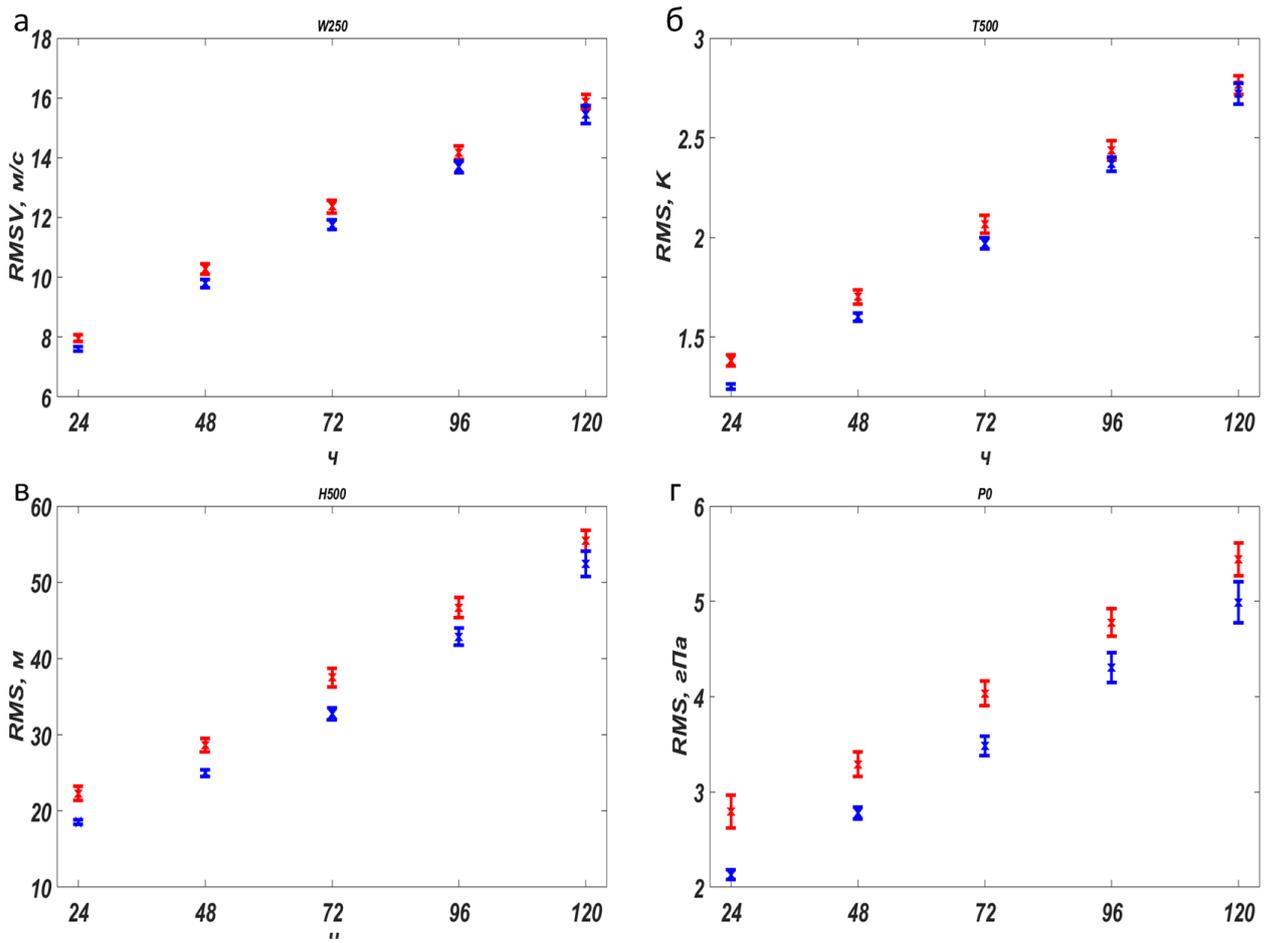


Рисунок 3.3 — Среднеквадратическая ошибка полей прогнозов и её доверительный интервал. Верификация по полям оперативного анализа Гидрометцентра России. Прогностические поля: а) вектор скорости ветра на поверхности 250 гПа; б) температура на поверхности 500 гПа; в) давление на уровне моря; г) высота поверхности 500 гПа.

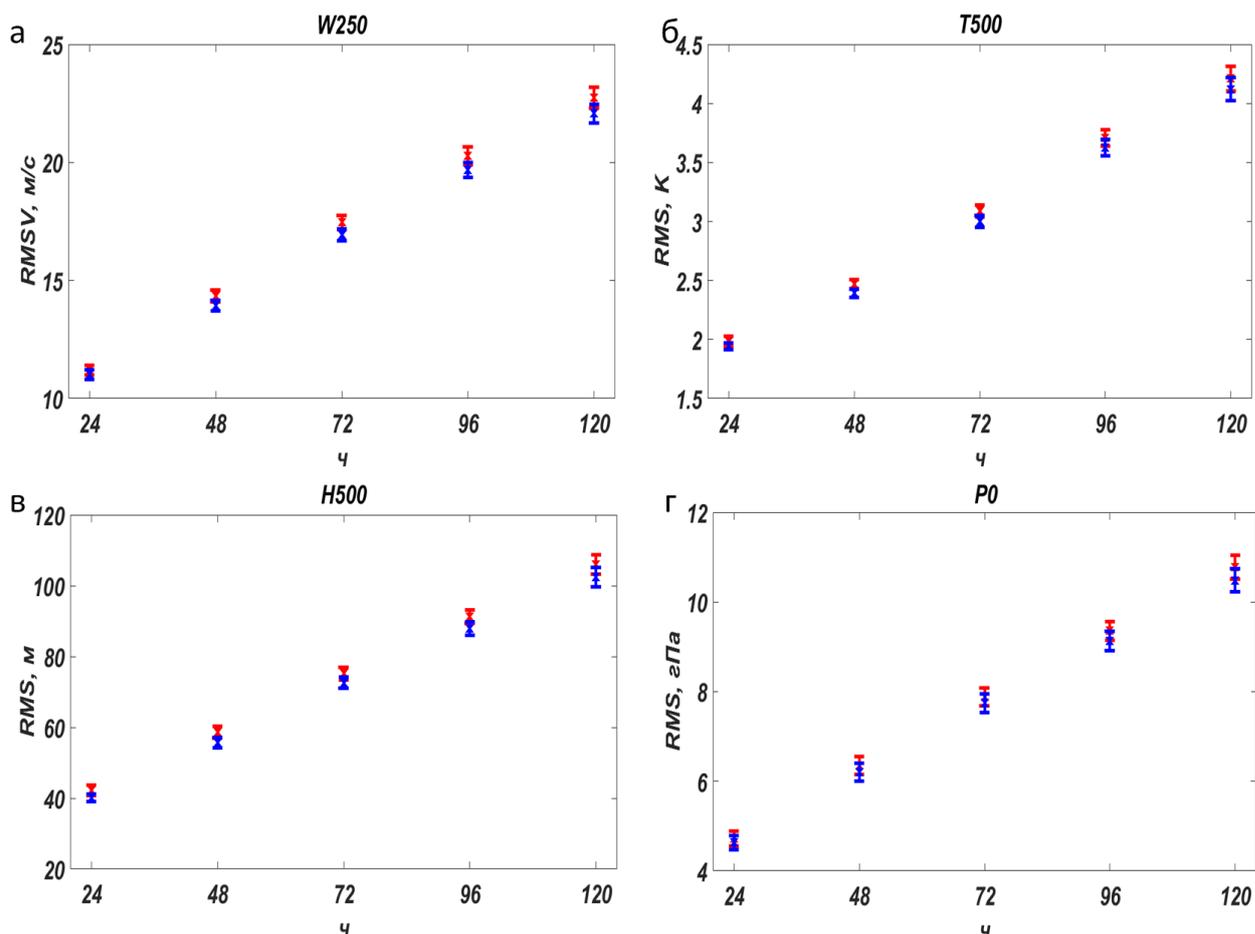


Рисунок 3.4 — Среднеквадратическая ошибка полей прогнозов и её доверительный интервал. Верификация по полям оперативного анализа Гидрометцентра России. Прогностические поля: а) вектор скорости ветра на поверхности 250 гПа; б) температура на поверхности 500 гПа; в) давление на уровне моря; г) высота поверхности 500 гПа.

### 3.5 Автокорреляция ошибок наблюдений AMV и методы её учёта

Во время проведения каждого из этапов формирования наблюдения возможно внесение ошибок в определение всех величин, содержащихся в наблюдении. Этот факт обуславливает сложную структуру общей ошибки, а также наличие пространственной и временной корреляции этих ошибок.

Например, используемые для определения высоты профили температуры из прогностической модели, обычно пространственно скоррелированы, а значит, будут скоррелированы и ошибки наблюдений. Кроме того, обнаружение похожих облачных структур в соседних сегментах может привести к похожим

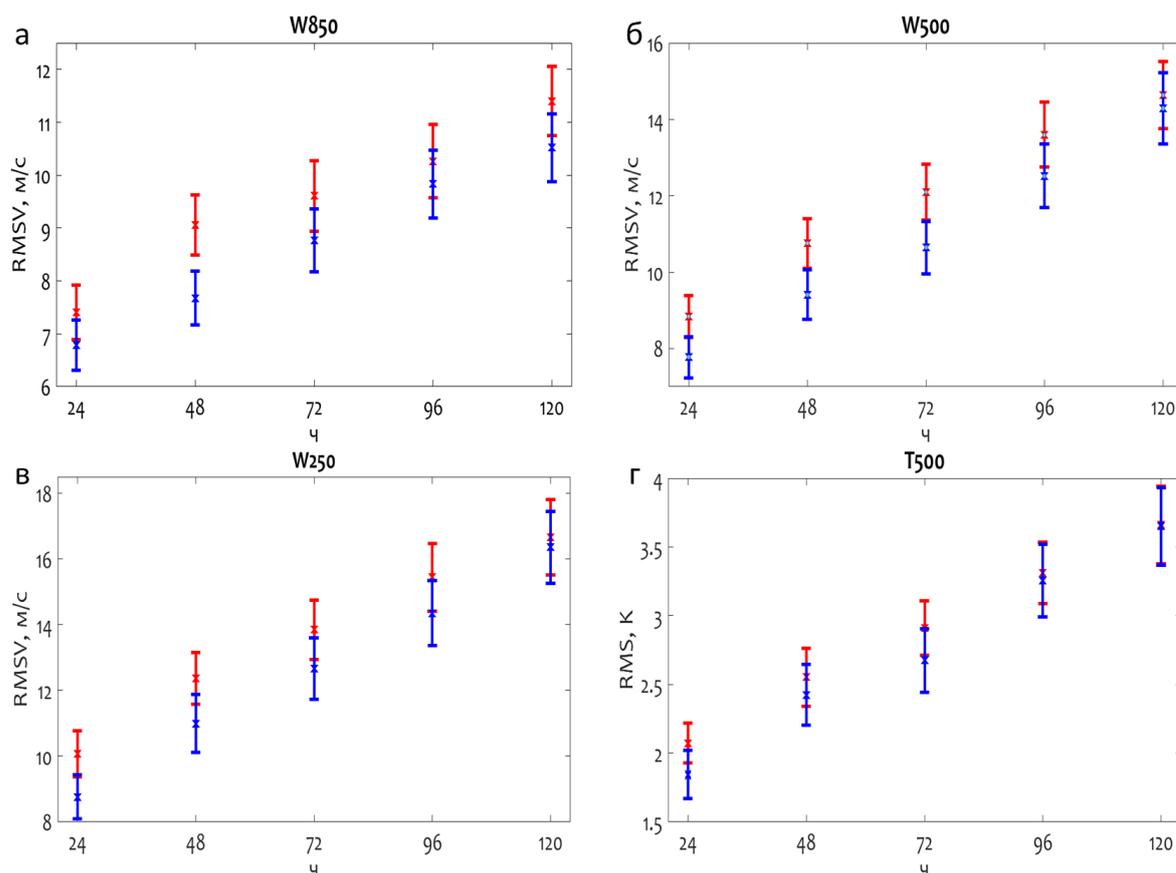


Рисунок 3.5 — Среднеквадратическая ошибка полей прогнозов и её доверительный интервал. Верификация по данным радиозондовых наблюдений. Прогностические поля: а) вектор скорости ветра на поверхности 850 гПа; б) вектор скорости ветра на поверхности 500 гПа; в) вектор скорости ветра на поверхности 250 гПа; г) температура на поверхности 500 гПа.

ошибкам обнаружения или вычисления высоты. Процедуры контроля качества также сильнее доверяют ветрам, близкими по скорости и направлению с ветрами в соседних точках, что увеличивает вероятность скоррелированных ошибок [44].

Исследование пространственной корреляции ошибок наблюдений AMV проводится в [44]. Наблюдения AMV сравнивались с наблюдениями радиозондов и рассчитывались корреляционные функции наблюдений AMV в предположении, что ошибки радиозондов и AMV не коррелированы. Результаты исследования показали, что величина автокорреляции ошибок наблюдений ветра AMV для разных спутников и каналов получения данных для близко расположенных наблюдений составляют от 0,25 до 0,5. При этом радиусы корреляций достигают 800 км.

На рисунке 3.6 приведены графики функций корреляций из [44, с. 710] для различных спутников в диапазоне водяного пара. Графики функций корреляций для различных спутников в инфракрасном диапазоне показывают похожие результаты.

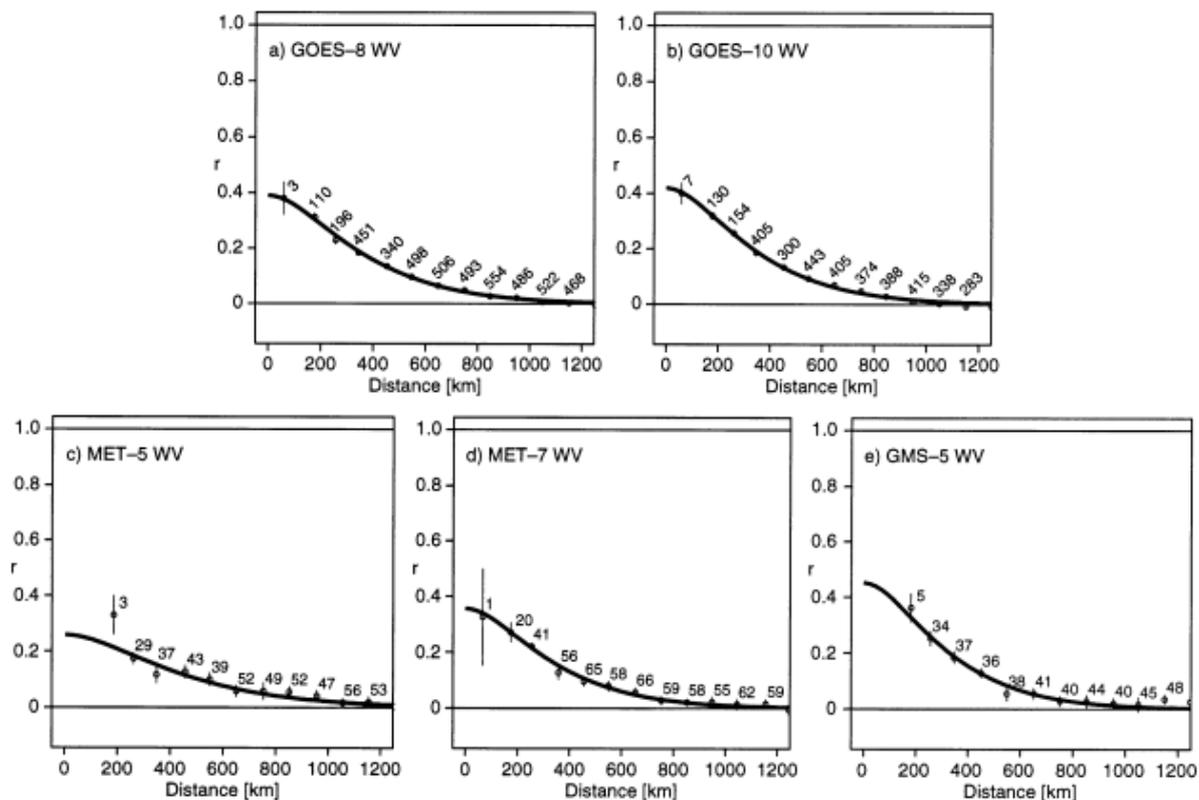


Рисунок 3.6 — Корреляции ошибок (AMV-радиозонд) в зависимости от расстояния для Северного полушария [44, с. 710]

Методы учёта корреляций ошибок наблюдений описаны в разделе 2.3. Простейшим способом является прореживание наблюдений путём расчёта или выбора в заданной пространственной ячейке одного единственного наблюдения. При этом размер ячейки выбирается так, чтобы расстояние между соседними наблюдениями превышало типичный радиус пространственной корреляции. Такими наблюдениями могут стать:

- осреднённые по ячейке (супернаблюдения);
- обладающие наиболее высоким индикатором качества, присвоенным производителем наблюдений;
- имеющие наименьшую разность с первым приближением;
- имеющие наименьшую вариацию скорости ветра в сравнении с наблюдениями в соседних точках и т.д.

При этом дисперсии ошибок наблюдений несколько завышаются, а наличие корреляций ошибок наблюдений игнорируется, то есть матрица  $\mathbf{R}$ , фигурирующая в формулах фильтра, полагается диагональной, что упрощает проводимые вычисления, т.к. не требует операций по обращению этой матрицы.

Алгоритм прореживания был реализован в программе, зарегистрированной в Роспатенте. Наилучшим считается наблюдение с наибольшим индикатором качества. В случае, если у двух или более индикатор совпадает, выбирается наблюдение, географические координаты которого ближе всего к центру ячейки.

Такой подход оказывается лучше, чем использование заведомо большого числа наблюдений с заведомо зависимыми ошибками в предположении, что они независимы (с диагональной матрицей  $\mathbf{R}$ ), но хуже, чем правильный учёт корреляций при большем количестве наблюдений [54; 119]. Так происходит, потому что исключение существенного количества наблюдений и отказ от учёта корреляции в их ошибках лишает систему части информации, которая могла бы быть использована более эффективно.

Невозможность использования «правильной» матрицы  $\mathbf{R}$ , учитывающей наличие автокорреляций в ошибках наблюдений, приводит к необходимости определения её элементов с помощью различных ковариационных функций. Одной из таких функций является предложенная в [44] авторегрессионная функция второго порядка (англ. Second Order Autoregressive function, SOAR). Её подробное описание приводится в разделе 3.5.2. Учёт недиагональных элементов более точно отражает пространственную структуру ковариаций ошибок наблюдений, что положительно сказывается в т.ч. и на мелкомасштабной составляющей ошибок полученных результатов.

Подробное описание этого эффекта и эксперименты по его выявлению приводятся в [108], где в качестве наблюдений использовались искусственно сгенерированные наблюдения, имеющие заданную матрицу ковариаций ошибок с недиагональными элементами отличными от нуля. Исследование свойств полей, получаемых при различных видах участвующей в усвоении матрицы  $\mathbf{R}$ , проводится с помощью спектрального анализа.

При использовании моделирования  $\mathbf{R}$  становится многодиагональной матрицей, что усложняет её обращение. Однако, при локализации наблюдений, описанной в разделе 1.3.5, размер локальной матрицы позволяет осуществлять об-

ращение довольно быстро. Потеря вычислительного времени на обращение не представляется недопустимо большой.

### 3.5.1 Влияние прореживания наблюдений

В соответствии с выводами об ошибках наблюдений, представленными в разделе 3.5, для уменьшения негативного влияния коррелированных ошибок в наборе усваиваемых данных AMV, было реализовано прореживание имеющегося набора наблюдений. Для этих целей физическое пространство было поделено на ячейки размером  $200 \times 200$  км по горизонтали и  $50 - 175$  гПа по вертикали. В каждой такой ячейке выбирается наблюдение, имеющее наибольшее значение QI (индикатора качества). При равенстве этих значений у двух или более наблюдений выбирается ближайшее по времени к дате анализа. Кроме того, в целях уменьшения набора усваиваемых данных и уменьшения возможной ошибки, вызванной тем, что некоторые наблюдения сделаны в далёкие от даты анализа моменты времени, в усвоении участвовали лишь те наблюдения AMV, которые попадали в окно  $\pm 1$  час от времени анализа. Матрица  $\mathbf{R}$  в описанных экспериментах по усвоению предполагалась диагональной, содержащей дисперсии ошибок наблюдений AMV, значения которых были взяты из [69].

Такое прореживание уменьшило количество обрабатываемых наблюдений с 250000 за один срок анализа до 50000. Глобальное покрытие наблюдениями при этом сохраняется, хотя и уменьшается пространственное разрешение. Рассчитанные по данным радиозондовых наблюдений ошибки прогнозов, стартовавших с полученных анализов, уменьшились на всех сроках от 1 до 5.

### 3.5.2 Влияние учёта корреляций ошибок наблюдений

Следующим этапом стала проверка влияния моделирования недиагональных элементов матрицы  $\mathbf{R}$  в схеме усвоения. Использование недиагональной матрицы  $\mathbf{R}$  в алгоритме LETKF требует её обращения, что сказывается на

общем времени выполнения параллельной программы, которая реализует усвоение данных.

Для локального набора данных AMV, которые попадают в установленный радиус влияния на анализ в точке модельной сетки, матрица  $\mathbf{R}'$  модерировалась в соответствии с матричным уравнением

$$\mathbf{R}' = \mathbf{D}\rho\mathbf{D}. \quad (3.4)$$

Здесь диагональная матрица, на диагонали которой стоят значения ошибок наблюдений, зависящие от расстояния между рассматриваемой точкой сетки и точкой наблюдения, попавшего в радиус локализации

$$\mathbf{D} = \begin{pmatrix} \sqrt{\sigma_1^2} & \dots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \dots & \sqrt{\sigma_{nloc}^2} \end{pmatrix}, \quad (3.5)$$

$nloc$  — количество попавших в радиус локализации наблюдений.

Элементы корреляционной матрицы  $\rho$  вычисляются с помощью предложенной в [44] т. н. авторегрессионной функции второго порядка (англ. Second Order Autoregressive function, SOAR):

$$\rho_{ij} = R_0 \left(1 + \frac{r_{ij}}{L}\right) e^{-\frac{r_{ij}}{L}}, i = \overline{1, nloc}, j = \overline{1, nloc}. \quad (3.6)$$

Здесь  $R_0 > 0$  — настроечный параметр, отвечающий за долю коррелированной части ошибки,  $L > 0$  — настроечный параметр, измеряемый в км и отвечающий за расстояние, на котором влияние последнего сказывается на величине корреляции,  $r_{ij}$  — расстояние на сфере между двумя наблюдениями. Первые два параметра варьируются исходя из диапазона, широты и высоты наблюдения.

Следует отметить, что внесение в схему усвоения дополнительного вычисления элементов матрицы  $\mathbf{R}$  и её последующего обращения для каждой точки модельной сетки существенно увеличивает общее количество вычислений в параллельной программе, реализующей алгоритм усвоения данных. Для того, чтобы оптимизировать время счёта программы анализа для одной точки модельной сетки, была применена блочная структура матрицы  $\mathbf{R}$ . Формирование элементов каждого из таких блоков, количество которых совпадает с количеством наблюдений AMV различного вида, а также последующее обращение та-

ких блоков в описываемой схеме проводится по отдельности для каждого блока. Такой подход в сочетании с хорошей оптимизацией встроенных в библиотеку LAPACK процедур линейной алгебры помогает добиться существенного снижения времени счёта параллельной программы.

Так, полное время, необходимое для работы реализующей алгоритм LETKF программы, которая написана на языке Фортран с использованием параллельных библиотек MPI (распараллеливание расчётной сетки по широтам) и OpenMP (распараллеливание по долготам), при использовании 992 ядер (248 процессов MPI, порождающих по 4 потока OpenMP) возрастает примерно в 2 раза с 5-6 минут до 10-12 минут. Это время в дальнейшем может быть несколько уменьшено путём подбора оптимального соотношения радиусов корреляции из уравнения 3.6 и радиусов локализации наблюдений AMV вокруг точек модельной сетки. Также общее время работы программы может быть уменьшено на несколько десятков процентов за счёт увеличения вычислительных ресурсов (увеличение количества потоков OpenMP). В настоящее время описываемая система усвоения работает в Гидрометцентре России в оперативном режиме для генерации начальных данных в составе системы ансамблевого прогноза, один запуск которой при расчёте прогнозов на 10 суток длится порядка полутора часов. Возрастание общего времени счёта на 6 минут не представляется критическим при повышении точности начальных данных и прогнозов.

Для оценки эффекта от применения моделирования ковариаций ошибок наблюдений при усвоении данных наблюдений AMV были проведены несколько серий экспериментов. В ходе каждого из проведённых экспериментов сравнивались среднеквадратические ошибки анализов или прогнозов, стартовавших с сгенерированных начальных данных.

В этих экспериментах использована версия модели ПЛАВ [23] с тем же горизонтальным разрешением, но с гибридной вертикальной координатой (96 уровней). Она применялась как для получения ансамбля полей первого приближения в системе усвоения, так и для расчёта прогнозов с заблаговременностью от 24 до 120 часов с дискретностью одни сутки.

В каждом из приведённых ниже экспериментов сравниваются результаты усвоения несколько различающихся наборов данных наблюдений AMV. В одном из них данные были прорежены путём выбора одного единственного наблюдения в ячейке размером  $100 \text{ км} \times 100 \text{ км} \times 10 \text{ гПа}$ , имеющего наибольший

индикатор качества, который присваивается каждому наблюдению на этапе его получения производителем. При усвоении этого набора матрица  $\mathbf{R}$  полагалась диагональной, дисперсии несколько завышены относительно другого набора (в зависимости от спутника, региона и высоты от 1,2 до 1,5 раза), в котором данные не прореживались. Радиусы локализации наблюдений вокруг точек модельной сетки полагались одинаковыми.

Для определения величины элементов корреляционной матрицы  $\rho$  использовались значения настроечных параметров функции SOAR, представленные в таблице 3. Значения этих параметров были взяты из работы [44]. Их подбор и оценка чувствительности точности анализов и прогнозов к их значениям в рамках данной работы не проводились.

Таблица 3 — Таблица значений настроечных параметров авторегрессионной функции второго порядка (SOAR)

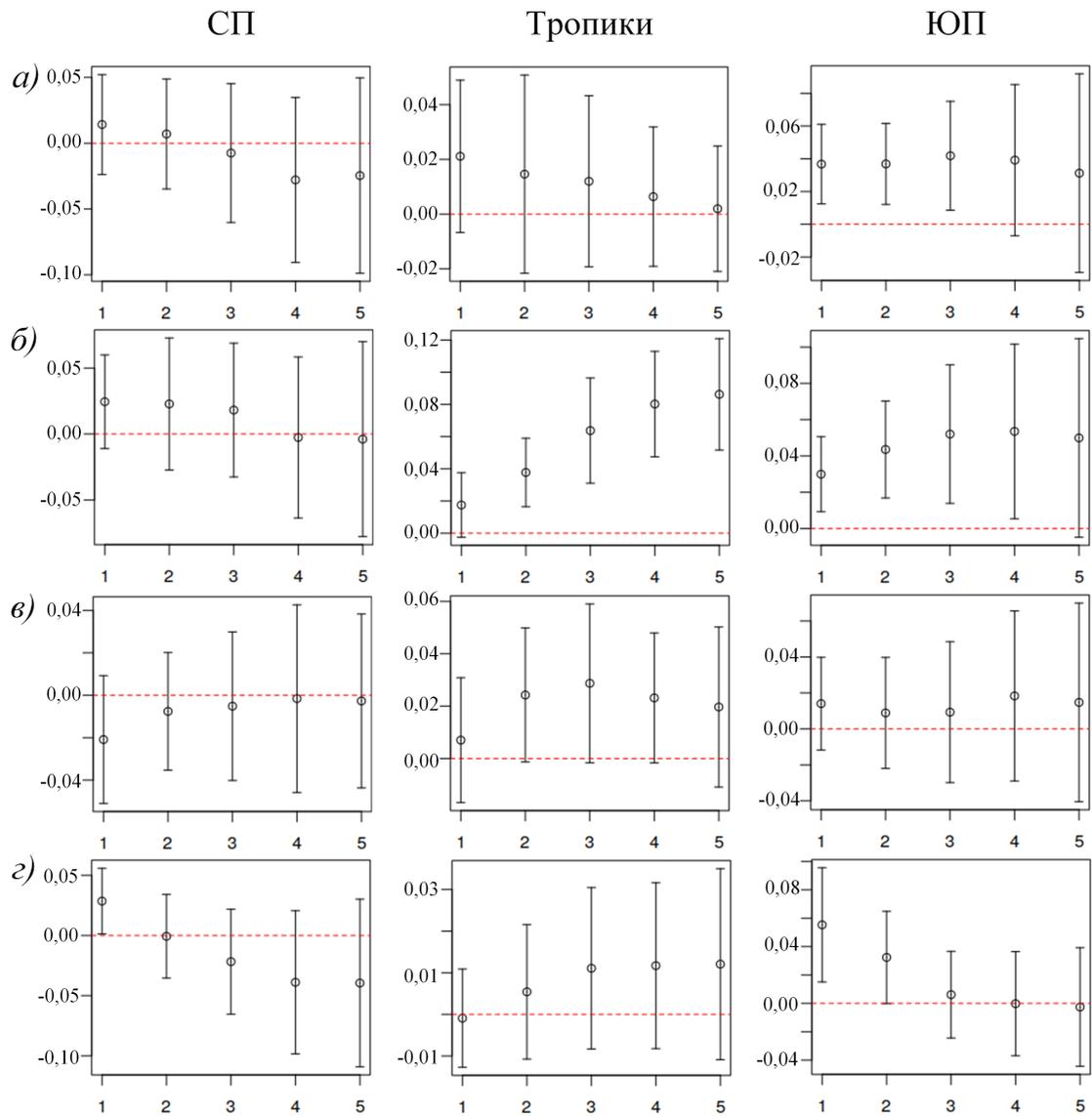
Спутник	Диапазон значений высоты	Внетропическая часть Северного полушария	Тропики	Внетропическая часть Южного полушария
GOES-13/15	Выше 400 гПа	$R_0 = 0,35$ $L = 200$ км	$R_0 = 0,27$ $L = 340$ км	$R_0 = 0,37$ $L = 170$ км
	От 400 до 700 гПа	$R_0 = 0,42$ $L = 210$ км	$R_0 = 0,29$ $L = 280$ км	$R_0 = 0,43$ $L = 190$ км
	Ниже 700 гПа	$R_0 = 0,40$ $L = 210$ км	$R_0 = 0,30$ $L = 310$ км	$R_0 = 0,41$ $L = 200$ км
METEOSAT-7	Выше 400 гПа	$R_0 = 0,29$ $L = 170$ км	$R_0 = 0,29$ $L = 180$ км	$R_0 = 0,31$ $L = 190$ км
	От 400 до 700 гПа	$R_0 = 0,44$ $L = 130$ км	$R_0 = 0,30$ $L = 190$ км	$R_0 = 0,35$ $L = 160$ км
	Ниже 700 гПа	$R_0 = 0,50$ $L = 210$ км	$R_0 = 0,29$ $L = 170$ км	$R_0 = 0,28$ $L = 150$ км
METEOSAT-9	Выше 400 гПа	$R_0 = 0,25$ $L = 140$ км	$R_0 = 0,29$ $L = 170$ км	$R_0 = 0,28$ $L = 150$ км
	От 400 до 700 гПа	$R_0 = 0,35$ $L = 130$ км	$R_0 = 0,31$ $L = 200$ км	$R_0 = 0,38$ $L = 170$ км
	Ниже 700 гПа	$R_0 = 0,37$ $L = 180$ км	$R_0 = 0,35$ $L = 190$ км	$R_0 = 0,37$ $L = 180$ км
HIMAWARI-7	Выше 400 гПа	$R_0 = 0,29$ $L = 320$ км	$R_0 = 0,27$ $L = 340$ км	$R_0 = 0,29$ $L = 320$ км
	От 400 до 700 гПа	$R_0 = 0,30$ $L = 320$ км	$R_0 = 0,35$ $L = 310$ км	$R_0 = 0,30$ $L = 350$ км
	Ниже 700 гПа	$R_0 = 0,35$ $L = 290$ км	$R_0 = 0,35$ $L = 300$ км	$R_0 = 0,35$ $L = 290$ км
LeoGeo	Выше 400 гПа	$R_0 = 0,46$ $L = 150$ км	— —	$R_0 = 0,40$ $L = 150$ км
	От 400 до 700 гПа	$R_0 = 0,32$ $L = 130$ км	— —	$R_0 = 0,35$ $L = 180$ км
	Ниже 700 гПа	$R_0 = 0,35$ $L = 130$ км	— —	$R_0 = 0,38$ $L = 200$ км

Сравнение результатов, полученных в ходе экспериментов по усвоению двух наборов наблюдений (с использованием предположения о корреляции и без его использования) позволяет сделать вывод о том, какой подход является более предпочтительным: простое прореживание с применением диагональной матрицы  $\mathbf{R}$  и искусственным завышением дисперсий или полный набор с моделированием корреляций ошибок наблюдений (в т. ч. недиагональных элементов матрицы  $\mathbf{R}$ ).

В первой серии экспериментов для каждого набора прогнозов с помощью описанной в разделе 3.2 системы верификации собственной разработки были посчитаны среднеквадратические ошибки прогнозов в период времени с января по март 2021 года относительно оперативного анализа Гидрометцентра России. Для осреднённых по регионам внетропических частей Северного (СП, n.hem) и Южного (ЮП, s.hem) полушарий, а также по тропическому региону (тропики, tropics), среднеквадратических ошибок прогнозов каждой из заблаговременностей в течение трёх расчётных месяцев формировались относительные изменения среднеквадратических ошибок из первой (сформирована из среднеквадратических ошибок прогнозов, стартовавших с начальных данных, полученных без использования предположения о наличии корреляций в ошибках AMV) и второй выборки (состоит из среднеквадратических ошибок прогнозов, начальные данные для которых получены при усвоении с предположением о корреляции ошибок в наблюдениях AMV).

На графиках рисунка 3.7 нанесены средние значения этого изменения и доверительный интервал с уровнем доверия 95 %. Ввиду малого объёма выборок (порядка 120 значений), для вычисления доверительных интервалов использовалась процедура Бутстреп (bootstrap). Из представленных графиков можно сделать вывод о том, что в целом использование предположения о наличии корреляций в ошибках и моделирование матрицы ковариации ошибок наблюдений  $\mathbf{R}$  помогает уменьшить среднеквадратическую ошибку прогноза. Средняя разница ошибок в выборках находится выше нулевого уровня в большинстве случаев, особенно в тропическом регионе и внетропической части Южного полушария. При этом наиболее значимый положительный эффект можно увидеть для значений скорости ветра в тропическом регионе. Среднеквадратическая ошибка прогнозов скорости ветра на уровне 500 гПа (W500) для больших заблаговременностей уменьшается на величину около 8%. Во внетропической

части Южного полушария среднеквадратическая ошибка скорости ветра также уменьшается. Судя по всему, такой эффект возникает вследствие меньшего количества контактных (более точных) наблюдений в этих областях, а значит и положительный эффект от более точного описания корреляций в ошибках наблюдений AMV сказывается сильнее.



Заблаговременность прогноза, сутки

Рисунок 3.7 — Разности и доверительные интервалы этих разностей, вычисленные для среднеквадратических ошибок прогнозов, стартовавших с начальных данных, подготовленных без усвоения AMV и при усвоении AMV: модуля скорости ветра на уровнях 250 гПа (а) и 500 гПа (б), геопотенциала изобарической поверхности 250 гПа (в) и температуры на уровне 500 гПа (г)

Кроме представленных выше, были проведены также и эксперименты, призванные оценить влияние применения схемы усвоения данных наблюдений AMV с учётом корреляции ошибок наблюдений на точность полей анализа. Для этого из числа усваиваемых наблюдений были исключены данные с радиозондов, а полученные поля анализов верифицировались с помощью исключённых данных для получения независимой оценки анализов. Полученные результаты представлены на рисунке 3.8.

Кружками различного цвета показаны результаты сравнения двух выборок состоящих из среднеквадратических ошибок полученных полей анализов на изобарических поверхностях 850 гПа, 500 гПа и 250 гПа относительно не участвующих в усвоении данных радиозондовых наблюдений. Усваивались данные наблюдений за период с января по март 2022 года, верифицировались поля анализов в сроки 00 часов всемирного скоординированного времени. Как и в первой серии экспериментов, начальные данные для первой выборки получены без использования коррелированных ошибок данных наблюдений AMV с прореженными наблюдениями и несколько завышенными дисперсиями. Для второй выборки начальные данные получены с использованием коррелированных ошибок AMV без прореживания. Аналогично схеме сравнения двух выборок в первой серии экспериментов, исследовались средние разности двух выборок, состоящие из ошибок анализов. Закрашенная часть круга показывает процентное отношение изменения среднеквадратической ошибки. Эффект тем сильнее и более выраженный, чем темнее и больше кружок на рисунке. Круг чёрного цвета показывает относительное изменение среднеквадратической ошибки величиной в 1%. Зелёным цветом показаны случаи уменьшения среднеквадратической ошибки, красным – увеличения. Яркий цвет показывает различие, при котором доверительный интервал целиком лежит выше (или ниже) нулевого уровня, неяркий цвет – интервал нулевой уровень пересекает.

Наибольший эффект виден для ошибок модуля скорости ветра на уровнях 250 и 500 гПа, где сосредоточено основное количество наблюдений AMV (рисунок 3.8). Из представленных графиков можно сделать вывод, что использование предположения о наличии корреляций в ошибках наблюдений уменьшает среднеквадратическую ошибку анализа (яркие зелёные кружки наблюдаются для ветра в тропической части, для большинства остальных переменных) При этом значимые улучшения для большинства переменных отсутствуют.

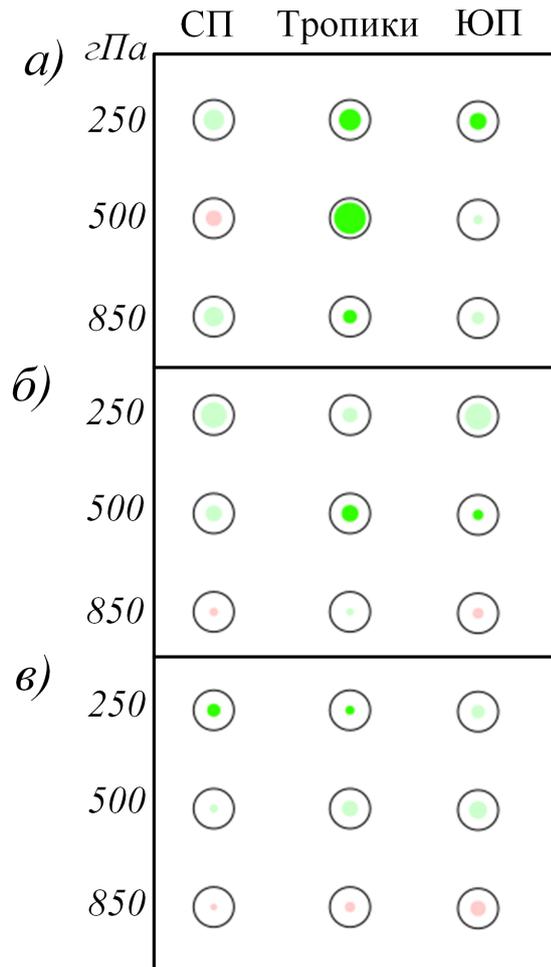


Рисунок 3.8 — Сравнение выборок, состоящих из среднеквадратических ошибок полей анализов относительно данных радиозондов для различных полей на стандартных синоптических уровнях: скорости ветра (а); температуры (б) и геопотенциала (в). Закрашенная часть круга показывает процентное отношение изменения среднеквадратической ошибки

В качестве более детальной пространственной верификации можно использовать спектральный анализ полей ошибок для оценки мелкомасштабной составляющей этих ошибок. Подробное описание подобных экспериментов приведено в [88], где в качестве наблюдений использовались искусственно сгенерированные наблюдения, имеющие заданную матрицу ковариаций ошибок с недиагональными элементами отличными от нуля. Исследование свойств полей, получаемых при различных видах участвующей в усвоении матрицы  $\mathbf{R}$ , проводится с помощью спектрального анализа.

Для оценки эффекта применения недиагональной матрицы  $\mathbf{R}$  в наших экспериментах применялась процедура, подобная описанной в [88]. Исследовались спектральные разложения полей ошибок скорости ветра для января 2021 г.

Полученные в обеих сериях поля анализов скорости ветра интерполировались на сетку  $1^\circ \times 1^\circ$ . Далее считалась разность этих полей с полями оперативного анализа Гидрометцентра России (рассчитываемого независимо), переинтерполированных на ту же сетку. Поле ошибки для каждого срока подвергалось спектральному преобразованию. Возведённые в квадрат спектральные коэффициенты осреднялись далее по зональным волновым числам для каждого полного волнового числа. Результирующая зависимость осреднялась по срокам получения анализа в эксперименте. Указанная процедура применялась к ошибкам ветра на стандартных уровнях.

На рисунке 3.9 предоставлены полученные таким образом графики для анализов ошибки модуля скорости ветра.

При использовании коррелированных ошибок максимальный отклик можно увидеть на поверхности 250 гПа, где находится абсолютное большинство наблюдений. Ошибки анализа ветра на этих высотах уменьшились практически по всем значениям волновых чисел. Наименьший отклик на применение коррелированных ошибок при усвоении наблюдений AMV можно увидеть для ветра на поверхности 850 гПа (на графиках не приводится). Уменьшение ошибки на больших номерах волновых чисел, свидетельствующее об уменьшении мелко-масштабной составляющей ошибки полученных анализов ветра, достигает для некоторых волновых чисел величины 8,5%.

### **3.6 Влияние включения данных AMV на результаты среднесрочных прогнозов**

Для оценки влияния данных AMV на результаты среднесрочных прогнозов между собой сравнивались осреднённые за май 2023 года среднеквадратические ошибки прогнозов во внетропических частях Северного и Южного полушарий, а также в тропиках. Расчёт прогнозов стартовал с начальных данных, при подготовке которых использовались и не использовались данные спутниковых наблюдений ветра AMV. На рисунках 3.10 – 3.12 представлены результаты этого сравнения. Доверительные интервалы посчитаны с применением t-критерия Стьюдента с уровнем доверия 95 %. Длина анализируемых выборок – 62.

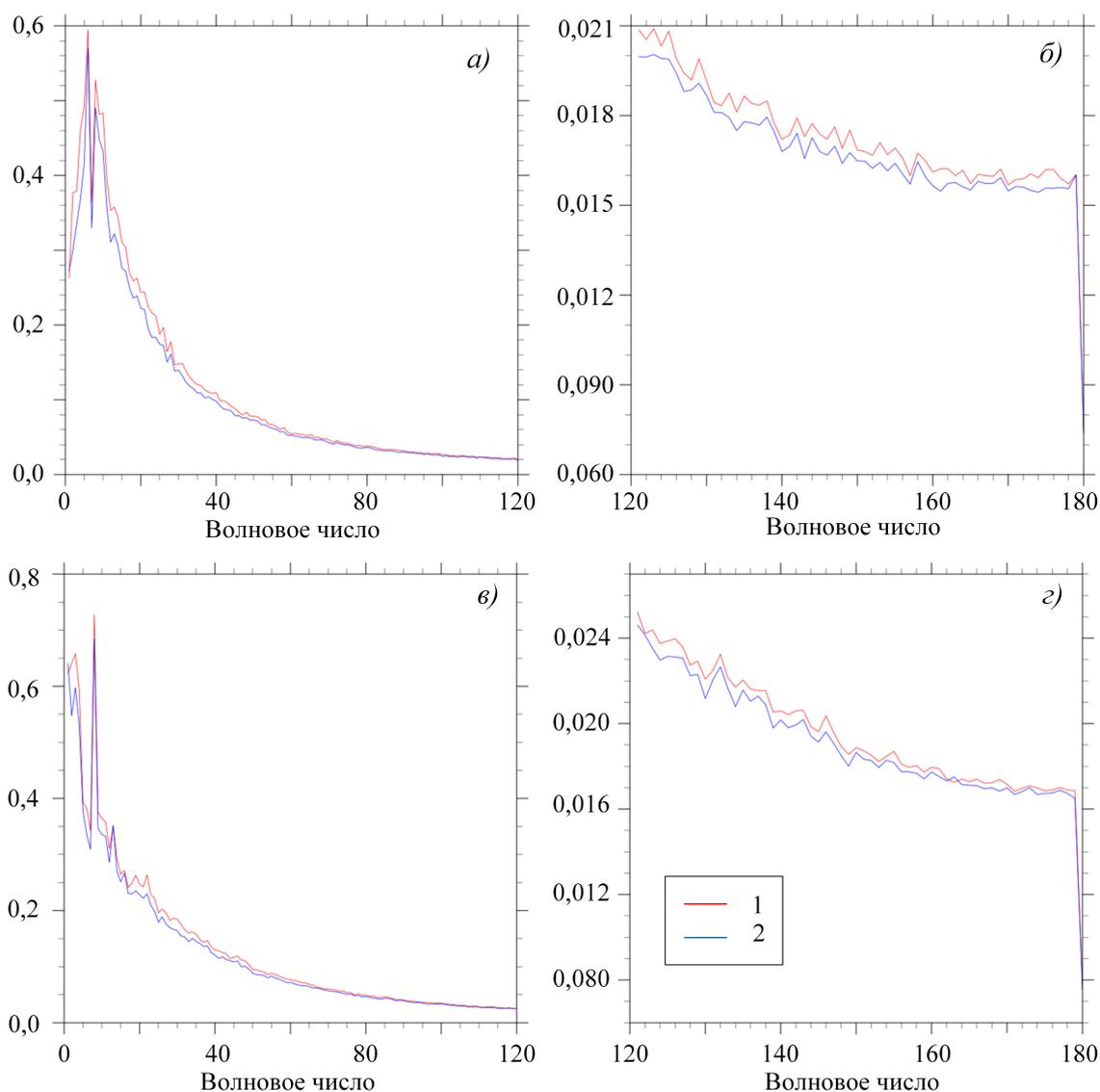


Рисунок 3.9 — Спектры средних ошибок скорости ветра для полей анализа на уровнях 500 гПа (нижний ряд) и 250 гПа (верхний ряд). Красная линия — спектр ошибки анализа, полученного без моделирования корреляций ошибок наблюдений AMV, синяя — применялось моделирование корреляции. Слева показаны графики для набора волновых чисел от 1 до 120, справа — та же кривая при волновых числах от 121 до 180

Значимость различия оценивалась с помощью сравнения ежедневной разности оценок аналогично экспериментам на рисунке 3.7 (зелёным цветом обведены столбцы, где изменения значимы в сторону уменьшения ошибки, красным — в сторону увеличения). Красным цветом показана ошибка прогнозов без усвоения данных AMV, синим — при усвоении AMV. Над столбцами показана величина относительного изменения ошибки в процентах.

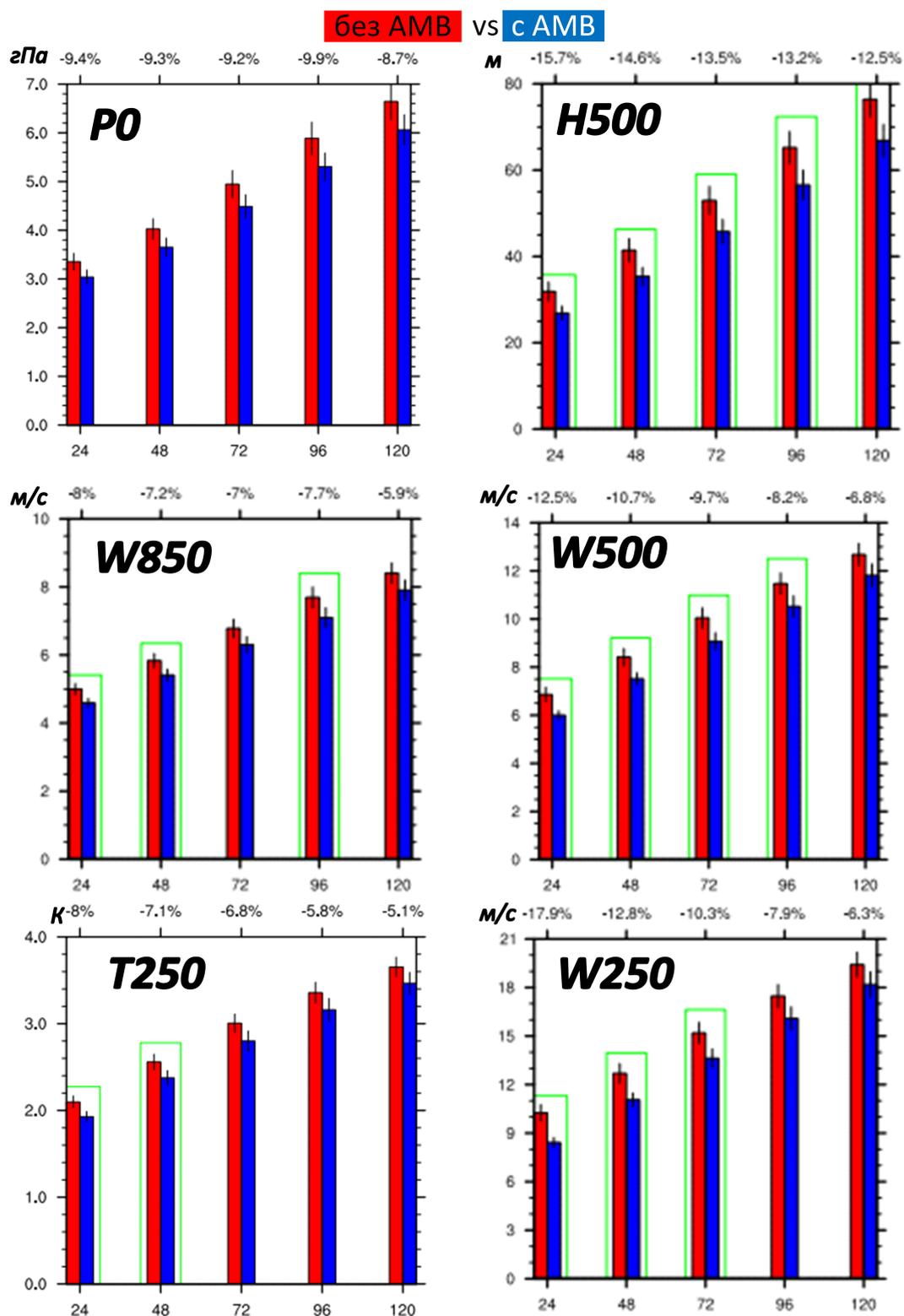


Рисунок 3.10 — Среднеквадратические ошибки прогнозов без усвоения AMV (красные столбцы) и при усвоении AMV (синие столбцы) в зависимости от заблаговременности прогноза во внетропической части Северного полушария.

Май 2023 года. Обозначения переменных приведены в разделе 3.2

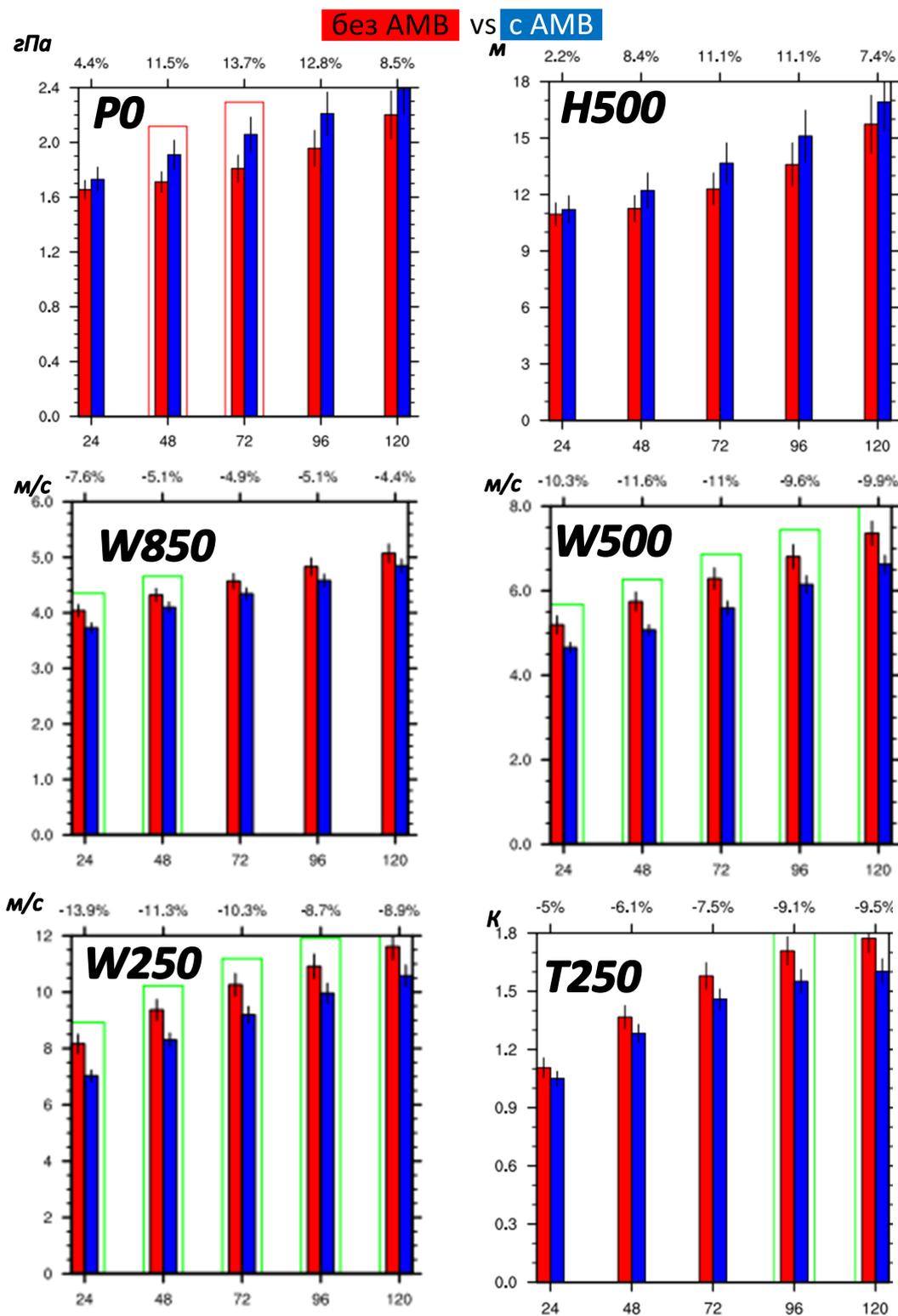


Рисунок 3.11 — Среднеквадратические ошибки прогнозов без усвоения AMV (красные столбцы) и при усвоении AMV (синие столбцы) в зависимости от заблаговременности прогноза в тропиках. Май 2023 года. Обозначения переменных приведены в разделе 3.2

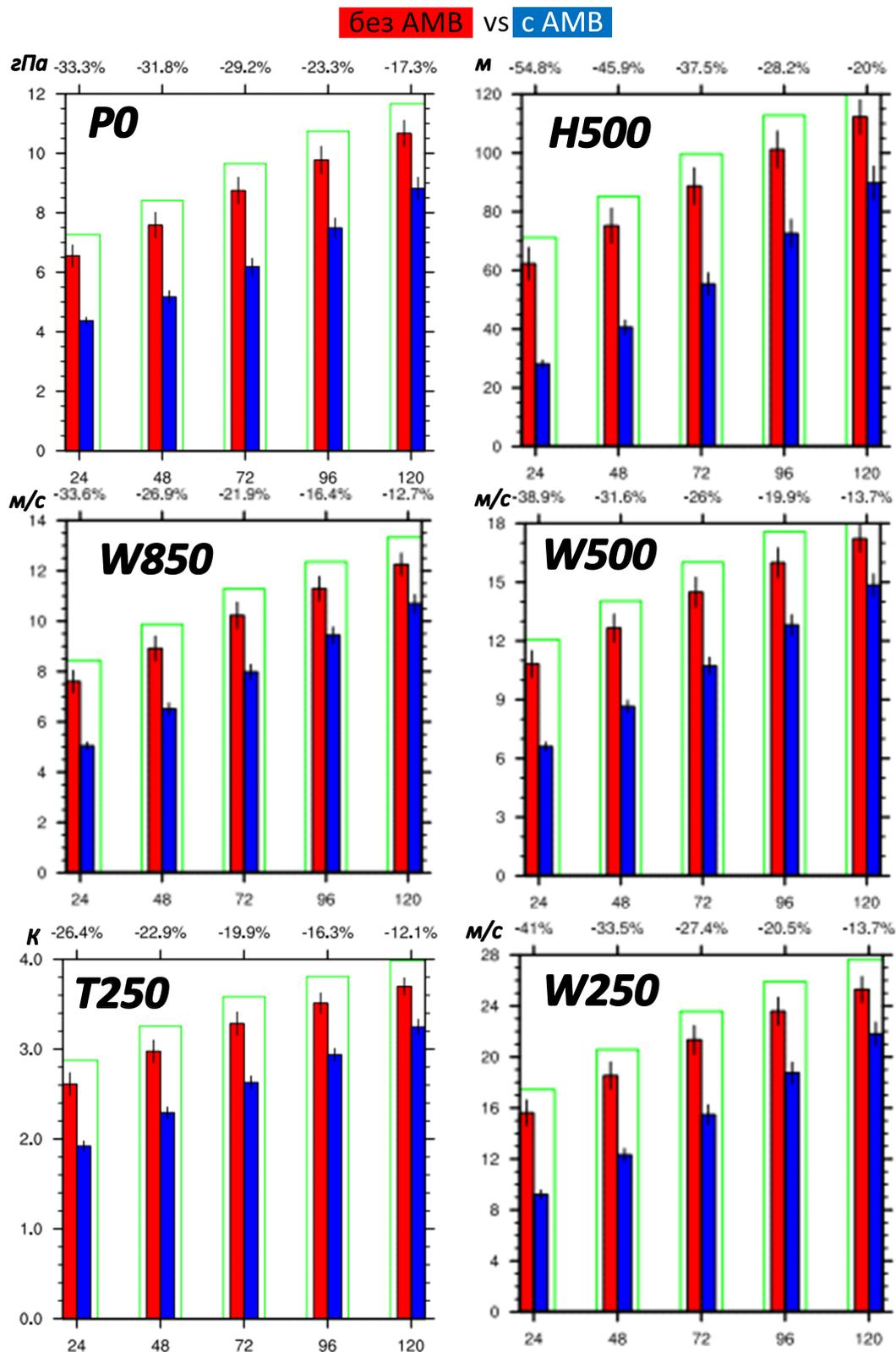


Рисунок 3.12 — Среднеквадратические ошибки прогнозов без усвоения AMV (красные столбцы) и при усвоении AMV (синие столбцы) в зависимости от заблаговременности прогноза во внетропической части Южного полушария. Май 2023 года. Обозначения переменных приведены в разделе 3.2

Максимальный эффект от применения данных AMV наблюдается во внетропической части Южного полушария (значимые уменьшения ошибки составили от 12 до 54 %). Этот факт можно объяснить тем, что в этом регионе очень мало других наблюдений, поэтому добавление дополнительного источника информации (данные наблюдений AMV) делает оценку состояния атмосферы более точной. Для внетропической части Северного полушария эффект от добавления наблюдений AMV менее заметен ввиду наличия большого количества наблюдений других типов, благодаря которым оценка состояния атмосферы изначально имеет высокую точность. При этом уменьшение ошибки при применении AMV наблюдается и в этом регионе (значимые уменьшения ошибки составили от 7 до 17 %). Увеличение ошибки можно отметить для некоторых переменных (давление на уровне моря, геопотенциал 500 гПа (H500)) в тропическом регионе. Это увеличение не всегда значимо. По всей видимости, требуется более точный подбор настроечных параметров в этом регионе. При этом значимые уменьшения ошибки наблюдаются для ветра (на величину от от 5 до 13 %).

### 3.7 Выводы к главе 3

В главе 3 приведены результаты усвоения данных спутниковых наблюдений AMV с использованием разработанной системы усвоения на основе LETKF. С помощью численных экспериментов показано, что учёт свойств данных спутниковых наблюдений ветра AMV повысил точность полей ансамбля начальных данных и прогнозов на их основе.

Применение переопределения высоты наблюдения уменьшает ошибки прогнозов, полученных с использованием старта модели с полученных в ходе усвоения начальных данных. Для большинства регионов и переменных уменьшение ошибки статистически значимо на величину до 3–4%. Наибольший положительный отклик на переопределение высоты показывают ошибки во внетропической части Северного полушария.

Для учёта автокорреляций в ошибках наблюдений ветра AMV можно эффективно применять схему на основе функции SOAR. Помимо уменьшения

ошибки анализа и прогноза применение этой схемы улучшает также и их мелко-масштабную структуру. В этом случае более точно учитываются связи ошибок наблюдений. При надлежащем подборе настроечных параметров  $\mathbf{R}_0$  и  $L$  (для различных спутников, каналов и регионов) функции SOAR, можно добиться уменьшения среднеквадратических ошибок прогнозов и полей начальных данных.

Ансамблевая система усвоения данных была подготовлена к использованию в системе ансамблевого прогноза.

## **Глава 4. Применение системы ансамблевого усвоения в глобальной системе ансамблевого среднесрочного прогноза**

В данной главе приводится описание и результаты оперативных испытаний разработанной при участии автора технологии глобального ансамблевого среднесрочного прогноза на основе модели ПЛАВ.

В разделе 4.1 приводится описание системы. Работы, описанные в подразделах 4.1.1 и 4.1.2, а также сведение всех компонентов системы ансамблевого прогноза в единый комплекс, выполнены автором данной диссертации. Результаты оперативных испытаний приведены в разделе 4.2. В заключительном разделе 4.3 содержатся выводы к текущей главе.

Результаты данной главы опубликованы в [30].

### **4.1 Технология глобального ансамблевого среднесрочного прогноза на основе ансамблевого усвоения данных**

Технология ансамблевого прогноза построена на основе ансамблевой системы усвоения, описанной в главах 2 и 3. Ансамблевая система усвоения формирует ансамбль начальных данных, который используется для получения ансамбля численных прогнозов в модели ПЛАВ.

Все компоненты системы ансамблевого прогноза (далее — САП) работают на одной и той же модельной сетке (горизонтальное разрешение  $0,9^\circ \times 0,72^\circ$  по долготе и широте соответственно и 96 уровней по вертикали с использованием гибридной вертикальной координаты).

Система ансамблевого прогноза работает циклически с шагом цикла 6 часов и состоит из следующих компонентов:

- блок подготовки стартового ансамбля на основе ансамблевой системы усвоения;
- блок центрирования;
- блок подготовки почвенных переменных;
- полулагранжева модель атмосферы (ПЛАВ),

– блок верификации.

В модели в качестве переменной влажности используется удельная влажность, в остальных блоках относительная влажность.

Для учёта неопределённостей, возникающих при параметрическом описании процессов подсеточного масштаба, в модели ПЛАВ была реализована комбинация методов стохастического возмущения параметров и тенденций параметризаций (см. [30]). Всего возмущается 27 параметров и 2 тенденции (температура и завихренность, тенденция влажности не возмущается, чтобы не нарушать законы сохранения). В модели был реализован генератор случайных двумерных полей, скоррелированных по пространству и времени. Для каждого из параметров и тенденций можно настроить амплитуду возмущения, масштаб корреляции по пространству и время декорреляции. Настройка амплитуды возмущений проводилась в соответствии с физическим смыслом параметров. Амплитуда параметров, которые не являются строго определёнными физическими константами, подбиралась эмпирически из соображений устойчивости работы модели. Более подробно учёт неопределённостей изложен в [30].

#### 4.1.1 Генерация начальных данных и центрирование

Генерация ансамбля начальных данных для старта ансамблевого прогноза начинается с запуска усвоения с помощью описанной в главах 2 и 3 системы. Для получения как ансамбля первых приближений, так и ансамбля среднесрочных прогнозов используется модель ПЛАВ, описание которой приведено в разделе 2.2. Возможен режим работы САП, при котором с полученных полей начальных данных стартует ансамблевый прогноз (автономный режим). Схема такого режима работы в цикле приведена на рисунке 4.1.

Однако большая часть спутниковых наблюдений, несущих значительный объем информации о состоянии атмосферы, в представленной системе усвоения в настоящее время никак не учитываются. Это приводит к повышенной погрешности анализов при автономной работе системы усвоения (особенно в Южном полушарии, где число усваиваемых наблюдений невелико). Поэтому полученный в ходе усвоения ансамбль подвергается процедуре т. н. «центриро-

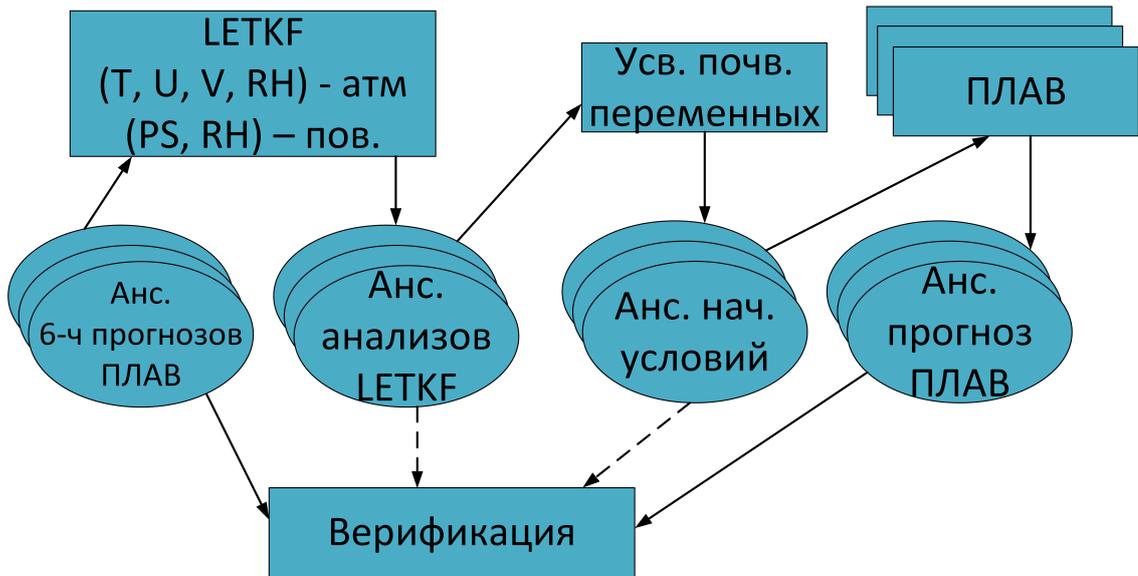


Рисунок 4.1 — Схема автономной работы в цикле системы ансамблевого прогноза

вания», которая применяется для повышения точности ансамблевого прогноза и состоит в использовании дополнительной информации – оперативных анализов Гидрометцентра России.

Суть центрирования состоит в следующем. Пусть  $x_1^a, \dots, x_{N_f}^a$  – ансамбль анализов, сгенерированный ансамблевой системой усвоения на основе LETKF,  $\bar{x}^a = \frac{\sum_i^{N_f} x_i^a}{N_f}$  – средний по ансамблю анализ. Пусть  $x^{op}$  – оперативный анализ Гидрометцентра. Тогда элементы центрированного ансамбля  $x_i^{ac}$  можно получить следующим образом:

$$x_i^{ac} = x_i^a + (x^{op} - \bar{x}^a). \quad (4.1)$$

Такой подход был предложен в системе MOGREPS [115]. В данной системе впервые центрирование было реализовано В.С. Рогутовым в [20], однако оно было реализовано на модельных уровнях. Описываемый же в данном разделе и применяемый в оперативной версии алгоритм использует в качестве «центрального» оперативный анализ Гидрометцентра  $x^{op}$ , который задан на изобарических поверхностях на сетке с шагом 0,5 градуса по широте и долготе.

Для использования в блоке центрирования оперативный анализ Гидрометцентра сначала интерполируется на модельную сетку по горизонтали. При этом среднее по ансамблю анализов LETKF интерполируется на  $p$ -уровни по вертикали. Далее ищется разность полученных полей, интерполируется на модельные гибридные уровни и прибавляется к каждому члену первоначального

ансамбля анализов LETKF. Таким образом, в блоке центрирования ансамбль анализов модифицируется так, чтобы средний по ансамблю анализ совпал с оперативным анализом Гидрометцентра России с сохранением отклонений отдельных анализов из ансамбля от среднего по первоначальному ансамблю анализа. В результате среднеквадратичная ошибка ансамблевого прогноза наряду с точностью численной модели определяется оперативным анализом Гидрометцентра России, а разбросы ансамблей — системой усвоения LETKF. Схема работы системы ансамблевого прогноза с центрированием на оперативные анализы Гидрометцентра приведена на рисунке 4.2.

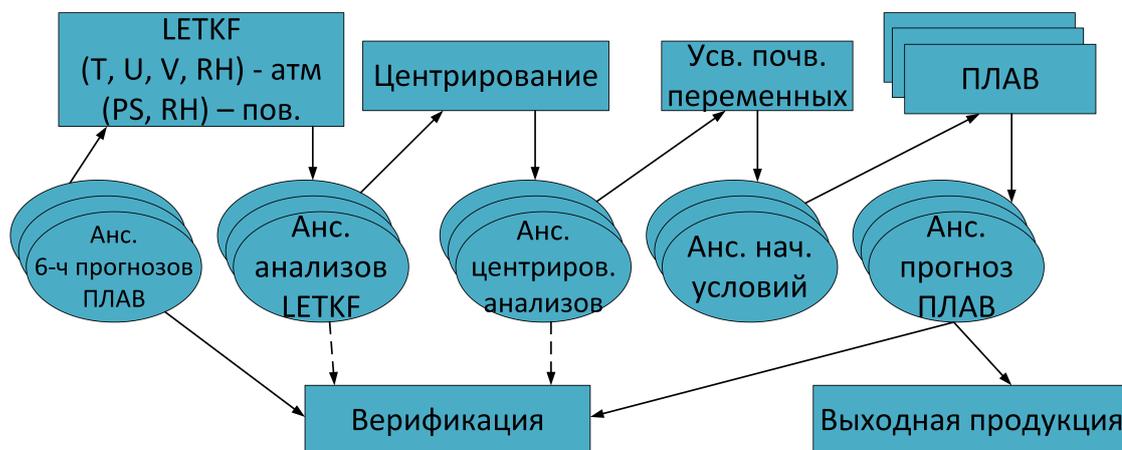


Рисунок 4.2 — Схема работы в цикле системы ансамблевого прогноза с центрированием на оперативные анализы Гидрометцентра России

#### 4.1.2 Верификация ансамблевых прогнозов

Блок верификации разработан для расчёта оценок успешности полученных ансамблей прогнозов и анализов по оперативному анализу Гидрометцентра России и данным наблюдений с наземных метеорологических станций (для верификации приземных прогностических полей и осадков). С августа 2022 года система верификации внедрена в оперативную практику Гидрометцентра России для обмена результатами работы системы ансамблевого прогноза с другими ведущими прогностическими центрами ВМО. Результаты ежемесячно отправляются в Ведущий центр верификации систем ансамблевого прогноза (<https://epsv.kishou.go.jp/EPsv/>).

Для традиционных показателей успешности детерминистических прогнозов, в качестве которых используются средние по ансамблю, оцениваются средние и среднеквадратические ошибки, корреляции аномалий, градиентные ошибки (для некоторых полей) аналогично разделу 3.2.

По данным оперативного анализа Гидрометцентра России верифицируются следующие поля:

- давление на уровне моря (PMSL или P0 на рисунках далее);
- геопотенциальные высоты изобарических поверхностей 850, 500, 250 гПа (соответственно H850, H500, H250);
- температура на поверхностях 850, 500, 250 гПа (соответственно T850, T500, T250);
- скорость ветра (зональная и меридиональная составляющие) на поверхностях 850, 500, 250 гПа (соответственно U850, U500, U250 и V850, V500, V250);
- модуль скорости ветра на поверхностях 850, 500, 250 гПа (соответственно W850, W500, W250).

Для оценки вероятностных свойств ансамблевого прогноза используются следующие показатели, определённые в наставлении [13]:

- разброс — стандартное отклонение членов ансамбля от среднего, поля разброса осредняются по тем же самым районам и переменным, которые использованы для среднеквадратической ошибки среднего по ансамблю прогноза (см. раздел 3.2):

$$spread_i = \sqrt{\frac{1}{m} \sum_{j=1}^m (F_{i,j} - F_i)^2}, \quad (4.2)$$

где  $F_{i,j}$  — ансамбль спрогнозированных значений параметра в  $i$ -ой точке,  $F_i$  — среднее по ансамблю спрогнозированных значений параметра в  $i$ -ой точке;

- показатели CRPS:

$$CRPS_i = \int_{-\infty}^{\infty} \left( \hat{P}_i(x) - \hat{O}_i(x) \right)^2 dx, \quad (4.3)$$

где  $\hat{P}_i(x) = \frac{1}{m} \sum_{j=1}^m H(x - F_{i,j})$  — кусочно-заданная кумулятивная функция распределения вероятности ансамбля прогнозов параметра,  $\hat{O}_i(x) = H(x - A_i)$  — кусочно-заданная кумулятивная функция распределения вероятности фактического значения параметра (по данным анализа или наблюдений),  $A_i$  — фак-

тическое значение (по данным анализа или наблюдений),  $H(x)$  – единичная функция Хевисайда;

– таблицы надёжности – показывают число сбывшихся и не сбывшихся прогнозов событий в градациях вероятностей, соответствующих количеству членов ансамбля, строятся по примеру таблицы 4, обозначения:  $FOR$  – количество членов, предсказывающих осуществление события;  $NFOR$  – количество членов, предсказывающих осуществление события;  $j$  – номер вероятностной градации, изменяется от 1 до  $m+1$ ;  $O_j = \sum \omega_i(O)_i$ ,  $O = 1$ , если в  $i$ -ой точке из  $j$ -ой градации событие прогнозировалось и наблюдалось в реальности,  $O = 0$  в противном случае;  $NO_j = \sum \omega_i(NO)_i$ ,  $NO = 1$ , если в  $i$ -ой точке из  $j$ -ой градации событие прогнозировалось, но не наблюдалось в реальности,  $NO = 0$  в противном случае;  $\omega_i = \cos\varphi_i$  – весовой коэффициент (равен единице в случае верификации по данным наблюдений);  $\varphi_i$  – географическая долгота точки сетки в случае верификации по данным оперативного анализа.

Таблица 4 – Пример заполнения таблицы надёжности

Номер вероятностной категории	Распределение членов ансамбля	Количество наступивших событий	Количество ненаступивших событий
1	$FOR = 0, NFOR = m$	$O_1$	$NO_1$
2	$FOR = 1, NFOR = m - 1$	$O_2$	$NO_2$
3	$FOR = 2, NFOR = m - 2$	$O_3$	$NO_3$
...	...	...	...
$j$	$FOR = j - 1, NFOR = m - j + 1$	$O_j$	$NO_j$
...	...	...	...
$m$	$FOR = m - 1, NFOR = 1$	$O_m$	$NO_m$
$m+1$	$FOR = m, NFOR = 0$	$O_{m+1}$	$NO_{m+1}$

На основе таблиц надёжности рассчитываются следующие показатели успешности ансамблевых прогнозов:

– Оценка Брайера (Brier Score):

$$BS_i = (P_i - Obs_i)^2, \quad (4.4)$$

где  $P_i$  – вероятность события,  $Obs_i$  – наблюдалось ли (равно 1) или не наблюдалось (равно 0) событие в данной точке. Оценка Брайера сильно зависят от повторяемости события, поэтому она обычно дополняется оценкой Брайера

по отношению к эталонному прогнозу BSS (Brier Skill Score):

$$BSS = 1 - \frac{BS}{BS_{ref}} \quad (4.5)$$

где  $BS_{ref}$  – оценка Брайера для эталонного (климатического) прогноза;

– Сравнительная оперативная характеристика (Relative operating characteristic – ROC score) – график зависимости  $HR_j$  от  $FAR_j$ , где

$$HR_j = \frac{\sum_{k=j}^{m+1} O_k}{\sum_{k=1}^{m+1} O_k}, \quad FAR_j = \frac{\sum_{k=j}^{m+1} NO_k}{\sum_{k=1}^{m+1} NO_k};$$

– Площадь под кривой ROC (ROCA, ROC Area) – площадь на графике, ограниченная кривой ROC.

Программа, реализующая блок верификации на ЭВМ, зарегистрирована в Роспатенте (Свидетельство о государственной регистрации 2017610851 от 18 января 2017 года).

На рисунке 4.3 представлены для примера графики кривых ROC. Показаны кривые для прогноза события «положительная аномалия T850 превышает одно климатическое стандартное отклонение во внетропической части Северного полушария» в декабре 2021 года. Слева показан график ROC по прогнозам предыдущей оперативной САП Гидрометцентра России (Т169), взятый с приведённого выше сайта. Справа – график ROC по прогнозам разработанной САП (ПЛАВ) во время оперативных испытаний, построенный с помощью описываемого блока ансамблевой верификации.

### 4.1.3 Результаты подключения данных ветра AMV

Данные наблюдений AMV оказывают существенное положительное влияние на точность среднего по ансамблю прогноза. На рисунках 4.4 – 4.5 представлены результаты сравнения осреднённых за май 2023 года (выборки состоят из 31 элементов каждая) среднеквадратических ошибок аналогично результатам на рисунках 3.10 – 3.12 для средних по ансамблю прогнозов с использованием данных AMV (синим) и без использования данных AMV (красным). Отличие от приведённых в главе 3 графиков состоит в том, что там приводятся средне-

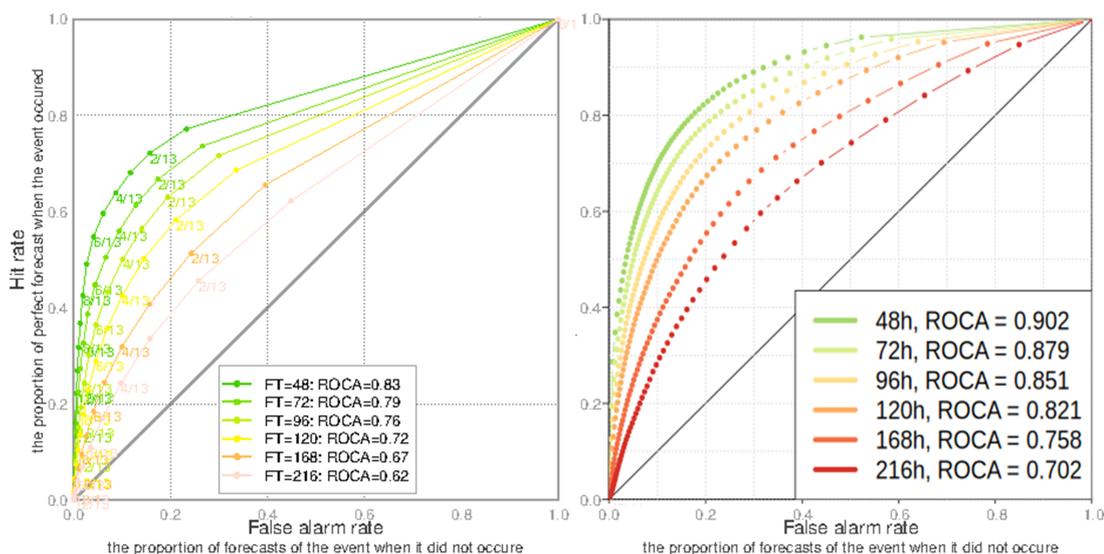


Рисунок 4.3 — Кривые ROC для прогноза события «положительная аномалия Т850 превышает одно климатическое стандартное отклонение». Внетропическая часть Северного полушария. Декабрь 2021 года. Т169 (слева) и ПЛАН (справа)

квадратические ошибки единственного детерминированного прогноза, а здесь — среднего по ансамблю прогнозов.

Статистически значимое уменьшение среднеквадратической ошибки во внетропической части Северного полушария достигает 14,6% (для прогнозов W250 на двое суток), для внетропической части Южного полушария достигает 10,9% (для прогнозов W250 на пять суток).

Также для оценки успешности использования данных AMV в разработанной системе ансамблевого прогноза по тем же самым прогнозам были рассчитаны непрерывные ранговые вероятностные оценки (анг. continuous ranked probability score, CRPS), которые широко используются в качестве одного из показателей успешности ансамблевых прогнозов. Результаты представлены на рисунке 4.6 в виде карточки показателей. Синим цветом и его яркостью показано уменьшение показателя CRPS и относительная величина (в процентах) этого улучшения. Красным — то же самое, но для увеличения ошибки.

Во внетропической части Южного полушария уменьшение показателя CRPS носят наиболее заметный характер. Наибольший положительный отклик на включение спутниковых данных ветра AMV можно увидеть для Н500 и Н850 (порядка 60%) в первые сутки прогноза. Во внетропической части Северного полушария и в тропиках можно увидеть слабый отклик на включение спутни-

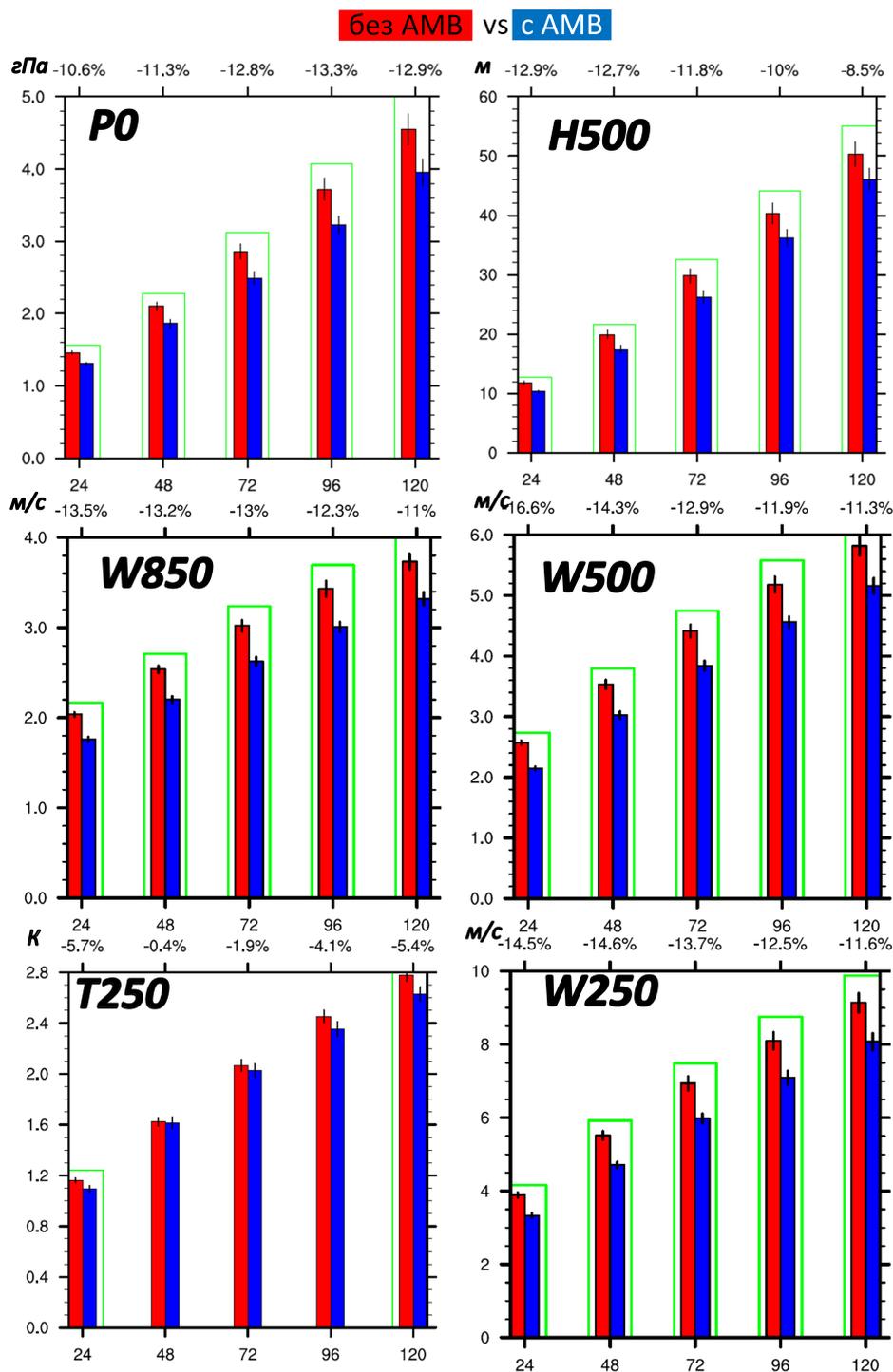


Рисунок 4.4 — Среднеквадратические ошибки средних по ансамблю прогнозов без использования AMV (красным) и при использовании AMV (синим) в зависимости от заблаговременности прогноза во внетропической части Северного полушария. Обозначения переменных приведены в разделе 4.1.2

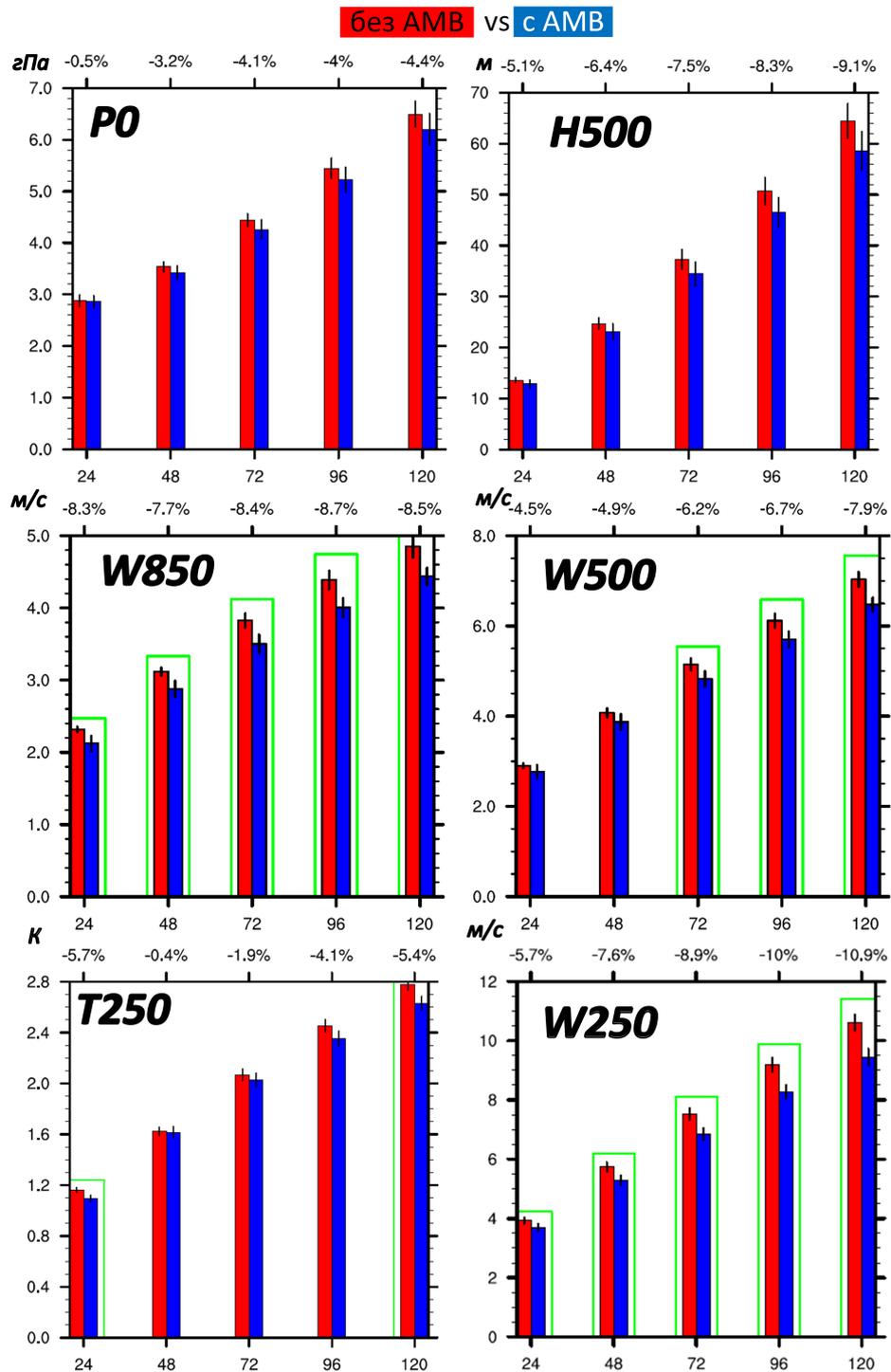


Рисунок 4.5 — Среднеквадратические ошибки средних по ансамблю прогнозов без использования AMV (красным) и при использовании AMV (синим) в зависимости от заблаговременности прогноза во внетропической части Южного полушария. Обозначения переменных приведены в разделе 4.1.2

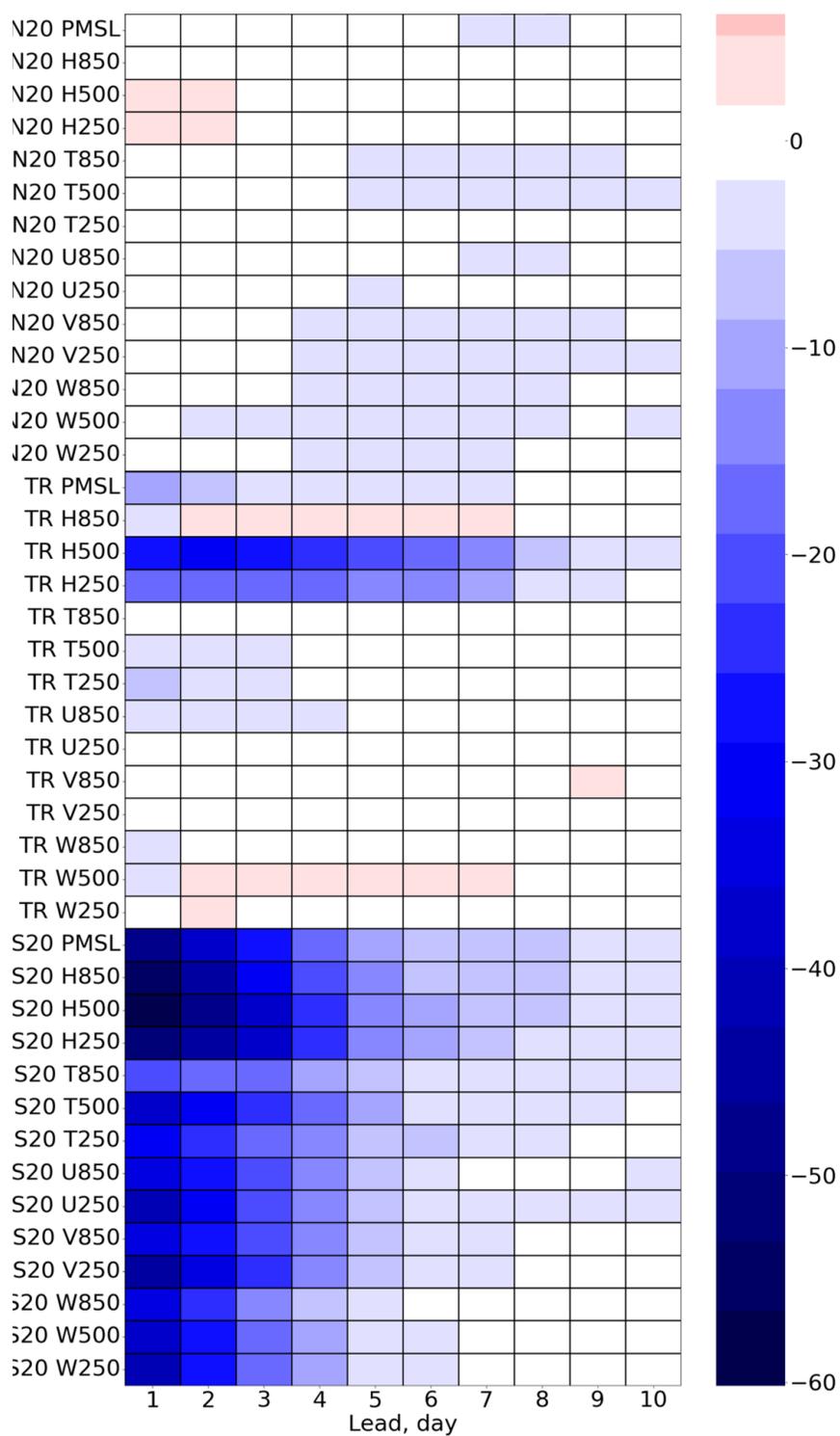


Рисунок 4.6 — Изменение показателей CRPS для ансамблевых прогнозов с усвоением спутниковых данных AMV и без их усвоения в зависимости от заблаговременности прогноза за май 2023 г. Обозначения на рисунке: N20 – внетропическая часть Северного полушария, Tr – тропики, S20 – внетропическая часть Южного полушария.

ковых данных ветра AMV. Вероятно, так происходит по причине малого количества других данных в Южном полушарии. Использование большого (по сравнению с другими видами) числа данных в этой части атмосферы увеличивает качество ансамбля начальных данных, что отражается в уменьшении показателя CRPS.

## 4.2 Результаты оперативных испытаний системы ансамблевого прогноза

Оперативные испытания системы ансамблевого прогноза проводились сотрудниками Лаборатории глобальных численных прогнозов с декабря 2020 года.

Оценка технологии ансамблевого прогнозирования на основе модели ПЛАВ (в версии с постоянным горизонтальным разрешением по пространству  $0,9^\circ \times 0,72^\circ$  с 96 вертикальными гибридными уровнями, 40 членов ансамбля, SLAV на рисунках ниже) была произведена для периода с августа 2021 по январь 2022 года. Среднесрочные прогнозы на 10 суток рассчитывались по исходным данным в срок 00 часов всемирного скоординированного времени. В сроки 06, 12, 18 часов всемирного скоординированного времени рассчитывался только ансамбль первых приближений (6-часовой ансамблевый прогноз). Результаты работы описываемой технологии сравнивались с аналогичными результатами следующих САП:

- UKMO [130] (the Met Office Global and Regional Ensemble Prediction System (MOGREPS), горизонтальное разрешение порядка 20 км, 70 вертикальных уровней, ансамбль состоит из 44 членов, UKMO на рисунках ниже)
- оперативная (на момент начала 2022 года) ансамблевая система Гидрометцентра [1] (горизонтальное разрешение порядка 70 км, 31 вертикальный уровень, ансамбль состоит 13 прогнозов спектральной модели T169L31 и 1 прогноза модели ПЛАВ20, EnsAst на рисунках ниже).

При выполнении испытаний верифицировались 9 стандартных полей:

1. PMSL — давление на уровне моря;
2. H500 — геопотенциал 500 гПа;

3. T2m — приземная температура (температура на высоте 2 метра над поверхностью);
4. T250 — температура на уровне 250 гПа;
5. T850 — температура на уровне 850 гПа;
6. U250 — зональная составляющая скорости ветра на уровне 250 гПа;
7. U850 — зональная составляющая скорости ветра на уровне 850 гПа;
8. V250 — меридиональная составляющая скорости ветра на уровне 250 гПа;
9. V850 — меридиональная составляющая скорости ветра на уровне 850 гПа.

Оценки проводились по полям оперативного объективного анализа Гидрометцентра России (на сетке с шагом 1,5 градуса по широте и долготе). Для каждого из полей рассчитывались оценки Брайера, сравнительные оперативные характеристики ROC и площади под ROC кривыми для Северного полушария (20 – 90 с. ш.), Южного полушария (90 – 20 ю. ш.) и в тропиках (20 ю. ш. – 20 с. ш.) для положительных аномалий более 2, 1,5 и 1 стандартного отклонения, отрицательных аномалий более 2, 1,5 и 1 стандартного отклонения. Описание методики расчёта ROC кривых и оценок Брайера приведено в [13].

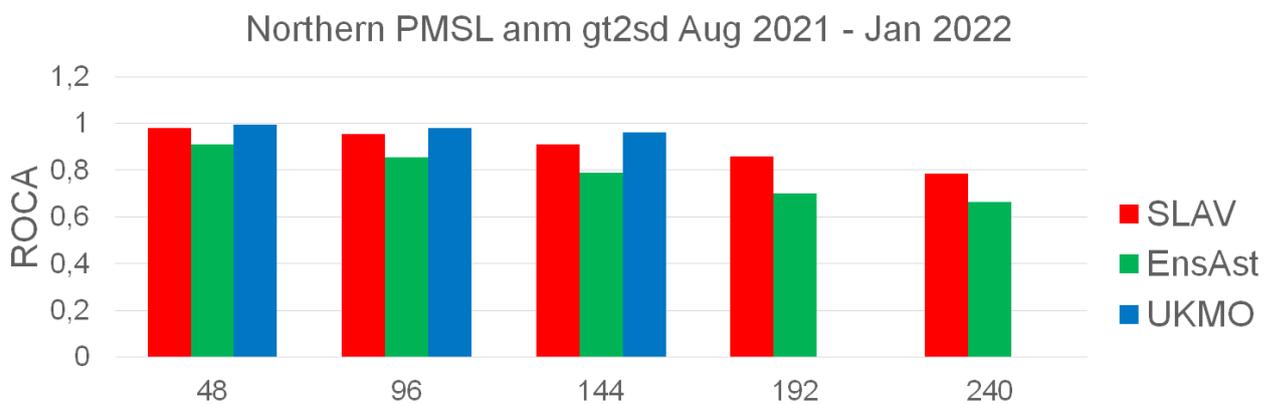


Рисунок 4.7 — Площадь под ROC кривой, посчитанной в период испытаний для положительной аномалии более 2 стандартных отклонений давления на уровне моря в Северном полушарии в зависимости от заблаговременности прогноза в часах

На рисунках 4.7 – 4.14 приведены площади под ROC кривыми (ROCA), полученные по результатам ансамблевых прогнозов полей давления на уровне моря (рисунки 4.7 – 4.10), геопотенциала 500 гПа (рисунки 4.11 – 4.14) в зависимости от заблаговременности прогноза (приведены результаты для заблаговременности прогнозов 2, 4, 6, 8, 10 суток). Результаты рассчитаны для Северного полушария. ROC кривые и площади под кривыми рассчитывались для поло-

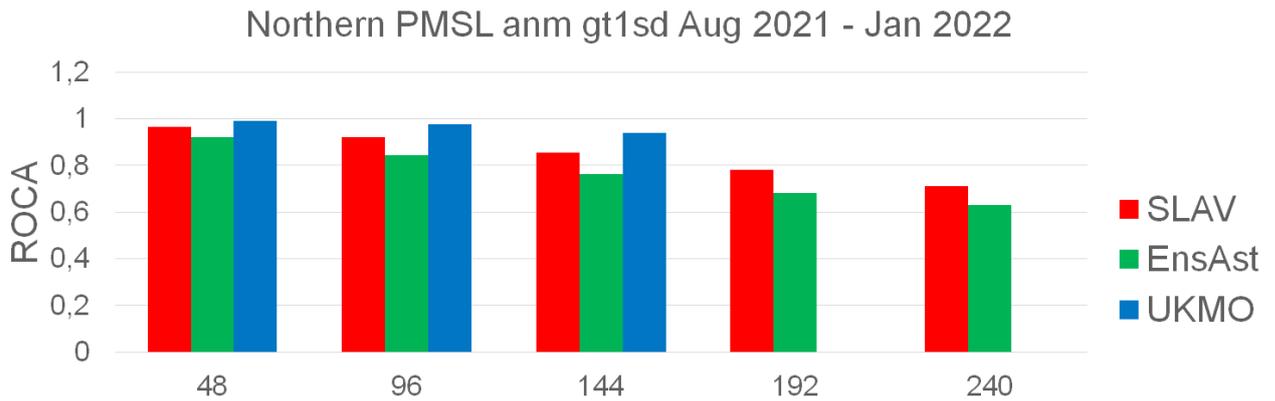


Рисунок 4.8 — Площадь под ROC кривой, посчитанной в период испытаний для положительной аномалии более 1 стандартного отклонения давления на уровне моря в Северном полушарии в зависимости от заблаговременности прогноза в часах

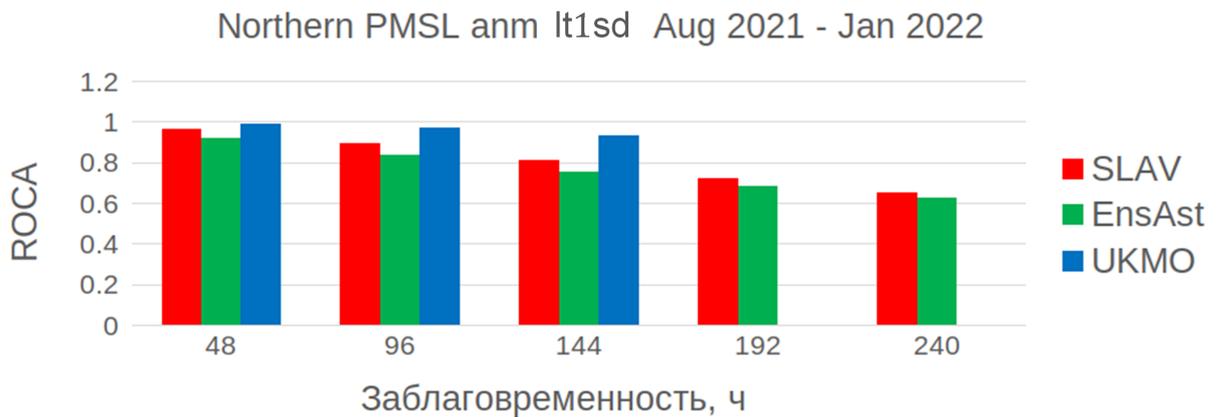


Рисунок 4.9 — Площадь под ROC кривой, посчитанной в период испытаний для отрицательной аномалии более 1 стандартного отклонения давления на уровне моря в Северном полушарии в зависимости от заблаговременности прогноза в часах

жительных аномалий более двух стандартных отклонений (рисунки 4.7 и 4.11), более одного стандартного отклонения (рисунки 4.8 и 4.12) и отрицательных аномалий менее двух стандартных отклонений (рисунки 4.9 и 4.13) и менее двух стандартного отклонения (рисунки 4.10 и 4.14).

Для сравнения вместе с результатами, полученными на испытаниях САП на основе модели ПЛАВ и ансамблевой системы усвоения на основе LETKF (на рисунках отмечена как SLAV, столбцы красного цвета) приведены площади под ROC кривыми этих же полей, полученные для САП на основе спектральной модели T169L31 [1] (EnsAst, зелёный цвет на рисунках) и площади под ROC

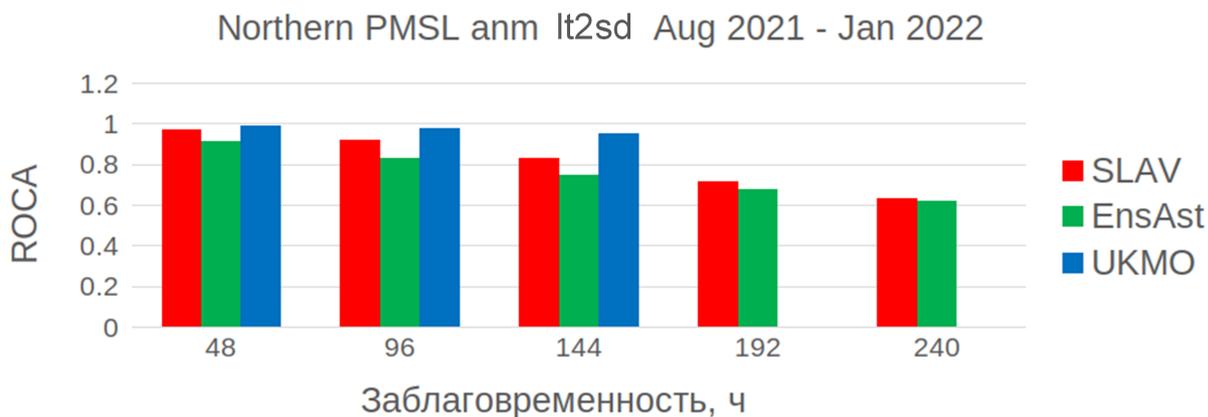


Рисунок 4.10 — Площадь под ROC кривой, посчитанной в период испытаний для отрицательных аномалий более 2 стандартных отклонений давления на уровне моря в Северном полушарии в зависимости от заблаговременности прогноза в часах

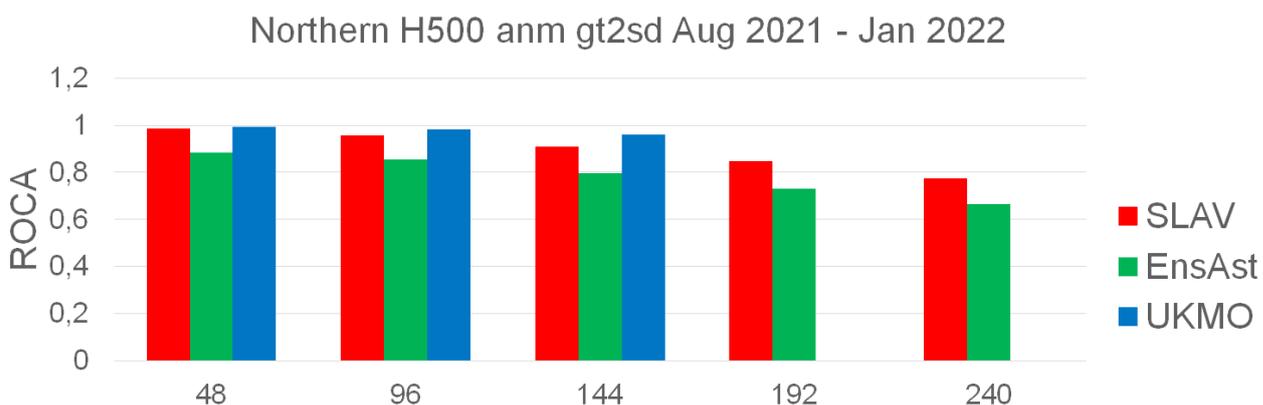


Рисунок 4.11 — Площадь под ROC кривой, посчитанной в период испытаний для положительных аномалий более 2 стандартных отклонений геопотенциала 500 гПа в Северном полушарии в зависимости от заблаговременности прогноза в часах

кривыми, полученные для системы ансамблевого прогноза UKMO (синий цвет на рисунках). Оценки для систем ансамблевого прогноза UKMO и EnsAst были взяты с сайта Главного центра верификации систем ансамблевого прогноза (<https://epsv.kishou.go.jp/EPsv/>).

Как видно из рисунков, успешность прогноза приземного давления системы ансамблевого прогноза на основе ПЛАВ в Северном полушарии заметно превосходит результаты системы ансамблевого прогноза на основе спектральной модели T169L31, но уступают системе UKMO для всех заблаговременностей прогнозов. Оценки, рассчитанные по прогнозам САП ПЛАВ для отрица-

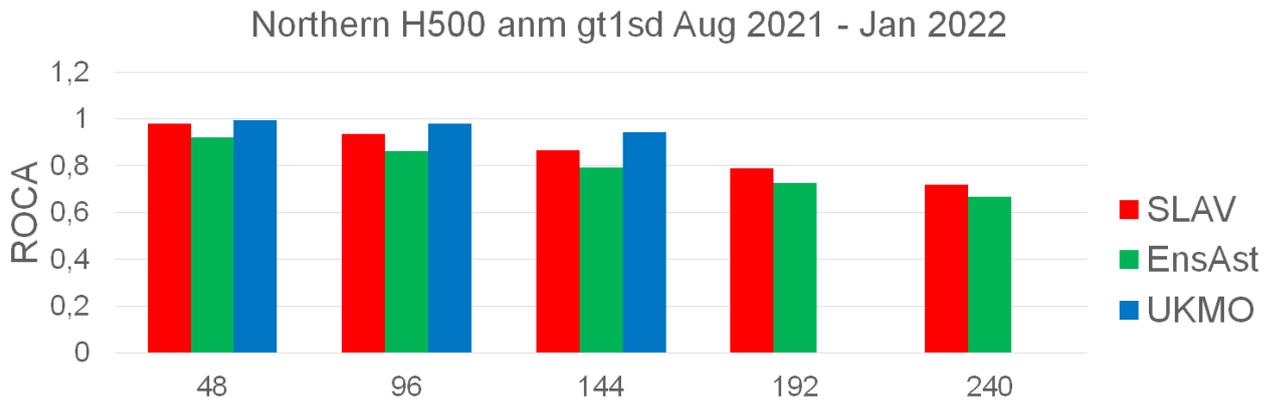


Рисунок 4.12 — Площадь под ROC кривой, посчитанной в период испытаний для положительных аномалий более 1 стандартного отклонения геопотенциала 500 гПа в Северном полушарии в зависимости от заблаговременности прогноза в часах

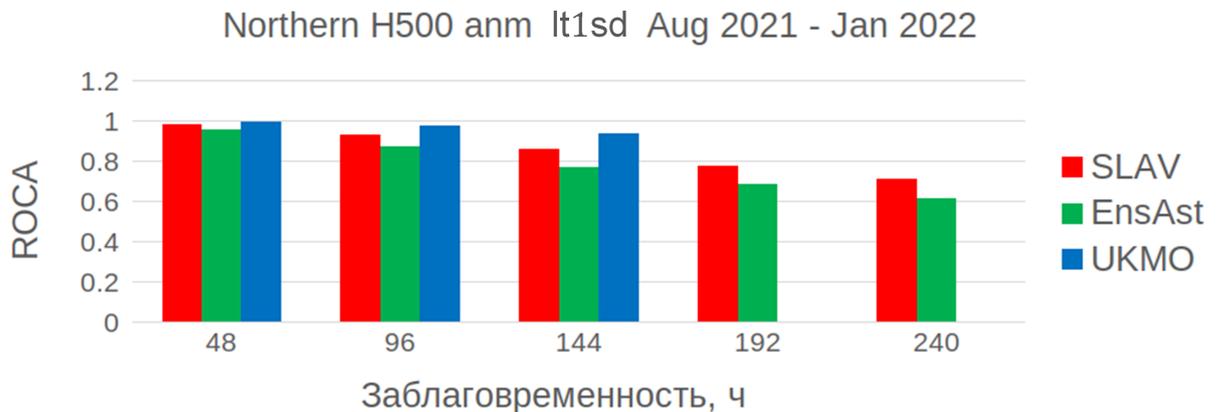


Рисунок 4.13 — Площадь под ROC кривой, посчитанной в период испытаний для отрицательных аномалий более 1 стандартного отклонения геопотенциала 500 гПа в Северном полушарии в зависимости от заблаговременности прогноза в часах

тельных аномалий, убывают быстрее, чем оценки прогнозов САП ПЛАВ, рассчитанные для положительных аномалий. При небольшой заблаговременности прогноза качество прогноза геопотенциала 500 гПа в САП ПЛАВ почти не уступает САП УКМО, но с ростом заблаговременности прогноза качество прогнозов САП ПЛАВ убывает быстрее.

На рисунках 4.15 и 4.16 приведены оценки Брайера, рассчитанные по результатам ансамблевых прогнозов полей давления на уровне моря в зависимости от заблаговременности прогноза (приведены результаты для заблаговременности прогнозов 2, 4, 6, 8, 10 суток — заблаговременности прогнозов приведены

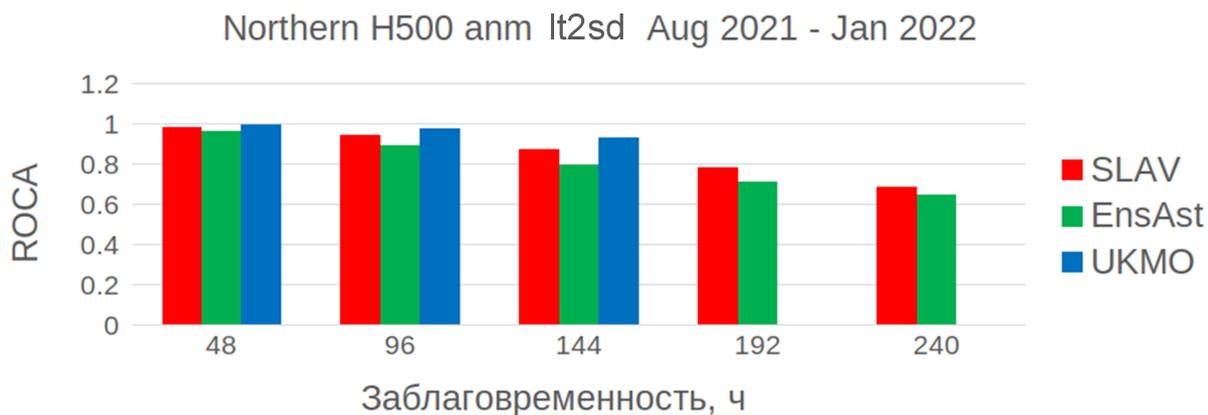


Рисунок 4.14 — Площадь под ROC кривой, посчитанной в период испытаний для отрицательных аномалий более 2 стандартных отклонений геопотенциала 500 гПа в Северном полушарии в зависимости от заблаговременности прогноза в часах

в часах на горизонтальной шкале). Для сравнения вместе с результатами, полученными на испытаниях описываемой в данной работе системы ансамблевого прогноза (на рисунках отмечена как SLAV, столбцы красного цвета), приведены оценки Брайера этих же полей, полученные для САП Гидрометцентра на основе спектральной модели [1] (EnsAst, зелёный цвет на рисунках) и оценки Брайера, полученные для системы ансамблевого прогноза UKMO (синий цвет на рисунках). Оценки Брайера для систем ансамблевого прогноза UKMO и EnsAst были взяты с сайта Главного центра верификации систем ансамблевого прогноза (<https://epsv.kishou.go.jp/EPsv/>).

Для заблаговременностей прогнозов 8 и 10 суток оценки Брайера системы ансамблевого прогноза UKMO не приведены, так как они отсутствуют на сайте центра верификации.

Как видно из рисунков, показатели Брайера поля приземного давления системы ансамблевого прогноза на основе модели ПЛАВ в Северном полушарии заметно превосходят результаты системы на основе спектральной модели T169L31, но уступают системе UKMO для всех заблаговременностей прогнозов.

На рисунке 4.17 приведены площади под ROC кривыми, полученные по результатам ансамблевых прогнозов в Южном полушарии (20 – 90 ю. ш.) в зависимости от заблаговременности прогноза (приведены результаты для заблаговременности прогнозов 2, 4, 6, 8, 10 суток). ROC кривые строились для положительных аномалий больше 1 стандартного отклонения (рисунок 4.17 а,

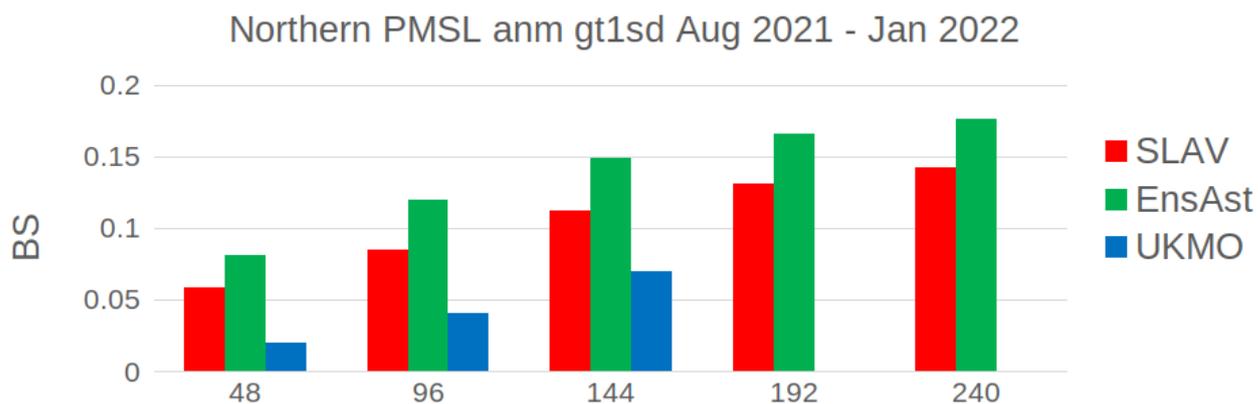


Рисунок 4.15 — Оценки Брайера положительных аномалий более 1 стандартного отклонения давления на уровне моря в Северном полушарии

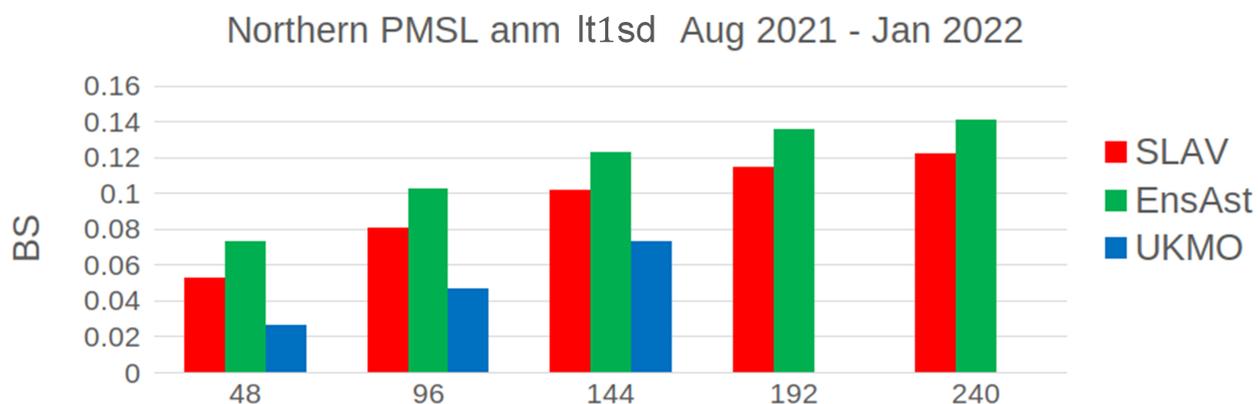


Рисунок 4.16 — Оценки Брайера отрицательных аномалий менее 1 стандартного отклонения давления на уровне моря в Северном полушарии

в, д), и отрицательных аномалий менее 1 стандартного отклонения (рисунок 4.17 б, г, е) для ансамблевых прогнозов полей приземного давления, температуры на уровне 850 гПа и геопотенциала 500 гПа. Для сравнения вместе с результатами, полученными на испытаниях описываемой в данной работе системы ансамблевого прогноза (на рисунках отмечена как SLAV, столбцы красного цвета), приведены площади под ROC кривыми этих же полей, полученные с использованием САП Гидрометцентра России на основе спектральной модели [1] (EnsAst, зелёный цвет на рисунках) и площади под ROC кривыми, полученные для системы ансамблевого прогноза UKMO (синий цвет на рисунках).

На рисунках видно, что в Южном полушарии ансамблевые прогнозы ПЛАВ превосходят ансамблевые прогнозы на основе спектральной модели T169L31, но уступают прогнозам UKMO (исключение — отрицательные аномалии менее 1 стандартного отклонения поля приземного давления (PMSL на

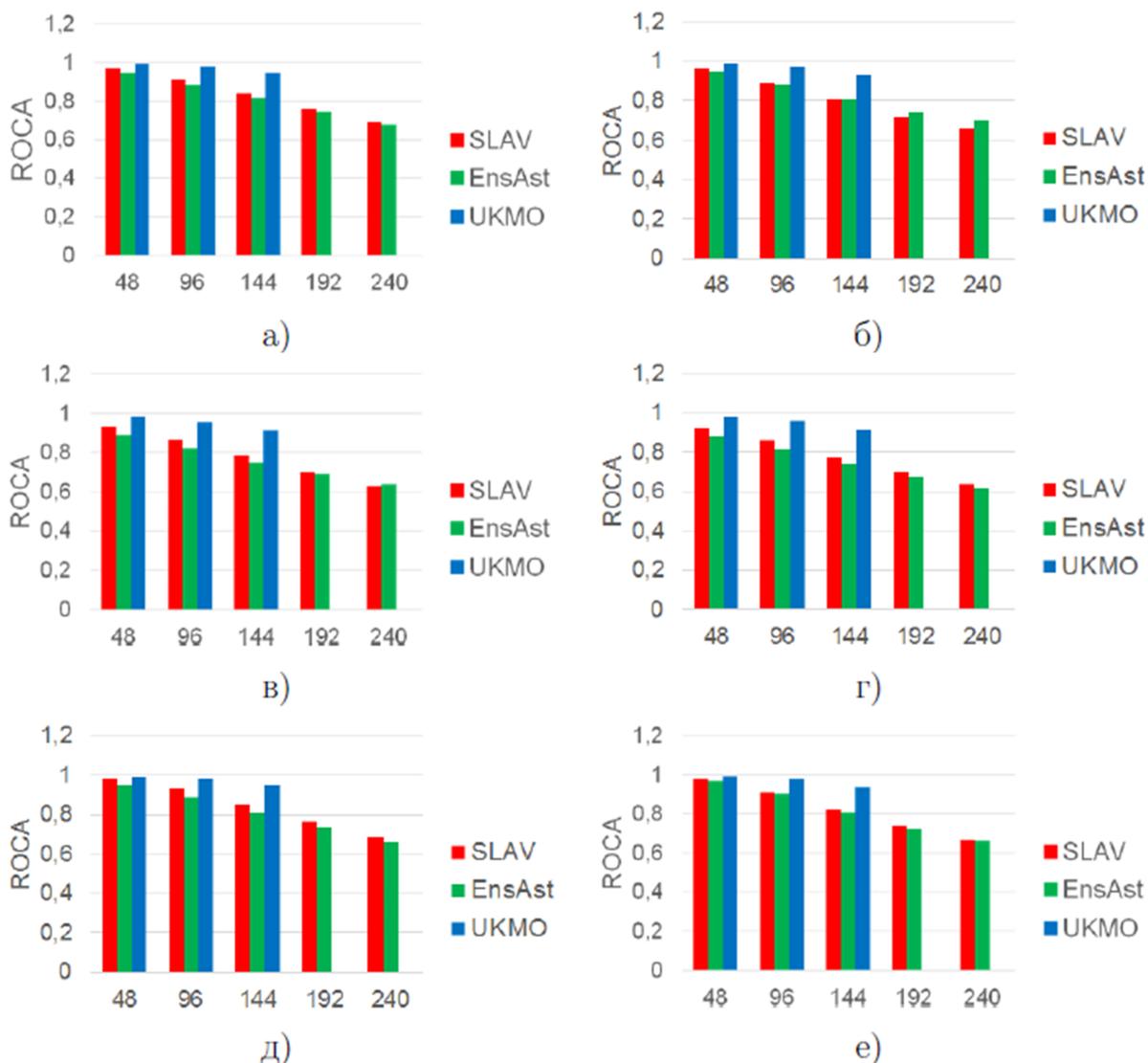


Рисунок 4.17 — Площади под ROC кривыми, рассчитанные в период испытаний для положительных аномалий более 1 стандартного отклонения (а, в, д) и отрицательных аномалий менее 1 стандартного отклонения (б, г, е) полей приземного давления (а, б), температуры на уровне 850гПа (в, г) и геопотенциала 500 гПа (д, е) в Южном полушарии по результатам ансамблевых прогнозов САП ПЛАВ (красные столбцы), САП на основе спектральной модели T169L31 (зелёные столбцы) и САП УКМО (синие столбцы). На горизонтальной шкале – заблаговременность прогноза в часах (рисунках) при заблаговременности прогноза 8 и 10 суток). Качество прогнозов ПЛАВ с ростом заблаговременности прогноза убывает быстрее, чем в системе УКМО. Для всех заблаговременностей и всех полей ансамблевый прогноз САП УКМО превосходит по качеству САП ПЛАВ. С ростом заблаговременности прогноза качество прогнозов САП ПЛАВ убывает быстрее. Наиболее за-

метна разница в качестве прогнозов для поля Т850. Этот недостаток может быть частично связан с недостаточным пространственным разрешением модели ПЛАВ2018. Качество прогнозов САП ПЛАВ убывает с ростом заблаговременности прогноза немного быстрее, чем САП на основе спектральной модели Т169L31.

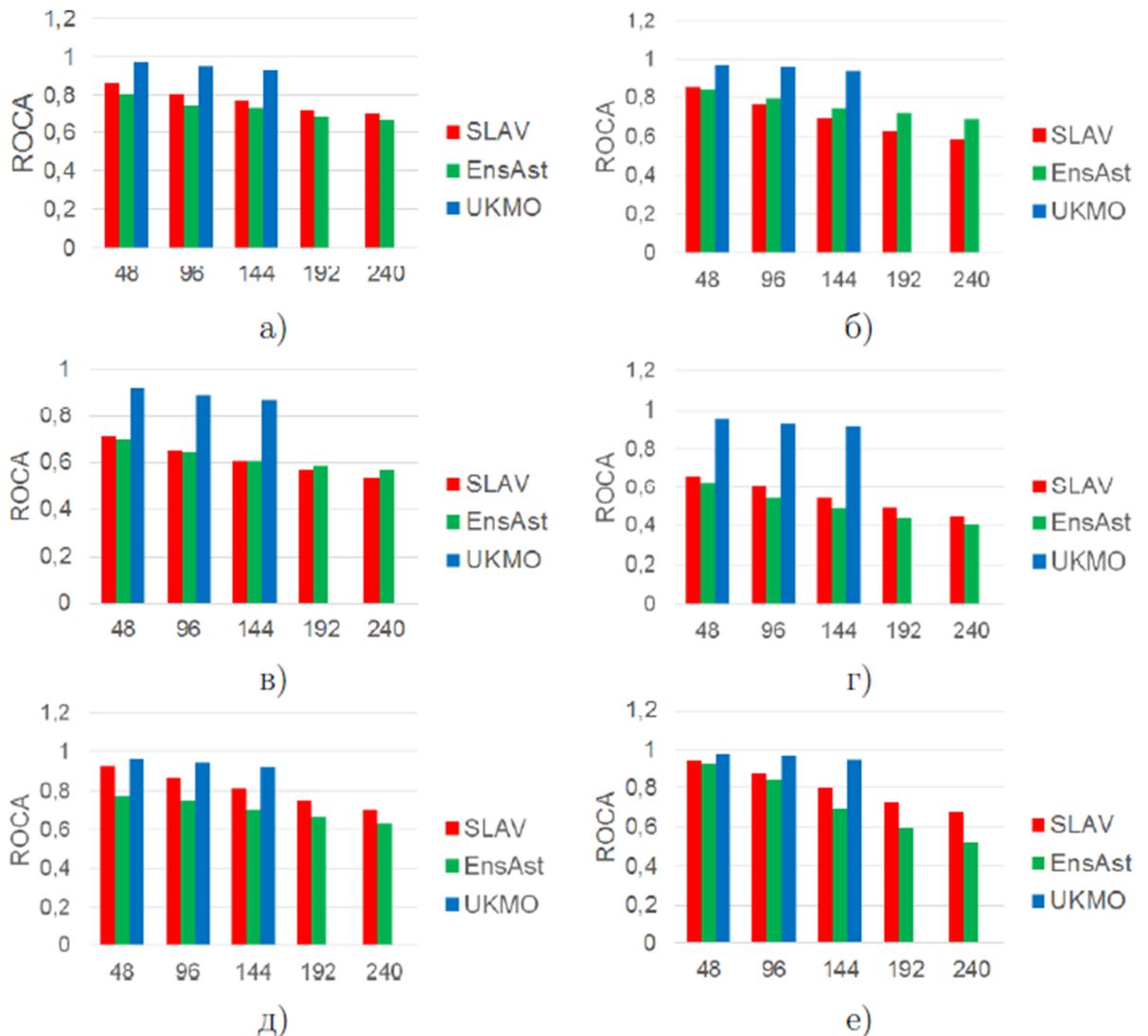


Рисунок 4.18 — Площади под ROC кривыми, рассчитанными в период испытаний для положительных аномалий величиной более 1 стандартного отклонения (а, в, д) и менее 1 стандартного отклонения (б, г, е) полей приземного давления (а, б), температуры на уровне 850гПа (в, г) и геопотенциала 500 гПа (д, е) в тропиках по результатам ансамблевых прогнозов САП ПЛАВ (красные столбцы), САП на основе спектральной модели Т169L31 (зелёные столбцы) и САП УКМО (синие столбцы). На горизонтальной шкале – заблаговременность прогноза в часах

На рисунке 4.18 приведены площади под ROC кривыми, полученные по результатам ансамблевых прогнозов в тропиках (20 ю. ш. – 20 с. ш.) в зависимости от заблаговременности прогноза (приведены результаты для заблаговременности прогнозов 2, 4, 6, 8, 10 суток — заблаговременности прогнозов приведены в часах на горизонтальной шкале). ROC кривые строились для положительных аномалий более 1 стандартного отклонения (рисунок 4.18 а, в, д) и отрицательных аномалий меньше 1 стандартного отклонения (рисунок 4.18 б, г, е) для ансамблевых прогнозов полей приземного давления, температуры на уровне 850 гПа и геопотенциала 500 гПа. Для сравнения вместе с результатами, полученными на испытаниях описываемой в данной работе системы ансамблевого прогноза (на рисунках отмечена как SLAV, столбцы красного цвета) приведены площади под ROC кривыми этих же полей, полученные с использованием САП Гидрометцентра России на основе спектральной модели [1] (EnsAst, зелёный цвет на рисунках) и площади под ROC кривыми, полученные для системы ансамблевого прогноза УКМО (синий цвет на рисунках).

Как видно на рисунках, в тропиках ансамблевые прогнозы САП на основе ПЛАВ заметно уступают прогнозам САП УКМО для всех заблаговременностей и всех приведённых переменных. Особенно заметна разница в качестве прогнозов поля T850, что может объясняться недостаточным пространственным разрешением модели ПЛАВ2018. Успешность (показатель ROCA) прогноза САП ПЛАВ поля H500 почти не уступает прогнозу САП УКМО при небольшой заблаговременности прогноза (меньше 2 суток), но качество прогноза САП ПЛАВ быстро убывает с ростом заблаговременности. Площади под ROC кривыми прогнозов САП ПЛАВ убывают быстрее, чем в САП на основе спектральной модели T169L31.

### 4.3 Выводы к главе 4

В главе 4 представлены работы по использованию ансамблевой системы усвоения данных на основе LETKF для реализации среднесрочного ансамблевого прогноза. В результате была создана система ансамблевого среднесрочного прогноза погоды с размером ансамбля несколько десятков, которая успешно

прошла оперативные испытания и внедрена в оперативную практику Гидрометцентра России.

В указанной системе ансамблевое усвоение применяется для генерации начальных данных. Недостаток применяемого для ансамблевого усвоения алгоритма LETKF, связанный с невозможностью использования во время усвоения наиболее полезных спутниковых данных наблюдений, преодолевается с помощью центрирования ансамбля на оперативный анализ Гидрометцентра России. При этом разброс в ансамбле прогнозов моделируется в ансамблевой системе усвоения.

Оперативные испытания показали высокую надёжность функционирования представленной системы ансамблевого прогноза погоды.

## Заключение

В диссертации разработана система ансамблевого усвоения данных, использующая доступные контактные наблюдения и спутниковые наблюдения ветра AMV. Для эффективного усвоения данных AMV переопределяется высота наблюдений и учитывается корреляции ошибок наблюдений AMV. Эта система применена в оперативной системе ансамблевого среднесрочного прогноза погоды Гидрометцентра России. Оперативные испытания, проведённые в период с августа 2021 по январь 2022 года, показали высокую надёжность функционирования описываемой технологии.

Основные результаты работы заключаются в следующем:

1. *Создана (совместно с соавторами) первая в России ансамблевая система усвоения данных наблюдений, работающая в оперативном режиме.* Создана и применена методика настройки ансамблевой системы усвоения.

2. *Учёт свойств данных спутниковых наблюдений ветра AMV повысил точность полей ансамбля начальных данных и среднесрочных прогнозов на их основе.* Этот результат достигнут за счёт переопределения высоты наблюдения AMV и моделирования недиагональной матрицы ковариации ошибок наблюдений AMV. Переопределение высоты выполняется с помощью вычисления коэффициента согласованности ветра. Моделирование матрицы ковариации ошибок наблюдений AMV позволяет более точно учитывать связи ошибок наблюдений для учёта корреляции в ошибках наблюдений AMV. Значимые уменьшения среднеквадратической ошибки составили от 12 до 54 % во внетропической части Северного полушария.

3. *На основе представленной ансамблевой системы усвоения разработана, испытана и внедрена (с соавторами) система ансамблевого среднесрочного прогноза погоды.* Для получения этого результата описанная система усвоения была адаптирована к применению в рамках системы ансамблевого прогноза погоды. По результатам оперативных испытаний указанная система ансамблевого прогноза была принята в качестве оперативной в Гидрометцентре России. Она показала лучшие результаты в сравнении с предыдущей системой и сравнимые с зарубежными аналогами.

В результате была решена важная научная задача, имеющая существенное значение для развития численного прогноза погоды в России — повышение качества среднесрочного ансамблевого прогноза. Представленная система ансамблевого среднесрочного прогноза погоды внедрена в оперативную практику Гидрометцентра России в 2022 году.

## Список литературы

1. *Астахова Е. Д., Бундель А. Ю., Багров А. Н., Розинкина И. А., Пономарева Т. Я., Рузанова И. В., Цветков В. И.* Система ансамблевого глобального прогноза метеорологических полей с заблаговременностью до 240 часов: результаты оперативных испытаний // Результаты испытания новых и усовершенствованных технологий, моделей и методов гидрометеорологических прогнозов. — 2016. — № 43. — С. 63—79.
2. *Беляев К. П., Тучкова Н. П., Курхнер И.* Метод коррекции модельных расчетов по данным измерений, основанный на диффузионном приближении, и его применения для анализа гидрофизических характеристик // Математическое моделирование. — 2009. — Т. 21, № 3. — С. 53—68.
3. *Богословский Н. Н., Толстых М. А., Шляева А. В.* Усвоение почвенных и приземных переменных в глобальной полулагранжевой модели прогноза погоды // Вычислительные технологии. — 2008. — Т. 13, S3.
4. *Гандин Л. С.* Автоматический контроль качества текущей метеорологической информации // Метеорология и гидрология. — 1969. — № 3. — С. 3—13.
5. *Гандин Л. С.* Объективный анализ метеорологических полей. — Ленинград : Гидрометеиздат, 1963. — 287 с.
6. *Калмыков В. В., Ибраев Р. А.* SMF-фреймворк для модели Земной системы высокого разрешения // Суперкомпьютерные дни в России. Труды международной конференции. Суперкомпьютерный консорциум университетов России. — 2015. — С. 34.
7. *Кокорев В. А., Шерстюков А. Б.* О метеорологических данных для изучения современных и будущих изменений климата на территории России // Арктика XXI век. Естественные науки. — 2015. — 2 (3).
8. *Мизяк В. Г., Алипова К. А., Толстых М. А., Рогутов В. С.* Результаты оперативных испытаний глобальной системы ансамблевого среднесрочного прогноза погоды на основе модели ПЛАВ // Результаты испытания новых и усовершенствованных технологий, моделей и методов гидрометеорологических прогнозов. — 2023. — № 50. — С. 44—65.

9. *Мизяк В. Г., Шляева А. В., Толстых М. А.* Использование данных спутниковых наблюдений ветра AMV в системе ансамблевого усвоения данных // Метеорология и гидрология. — 2016. — № 6. — С. 87—99.
10. *Мизяк В. Г., Шляева А. В., Толстых М. А.* Параллельная реализация локального ансамблевого фильтра Калмана для усвоения атмосферных данных // Инженерный журнал: наука и инновации. — 2013. — № 6.
11. *Мизяк В. Г., Шляева А. В., Толстых М. А.* Учет скоррелированности ошибок спутниковых данных наблюдений AMV в ансамблевой системе усвоения данных // Метеорология и гидрология. — 2023. — № 3. — С. 20—31.
12. *Михайлов Г. М., Беляев К. П., Тучкова Н. П., Сальников А. Н.* Ансамблевые эксперименты по усвоению данных наблюдений уровня океана в модель MPI-ESM // Научный сервис в сети Интернет. — 2016. — С. 259—263.
13. Наставление по Глобальной системе обработки данных и прогнозирования. — Женева : Всемирная Метеорологическая Организация, 2019. — 166 с.
14. Наставление по кодам. — Женева : Всемирная Метеорологическая Организация, 2019. — 266 с.
15. *Пененко В. В., Образцов Н. Н.* Вариационный метод согласования полей метеорологических элементов // Метеорология и гидрология. — 1976. — Т. 13, № 11. — С. 1.
16. *Пененко В. В.* Методы численного моделирования атмосферных процессов. — Гидрометеиздат, 1981.
17. Программа верификации ансамблевых прогнозов модели ПЛАВ : заявка RU 2017610851 Рос. Федерация / В. Г. Мизяк (Россия) ; Ф. государственное бюджетное учреждение «Гидрометеорологический научно-исследовательский центр Российской Федерации». — № 2016662860 ; заявл. 18.01.2017 ; опубл. 18.01.2017, приоритет (Рос. Федерация).
18. *Робер А.* Полуявный метод. Т. 2. — Л.: Гидрометеиздат, 1982. — С. 302—315.

19. *Рогутов В. С., Толстых М. А.* Усвоение и коррекция данных радиозондовых наблюдений за влажностью в системе усвоения данных на основе локального ансамблевого фильтра Калмана // *Метеорология и гидрология*. — М., 2015. — № 4. — С. 32—45.
20. *Рогутов В. С., Толстых М. А., Мизяк В. Г.* Система ансамблевого прогноза на основе локального ансамблевого фильтра Калмана // *Труды Гидрометеорологического научно-исследовательского центра Российской Федерации*. — 2017. — № 364. — С. 5—19.
21. *Толстых М. А.* Глобальная полулагранжева модель численного прогноза погоды. — М.; Обнинск : Эксмо: МИРБИС, 2010. — 111 с.
22. *Толстых М. А., Богословский Н. Н., Шляева А. В., Мизяк В. Г.* Оперативная технология расчета глобальных прогнозов с помощью полулагранжевой модели атмосферы ПЛАВ // *Труды Гидрометцентра России*. — 2011. — № 346. — С. 170—180.
23. *Толстых М. А., Фадеев Р. Ю., Шляева А. В., Мизяк В. Г., Рогутов В. С., Богословский Н. Н., Гойман Г. С., Махнорылова С., Юрова А. Ю., Шашкин В. В.* Система моделирования атмосферы для бесшовного прогноза. — Федеральное государственное бюджетное учреждение Гидрометеорологический научно-исследовательский центр Российской Федерации, 2017.
24. *Толстых М. А., Фадеев Р. Ю., Шашкин В. В., Травова С. В., Гойман 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*. — 2019. — № 4. — С. 77—91.
25. *Фадеев Р. Ю., Толстых М. А., Володин Е. М.* Климатическая версия модели атмосферы ПЛАВ: разработка и первые результаты // *Метеорология и гидрология*. — 2019. — № 1. — С. 22—35.
26. *Фадеев Р. Ю., Шашкин В. В., Толстых М. А., Травова С. В., Мизяк В. Г., Рогутов В. С., Алипова К. А.* Развитие системы долгосрочного прогноза Гидрометцентра России в 2020 году // *Гидрометеорологические исследования и прогнозы*. — 2021. — № 1. — С. 58—72.

27. *Цырульников М. Д., Свиренко П. И., Горин В. Е., Горбунов М. Е., Ордин А. Л., Багров А. Н.* Новая технология объективного анализа на основе схемы 3d-VAR // Информационный сборник № 39. Результаты испытания новых и усовершенствованных технологий, моделей и методов гидрометеорологических прогнозов/под редакцией канд. геогр. наук ГК Веселовой, канд. геогр. наук АА Алексеевой.—М., Обнинск: ИГ–СОЦИН, 2012.—132 стр. — 2012. — С. 7.
28. *Шашкин В. В., Толстых М. А., Иванова А. Р., Скриптунова Е. Н.* Версия модели атмосферы ПЛАВ в гибридной  $\sigma$ -р-системе координат по вертикали // Метеорология и гидрология. — 2017. — № 9. — С. 24–35.
29. *Шляева А. В., Мизяк В. Г., Толстых М. А.* Экспериментальная система усвоения данных на основе локального ансамблевого фильтра Калмана для глобальной модели атмосферы // Труды Гидрометцентра России. — 2012. — № 348.
30. *Alipova K. A., Goyman G. S., Tolstykh M. A., Mzyak V. G., Rogutov V. S.* Stochastic perturbation of tendencies and parameters of parameterizations in the global ensemble prediction system based on the SL-AV model // Russian Journal of Numerical Analysis and Mathematical Modelling. — 2022. — Т. 37, № 6. — С. 331–347.
31. *Anderson E., Järvinen H.* Variational quality control // Quarterly Journal of the Royal Meteorological Society. — 1999. — Vol. 125, no. 554. — P. 697–722.
32. *Anderson J. L.* An Ensemble Adjustment Kalman Filter for Data Assimilation // Monthly Weather Review. — 2001. — Dec. — Vol. 129, no. 12. — P. 2884–2903.
33. *Apte A., Hairer M., Stuart A., Voss J.* Sampling the posterior: An approach to non-Gaussian data assimilation // Physica D: Nonlinear Phenomena. — 2007. — Vol. 230, no. 1. — P. 50–64.
34. *Aravéquia J. A., Szunyogh I., Fertig E. J., Kalnay E., Kuhl D., Kostelich E. J.* Evaluation of a strategy for the assimilation of satellite radiance observations with the local ensemble transform Kalman filter // Monthly Weather Review. — 2011. — Vol. 139, no. 6. — P. 1932–1951.

35. *Bazile E., El Haiti M., Bogatchev A., Spiridonov V.* Improvement of the snow parameterization in ARPEGE/ALADIN // Proceedings of SRNWP/HIRLAM Workshop on surface processes, turbulence and mountain effects. T. 22. — 2001. — C. 24.
36. *Bergthórsson P., Döös B. R.* Numerical Weather Map Analysis1 // *Tellus*. — 1955. — Vol. 7, no. 3. — P. 329–340.
37. *Bishop C. H., Etherton B. J., Majumdar S. J.* Adaptive Sampling with the Ensemble Transform Kalman Filter. Part I: Theoretical Aspects // *Monthly Weather Review*. — 2001. — Mar. — Vol. 129, no. 3. — P. 420–436.
38. *Bonavita M., Holm E., Isaksen L., Fisher M.* The evolution of the ECMWF hybrid data assimilation system. — 2014.
39. *Bonavita M., Torrisi L., Marcucci F.* Ensemble data assimilation with the CNMCA regional forecasting system // *Quarterly Journal of the Royal Meteorological Society*. — 2010. — Vol. 136, no. 646. — P. 132–145.
40. *Bonavita M., Torrisi L., Marcucci F.* The ensemble Kalman filter in an operational regional NWP system: preliminary results with real observations // *Quarterly Journal of the Royal Meteorological Society*. — 2008. — Vol. 134, no. 636. — P. 1733–1744.
41. *Boone A., Masson V., Meyers T., Noilhan J.* The influence of the inclusion of soil freezing on simulations by a soil–vegetation–atmosphere transfer scheme // *Journal of Applied Meteorology and Climatology*. — 2000. — T. 39, № 9. — C. 1544–1569.
42. *Borde R.* Atmospheric motion vectors: status of operational product and ongoing development work // Proceedings of the 9th International Wind Workshop, Annapolis, Maryland, USA. — Presentation at International Symposium on Data Assimilation, Offenbach, Germany. 2012.
43. *Bormann N., Bauer P.* Estimates of spatial and interchannel observation-error characteristics for current sounder radiances for numerical weather prediction. I: Methods and application to ATOVS data // *Quarterly Journal of the Royal Meteorological Society*. — 2010. — Vol. 136, no. 649. — P. 1036–1050.

44. *Bormann N., Saarinen S., Kelly G., Thépaut J.-N.* The spatial structure of observation errors in atmospheric motion vectors from geostationary satellite data // *Monthly Weather Review*. — 2003. — Vol. 131, no. 4. — P. 706–718.
45. *Brankart J.-M., Cosme E., Testut C.-E., Brasseur P., Verron J.* Efficient Adaptive Error Parameterizations for Square Root or Ensemble Kalman Filters: Application to the Control of Ocean Mesoscale Signals // *Mon. Weather Rev.* — 2010. — Vol. 138, no. 3. — P. 932–950.
46. *Buizza R., Miller M., Palmer T. N.* Stochastic representation of model uncertainties in the ECMWF Ensemble Prediction System // *Quarterly Journal of the Royal Meteorological Society*. — 1999. — Vol. 125, no. 560. — P. 2887–2908.
47. *Cardinali C.* Monitoring the observation impact on the short-range forecast // *Quarterly Journal of the Royal Meteorological Society*. — 2009. — Vol. 135, no. 638. — P. 239–250.
48. *Charney J. G., Fjörtoft R., Neumann J. v.* Numerical integration of the barotropic vorticity equation // *Tellus*. — 1950. — Vol. 2, no. 4. — P. 237–254.
49. *Chou M.-D., Suarez M. J.* A solar radiation parameterization for atmospheric studies : TEX. ОТЧ. — 1999.
50. *Collins W. G.* The operational complex quality control of radiosonde heights and temperatures at the National Centers for Environmental Prediction. Part I: Description of the method // *Journal of applied meteorology*. — 2001. — Vol. 40, no. 2. — P. 137–151.
51. *Courtier P., Thépaut J.-N., Hollingsworth A.* A strategy for operational implementation of 4D-Var, using an incremental approach // *Quarterly Journal of the Royal Meteorological Society*. — 1994. — Vol. 120, no. 519. — P. 1367–1387.
52. *Cressman G. P.* An operational objective analysis system // *Mon. Wea. Rev.* — 1959. — Vol. 87, no. 10. — P. 367–374.
53. *Daley R.* Atmospheric data analysis. — Cambridge university press, 1993.

54. *Desroziers G.* Observation error specifications // Advanced Data Assimilation for Geosciences: Lecture Notes of the Les Houches School of Physics: Special Issue, June 2012. — 2014. — P. 209.
55. *Dimet F.-X. L., Talagrand O.* Variational algorithms for analysis and assimilation of meteorological observations: theoretical aspects // *Tellus A.* — 1986. — Vol. 38, no. 2. — P. 97–110.
56. *English S., McNally T., Bormann N., Salonen K., Matricardi M., Horanyi A., Rennie M., Janisková M., Di Michele S., Geer A., [et al.].* Impact of satellite data : tech. rep. / ECMWF Technical Memorandum. — 2013.
57. *Evensen G.* Advanced data assimilation for strongly nonlinear dynamics // *Monthly weather review.* — 1997. — Vol. 125, no. 6. — P. 1342–1354.
58. *Evensen G.* The Ensemble Kalman Filter: theoretical formulation and practical implementation // *Ocean Dynamics.* — 2003. — Nov. — Vol. 53, no. 4. — P. 343–367.
59. *Eyre J., Bell W., Cotton J., English S., Forsythe M., Healy S., Pavein E.* Assimilation of satellite data in numerical weather prediction. Part II: Recent years // *Quarterly Journal of the Royal Meteorological Society.* — 2022. — T. 148, № 743. — C. 521–556.
60. *Forsythe M., Saunders R.* AMV errors: a new approach in NWP // *Proceedings of the 9th International Wind Workshop, Annapolis, Maryland, USA.* — Citeseer. 2008. — P. 14–18.
61. *Gandin L. S.* Complex quality control of meteorological observations // *Monthly Weather Review.* — 1988. — Vol. 116, no. 5. — P. 1137–1156.
62. *Gaspari G., Cohn S. E.* Construction of correlation functions in two and three dimensions // *Quarterly Journal of the Royal Meteorological Society.* — 1999. — Vol. 125, no. 554. — P. 723–757.
63. *Giard D., Bazile E.* Implementation of a new assimilation scheme for soil and surface variables in a global NWP model // *Monthly Weather Review.* — 2000. — No. 1991. — P. 997–1015.
64. *Gorin V. E., Tsyrunikov M. D.* Estimation of multivariate observation-error statistics for AMSU-A data // *Monthly weather review.* — 2011. — T. 139, № 12. — C. 3765–3780.

65. *Greybush S. J., Kalnay E., Miyoshi T., Ide K., Hunt B. R.* Balance and Ensemble Kalman Filter Localization Techniques // Monthly Weather Review. — 2011. — Feb. — Vol. 139, no. 2. — P. 511–522.
66. *Hamill T. M., Whitaker J. S., Snyder C.* Distance-Dependent Filtering of Background Error Covariance Estimates in an Ensemble Kalman Filter // Monthly Weather Review. — 2001. — Nov. — Vol. 129, no. 11. — P. 2776–2790.
67. *Hamill T. M., Snyder C.* A hybrid ensemble Kalman filter–3D variational analysis scheme // Monthly Weather Review. — 2000. — Vol. 128, no. 8. — P. 2905–2919.
68. *Harlim J., Hunt B. R.* Four-dimensional local ensemble transform Kalman filter: numerical experiments with a global circulation model // Tellus A. — 2007. — Vol. 59, no. 5. — P. 731–748.
69. *Hernandez-Carrascal A., Bormann N., Borde R., Lutz H., Otkin J., Wanzong S.* Atmospheric Motion Vectors from model simulations. Part I: Methods and characterisation as single-level estimates of wind : tech. rep. / ECMWF Technical Memo. — 2012.
70. *Houtekamer P. L., Mitchell H. L.* A Sequential Ensemble Kalman Filter for Atmospheric Data Assimilation // Monthly Weather Review. — 2001. — Jan. — Vol. 129, no. 1. — P. 123–137.
71. *Houtekamer P. L., Mitchell H. L.* Data Assimilation Using an Ensemble Kalman Filter Technique // Monthly Weather Review. — 1998. — Mar. — Vol. 126, no. 3. — P. 796–811.
72. *Houtekamer P. L., Mitchell H. L.* Ensemble Kalman filtering // Quarterly Journal of the Royal Meteorological Society. — 2005. — Oct. — Vol. 131, no. 613. — P. 3269–3289.
73. *Houtekamer P. L., Mitchell H. L., Deng X.* Model Error Representation in an Operational Ensemble Kalman Filter // Monthly Weather Review. — 2009. — July. — Vol. 137, no. 7. — P. 2126–2143.

74. *Hunt B. R., Kalnay E., Kostelich E. J., Ott E., Patil D. J., Sauer T., Szunyogh I., Yorke J. a., Zimin A. V., Corazza M.* A local ensemble Kalman filter for atmospheric data assimilation // *Tellus A.* — 2004. — Aug. — Vol. 56, no. 4. — P. 273–277.
75. *Hunt B. R., Kostelich E. J., Szunyogh I.* Efficient data assimilation for spatiotemporal chaos: A local ensemble transform Kalman filter // *Physica D: Nonlinear Phenomena.* — 2007. — June. — Vol. 230, no. 1/2. — P. 112–126.
76. *Isaksen L., Bonavita M., Buizza R., Fisher M., Haseler J., Leutbecher M., Raynaud L.* Ensemble of data assimilations at ECMWF : Technical Memorandum. — 12/2010. — P. 45.
77. *Kalman R. E.* A new approach to linear filtering and prediction problems // *Journal of basic Engineering.* — 1960. — Vol. 82, no. 1. — P. 35–45.
78. *Kalnay E.* Atmospheric modeling, data assimilation and predictability. — Cambridge university press, 2003.
79. *Kalnay E., Li H., Miyoshi T., YANG S.-C., BALLABRERA-POY J.* 4-D-Var or ensemble Kalman filter? // *Tellus A.* — 2007. — Vol. 59, no. 5. — P. 758–773.
80. *Kalnay E., Li H., Miyoshi T., Yang S.-C., Ballabrera-Poy J.* 4-D-Var or ensemble Kalman filter? // *Tellus A.* — 2007. — Oct. — Vol. 59, no. 5. — P. 758–773.
81. *Kaurkin M., Ibrayev R., Koromyslov A.* EnOI-based data assimilation technology for satellite observations and ARGO float measurements in a high resolution global ocean model using the CMF platform // *Communications in Computer and Information Science.* — 2016.
82. *Keppenne C. L., Rienecker M. M., Applications S., Corporation I., Space G.* Initial Testing of a Massively Parallel Ensemble Kalman Filter with the Poseidon Isopycnal Ocean General Circulation Model // *Mon. Weather Rev.* — 2002. — Dec. — Vol. 130, no. 12. — P. 2951–2965.

83. *Kleist D. T., Ide K.* An OSSE-Based Evaluation of Hybrid Variational–Ensemble Data Assimilation for the NCEP GFS. Part II: 4D-EnVar and Hybrid Variants // *Mon. Weather Rev.* — 2015. — Vol. 143, no. 2. — P. 452–470.
84. *Lahoz W., Khattatov B., Menard R.* Data assimilation: making sense of observations. — Springer Science & Business Media, 2010.
85. *Leeuwen P. J. van.* Nonlinear data assimilation in geosciences: an extremely efficient particle filter // *Quarterly Journal of the Royal Meteorological Society.* — 2010. — Vol. 136, no. 653. — P. 1991–1999.
86. *Lions J.-L.* Contrôle optimal de systèmes gouvernés par des équations aux dérivées partielles. — 1968.
87. *Liu C., Xiao Q., Wang B.* An ensemble-based four-dimensional variational data assimilation scheme. Part I: Technical formulation and preliminary test // *Monthly Weather Review.* — 2008. — Vol. 136, no. 9. — P. 3363–3373.
88. *Liu Z.-Q., Rabier F.* The potential of high-density observations for numerical weather prediction: A study with simulated observations // *Quarterly Journal of the Royal Meteorological Society: A journal of the atmospheric sciences, applied meteorology and physical oceanography.* — 2003. — T. 129, № 594. — C. 3013–3035.
89. *Lorenc A. C., Hammon O.* Objective quality control of observations using Bayesian methods. Theory, and a practical implementation // *Quarterly Journal of the Royal Meteorological Society.* — 1988. — Vol. 114, no. 480. — P. 515–543.
90. *Lorenz E. N.* Predictability: A problem partly solved // *Proc. Seminar on predictability.* Vol. 1. — 1996.
91. *Marchuk G., Penenko V.* Application of optimization methods to the problem of mathematical simulation of atmospheric processes and environment. — 1979.
92. *Medeiros M. S., Herdies D. L., Aravequia J. A., Souza S. S., [et al.].* The Impact of Assimilation with the Inclusion of Amsu-A Radiances in 4D-LETKF/AGCM System // *AGU Fall Meeting Abstracts.* — 2010.

93. *Miyoshi T., Kunii M.* The Local Ensemble Transform Kalman Filter with the Weather Research and Forecasting Model: Experiments with Real Observations // Pure and Applied Geophysics. — 2011. — July. — Vol. 169, no. 3. — P. 321–333.
94. *Miyoshi T., Sato Y.* Assimilating Satellite Radiances with a Local Ensemble Transform Kalman Filter (LETKF) Applied to the JMA Global Model (GSM) // Sola. — 2007. — Vol. 3. — P. 37–40.
95. *Miyoshi T., Sato Y., Kadowaki T.* Ensemble Kalman Filter and 4D-Var Intercomparison with the Japanese Operational Global Analysis and Prediction System // Monthly Weather Review. — 2010. — July. — Vol. 138, no. 7. — P. 2846–2866.
96. *Mlawer E. J., Taubman S. J., Brown P. D., Iacono M. J., Clough S. A.* Radiative transfer for inhomogeneous atmospheres: RRTM, a validated correlated-k model for the longwave // Journal of Geophysical Research: Atmospheres. — 1997. — T. 102, № D14. — C. 16663–16682.
97. *Noilhan J., Mahfouf J.-F.* The ISBA land surface parameterisation scheme // Global and planetary Change. — 1996. — T. 13, № 1–4. — C. 145–159.
98. *Oczkowski M., Szunyogh I., Patil D.* Mechanisms for the development of locally low-dimensional atmospheric dynamics // Journal of the atmospheric sciences. — 2005. — Vol. 62, no. 4. — P. 1135–1156.
99. *Oke P. R., Sakov P., Corney S. P.* Impacts of localisation in the EnKF and EnOI: experiments with a small model // Ocean Dynamics. — 2006. — Sept. — Vol. 57, no. 1. — P. 32–45.
100. *Ott E., Hunt B. R., Szunyogh I., Zimin A. v. A., Kostelich E. J., Corazza M., Kalnay E., Patil D. J., Yorke J. A.* A Local Ensemble Kalman Filter for Atmospheric Data Assimilation // Tellus A. — 2004. — Oct. — Vol. A 56, April. — P. 415–428.
101. *Ott E., Patil D., Kalnay E., Corazza M., Szunyogh I., Hunt B., Yorke J. A.* Exploiting local low dimensionality of the atmospheric dynamics for efficient ensemble Kalman filtering : tech. rep. — 2002.

102. *Palmer T. N., Buizza R., Doblas-Reyes F., Jung T., Leutbecher M., Shutts G., Steinheimer M., Weisheimer A.* Stochastic parametrization and model uncertainty // ECMWF, Shinfield Park, Reading RG2-9AX, UK, ECMWF Research Department Technical Memorandum. — 2009. — No. 598. — P. 42.
103. *Parrish D. F., Derber J. C.* The National Meteorological Center's spectral statistical-interpolation analysis system // Monthly Weather Review. — 1992. — Vol. 120, no. 8. — P. 1747–1763.
104. *Patil D. J., Hunt B. R., Kalnay E., Yorke J. A., Ott E.* Local low dimensionality of atmospheric dynamics // Physical Review Letters. — 2001. — Vol. 86, no. 26. — P. 5878.
105. *Patil D., Hunt B., Kalnay E., Yorke J., Ott E.* Local Low Dimensionality of Atmospheric Dynamics // Physical Review Letters. — 2001. — June. — Vol. 86, no. 26. — P. 5878–5881.
106. *Perianez A., Reich H., Potthast R.* Adaptive localization for ensemble methods in data assimilation // International Conference on Ensemble Methods in Geophysical Sciences. — 2012.
107. *Rabier F., Järvinen H., Klinker E., Mahfouf J.-F., Simmons A.* The ECMWF operational implementation of four-dimensional variational assimilation. I: Experimental results with simplified physics // Quarterly Journal of the Royal Meteorological Society. — 2000. — Vol. 126, no. 564. — P. 1143–1170.
108. *Rainwater S., Bishop C. H., Campbell W. F.* The benefits of correlated observation errors for small scales // Quarterly Journal of the Royal Meteorological Society. — 2015. — T. 141, № 693. — C. 3439–3445.
109. *Richardson L.* Weather Prediction by Numerical Process. — 1922.
110. *Sakov P., Bertino L.* Relation between two common localisation methods for the EnKF // Computational Geosciences. — 2010. — July. — Vol. 15, no. 2. — P. 225–237.
111. *Sakov P., Oke P. R.* Implications of the Form of the Ensemble Transformation in the Ensemble Square Root Filters // Monthly Weather Review. — 2008. — Mar. — Vol. 136, no. 3. — P. 1042–1053.

112. *Salonen K., Bormann N.* Accounting for the characteristics of AMV observation errors in data assimilation // Proceedings of the 2011 EUMETSAT Meteorological Satellite Conference, Oslo, Norway. — 2011. — P. 5–9.
113. *Salonen K., Cotton J., Bormann N., Forsythe M.* Characterising AMV height assignment error by comparing best-fit pressure statistics from the Met Office and ECMWF system // Proceedings of the 11th International Wind Workshop, Auckland, New Zealand. — 2012. — P. 20–24.
114. *Salonen K., Cotton J., Bormann N., Forsythe M.* Characterizing AMV Height-Assignment Error by Comparing Best-Fit Pressure Statistics from the Met Office and ECMWF Data Assimilation Systems // Journal of Applied Meteorology and Climatology. — 2015. — Vol. 54, no. 1. — P. 225–242.
115. *Schellekens J., Weerts A. H., Moore R. J., Pierce C. E., Hildon S.* The use of MOGREPS ensemble rainfall forecasts in operational flood forecasting systems across England and Wales // Advances in Geosciences. — 2011. — T. 29. — C. 77–84.
116. *Shlyayeva A., Tolstykh M., Mizyak V., Rogutov V.* Local ensemble transform Kalman filter data assimilation system for the global semi-Lagrangian atmospheric model // Russian Journal of Numerical Analysis and Mathematical Modelling. — 2013. — Jan. — Vol. 28, no. 4.
117. *Shutts G.* A kinetic energy backscatter algorithm for use in ensemble prediction systems // Quarterly Journal of the Royal Meteorological Society. — 2005. — Oct. — Vol. 131, no. 612. — P. 3079–3102.
118. *Staniforth A., Côté J.* Semi-Lagrangian integration schemes for atmospheric models—A review // Monthly weather review. — 1991. — T. 119, № 9. — C. 2206–2223.
119. *Stewart L., Dance S., Nichols N.* Data assimilation with correlated observation errors: analysis accuracy with approximate error covariance matrices // School of Mathematics and Statistics. — 2012.
120. *Stewart L. M., Dance S., Nichols N.* Correlated observation errors in data assimilation // International journal for numerical methods in fluids. — 2008. — Vol. 56, no. 8. — P. 1521–1527.

121. *Szunyogh I., Kostelich E. J., Gyarmati G., Kalnay E., Hunt B. R., Ott E., Satterfield E., Yorke J. A.* A local ensemble transform Kalman filter data assimilation system for the NCEP global model // *Tellus A.* — 2007. — Sept. — P. 1–18.
122. *Tarasova T., Fomin B.* The use of new parameterizations for gaseous absorption in the CLIRAD-SW solar radiation code for models // *Journal of Atmospheric and Oceanic Technology.* — 2007. — T. 24, № 6. — С. 1157–1162.
123. *Tavolato C., Isaksen L.* Data Usage and Quality Control for ERA-40, ERA-Interim and the Operational ECMWF Data Assimilation System. — European Centre for Medium Range Weather Forecasts, 2011.
124. *Termonia P., Fischer C., Bazile E., Bouyssel F., Brožková R., Bénard P., Bochenek B., Degrauwe D., Derková M., El Khatib R.* [и др.]. The ALADIN System and its canonical model configurations AROME CY41T1 and ALARO CY40T1 // *Geoscientific Model Development.* — 2018. — T. 11, № 1. — С. 257–281.
125. *Tippett M. K., Anderson J. L., Bishop C. H., Hamill T. M., Whitaker J. S.* Ensemble Square Root Filters // *Monthly Weather Review.* — 2003. — ИЮЛЬ. — Т. 131, № 7. — С. 1485–1490.
126. *Tolstykh M., Shashkin V., Fadeev R., Goyman G.* Vorticity-divergence semi-Lagrangian global atmospheric model SL-AV20: dynamical core // *Geoscientific Model Development.* — 2017. — T. 10, № 5. — С. 1961–1983.
127. *Tsyrlunikov M., Sotskiy A.* A locally stationary convolutional ensemble filter // *arXiv preprint arXiv:2306.14318.* — 2023.
128. *Tuan Pham D., Verron J., Christine Roubaud M.* A singular evolutive extended Kalman filter for data assimilation in oceanography // *J. Mar. Syst.* — 1998. — Vol. 16, no. 3/4. — P. 323–340.
129. *Waller J. A., Dance S. L., Lawless A. S., Nichols N. K., Eyre J.* Representativity error for temperature and humidity using the Met Office high-resolution model // *Quarterly Journal of the Royal Meteorological Society.* — 2014. — Vol. 140, no. 681. — P. 1189–1197.

130. *Walters D., Baran A. J., Boutle I., Brooks M., Earnshaw P., Edwards J., Furtado K., Hill P., Lock A., Manners J., [et al.]*. The Met Office Unified Model global atmosphere 7.0/7.1 and JULES global land 7.0 configurations // *Geoscientific Model Development*. — 2019. — Vol. 12, no. 5. — P. 1909–1963.
131. *Wang X., Barker D. M., Snyder C., Hamill T. M.* A hybrid ETKF–3DVAR data assimilation scheme for the WRF model. Part I: Observing system simulation experiment // *Monthly Weather Review*. — 2008. — Vol. 136, no. 12. — P. 5116–5131.
132. *Wang X., Barker D. M., Snyder C., Hamill T. M.* A hybrid ETKF–3DVAR data assimilation scheme for the WRF model. Part II: Real observation experiments // *Monthly Weather Review*. — 2008. — Vol. 136, no. 12. — P. 5132–5147.
133. *Wang X., Hamill T. M., Whitaker J. S., Bishop C. H., Wang X.* A comparison of hybrid ensemble transform Kalman filter–OI and ensemble square-root filter analysis schemes // *Mon. Wea. Rev.* — 2006.
134. *Wang X., Snyder C., Hamill T. M.* On the theoretical equivalence of differently proposed ensemble–3DVAR hybrid analysis schemes // *Monthly weather review*. — 2007. — Vol. 135, no. 1. — P. 222–227.
135. *Weissmann M., Folger K., Lange H.* Height correction of atmospheric motion vectors using airborne lidar observations // *Journal of Applied Meteorology and Climatology*. — 2013. — Vol. 52, no. 8. — P. 1868–1877.
136. *Whitaker J. S., Compo G. P., Wei X., Hamill T. M.* Reanalysis without radiosondes using ensemble data assimilation // *Monthly Weather Review*. — 2004. — May. — Vol. 132, no. 5. — P. 1190–1200.
137. *Whitaker J. S., Hamill T. M.* Ensemble Data Assimilation without Perturbed Observations // *Monthly Weather Review*. — 2002. — July. — Vol. 130, no. 7. — P. 1913–1924.
138. *Whitaker J. S., Hamill T. M., Wei X., Song Y., Toth Z.* Ensemble data assimilation with the NCEP global forecast system // *Monthly Weather Review*. — 2008. — Vol. 136, no. 2. — P. 463–482.

139. *Wiener N., Masani P.* The prediction theory of multivariate stochastic processes // *Acta Mathematica*. — 1957. — Vol. 98, no. 1. — P. 111–150.
140. *Wilks D. S.* Statistical methods in the atmospheric sciences. T. 100. — Academic press, 2011.
141. *Yang C., Lu Q., Zhang P.* A study on height reassignment for the AMV products of the FY-2C satellite // *Acta Meteorologica Sinica*. — 2012. — Vol. 26, no. 5. — P. 614–628.
142. *Yang S.-C., Corazza M., Carrassi A., Kalnay E., Miyoshi T.* Comparison of ensemble-based and variational-based data assimilation schemes in a quasi-geostrophic model // *Mon. Wea. Rev.* — 2009. — Vol. 137. — P. 639–709.
143. *Zhang F., Zhang M., Hansen J.* Coupling ensemble Kalman filter with four-dimensional variational data assimilation // *Adv. Atmos. Sci.* — 2009. — Vol. 26, no. 1. — P. 1–8.
144. *Zupanski M.* Maximum likelihood ensemble filter: Theoretical aspects // *Monthly Weather Review*. — 2005. — Vol. 133, no. 6. — P. 1710–1726.

