

ФЕДЕРАЛЬНОЕ ГОСУДАРСТВЕННОЕ БЮДЖЕТНОЕ УЧРЕЖДЕНИЕ
НАУКИ ИНСТИТ ФИЗИКИ АТМОСФЕРЫ ИМ. А.М. ОБУХОВА
РОССИЙСКОЙ АКАДЕМИИ НАУК

На правах рукописи

Бабанов Борис Андреевич

**Режимы крупномасштабной атмосферной циркуляции в регионах Евро-
Атлантики и Северной Евразии в условиях меняющегося климата**

Специальность 1.6.18 – Науки об атмосфере и климате

Диссертация на соискание ученой степени
кандидата физико-математических наук

Научный руководитель:
доктор физико-математических наук
академик РАН
Семенов Владимир Анатольевич

Москва, 2025

ОГЛАВЛЕНИЕ

ВВЕДЕНИЕ	4
ГЛАВА 1. СРАВНЕНИЕ МЕТОДОВ КЛАСТЕРНОГО АНАЛИЗА ДЛЯ ИДЕНТИФИКАЦИИ ПОГОДНЫХ РЕЖИМОВ В ЕВРО-АТЛАНТИЧЕСКОМ СЕКТОРЕ.....	13
1.1 Ранние объективные методы типизации атмосферных течений.....	14
1.2 Современные методы кластерного анализа.....	17
1.2.1 Иерархическая кластеризация Уорда	19
1.2.2 k-means	21
1.2.3 Модель гауссовой смеси	24
1.2.4 Самоорганизующиеся карты Кохонена	27
1.3 Определение числа кластеров	30
1.4. Идентификация погодных режимов в ЕАТ.....	37
1.4.1. Предобработка данных и кластеризация	37
1.4.2. Поля погодных режимов в ЕАТ	42
1.5. Выводы	46
ГЛАВА 2. ХАРАКТЕРИСТИКИ ПОГОДНЫХ РЕЖИМОВ В РЕГИОНАХ ЕВРО-АТЛАНТИКЕ И СЕВЕРНОЙ ЕВРАЗИИ В ЗИМНИЙ И ЛЕТНИЙ СЕЗОНЫ.....	48
2.1. Виды временных характеристик погодных режимов	48
2.2. Временные характеристики погодных режимов в ЕАТ.....	50
2.3. Идентификация погодных режимов в Северной Евразии.....	59
2.3.1. Предобработка данных.....	59
2.3.2. Выбор числа режимов.....	61
2.3.3. Анализ временных характеристик режимов в Северной Евразии	64
2.4. Пространственные структуры погодных режимов в Северной Евразии	66
2.5. Временные характеристики погодных режимов в Северной Евразии	67
2.6. Выводы	74
ГЛАВА 3. ВОСПРОИЗВЕДЕНИЕ ЕВРО-АТЛАНТИЧЕСКИХ ПОГОДНЫХ РЕЖИМОВ В МОДЕЛЯХ КЛИМАТА.	76

3.1. Данные и методы	77
3.2. Воспроизведение полей погодных режимов	79
3.3. Воспроизведение временных характеристик погодных режимов ..	83
3.3.1. Матрицы вероятностей переходов погодных режимов	89
3.4. Выводы	92
ГЛАВА 4. ПРЕДСКАЗУЕМОСТЬ ПОГОДНЫХ РЕЖИМОВ ВЕДУЩИМИ МОДАМИ КЛИМАТИЧЕСКОЙ ИЗМЕНЧИВОСТИ И ВЛИЯНИЕ РЕЖИМОВ НА ЭКСТРЕМАЛЬНЫЕ ПОГОДНЫЕ ЯВЛЕНИЯ.	93
4.1. Связь погодных режимов в Северной Евразии с Nino3.4 и колебаниями площади морского льда	93
4.2. Связь погодных режимов в Северной Евразии с экстремальными аномалиями температур и осадков	97
4.3. Современные погодные режимы в Евро-Атлантике и их влияние на экстремальные аномалии температур, осадков и ветра	101
4.4. Связь интенсивности осадков в Крыму с летними погодными режимами на европейском континенте	108
4.4.1. Идентификация погодных режимов в Европе	108
4.4.2. Анализ связи режимов в Европе с осадками в ЮБК	111
4.5 Выводы	115
ЗАКЛЮЧЕНИЕ	116
СПИСОК ЛИТЕРАТУРЫ.....	118

ВВЕДЕНИЕ

Крупномасштабная циркуляция атмосферы во внетропических широтах Северного полушария является одним из центральных объектов изучения климатологии и метеорологии. Существует множество подходов к изучению крупномасштабных атмосферных структур, в том числе статистический анализ данных наблюдений и гидродинамическое моделирование. Одним из подходов к изучению крупномасштабной циркуляции атмосферы является выделение ограниченного числа характерных структур пространственной изменчивости [1, 2], имеющих выраженные особенности, позволяющие отделять эти структуры друг от друга.

При классификации атмосферных течений фактически непрерывно меняющиеся состояния атмосферы [3], уникальные в каждый момент времени, в результате некоторой процедуры типизации приписываются одному из выделенных типов [4]. Состояния атмосферы, отнесенные в результате типизации на основе выбранных исследователем критериев к одному и тому же типу течения, рассматриваются как подобные, и уже не анализируются отдельно, а лишь как частные проявления одного из выделенных типов. Анализ всего множества состояний атмосферы при таком подходе сводится к исследованию свойств ограниченного числа выделенных типов течений [5], включая их пространственные и временные характеристики, а эволюция атмосферы рассматривается как череда сменяющих друг друга квазистационарных типов течений.

Активные исследования по типизации крупномасштабных циркуляционных процессов в атмосфере начались с середины XX века [6, 7, 8]. Критерии при оценке числа типов и способе их разделения в ранних работах основывались на некоторых, не всегда очевидных, экспертных оценках исследователей, из-за чего в последствии такие типизации были названы «субъективными» [9]. В настоящее время чисто субъективные типизации не используются, но иногда пересматриваются с использованием численных методов [10]. Благодаря распространению и росту мощностей компьютеров со второй половины XX века [11] начали разрабатываться и рутинно применяться численные классификации, основанные на компьютерной обработке больших объёмов данных и названные впоследствии «объективными» [12].

Среди объективных методов типизации атмосферных течений существует несколько подходов к тому, как и что классифицировать для

разделения множества крупномасштабных атмосферных течений на отдельные типы. Распространенным методом объективных классификаций атмосферных течений является идентификация крупномасштабных циркуляционных режимов [13], по-другому называемых погодными режимами [14]. При «режимном подходе» низкочастотная изменчивость атмосферы на крупных пространственных масштабах – всего Северного полушария [15] или масштабов отдельных континентов [16] – рассматривается в виде набора из небольшого числа квазистационарных повторяющихся пространственных структур, называемых погодными режимами, в каждом из которых атмосфера пребывает в течение некоторого длительного, относительно синоптических масштабов, времени, обычно от одной до нескольких недель [17], а затем сменяется другим режимом.

Идентификация погодных режимов в большинстве современных исследований [9] выполняется на основе методов кластерного анализа, например, метода *k*-means [18], которые применяются, как правило, к среднесуточным полям высоты геопотенциальной поверхности на различных уровнях, обычно на уровне 500 гектопаскалей [19] (*z*500), или давления на уровне моря [20] (SLP). В результате процедуры кластеризации каждое, или большинство, из суточных полей *z*500 и/или SLP приписывается к одному из нескольких *k* погодных режимов, при этом большую трудность представляет тот факт, что число режимов *k* заранее неизвестно, и его определение в рамках режимного подхода представляет собой отдельную непростую задачу [21].

Наиболее часто режимный подход, при использовании его на континентальных масштабах, применяется для выделения режимов в областях, называемых Северная Атлантика или Евро-Атлантика (ЕАТ), охватывающих всю протяженность по долготе северной внетропической части Атлантического океана и Европы. Режимный подход для исследования циркуляции атмосферы в регионе ЕАТ обычно применяется в зимний период, который, относительно других сезонов, характеризуется повышенными аномалиями полей SLP или *z*500, благодаря чему зимние погодные режимы оказываются более выраженными. Ведущей модой низкочастотной изменчивости SLP в ЕАТ представляет собой диполь давления между регионами вблизи Исландии и Азорских островов, названный Северо-Атлантическим колебанием (NAO – North Atlantic Oscillation) [22].

При идентификации небольшого числа *k* режимов (обычно от 4 до 6) в зимний период в области ЕАТ, как правило, всегда появляются два режима, которые можно ассоциировать с положительной и отрицательной фазами

колебания NAO – NAO+ и NAO-. Оставшиеся два режима, при $k = 4$, обычно представляют собой положительную фазу Скандинавской осцилляции, называемую Скандинавский блокинг (SB – Scandinavian Blocking), и режим с положительной аномалией z500 над Северной Атлантикой (AR - Atlantic Ridge), см., например, [19, 23]. При увеличении числа режимов снижается их относительная повторяемость, но растёт выраженность аномалий режимов, благодаря чему они более точно описывают реальное состояние атмосферы. Например, при использовании $k = 6$ добавляются режимы, ассоциированные с отрицательными фазами SB и AR, то есть SB «расщепляется» на SB+ и SB-, а AR на AR+ и AR- [24].

Помимо региона Евро-Атлантики, на континентальных масштабах режимный подход в Северном полушарии так же применялся к идентификации режимов в регионах Северной Америки [18, 25], Восточной Азии [26] и северной части Тихого океана [27], однако область Северной Евразии, занимаемая Россией, остается малоисследованной в рамках режимного подхода. Применение режимного подхода в Северной Евразии (СЕ) позволило бы развить понимание крупномасштабных циркуляционных процессов, непосредственно влияющих на погоду, включая её экстремальные проявления, на территории России. Известно, что погодные режимы влияют на распределение опасных экстремальных погодных явлений [28], включая волны жары и холода [29], сильные осадки и ветра [30].

Важной особенностью погодных режимов является их время жизни, часто превышающее временной масштаб одной-двух недель [31], который полагают в качестве предела предсказуемости погоды (≈ 10 суток), связанного со стохастической природой динамики атмосферы и экспоненциальным нарастанием со временем малых начальных возмущений состояния атмосферы [32]. Исследование предсказуемости погоды в рамках эволюции погодных режимов потенциально могло бы повысить точность предсказуемости за пределами этого теоретического ограничения. Например, для наиболее изученной в рамках режимного подхода области ЕАТ известно, что повторяемость погодных режимов NAO+ и NAO- модулируется фазами осцилляции Маддена – Джулиана в Тихом океане за 1-2 недели до их наступления [19].

Цель данной работы – сравнить известные методы кластерного анализа, использующиеся для идентификации погодных режимов; получить количественные оценки и тренды пространственных и временных характеристик погодных режимов в регионах ЕАТ и СЕ в зимний и летний

периоды по данным реанализа; оценить воспроизводимость характеристик режимов в климатических моделях; исследовать предсказуемость режимов с помощью ведущих мод изменчивости климатической системы и влияние погодных режимов на вероятности экстремальных погодных явлений.

Для достижения этой цели были поставлены следующие **задачи**:

1. Проанализировать используемые в научной литературе методы кластерного анализа для идентификации погодных режимов и произвести количественное сравнение характеристик погодных режимов, полученных различными методами, с целью выбора наиболее оптимального метода кластеризации для дальнейшего анализа погодных режимов.

2. Идентифицировать погодные режимы с помощью выбранного оптимального метода кластеризации для регионов ЕАТ и СЕ в зимний и летний периоды по данным реанализа и произвести сравнительный анализ пространственных и временных характеристик погодных режимов в этих регионах, а также их долгосрочных трендов.

3. Исследовать пространственно-временные характеристики погодных режимов по данным климатических моделей и сравнить их с характеристиками погодных режимов, полученных по данным реанализа.

4. Оценить изменение повторяемости погодных режимов в зависимости от фаз ведущих мод естественной изменчивости климатической системы в предшествующие сезоны, таких как колебание температуры поверхности в области Тихого океана (Эль-Ниньо – Южное колебание) и вариации площади морского льда в Северном полушарии.

5. Количественно оценить частоту событий с экстремально высокими и низкими аномалиями температур, осадков и скорости ветра в зависимости от наличия тех или иных погодных режимов в регионах ЕАТ и СЕ.

Основные положения, выносимые на защиту:

1. Результаты сравнения четырех наиболее распространённых методов кластерного анализа, использующихся для идентификации погодных режимов, указывающие на то, что k-means является оптимальным методом, позволяющим выделять погодные режимы в Евро-Атлантике с наименьшими внутрикластерными расстояниями и наибольшими пространственными корреляциями с полями погодных режимов, полученных другими методами.

2. Получены пространственные структуры, количественные оценки повторяемости, продолжительности и матрицы переходов зимних и летних погодных режимов в Евро-Атлантике и Северной Евразии за 1940-2022 гг. по данным реанализа ERA5, в том числе статистически значимые тренды сезонной повторяемости режимов и значимые по частоте переходы между режимами.

3. Количественные оценки способности моделей климата воспроизводить поля и временные характеристики погодных режимов в Евро-Атлантике в экспериментах historical. Большинство исследованных моделей воспроизводят пространственные структуры и особенности временных характеристик погодных режимов, полученных по реанализу ERA5; зимние погодные режимы в моделях воспроизводятся лучше, чем летние.

4. Выявленные связи повторяемости погодных режимов в Северной Евразии с фазами явления Эль-Ниньо и площадью морского льда в Северном полушарии в предшествующие сезоны, включая статистически значимую связь повторяемости режимов уральского максимума (WUH и SUH) с отрицательными аномалиями площади морского льда.

5. Количественные характеристики связи погодных режимов в Евро-атлантическом и Северо-Евразийском регионах с вероятностями сильных аномалий температуры, осадков и ветра над континентами Северного полушария, включая значимую связь летнего режима уральского максимума в СЕ с ростом вероятности жарких дней на европейской территории России в 2-3 раза и зимнего режима NAO- в ЕАТ с кратным ростом аномально-холодных дней на территории Европы.

Научная новизна исследования. В работе впервые произведено количественное сравнение характеристик погодных режимов в Евро-Атлантике, полученных несколькими различными методами кластерного анализа. Впервые в рамках режимного подхода получены типы атмосферной циркуляции отдельно для региона Северной Евразии в зимний и летний

периоды и представлены их пространственные и временные характеристики, включая многолетние тренды и особенности переходов. Оценена воспроизводимость пространственно-временных характеристик не только зимних, но и летних погодных режимов в Евро-Атлантике в современных климатических моделях, включая вероятности переходов. Получены оценки изменения повторяемости погодных режимов в Северной Евразии в зависимости от фаз Эль-Ниньо и аномалий площади морского льда в Северном полушарии с заблаговременностью в несколько месяцев. В работе впервые получено пространственное распределение вероятностей экстремальных аномалий температур, осадков и скорости ветра в зависимости от погодных режимов Северной Евразии и режимов циркуляции в европейском регионе с экстремальными осадками на Крымском полуострове.

Теоретическая и практическая значимость. В работе представлен анализ крупномасштабной атмосферной циркуляции над Евро-Атлантикой и Северной Евразией в рамках режимного подхода с применением кластерного анализа. Режимный подход позволяет по-новому взглянуть на климатические процессы в атмосфере и их долгосрочные изменения, представляя атмосферу как набор из различных чередующихся квазистационарных состояний, характеристики которых при этом испытывают долгосрочные изменения, что может рассматриваться как проявление современных изменений климата.

Проведенный анализ погодных режимов в атмосфере над обширными территориями, влияющими на погоду, включая её экстремальные проявления, на территории России, показывает значимость режимного подхода в исследовании механизмов формирования экстремальных погодных явлений. Анализ воспроизводимости характеристик погодных режимов Евро-Атлантике и Северной Евразии в климатических моделях может использоваться в качестве диагностики способности моделей воспроизводить крупномасштабную циркуляцию атмосферы. Оценки предсказуемости режимов ведущими модами климатической изменчивости могут использоваться для повышения точности предсказуемости погоды на субсезонных масштабах.

Степень достоверности. Достоверность полученных результатов обоснована использованием современного высокоточного климатического реанализа ERA5 и множества современных климатических моделей из проекта CMIP6. Количественные оценки, включая их степени неопределенности и статистические значимости, проверены статистическими методами, включая бутстрэппинг, тесты значимости трендов, включая тест Манн-Кендалла и

другие, генерации случайных подвыборок на основе реальной выборки данных и прочие статистические методы оценки значимости результатов.

Характеристики и структуры погодных режимов, полученных автором, сравнивались с результатами аналогичных известных исследований с использованием режимного подхода, с которыми получено качественное совпадение [19, 23, 24, 27], а количественные расхождения отдельных характеристик режимов основываются на обоснованных различиях в методике кластеризации режимов и временных интервалах, различающихся в отдельных исследованиях.

Апробация работы. Результаты диссертационного исследования представлялись на следующих отечественных и международных научных конференциях и семинарах:

«Собственное излучение, структура и динамика средней и верхней атмосферы» посвященная памяти А. И. Семенова и Н. Н. Шефова // Москва, 22-23 ноября 2021; «Турбулентность, динамика атмосферы и климата», посвященная памяти академика Александра Михайловича Обухова // Москва, 22–24 ноября 2022; XX научная школа «Нелинейные волны – 2022» // Нижний Новгород, ИПФ РАН, 2022; Международная научно-практическая конференция «Фундаментальные и прикладные аспекты геологии, геофизики и геоэкологии с использованием современных информационных технологий» // Майкоп, 2023; Вычислительно-информационные технологии для наук об окружающей среде CITES-2023 // Москва, 2023; Russian-China workshop: 15th Conference of Atmospheric Physics, Climate and Environment // Moscow, September 11, 2023; V Международная научная конференция «Наука будущего» // Орёл, 2023; WCRP Open Science Conference 2023 // Kigali, Rwanda, 2023; «Национальная модель Земной системы: теория, технологии и результаты» // Сириус, 20-22 февраля 2024; Семинар в научном фонде «Международный центр по окружающей среде и дистанционному зондированию имени Нансена» // Санкт-Петербург, 5 марта 2024; Школа «Климатическая система Земли: диагностика, моделирование и прогноз» научно-технологического университета «Сириус» // Сириус, 14-19 октября 2024; Доклады на семинарах и ученых советах ИФА РАН о методах кластерного анализа и погодных режимах // Москва ноябрь 2023, Москва 11.04.2024, Кисловодск 17.05.2024.

Личный вклад автора. Все научные результаты, представленные в диссертационной работе, включая сравнение методов кластерного анализа и последующего исследования полученных с помощью них характеристик

погодных режимов, получены автором лично или соавторстве с научным руководителем д.ф.-м.н. Семеновым В. А. Автору принадлежит ведущая роль в постановке задач диссертационного исследования, подготовке публикаций полученных результатов в научных журналах, а также их представлении на конференциях и семинарах. Текст диссертации полностью написан автором. Иллюстрации, кроме случаев, когда в подписи приведена ссылка, получены автором. Для обработки данных, расчета результатов и построения графиков и прочих иллюстраций, использованы программы, написанные автором лично для среды MATLAB, либо их комбинация с пакетом программ «Climate Data Toolbox» [33], находящихся в свободном доступе.

Основные публикации. По результатам диссертационной работы опубликовано 3 статьи в рецензируемых научных изданиях, рекомендованных ВАК и входящих в базы данных SCOPUS, Web of Science или Russian Science Citation Index (RSCI):

1. **Бабанов Б. А.**, Семенов В. А., Мохов И. И. Сравнение различных методов кластеризации для определения погодных режимов в Евро-Атлантическом секторе в зимний и летний сезоны // Известия Российской академии наук. Физика атмосферы и океана. – 2023. – Т. 59, № 6. – С. 686-706. – DOI 10.31857/S0002351523060020. – EDN OSUIJZ.
2. **Б. А. Бабанов**, Семенов В. А., Акперов М. Г., Мохов И. И., Keenlyside N. S. Повторяемость зимних режимов атмосферной циркуляции в Евро-Атлантическом регионе и связанные с ними экстремальные погодно-климатические аномалии в Северном полушарии // Оптика атмосферы и океана. – 2023. – Т. 36, № 4(411). – С. 304-312. – DOI 10.15372/AOO20230407. – EDN GJSUQM.
3. **Babanov B. A.**, Semenov V. A. Weather Regimes in Northern Eurasia: Statistics, Predictability and Associated Weather Anomalies // Atmosphere. – 2024. – Vol. 15, No. 11. – P. 1392. – DOI 10.3390/atmos15111392. – EDN RSXBRA.

Благодарности. Автор выражает глубокую благодарность своему научному руководителю, академику РАН, доктору физико-математических наук, Семенову Владимиру Анатольевичу. Автор также признателен старшим коллегам и соавторам публикаций за поддержку, помощь и советы при написании диссертации – старшему научному сотруднику ИФА РАН, кандидату физико-математических наук, Мирсеиду Акперову, главному научному сотруднику ИФА РАН и доктору физико-математических наук Елисееву Алексею Викторовичу, академику РАН и доктору физико-математических наук, главному научному сотруднику ИФА РАН, Мохову Игорю Ивановичу. Автор выражает благодарность всем коллегам из ИФА РАН, в особенности сотрудникам лабораторий теории климата и лаборатории предсказуемости климата.

Содержание и объём диссертации. Диссертация состоит из введения, четырех глав, заключения и списка литературы. Общий объём диссертации составляет 126 страниц, включая 27 рисунков, 18 таблиц. Список литературы содержит 106 наименований.

ГЛАВА 1. СРАВНЕНИЕ МЕТОДОВ КЛАСТЕРНОГО АНАЛИЗА ДЛЯ ИДЕНТИФИКАЦИИ ПОГОДНЫХ РЕЖИМОВ В ЕВРО-АТЛАНТИЧЕСКОМ СЕКТОРЕ

Процесс любой классификации состоит из двух основных этапов: определения типов и отнесения анализируемых данных к выделенным типам. При определении типов существует следующее существенное различие между подходами: 1) в одних подходах типы определяются до этапа присвоения к ним данных, их можно назвать априорными; 2) в других типы выделяются и последовательно меняются от начального предположения к окончательному результату в ходе самого процесса классификации.

При априорном определении типов атмосферной циркуляции можно руководствоваться либо экспертными знаниями (эмпирическим правилом), как в каталоге Гесса-Брезовски [34], или физическими и пространственными соображениями, такими как направление воздушного потока, его сила и степень завихренности, как в каталоге Лэмба [35]. Априорные классификации требуют выбора, который в значительной степени произволен, поэтому их иногда называют субъективными. Помимо классификаций Гесса-Брезовски и Лэмба, широко известны две субъективные классификации, разработанные в СССР в середине XX столетия: типизация Г. Я. Вангенгейма [8], развитая впоследствии А. А. Гирсом [36], и типизация Б. Л. Дзердзеевского [7, 37].

Классификации, в которых и типы, и присвоение данных к типам, определяются с помощью некоторой численной алгоритмической процедуры, обычно называются объективными. Стоит отметить, что слово «объективный» не подразумевает объективность всего процесса классификации, поскольку объективные процедуры предполагают принятие исследователем субъективных решений, которые могут существенно повлиять на результаты классификации.

В данной главе представлено теоретическое описание методов кластерного анализа, используемых в качестве современного способа объективных типизаций в задачах идентификации погодных режимов, приведены способы определения оптимального числа кластеров, то есть погодных режимов, и на примере региона Евро-Атлантики будет приведено сопоставление результатов кластеризации, полученных разными методами кластерного анализа.

1.1 Ранние объективные методы типизации атмосферных течений

Относительно простым подходом при объективной классификации является расчет и последующий анализ *матрицы взаимных расстояний* между полями циркуляции, представленных в виде суточных полей SLP или z500. Взаимные расстояния обычно измеряются либо с помощью пространственной корреляции между полями, либо с помощью квадрата Евклидова расстояния. «Квадрат» удобен тем, что при прочих равных на его расчет затрачивается на одну процедуру меньше – не нужно считать квадратный корень из суммы квадратов. Пусть поля циркуляции атмосферы представлены в виде среднесуточных полей $z500(lon, lat, time)$ или $SLP(lon, lat, time)$ на некоторой регулярной сетке $[lon \times lat]$ с одинаковым шагом в градусах. Перед расчетом взаимных расстояний поля z500 (SLP) нормируют по пространству для учета разницы площадей, приходящихся на ячейки на разных широтах:

$$z500_w(i, j, t) = z500(i, j, t) * \sqrt{w(i, j)}, \quad (1.1)$$

$$\text{где } w'(i, j) = A * \cos(lat(j)), \quad (1.2)$$

$$w(i, j) = \frac{w'(i, j)}{\sum_{i, j} w'(i, j)} \quad (1.3)$$

где A – любая постоянная. Для сохранения физического смысла, A можно подобрать так, чтобы сумма $z500_w(i, j, t)$ по пространству (по $i \in [lon]$, $j \in [lat]$) была равна средней величине $z500(i, j, t)$ по пространству. Такой подбор A выполняется с помощью формулы (1.3), но не является обязательным, по существу нормировка происходит на квадратный корень из косинуса широты. Квадратный корень в формуле (1.1) впоследствии «уходит», так как при расчете взаимных расстояний используются величины, пропорциональные квадрату z500. Вообще говоря, нормировки могут быть и более изощренными, в зависимости от целей исследования, например, концентрирующими сигнал z500 в области северных широт, как в [38]. Автор лишь хочет отметить, что схождение линий меридианов к полюсам в той или иной степени должно учитываться при анализе пространственных переменных. После нормировки взаимное расстояние между двумя полями можно рассчитать по формуле квадрата Евклидова расстояния $d^2(t_1, t_2)$ или пространственной корреляции $r(t_1, t_2)$:

$$d^2(t_1, t_2) = \sum_{i, j} (z500_w(i, j, t_1) - z500_w(i, j, t_2))^2 \quad (1.4)$$

$$r(t_1, t_2) = \frac{\sum_{i,j} ((z500_w(i, j, t_1) - \overline{z500_w(t_1)}) * (z500_w(i, j, t_2) - \overline{z500_w(t_2)}))}{\sqrt{\sum_{i,j} (z500_w(i, j, t_1) - \overline{z500_w(t_1)})^2 * \sum_{i,j} (z500_w(i, j, t_2) - \overline{z500_w(t_2)})^2}}, \quad (1.5)$$

$$\text{где } \overline{z500_w(t)} = \frac{\sum_{i,j} z500_w(i, j)}{\sum_{i,j} 1}, \quad (1.6)$$

то есть $\overline{z500_w(t)}$ – это среднее по пространству (по узлам *lon x lat* сетки) значение нормированной *z500* в исследуемой области в момент времени *t*.

Основная идея типизации на основе матрицы взаимных расстояний $TD(t_1, t_2) = r(t_1, t_2)$ или $TD(t_1, t_2) = d^2(t_1, t_2)$ в том, что среди значений этой матрицы найдётся такое поле *t*, которое имеет наибольшую среднюю пространственную корреляцию (или наименьшее среднее Евклидово расстояние) со всеми остальными полями. Такое *ключевое поле*, а также все подобные поля, имеющие пространственную корреляцию с ключевым полем выше (или Евклидово расстояние – ниже) некоторого выбранного исследователем *порога*, объявляются представителями первого типа циркуляции. Поля, отнесенные к первому типу, удаляются из матрицы взаимных расстояний, и процедура поиска ключевого поля и подобных ему полей, корреляция которых с ним выше заданного порога, повторяется заново. Алгоритм повторяется до тех пор, пока в уменьшенной матрице взаимных расстояний не останется величин попарных корреляций, превышающих установленный порог. Недостатком классификаций на основе анализа матрицы взаимных расстояний является их тенденция к получению одного непропорционально огромного кластера, сопровождаемого множеством мелких.

Одной из первых работ по объективной классификации типов атмосферных течений путем анализа матрицы взаимных расстояний является статья И. Лунда 1963 года [11], в которой 445 суточных полей SLP, полученных путем агрегации данных станционных наблюдений в области восточного побережья Северной Америки в зимние месяцы за 1949 – 1953 гг., классифицировались на основе взаимной пространственной корреляции. В работе использовался порог пространственной корреляции $r(t_1, t_2) \geq 0.70$, и с помощью процедуры, описанной в абзаце выше, выделено 10 типов течений. В результате типизации в первые 2 типа было отнесено 36% и 21% из исследуемых полей (162 и 95 полей, соответственно), а на последние 5 приходится не более чем по 2-3% из исследованного набора полей SLP, при этом 10.5% (47 полей) остались не типизированы. Следует ожидать, что при

увеличении порога корреляции r населенность первых типов снизится, однако сами типы будут менее различимы друг от друга, и увеличится доля неклассифицированных полей, а при уменьшении порога r первые несколько типов будут занимать слишком большую долю от общего числа полей, при этом оба варианта нежелательны.

По мере развития численных подходов к анализу данных, простой анализ матрицы взаимных расстояний для объективной типизации атмосферных течений уступил более продвинутым методам, таким как разложение данных на эмпирические ортогональные функции (ЭОФ) и анализ главных компонент (PCA, Principal Component Analysis) и современным методам кластерного анализа [39], включая: иерархическую кластеризацию Уорда, k-means, модели Гауссовой смеси и самоорганизующиеся карты Кохонена. Описание этих четырех методов будет представлено в подразделах главы 1 ниже.

В общем виде любой кластеризуемый набор однородных данных можно представить в виде N векторов размерности L :

$$\mathbf{X}_i = \{x_{i1}, x_{i2}, \dots, x_{il}, \dots, x_{iL}\}, \quad i = [1, \dots, N], \quad l = [1, \dots, L] \quad (1.7)$$

где под размерностью L понимается количество однородных нормированных обезразмеренных чисел, которыми описывается каждый из объектов входного набора данных. Квадрат Евклидова расстояния, наиболее часто используемый в качестве метрики расстояния между двумя объектами в естественно-научных задачах, будет записываться так:

$$d^2(\mathbf{X}_i, \mathbf{X}_j) = \|\mathbf{X}_i - \mathbf{X}_j\|^2 = \sum_{l=1}^L (x_{il} - x_{jl})^2, \quad (1.8)$$

Для суточных полей z500 или SLP в качестве индекса векторов $i = [1, \dots, N]$ выше использовался символ t , так как входные объекты можно индексировать их датой, а в качестве индекса размерности $l = [1, \dots, L]$ использовались символы i и j , указывающие на номер широт и долгот, на которых заданы поля, поэтому частный случай формулы (1.8) применительно к 2D-полям z500 представлен формулой (1.4), где в аргументах функции расстояния $d^2(t_1, t_2)$ для краткости вместо векторов, т.е. полей z500(lon, lat, time), указываются лишь их временные индексы.

Стоит отметить, что в большинстве кластеризующих компьютерных программ входной набор данных подается именно в виде набора из векторов,

то есть 2D-матрицы, каждая строка которой это одна из N кластеризуемых точек (векторов), представленных в виде формулы (1.7), а каждый столбец — это координаты каждой из N точек по одной из L размерностей. Перед передачей полей z500 или SLP кластеризуемым программам, все их значения вдоль широт и долгот сворачиваются в строку, а каждая строка соответствует определенной дате t из анализируемого временного интервала.

В дальнейшем в тексте данной работы, если контекстом не будет явно указано иное, любой отдельный объект из кластеризуемого набора данных, будет именоваться точкой, вектором или полем, эти три термина будут иметь эквивалентный смысл. Точки и другие многомерные объекты, так или иначе представимые в виде строки из чисел, как в формуле (1.7), будут выделяться жирным шрифтом.

1.2 Современные методы кластерного анализа

Подходы кластерного анализа можно разделить на иерархические и неиерархические. Ключевое отличие этих подходов в том, что в неиерархических кластеризация производится единожды с предзаданным числом кластеров, а в иерархических несколько раз подряд, в результате чего пользователь имеет несколько иерархических уровней, состоящих из вложенных друг в друга кластеров, среди которых он может выбрать необходимый ему для дальнейшего анализа уровень или анализировать всё иерархическое «дерево» (дендрограмму) целиком.

Иерархические подходы, в свою очередь, тоже подразделяются на 2 класса, агломеративные и дивизивные [40]: 1) в агломеративных подходах крупные кластеры создаются путём объединения меньших кластеров, при этом кластеризация начинается с того, что каждый объект в наборе данных объявляется отдельным кластером, и по мере работы алгоритма общее число кластеров уменьшается путём их объединения. Такой алгоритм продолжается либо вплоть до объединения набора данных в один единственный кластер, либо пока пользователь кластеризующей программы не укажет, на каком количестве кластеров следует остановиться (возможны и другие критерии остановки работы алгоритма). Агломеративные подходы по-другому называют кластеризацией «снизу-вверх» или «от листьев к корням»; 2) в случае дивизивных подходов, еще называемых «сверху-вниз» или «от корней к листьям», новые кластеры создаются путём их деления на меньшие

подкластеры. Процесс начинается с того, что все объекты из набора данных представляют собой один единственный кластер, делящийся согласно некоторым численным критериям. Дивизивный алгоритм может продолжаться до тех пор, пока количество кластеров не достигнет количества объектов из кластеризуемого набора данных, либо вплоть до выполнения другого критерия, выбранного пользователем.

Схематически различие двух классов иерархической кластеризации представлено на рисунке 1.1. В отличие от агломеративных методов, дивизивная иерархическая кластеризация является скорее подходом к кластерному анализу данных, чем собственным отдельным методом, так как при разделении кластеров на подкластеры могут использоваться различные алгоритмы, в том числе неиерархические, например, k-means или анализ главных компонент (РСА) и прочие. Для агломеративных кластеризаций, напротив, разработано множество собственных алгоритмов, различающихся способами связывания («linkage») малых кластеров в более крупные.



Рисунок 1.1. Схематическое представление работы агломеративных и дивизивных методов иерархических кластеризаций [41].

В последующих подразделах, где приведено описание алгоритмов кластеризации отдельных методов, использующихся для идентификации погодных режимов, кластеризация по методу Уорда (подраздел 1.2.1) относится к агломеративным иерархическим кластеризациям, а остальные три метода (подразделы 1.2.2 – 1.2.4): k-means, модель гауссовой смеси и самоорганизующиеся карты Кохонена – к неиерархическим.

1.2.1 Иерархическая кластеризация Уорда

Различные методы агломеративной кластеризации отличаются критериями для отбора и связывания пар кластеров в более крупные, они называются критериями связывания (linkage criteria) и по сути представляют собой способ расчета *расстояний между кластерами*, которое не следует путать с *расстоянием между точками* из набора данных (обычно рассчитываемого по формуле (1.8)). Под точкой из набора данных может пониматься любой объект, заданный в многомерном пространстве, в том числе суточное поле z500, размерность которого определяется числом узлов сетки. Любой метод агломеративной кластеризации можно представить в виде следующего итеративного алгоритма:

1. Представление набора данных в виде ограниченного набора кластеров, каждый из которых содержит часть анализируемых точек из набора данных. Расчет матрицы попарных расстояний между кластерами согласно выбранному критерию связывания.
2. Нахождение пары кластеров «А» и «Б» с *минимальным* взаимным расстоянием, объединение этой пары в новый кластер «АБ» и удаление кластеров «А» и «Б» как отдельных из анализируемого набора кластеров.
3. Повторение пунктов 1-2 до объединения набора кластеров в один единственный кластер.

В зависимости от природы набора данных могут использоваться различные критерии отбора пар объединяемых кластеров, такие как single linkage, complete linkage, average linkage, centroids linkage, и наконец, Ward's linkage:

1. Single linkage – расстояние между кластерами считается как ближайшее попарное расстояние между двумя точками из двух разных кластеров, первая точка в одном кластере, вторая – в другом (аналогично для остальных методов). Расстояния между точками, в свою очередь, считаются на основе выбранной метрики расстояния, то есть, как правило, Евклидовой (формула (1.8)).
2. Complete linkage – расстояние между кластерами считается как самое дальнее из попарных расстояний между точками из двух кластеров.

3. Average linkage – расстояние между кластерами считается как среднее попарное расстояние между всеми парами точек из двух кластеров.

4. Centroids linkage – расстояние между кластерами считается как расстояние между центроидами этих кластеров.

5. Ward's linkage [42] – расстояние между кластерами считается как разница между внутрикластерной дисперсией после потенциального объединения и суммой внутрикластерных дисперсий до объединения:

$$\Delta = \sum_{X_i \in AB} (X_i - \overline{ab})^2 - \left(\sum_{X_i \in A} (X_i - \overline{a})^2 + \sum_{X_i \in B} (X_i - \overline{b})^2 \right), \quad (1.9)$$

где X_i – точка из набора данных, например, нормированное суточное поле $z500(lon, lat, i)_w$, а «А», «В» и «АВ» это пара кластеров до объединения и объединенный кластер, «а», «b», и «ab» соответствующие этим кластерам центроиды. Стоит отметить, что $\Delta > 0$ для любой пары непустых кластеров.

Преимущество критерия связывания Уорда (Ward's linkage) по сравнению с другими методами в том, что в результате итеративной работы алгоритма вплоть до его остановки минимизируются внутрикластерные дисперсии, благодаря чему итоговые оставшиеся K кластеров получаются более выраженными по сравнению с применением других критериев linkage. Свойство минимизации суммы внутрикластерной дисперсии V делает иерархическую кластеризацию Уорда похожей на метод кластерного анализа k-means. Итоговую V при остановке алгоритма на выбранном количестве K кластеров можно рассчитать по формуле:

$$V(K) = \sum_{k=1}^K \sum_{i \in k} d^2(X_i, C_k), \quad (1.10)$$

$$C_k = \frac{\sum_{i \in k} X_i}{\sum_{i \in k} 1}, \quad (1.11)$$

где C_k – это так называемый центроид k -го из K кластеров. Если под X_i подразумевать взвешенные поля $z500_w(lon, lat, t)$, то $C_k = C(lon, lat, k)$ будет средним полем из всех суточных полей $t \in k$, попавших на некотором этапе алгоритма кластеризации в один и тот же k -ый кластер.

Стоит отметить, что так как в качестве метрики используется квадрат Евклидова расстояния d^2 (формула 1.8), то внутрикластерная дисперсия и сумма внутрикластерных расстояний представляют собой одну и ту же

величину с разницей до постоянной. Здесь и далее по тексту диссертации термины «сумма внутрикластерных дисперсий» и «сумма внутрикластерных расстояний» будут использоваться эквивалентно. Ими будет называться величина, рассчитанная по формуле 1.10, без деления на населенности кластера $1 / (N_k - 1)$, то есть это не дисперсии в строго математическом смысле.

1.2.2 k-means

В последние десятилетия метод кластерного анализа k-means (КМ), относящийся к неиерархическим, оказался наиболее часто используемым в задачах классификации атмосферных течений [9], он активно использовался в статьях с применением режимного подхода с конца XX столетия [43], в середине «нулевых» [19, 20], «десятих» [23] и вплоть до настоящего времени [24, 25, 27]. Преимущество метода заключается в его относительной простоте и минимуме входных параметров, требуемых от исследователя для начала работы алгоритма. Если пользователю иерархической кластеризации следует провести некоторый анализ для выбора предпочтительного критерия связывания linkage, то для начала k-means нужно лишь задать требуемое число кластеров K и способ случайного задания начальных положений центроидов.

Алгоритм кластеризации k-means состоит из следующих шагов:

1. Некоторым способом с привлечением рандомизации на основе кластеризуемой выборки данных X_i задаются начальные предполагаемые положения K ($k = 1:K$) центроидов C_k будущих кластеров.
2. Рассчитываются расстояния между каждой точкой из набора данных X_i и каждым из заданных K центроидов C_k . Каждая точка из набора данных присваивается тому кластеру k , расстояние до центроида которого $d^2(X_i, C_k)$ у неё минимально (формула (1.8)).
3. Координаты каждого центроида C_k пересчитывается как «центр масс» точек (формула (1.11)), присвоенных кластеру k на предыдущем шаге.
4. Пункты 2 и 3 последовательно повторяются до тех пор, пока координаты всех центроидов не перестанут меняться или, что эквивалентно, все точки не перестанут менять кластеры.

В процессе работы алгоритма k-means с каждым повторением шагов 2 и 3 последовательно убывает сумма внутрикластерных дисперсий V (формула (1.10)) и сходится к некоторому итоговому значению. Алгоритм k-means

называют детерминированным, так как его итоговый результат зависит исключительно от задания начальных положений кластеров, поэтому их следует задавать, ориентируясь на распределение входного набора данных, однако, из-за того что разные начальные координаты центроидов приводят к разным итоговым V , в процесс инициализации координат C вносят элемент случайности с целью отбора такого решения k-means, чтобы итоговое V было минимальным, что будет соответствовать максимальной близости точек внутри кластеров и их удалению между разными кластерами.

Теоретически, для любого входного набора данных с фиксированными параметрами существует такая минимальная величина V , которую называют глобальным оптимумом, к которой устремляется любое V , фактически получаемое по завершению алгоритма k-means. Итоговое V , полученное одним прогоном k-means, зависящее от задания начальных положений, называют локальным оптимумом. Локальный оптимум V меньше (либо равен) глобального оптимума. Для получения решений, максимально приближенных к глобальному оптимуму, существуют оптимизации оригинального алгоритма k-means: 1) способ оптимальной рандомизированной инициализации центроидов, названный «kmeans++» [44], и 2) способ внесения рандомизации в процесс присвоения точек кластерам (шаг 2 оригинального алгоритма) под названием «simulated annealing» или метод «имитации отжига» [45].

Инициализация «k-means++» позволяет рандомизированно задать центроиды таким образом, чтобы они были распределены вдоль всего пространства набора данных, но при этом удалены друг от друга. Алгоритм «k-means++» состоит из следующих шагов:

1. Положение первого центроида C_1 выбирается случайно с координатами одной из точек X_i кластеризуемого набора данных.
2. Вторым центроид C_2 выбирается по координатам одной из оставшихся точек с вероятностью $\sim d^2(X_i, C_1)$ (формула (1.8)). Вероятности выбора второго центроида нормируются так, чтобы в сумме по оставшимся точкам вероятность равнялась 1.
3. Если $K > 2$, то каждый последующий k -ый центроид ($k \leq K$) инициализируются случайно по координатам точек с вероятностью, пропорциональной $\min_{k'=1:k-1} d^2(X_i, C_{k'})$, то есть пропорционально расстоянию от точки до ближайшего из уже заданных центроидов.

В работе [44], где предложен этот алгоритм, показано, что такая инициализация центроидов позволяет алгоритму сходиться с меньшими V за меньшее число итераций по сравнению с чисто случайной инициализацией, при которой не используется взвешивание вероятностей по $d^2(\mathbf{X}_i, \mathbf{C}_k)$.

Для устранения детерминированности k-means, то есть зависимости итогового решения исключительно от инициализации центроидов, существует оптимизация «simulated annealing» [45, 46]. Во время работы алгоритма на шаге присвоения точек кластерам точки могут быть отнесены не к кластеру ближайшего центроида, а к любому другому с вероятностью P , обратно пропорциональной удаленности от точки до центроида этого кластера:

$$P_{k,iter} = \exp\left(\frac{d_{old}^2 - d_{k,new}^2}{T_{iter}}\right), \quad (1.12)$$

где d_{old}^2 – расстояние от точки до центроида, к которому она была приписана на предыдущей итерации алгоритма, $d_{k,new}^2$ – расстояние от точки до k -го центроида, к которому она потенциально будет приписана по итогу текущей итерации алгоритма. T_{iter} – параметр, убывающий с каждой последующей итерацией – $T_{iter+1} = CT_{iter}$, где C – так называемый коэффициент охлаждения, очень близкий к единице и подбираемый вручную ($0.95 \lesssim C < 1$).

Коэффициент охлаждения C подбирается таким образом, чтобы в начале работы алгоритма точки часто приписывались к «неправильным», т. е. не ближайшим, кластерам с целью обойти решения с локальными оптимумами с высокими значениями V , но с ростом числа итераций сходились к итоговому решению, когда вероятности $P_{k,iter}$ быть присвоенной к не ближайшему кластеру станут настолько малы, что перестанут реализовываться, а алгоритм остановится по тому же критерию, что и оригинальный k-means. Стоит отметить, что при $d_{old}^2 = d_{k,new}^2$ вероятность P максимальна, то есть вероятность для точки остаться в ближайшем кластере всегда выше, чем в более удаленном. По формуле (1.12) такая вероятность равна 1, но все вероятности нормируются так ($P_{k,iter} = P_{k,iter} / \sum_k P_{k,iter}$), чтобы для каждой точки сумма вероятностей быть приписанной к любому из $k=1:K$ кластеров равнялась 1.

В работе по кластеризации суточных полей давления в Евро-Атлантике показано [46], что использование оптимизации «simulated annealing» повышает воспроизводимость и лучше минимизирует V итоговых решений,

устремляя их к глобальному оптимуму, по сравнению с оригинальным алгоритмом k-means.

1.2.3 Модель гауссовой смеси

При кластерном анализе размерность пространства определяется минимальным общим набором чисел, которыми определяется каждая точка из набора данных. Так, для набора «точек», определяющих крупномасштабную циркуляцию атмосферы, например, суточных полей $z500$ или SLP, их размерность будет равна общему числу узлов сетки, на которых эти поля заданы, то есть $L = L_{lon} * L_{lat}$, где L_{lon} и L_{lat} – это число шагов в сетке по долготе и широте соответственно. Об учете разности площадей, соответствующих узлам сетки на разных широтах, написано в разделе 1.1. В общем случае, каждая из размерностей $l = 1:L$, в которых определены «точки» из набора данных, может иметь разную физическую (или нефизическую) природу, например, разную *физическую* размерность – «м», «м²/с²», «Па» и т.п. В случае разнородности данных, перед кластеризацией их нормируют на стандартное отклонение для перевода всех $l = 1:L$ в безразмерные. При кластеризации погодных режимов, как правило, используется одна и та же переменная постоянной природы (высота геопотенциала или SLP), поэтому такая нормировка чаще опускается.

Одним из подходов при кластеризации многомерных данных является исследование их плотности распределения. При таком подходе в пространстве кластеризуемого набора данных ищутся локальные максимумы плотности распределения, представляющие из себя такие точки, которые имеют максимальное число соседей (других точек) в некотором радиусе R многомерной сферы, подобранном исследователем эмпирически, размерность сферы N при этом равна размерности исследуемого набора данных. Такой метод использовался в том числе при выделении погодных режимов в Северном полушарии в работе Molteni et al., 1990 года [47]. Развитием метода поиска локальных максимумов функций плотности распределения можно считать метод кластерного анализа под названием «DBSCAN» (Density-based spatial clustering of applications with noise) [48], особенность и преимущества которого состоят в способности выделять в пространстве «продолговатые» (вдоль какой-нибудь из множества размерностей) кластеры и отделять шум из набора данных, а также малом числе входных параметров перед кластеризацией, всего 2 параметра – радиус R и минимальное число соседей

для формирования кластера. К сожалению, из-за равномерности распределения аномалий z500 и SLP в их фазовом пространстве, методы на основе плотности склонны к формированию одного перенаселенного кластера, сопровождаемого множеством малых (так же, как при кластеризации на основе матриц взаимных расстояний). Например, в работе Molteni et al. [47], из 576 полей лишь 359 были кластеризованы в погодные режимы, а в первый из шести погодных режимов попало 261 поле, то есть 73% от кластеризованных и 45% от всех полей оказалось в одном кластере.

Частым предположением о природе исследуемого набора данных служит гипотеза, что данные распределены нормально. В случае кластеризации, когда исследователь полагает, что набор данных состоит из нескольких хорошо различимых групп, можно также предположить, что объекты внутри каждой из этих групп также распределены нормально [49]. Кластеризация гауссовой смесью (GM), в отличие от k-means и иерархической кластеризации Уорда (HW), относится к типу «мягких кластеризаций» (soft clustering) [50], когда по результатам работы алгоритма каждый объект приписывается не к одному определенному кластеру, а к каждому из них с некоторой вероятностью. У пользователя алгоритмов soft clustering, в отличие от hard clustering (k-means и HW), появляется свобода выбора, что вообще считать кластером – совокупность объектов, попавших в кластер с наибольшей вероятностью, либо лишь теоретические вероятностные распределение и их параметры. Например, исследователь по результатам кластеризации может ограничиться тем, чтобы присвоить каждый объект из набора данных исключительно тому кластеру, к которому он принадлежит с наибольшей вероятностью, превратив таким образом «мягкую» кластеризацию в «твердую».

При кластеризации методом GM в пространстве набора данных инициализируется K многомерных нормальных распределений, то есть кластеров, для каждого из которых каким-либо способом задаются (угадываются) параметры – среднее и матрица ковариаций [51]. Среднее при этом можно считать аналогом центроида кластера. Также для каждого распределения задаются веса, которые можно трактовать как вероятностный аналог доли точек из всей выборки, принадлежащих данному кластеру. Далее, в процессе так называемого EM-алгоритма (Expectation-Maximization algorithm) [52], основанного на теореме Байеса, и который, вообще говоря, можно использовать и для смеси других (негауссовых) распределений, средние, матрицы ковариаций и веса кластеров итеративно пересчитываются. EM-алгоритм состоит из следующих шагов:

1. Инициализируются средние μ_k , матрицы ковариаций Σ_k и веса кластеров w_k . Аналогично k-means, значения инициализируются рандомизированно, чтобы можно было прогнать алгоритм несколько раз и сравнить результаты. Для инициализации μ_k можно использовать, например, способ «k-means++», веса задаются равномерными, а матрицы Σ_k диагональными, элементы которых равны СКО точек X_i из набора данных.

2. Для каждой точки $X_i, i = (1 \dots N)$ из набора данных (N – общее число точек набора данных) пересчитывается вероятность её принадлежности к одному из кластеров (многомерных нормальных распределений):

$$P_{X_i}(B_k) = \frac{P(B_k) * P_{B_k}(X_i)}{P(X_i)} = \frac{w_k * \rho_{B_k}(X_i)}{\sum_{k=1}^K w_k * \rho_{B_k}(X_i)}, \quad (1.13)$$

где B_k – одно из K многомерных нормальных распределений, ρ_{B_k} – функция плотности вероятности этого распределения, для многомерного нормального распределения определяемая по формуле:

$$\rho_{B_k}(X_i) = \frac{1}{(2\pi)^{\frac{L}{2}} * |\Sigma_k|^{\frac{1}{2}}} * \exp\left(\frac{-1}{2} * (X_i - \mu_k)^T \Sigma_k^{-1} (X_i - \mu_k)\right), \quad (1.14)$$

где L – размерность данных, μ_k – среднее, Σ_k – матрица ковариаций, $|\Sigma_k|$ и Σ_k^{-1} – определитель и обратная матрица матрицы ковариаций.

3. По вероятностям принадлежности точек кластерам $P_{X_i}(B_k)$ пересчитываются параметры распределений w_k, μ_k и Σ_k :

$$w_k = \frac{1}{N} \sum_{i=1}^N P_{X_i}(B_k); \quad (1.15)$$

$$\mu_k = \frac{\sum_{i=1}^N (P_{X_i}(B_k) * X_i)}{\sum_{i=1}^N P_{X_i}(B_k)}; \quad (1.16)$$

$$\Sigma_k = \frac{\sum_{i=1}^N (P_{X_i}(B_k) * (X_i - \mu_k)(X_i - \mu_k)^T)}{\sum_{i=1}^N P_{X_i}(B_k)} \quad (1.17)$$

4. Шаги 2 и 3 ЕМ-алгоритма продолжаются до тех пор, пока алгоритм не сойдется, то есть параметры распределений w_k, μ_k и Σ_k или функция правдоподобия – сумма логарифмов вероятностей $\sum_i \log P_{X_i}(B_k)$ из формулы (10.13) – не перестанут меняться выше некоторой малой ошибки.

Среди преимуществ метода GM можно отметить способность создания кластеров «эллиптической» формы, в отличие от методов k-means и HW, в которых из-за минимизации внутрикластерных дисперсий, при расчете которых дисперсии каждой из размерностей складываются в одно число, кластеры получаются преимущественно «круглые», то есть кластеризация в них происходит по всем переменным независимо от параметров их одномерных распределений.

Для GM аналогом C_k из алгоритмов HW и k-means служит μ_k , однако процессе EM-алгоритма происходит максимизация так называемой функции правдоподобия («log likelihood») – логарифмической суммы вероятностей принадлежности точек многомерным нормальным распределениям с заданными параметрами, а не минимизация V (формула (1.10)), то есть подбираются такие параметры K многомерных нормальных распределений, которые с наибольшей вероятностью описывают набор входных данных, что позволяет получать кластеры, имеющие вытянутую («эллиптическую») форму вдоль тех размерностей пространства набора данных, в которых выше дисперсия. При режимном подходе к анализу атмосферной циркуляции метод кластеризации GM использовался, например, в работах [53, 54], где с помощью него выделялись погодные режимы во внетропических зонах Северного полушария.

1.2.4 Самоорганизующиеся карты Кохонена

Самоорганизующиеся карты Кохонена (Self-organizing maps, SOM) [55] представляют собой метод кластерного анализа, основанный на нейронных сетях. Данный метод позволяет получать кластеры из многомерного набора данных в таком виде, что их удобно визуализировать на двумерной карте, кластеры в которой отсортированы друг относительно друга (отсюда «самоорганизующиеся») на основе выбранной метрики расстояния (обычно, Евклидовой).

Так как SOM входит в раздел методов машинного обучения на основе нейронных сетей, для него имеется своя сложившаяся терминология – кластеры называются «нейронами». Перед запуском алгоритма кластеризации нейроны задаются в своём собственном двумерном пространстве, независимом от набора данных, то есть на нейронной сети или карте. На нейронной сети определяются расстояния между нейронами или, иначе

говоря, мера соседства нейронов, которая используется в процессе кластеризации для их взаимного «обучения». Центроиды, то есть координаты центров будущих кластеров, называются «векторами веса» или просто «весами» нейронов, итерации алгоритма называются «эпохами».

Нейроны представляют из себя объекты, состоящие из двух векторов, заданных в разных пространствах – вектора веса, заданного в пространстве набора данных, и координат, заданных в отдельном (как правило, двумерном) пространстве, определяющих положение нейрона на карте нейронной сети относительно других нейронов. В процессе работы алгоритма векторы веса нейронов «натягиваются» на «облако» набора входных данных, при этом векторы веса нейронов, расположенных рядом друг с другом на карте нейронной сети, «притягиваются», то есть обновляют значения векторов веса относительно входных данных, к соседним областям в пространстве набора данных.

Алгоритм SOM состоит из следующих шагов:

1. Инициализация нейронов и их весов на двумерной нейронной сети размера $K = M \times N$, где M и N – дискретные размеры нейронной сети по горизонтали и вертикали. Нейроны могут взаимно располагаться в виде квадратной, шестиугольной или случайной сетки, от чего будет зависеть расстояние между соседствующими нейронами. Веса нейронов задаются (инициализируются) некоторым случайным образом в диапазоне значений набора входных данных, то есть в диапазоне координат, которые принимают точки в пространстве входных данных.

2. Из набора данных выбирается случайная точка. Среди кластеров (нейронов) ищется так называемый «нейрон-победитель», вес которого на основе выбранной метрики (Евклидовой) будет ближе всего к данной точке:

$$\forall k \in [1, \dots, K] \exists c \in [1, \dots, K]: d(\mathbf{X}_i, \mathbf{W}_c(t)) \leq d(\mathbf{X}_i, \mathbf{W}_k(t)), \quad (1.18)$$

где $\mathbf{X}_i$ – случайно выбранная точка из набора данных, $\mathbf{W}_c(t)$ – вес нейрона-победителя на эпохе t алгоритма, $\mathbf{W}_k(t)$ – веса остальных нейронов.

3. Веса нейронов обновляются на основе расстояния от их весов до точки $\mathbf{X}_i$ и функции обучения, зависящей от эпохи и взаимного расстояния нейронов:

$$\mathbf{W}_k(t+1) = \mathbf{W}_k(t) + h_{ck}(t) * (\mathbf{X}_i - \mathbf{W}_k(t)); \quad (1.19)$$

$$h_{ck}(t) = \alpha(t) * \exp\left(\frac{-\|r_c - r_k\|^2}{2\sigma^2(t)}\right), \quad (1.20)$$

где $W_k(t+1)$ – обновленный вес k -го нейрона; $(X_i - W_k(t))$ – вектор, направленный от веса нейрона в сторону выбранной точки X_i ; $h_{ck}(t)$ – функция обучения; $\|r_c - r_k\|$ – расстояние между нейронами в пространстве нейронной сети, r – координаты нейронов на двумерной сетке $K = M \times N$; $\alpha(t)$ и $\sigma(t)$ – обучающие коэффициенты, монотонно убывающие с номером эпохи таким образом, чтобы в начале работы алгоритма веса нейронов быстро обновлялись, а веса соседствующих нейронов притягивались к соседним областям пространства входных данных, но по мере работы алгоритма (с ростом номера эпохи t) взаимное «обучение» затухало (под обучением имеются в виду обновления весов нейронов согласно входным данным), и алгоритм сходился.

4. Пункты 2 и 3 алгоритма продолжаются до достижения некоторого критерия – либо пока не будет достигнут заранее выбранный номер эпохи $t_{\max}$, либо пока «ошибка карты» не перестанет существенно убывать: $\Delta V = V(t+1) - V(t) \leq \varepsilon$; V – внутрикластерная дисперсия, для которой в качестве C_k используются веса нейронов $W_k(t)$.

Примеры использования нейронных сетей для выделения режимов крупномасштабной атмосферной циркуляции можно найти в работах [56, 57]. В работе Polo et al. 2011 года [56] делается вывод, что КМ и SOM показывают похожие результаты, что также является одним из выводов представленной работы.

Одно из основных преимуществ SOM состоит в автоматической сортировке кластеров друг относительно друга, то есть близкие по Евклидовому расстоянию кластеры будут располагаться рядом на нейронной карте. Например, на двумерной нейронной карте $M \times N = 5 \times 5$ (итого, $K=25$ нейронов, то есть кластеров), кластеры с координатами $\{1,1\}$, $\{1,2\}$ и $\{2,1\}$ будут похожи друг на друга и иметь схожие характеристики, а кластеры с координатами $\{1,1\}$ и $\{5,5\}$ будут максимально различаться (по метрике Евклидова расстояния). Для методов КМ и НВ такая взаимная сортировка кластеров потребовала бы отдельной процедуры, то есть при кластеризации с таким же $K=25$ кластеры с номерами 1 и 2 по результатам КМ и НВ могли бы располагаться где угодно (в пространстве набора данных) относительно друг друга. При использовании SOM обычно выбирают большие K , т. к. для небольших K , например < 10 , взаимная сортировка, как правило, не требуется.

Общим свойством вышеописанных методов кластеризации: НВ, КМ, ГМ, SOM является итеративность их алгоритмов. При этом в то время как НВ не требует инициализации и всегда производит единственный результат, результаты алгоритмов КМ, ГМ и SOM зависят от инициализации (в случае

simulated annealing для k-means еще и случайных переназначений в процессе), поэтому при использовании этих алгоритмов рекомендуется множественный прогон для получения наилучшего воспроизводимого результата, обладающего минимальной внутрикластерной дисперсией (для КМ и SOM) или наибольшим логарифмической функцией правдоподобия полученных параметров многомерных распределений (для GM).

1.3 Определение числа кластеров

Вышеперечисленные методы, кроме HW, перед началом работы требуют от исследователя заранее выбрать количество кластеров K , и производят кластеризацию согласно выбранному числу. Метод HW в результате создания системы вложенных кластеров позволяет выбрать пользователю решение с любым заданным числом кластеров, однако так же не дает ответа, какое K является предпочтительным. При прочих равных, исследователь заинтересован в выборе меньшего K , потому что такой выбор удовлетворяет формулировке кластеризации как «разбиению большого набора данных на малое число групп», но под «прочими равными» подразумеваются такие кластеризации, которые описывали бы кластеризуемый набор данных наилучшим образом, то есть имели наименьшую внутрикластерную дисперсию V .

Для определения оптимального количества кластеров существует несколько достаточно простых универсальных методов, например, «метод локтя» или «силуэта» [58], а также более сложных методов, разработанных специально в рамках задач идентификации погодных режимов, например, индексы классифицируемости и воспроизводимости, представленные в работе [43]. Применительно к задачам выделения погодных режимов через кластеризацию гладких суточных полей циркуляции атмосферы, простые методы, в которых оптимальное K определяется максимумом или точкой перегиба функции некоторого параметра от K , как правило, не дают явного и единственного K [21], однако, при хорошей «кластеризуемости» набора данных, этих методов будет достаточно.

1) Метод локтя

Метод «локтя» («Elbow method») является самым простым. Для его применения данные кластеризуются для каждого потенциального K в некотором интервале от 2 до $K_{\max}$, и для итоговых решений считаются

внутрикластерные дисперсии V . Строится график функции $V(K)$ (рассчитанной по формуле (1.10)), для которой ищется «локоть» - точка перегиба, такое K , после которого $V(K)$ перестаёт заметно убывать, то есть последующее добавление новых кластеров не приводит к явному улучшению качества кластеризации. Стоит отметить, что $V(K)$ это, в любом случае, монотонно убывающая функция, которая стремится к 0 при K стремящемся к N . Метод локтя работает тем лучше, чем более кучно распределены данные, что проиллюстрировано на рисунке 1.2 по случайно сгенерированным данным. Видно, что, при повышении СКО кластеров с 1 до 2, в исследуемой выборке точка перегиба в графике $V(K)$ пропадает, кластеры перестают быть различимы как визуально, так и в виде точки перегиба графика $V(K)$.

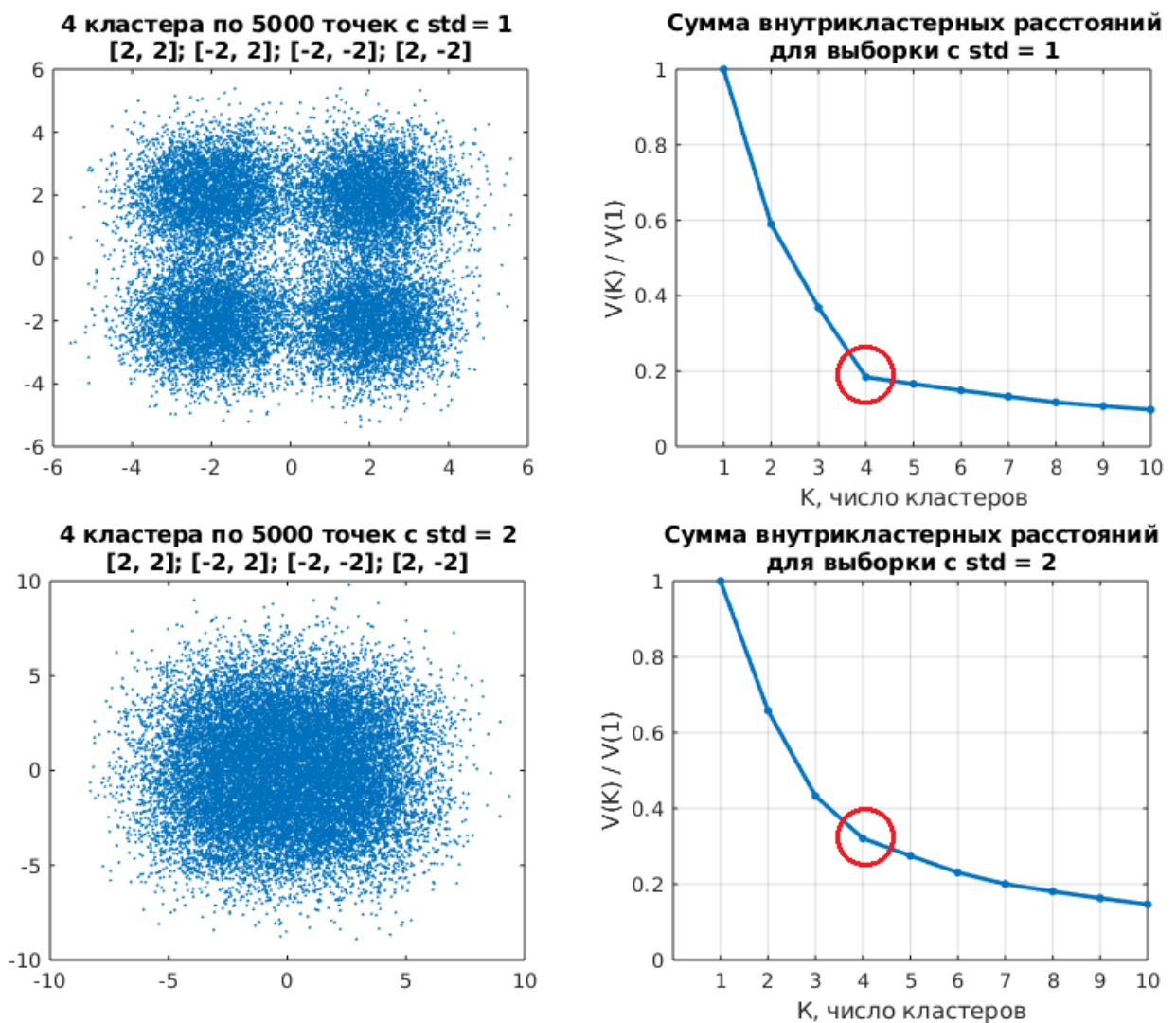


Рисунок 1.2. Работа метода локтя на примере выборок из четырех кластеров, составленных из двумерных нормальных распределений с центрами в [2, 2]; [-2, 2]; [-2, 2]; [2, -2] по 5000 точек в каждом. Сверху СКО(х)=СКО(у)=1, снизу СКО(х)=СКО(у)=2. Справа графики $V(K)$, нормированные на $V(1)$, иллюстрирующие, что при росте СКО точка перегиба исчезает.

2) Метод силуэта

При методе силуэта исследуется не «кластеризуемость» всей выборки целиком с помощью суммы внутрикластерных дисперсий $V(K)$, а считается так называемый коэффициент силуэта $S(i)$, имеющий разное значение для каждой из точек выборки. Коэффициент силуэта считается по формуле:

$$S(i) = \frac{b(i) - a(i)}{\max\{a(i), b(i)\}}; \quad (1.21)$$

$$a(i) = \frac{1}{N_k - 1} \sum_{j \in k, j \neq i} d^2(X_i, X_j) \quad (1.22)$$

$$b(i) = \min_{k' \neq k} \left\{ \frac{1}{N_{k'}} \sum_{j \in k'} d^2(X_i, X_j) \right\} \quad (1.23)$$

где N_k – число точек в кластере k ; $a(i)$ – среднее расстояние от точки i до остальных точек из того же кластера; $b(i)$ – среднее расстояние от точки i из кластера k до точек из ближайшего для неё соседнего кластера k' , то есть вначале для каждой точки $i \in k$ считается $b(i)$ по всем $k' \neq k$, а затем уже выбирается такое k' , которое для данной точки будет давать минимальное $b(i)$; $S(i)$ – «силуэт» точки i , меняющаяся в пределах от -1 до 1, причем чем ближе она к 1 (желательное значение), тем ближе точка к центроиду своего кластера и дальше от точек из других кластеров. Посчитав среднее $S(i)$ по i для каждой итоговой кластеризации для потенциальных K , строится график функции $f(K) = \langle S_K(i) \rangle$. В отличие от метода локтя, при котором ищется точка перегиба функции $V(K)$, в данном случае на графике ищется максимум $\langle S_K(i) \rangle$ от K .

На рисунке 1.3 представлен график $\langle S_K(i) \rangle$ для тех же сгенерированных выборок, что использовались на рисунке 1.2 для иллюстрации метода локтя. Видно, что при большой дисперсии данных (СКО = 2) изначально заложенные 4 кластера с центрами в точках $\{[2, 2]; [-2, 2]; [-2, -2]; [2, -2]\}$ перестают обнаруживаться с помощью метода силуэта, а максимум указывает на число кластеров $K = 3$. На примере двух простых методов показано, что для выборок с высокими внутрикластерными СКО, оптимальное число кластеров перестает обнаруживаться с помощью простых методов, использующих лишь фактически данную выборку данных и одно решение кластеризующей программы для каждого K . Ниже будут описаны методы, подразумевающие использование случайно сгенерированных рядов на основе оригинальной выборки данных и её половинчатых подвыборок для сопоставления

результатов кластеризации оригинальных данных со случайно сгенерированными.

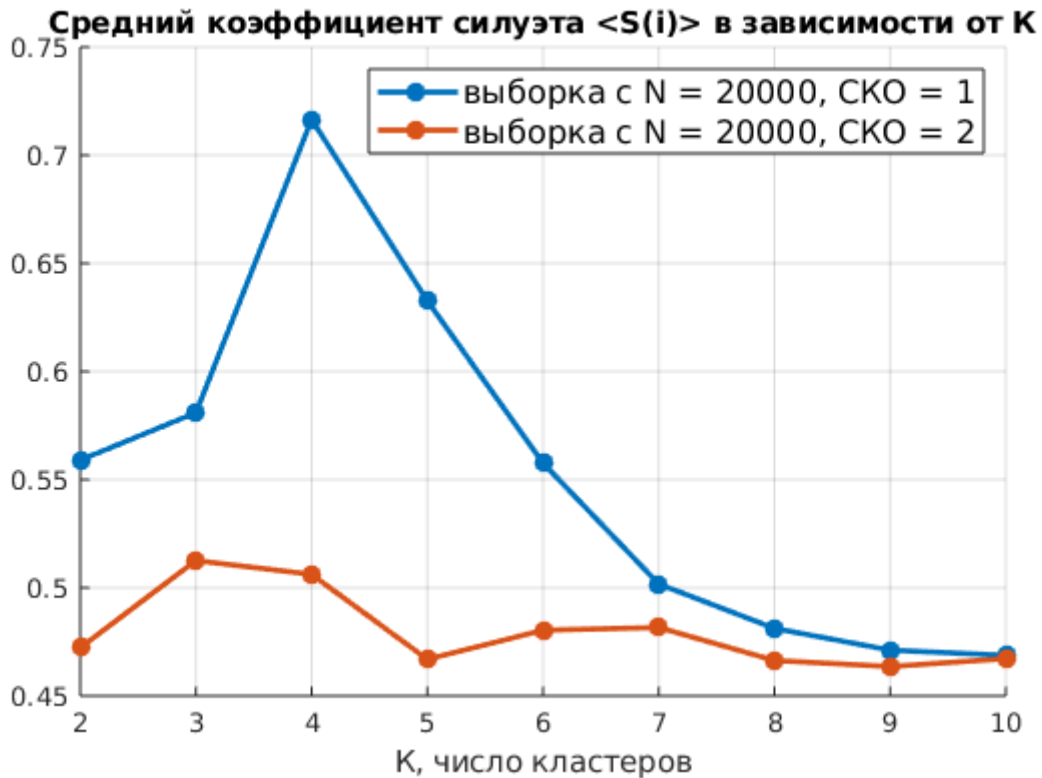


Рисунок 1.3. Значения среднего коэффициента силуэта для сгенерированной выборки из четырех кластеров с низким и высоким СКО (использованы те же две выборки, что представлены на рис. 1.2). Для выборки с низким внутрикластерным СКО явно выделяется 4 кластера, как оптимальные, а для выборки с высоким внутрикластерным СКО максимум $\langle S_K(i) \rangle$ приходится на $K = 3$, а функция слабо меняется с изменением K .

3) Критерий отношения дисперсий

Еще одним методом, в котором ищется максимум функции в зависимости от K , является критерий отношения дисперсий (Variance Ratio Criterion, VR) или индекс Калински-Харабаша [59], измеряющий отношение межкластерной дисперсии к внутрикластерной, нормированные на число их степеней свободы:

$$VR = \frac{BCSS/(K - 1)}{WCSS/(N - K)}, \quad (1.24)$$

$$BCSS = \sum_{k=1}^K N_k * d^2(\mathbf{C}_k, \mathbf{C}), \quad (1.25)$$

$$WCSS = \sum_{k=1}^K \sum_{i \in k} d^2(X_i, C_k) = V(K), \quad (1.26)$$

BCSS (Between-cluster sum of squares) это межкластерная дисперсия, в которой N_k – число точек в k -ом (из K) кластере, C без индекса – это единственный центроид при $K=1$, то есть среднее по всем X_i . *WCSS* (Within-cluster sum of squares) это внутрикластерная дисперсия $V(K)$ (формула 1.10) под другим обозначением. При оптимальном числе кластеров K , кластеризованные данные будут иметь высокую межкластерную дисперсию (*BCSS*) и низкую внутрикластерную дисперсию (*WCSS* или просто V).

Для индекса отношения дисперсии можно использовать не абсолютные значения VR , а сравнивать их со случайно сгенерированными выборками, построенными на основе оригинальной выборки данных, как это сделано в [60], чтобы найти такое K , для которого оригинальная выборка данных кластеризуется лучше, чем случайно сгенерированные выборки аналогичной длины, имеющие такое же среднее и СКО, что и оригинальная выборка.

Ниже будут приведены методы, представленные в работе [43], которые лучше подходят для гладких полей $z500$ или SLP , плотно и равномерно распределенных в фазовом пространстве главных компонент их эмпирических ортогональных функций. Для этих методов используется не только оригинальная выборка данных с единственным решением от, например, k -means, но используются несколько решений k -means для каждого K и множество сгенерированных выборок, имеющих идентичные моменты распределений с оригинальной выборкой.

4) Индекс классифицируемости

Индекс классифицируемости использовался в работе [43] для выделения погодных режимов посредством кластеризации полей высоты геопотенциала на уровне 700 мбар. Идея этого индекса состоит в том, что, при оптимальном K , различные решения, полученные при кластеризации методом КМ с разными инициализациями, более похожи друг на друга, чем при неоптимальных K .

Допустим, что для данного K методом k -means по разным инициализациям посчитано M различных решений. Для каждой пары решений P, Q можно посчитать матрицу попарных корреляций центроидов кластеров:

$$A_{ij}(P, Q) = \text{corr}(C_i(P), C_j(Q)), \quad (1.27)$$

где C – один из K центроидов данного решения (формула (1.11)), а индексы

указывают на номер кластера ($i = 1:K, j = 1:K$). Далее, для каждой строки i матрицы A_{ij} ищется максимум по j :

$$A'_i(P, Q) = \max_j A_{ij}(P, Q), \quad (1.28)$$

то есть $A'_i(P, Q)$ – это лучшие попарные корреляции для каждого из центроидов кластеров в решении P , найденные среди центроидов кластеров в решении Q , то есть корреляции пар центроидов похожих, аналогичных кластеров. Далее, среди пар аналогичных центроидов ищется «худшая» пара, то есть пара центроидов с минимальной корреляцией:

$$c(P, Q) = \min_i A'_i(P, Q) = \min_i \left(\max_j A_{ij}(P, Q) \right) \quad (1.29)$$

Корреляция этой пары и будет характеризовать схожесть двух различных решений (P и Q , в данном случае). Стоит отметить, что в общем случае $c(P, Q) \neq c(Q, P)$, например, неравенство будет в случае, когда для разных центроидов i и i' из решения P находится один и тот же центроид j решения Q с максимальной корреляцией. Посчитав среднее $c(P_m(K), P_{m'}(K))$ для всех имеющихся пар решений, получим индекс классифицируемости:

$$c^*(K) = \frac{1}{M(M-1)} \sum_{1 \leq m \neq m' \leq M} c(P_m(K), P_{m'}(K)). \quad (1.30)$$

Для оценки значимости, полученный индекс классифицируемости $c^*(K)$ сравнивают с таковым, полученным по случайным сгенерированным выборкам. Предполагается, что разные решения при оптимальном K в оригинальной выборке данных будут более похожи и поэтому иметь бóльшие $c^*(K)$ по сравнению с таковыми у сгенерированных выборок. Выборки генерируются как марковские процессы, имеющие такую же автокорреляцию 1-го порядка и ковариацию, что и оригинальные данные. Получаем $N_{\text{ген}}$ индексов $c^*_{\text{ген}}(K, N)$, которые сортируются по возрастанию. Долю выборок, $c^*_{\text{ген}}(K, N)$ которых оказалось меньше $c^*(K)$ оригинальной выборки, можно считать уровнем значимости ($1 - p$ -уровень) индекса классифицируемости. Выбирается такое наименьшее K , для которого уровень значимости оказался выше уровня, заданного исследователем. При использовании данного индекса рекомендуется использовать как можно большие M и $N_{\text{ген}}$ (50 и 100 в оригинальной статье [43]).

5) Индекс воспроизводимости

Идея индекса воспроизводимости, также описанного в [43] (Michelangeli et al., 1995), заключается в том, что если для определенного K набор данных «хорошо» кластеризуется, то случайные половинчатые подвыборки этого набора данных будут сходиться к тем же решениям (то есть кластеризациям с теми же координатами центроидов), что и решения на основе оригинальной выборки.

Методом КМ создаётся M различных решений, полученных с помощью разных инициализаций. Среди них ищется *эталонное решение* – такое $m \in M$ решение, для которого будет максимально величина $c(P_m, P_{m'})$ (формула (1.29)), усредненная по всем остальным решениям m' :

$$\max_m \frac{\sum_{m' \neq m} c(P_m, P_{m'})}{M - 1}, \quad (28)$$

то есть ищется m -ое решение, центроиды которого в среднем больше всего похожи на аналогичные центроиды остальных решений.

Далее, создаются R подвыборок, содержащих 50% случайно выбранных точек из оригинальной выборки данных. Для каждой из подвыборок, таким же образом, как для оригинальной выборки, проводится M кластеризаций и аналогичным образом выбирается эталонное решение. Методом попарных корреляций центроидов кластеров, аналогично индексу классифицируемости, для каждой пары эталонных решений оригинальной выборки P_m и одной из подвыборок P_{mr} считается $A_i'(P_m, P_{mr}) = \max_j A_{ij}(P_m, P_{mr})$ (формула (1.28)) и осредняется по всем $r = 1 \dots R$. Данная процедура проводится для каждого рассматриваемого K . Оптимальным можно считать такое K , для которого $\langle A_i'(P_m, P_{mr}) \rangle_r$, то есть средняя по всем подвыборкам попарная корреляция центроидов эталонного решения оригинальной выборки и эталонных решений подвыборок, будет выше некоторого заданного уровня значимости для каждого $i = 1 \dots K$ центроида эталонного решения оригинальной выборки.

Таким образом, расчетом индекса воспроизводимости происходит проверка того, чтобы каждый из $i = 1 \dots K$ центроидов кластеризаций оригинальной выборки хорошо воспроизводился в кластеризациях случайных половинчатых подвыборок. Для каждого потенциального K индекс воспроизводимости будет давать K чисел (по числу центроидов), характеризующих оптимальность выбранного числа кластеров. Для простоты анализа их можно усреднить, чтобы характеризовать «оптимальность» данного исследуемого числа кластеров K единственным числом.

Вопрос количества погодных режимов, остается одним из ключевых при режимном подходе в исследованиях крупномасштабной атмосферной циркуляции [21, 43, 46]. Оптимальное K , если оно существует, зависит как от выбранного метода кластеризации, так и от свойств набора данных – исследуемой характеристики, определяющей циркуляцию ($z500$, SLP или другой переменной), выбранной области, продолжительности временного ряда, процедуры предобработки данных и т. д. [24].

1.4. Идентификация погодных режимов в ЕАТ

Несмотря на наличие большого числа работ по выделению погодных режимов в ЕАТ, в подавляющем большинстве из них режимы выделялись для зимних месяцев [19, 23, 24, 27, 43], иногда включая ноябрь и март [61], когда суточные аномалии полей давления или высоты выбранной изобарической поверхности наиболее сильны, что позволяет выделять хорошо различимые кластеры, то есть кластеры с относительно высоким *межкластерным* и низким *внутрикластерным* расстояниями. Работ по определению погодных режимов в ЕАТ в летние месяцы относительно немного [62], при этом основное внимание в них уделяется именно летним фазам Североатлантического колебания (SNAO) [63]. В данной главе, сравнение результатов различных методов кластерного анализа при выделении погодных режимов в ЕАТ будет проводиться не только для зимнего сезона, но и для летнего, что позволит сопоставить различие режимов в оба сезона при их выделении одним и тем же методом.

1.4.1. Предобработка данных и кластеризация

1) Данные и предобработка

Для идентификации режимов использовались поля высоты геопотенциала на уровне 500 гПа ($z500$) по данным из реанализа ERA5 [64] с 1940 по 2022 гг. Поля были осреднены до суточных и интерполированы на сетку с шагом в 1 градус для удобства. Для выделения режимов в Евро-Атлантике был выбран сектор от 80 з. д. – 40 в. д. по долготе и 30 с. ш. – 80 с. ш. по широте. Аналогичная область с отклонениями $\pm 10^\circ$ чаще всего использовалась для выделения погодных режимов в ЕАТ [19, 23, 24, 27].

Перед кластеризацией были получены поля аномалий путем вычитания из суточных полей сезонного хода, который считался осреднением суточных полей за одни и те же календарные даты разных лет за весь период (1940 – 2022 гг.), сглаженные скользящим средним в 5 суток. Для удаления высокочастотных синоптических волн к полям аномалий применялся фильтр Баттерворта низких частот [65] с периодом фильтрации 10 суток. Из фильтрованных суточных полей аномалий высоты геопотенциала выбирались зимние (декабрь, январь, февраль) и летние (июнь, июль, август) месяцы для кластеризации зимних и летних режимов соответственно. За исследуемый 83-летний период 1940 – 2022 гг. для зимних месяцев имеется 7491 суточных полей, для летних – 7636.

Для сглаживания пространственной дисперсии полей и уменьшения размерности данных фильтрованные суточные поля аномалий высоты геопотенциала (далее, поля аномалий) раскладывались на эмпирические ортогональные функции (ЭОФ). Перед разложением на ЭОФ поля аномалий были взвешены на корень из площади для уравнивания вклада в дисперсию узлов сетки полей на разных широтах. Для зимних полей было отобрано 13 первых ЭОФ, объясняющих 94,6% дисперсии, для летних полей 20 первых ЭОФ, объясняющих 94,8% дисперсии. Такие числа ЭОФ выбраны, чтобы они объясняли примерно по 95% дисперсии в обоих случаях. Непосредственно кластеризации подвергались временные ряды главных компонент ЭОФ. Стоит отметить, в разных работах с выделением режимов атмосферной циркуляции в ЕАТ перед кластеризацией используют разное число ЭОФ, например, 4 в работах (Dawson et al., 2012 [23]) и (Fabiano et al., 2020 [27]) или 14 в (Cassou, 2008 [19]), (Charlton-Perez et al., 2018 [66]). Как правило, авторы отмечают, что результаты кластеризации не сильно меняются при увеличении числа ЭОФ. В работе (Falkena et al., 2020 [24]) исследуются различия в погодных режимах, полученных методом k-means (КМ), в зависимости от способа предобработки данных – использования (или неиспользования) временных фильтров и разложения на ЭОФ. Утверждается, что режимы, полученные при разложении данных на 10, 15 и 20 первых ЭОФ и без разложения (оригинальные данные), практически не отличаются и имеют схожие повторяемости и вероятности переходов, однако предпочтение отдаётся большему числу ЭОФ.

2) Выбор количества режимов

Была проведена оценка оптимального количества режимов на интервале K от 2 до 10. Как было отмечено выше, простые методы проверки

оптимального числа кластеров (метод локтя и Силуэт) не дают четкого ответа. Графики $V(K)$ и $\langle S(i) \rangle$ от K представляют собой убывающие функции, однако для зимних режимов наблюдается слабый перегиб $V(K)$ и локальный максимум $\langle S(i) \rangle$ при $k = 4$. Для летних режимов аналогичной картины не наблюдается, кривые $V(K)$ и $\langle S(i) \rangle$ от K плавно убывают.

Для оценки оптимального числа режимов использовались в том числе индексы классифицируемости и воспроизводимости. Для подсчета индекса классифицируемости было сгенерировано 400 случайных выборок с одинаковыми дисперсиями и автокорреляцией 1-го порядка, как у оригинальных временных рядов главных компонент, и 400 случайных подвыборок для подсчета индекса воспроизводимости. Для зимних режимов индексы показывают локальные пики для $K = 4$ и $K = 6$ (см. рис. 1.4), что в целом соответствует результатам работы (Falkena et al., 2020 [24]), но ни для одного K индекс классифицируемости оригинальной выборки не оказался выше, чем у 95% сгенерированных выборок. Для летних режимов локальный пик индексов наблюдается при $K = 3$ (см. рис. 1.5).

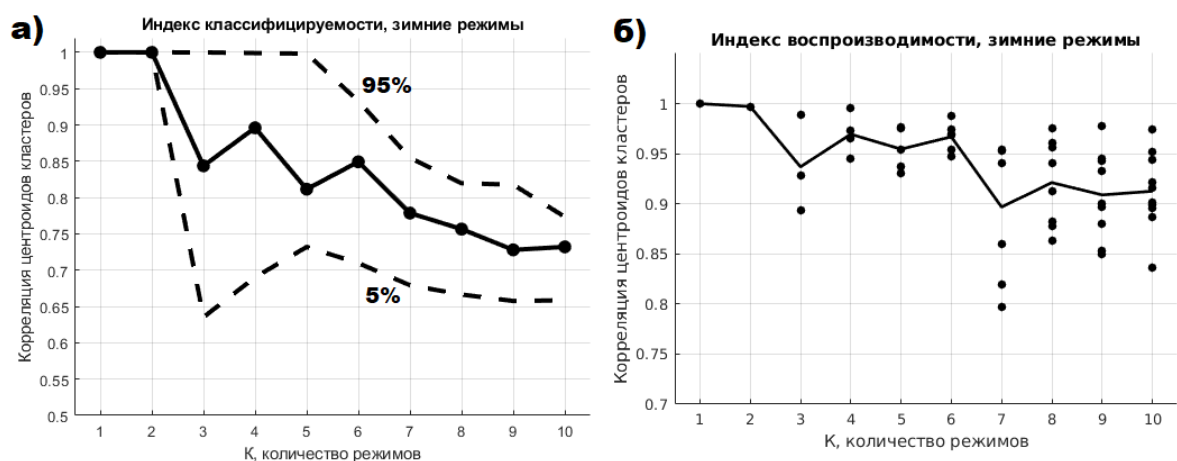


Рисунок 1.4. Индексы классифицируемости (а) и воспроизводимости (б) для зимних режимов. а) пунктирными линиями обозначены 5% (20-ая выборка из 400) и 95% (380-ая выборка из 400) уровни индекса сгенерированных выборок, сплошной линией с точками значение индекса для оригинальной выборки; б) сплошной линией среднее по всем режимам значение индекса воспроизводимости, точками обозначено значение индекса воспроизводимости для каждого из $k=1:K$ режимов в отдельности.

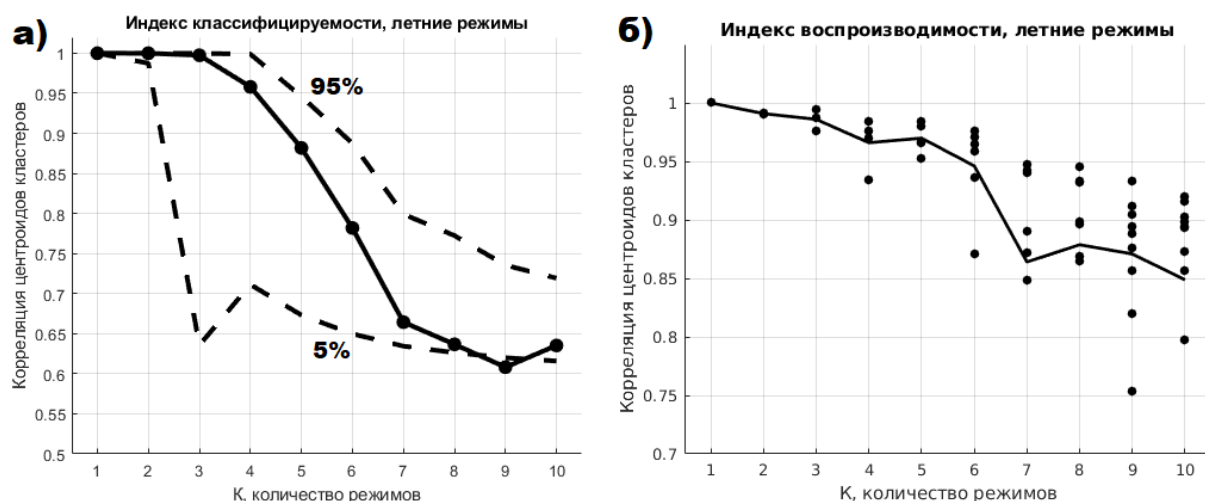


Рисунок 1.5. То же, что на рисунке 1, но для летних режимов.

Помимо индексов классифицируемости и воспроизводимости, были рассчитаны значения индекса VR Калински-Харабаша для оригинальной выборки и 400 случайно сгенерированных выборок (аналогично индексу классифицируемости), и сопоставлены между собой. Результаты представлены на рисунке 1.6. Значимость на рисунке показывает, какую долю индексов VR сгенерированных выборок превышает VR, полученная по оригинальной выборке данных, то есть на сколько оригинальные данные кластеризуются лучше (имеют выше межкластерную дисперсию и ниже внутрикластерную дисперсию), чем сгенерированные выборки красного шума с таким же средним, СКО и автокорреляцией lag-1.

Из рисунка 1.6 видно, что все K для зимнего сезона, начиная с $K = 4$, являются значимыми на уровне выше 95%, то есть кластеризуются лучше броуновского шума. При прочих равных, меньшее K является более предпочтительным, так как позволяет проще описать систему, например, число парных связей между режимами растёт $\sim K^2$. Для летних месяцев значимых K не обнаружено, все лежат ниже уровня 80%, то есть кластеризуются не лучше, чем красный шум, сгенерированный на основе оригинальной выборки.

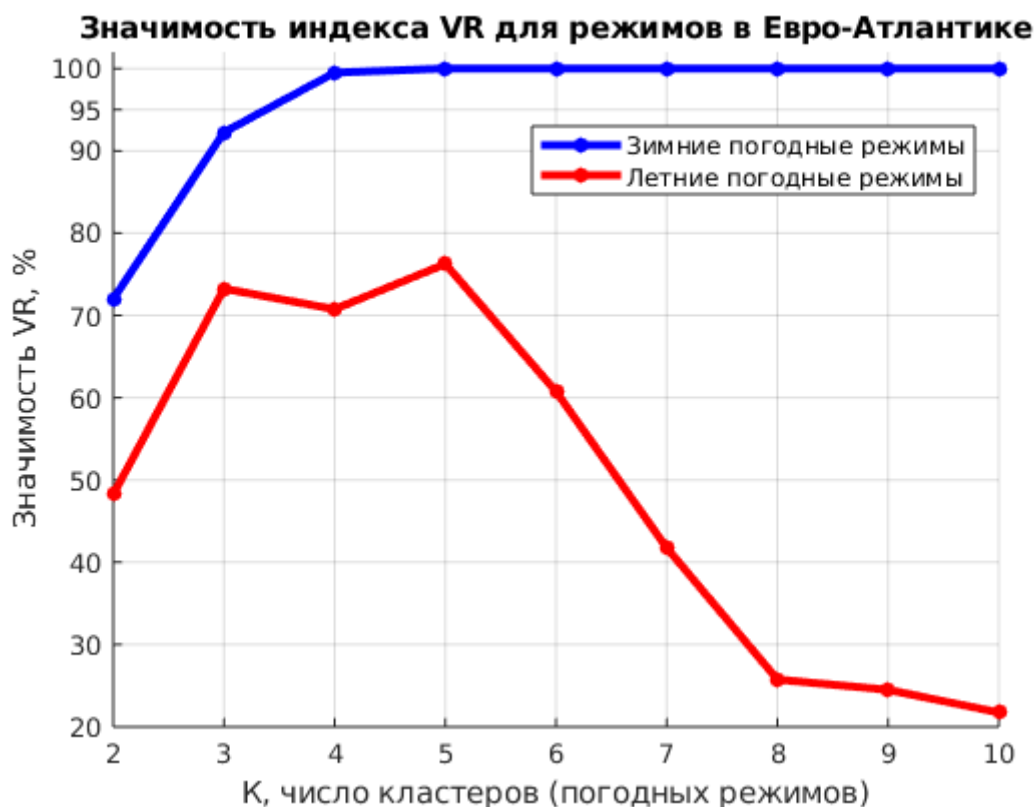


Рисунок 1.6. Значимость индекса VR для зимних и летних погодных режимов, рассчитанная как доля сгенерированных выборок, значения VR для которых оказались ниже, чем для оригинальной выборки временных рядов главных компонент ЭОФ-разложения полей z500.

В целом, использованные методы, кроме индекса VR Калински-Харабаша отношения дисперсий VR, указывающего на $K = 4$ для зимы, не дали однозначного ответа об оптимальном количестве режимов. Так как в большинстве исследований для выделения зимних режимов в Евро-Атлантике используется $K = 4$ [19, 23, 27, 43, 66], было решено выбрать это число для возможности сопоставления с результатами других авторов. Для летних режимов кластеризация так же проводилась с $K = 4$ для возможности сравнения полученных летних режимов с зимними. Стоит отметить, что существуют работы, в которых в ЕАТ зимой выделяется 6 режимов [24], в таком случае режимы NAO+, NAO- остаются визуально неизменными (поля центроидов этих режимов), а у режимов SB и AR появляются «противоположные», SB- и AR-.

Обработка данных и их последующая кластеризация выполнялись в программной среде MATLAB с использованием как встроенного функционала, так и дополнительных программ, разработанных в том числе автором диссертации. Предобработка данных (удаление сезонного хода, фильтрация по времени, разложение на ЭОФ) выполнялась как с

использованием встроенных функций, так и функций из пакета «Climate Data Toolbox for MATLAB» (Greene et al., 2019, [33]) («filt1», «eof» и другие). Для кластеризации КМ с использованием оптимизаций «kmeans++» и «simulated annealing», а также для подсчета индексов классифицируемости и воспроизводимости использовались написанные автором диссертации программы. Для остальных методов кластеризации – GM, HW и SOM – использовались встроенные в MATLAB функции («fitgmdist», «cluster», «linkage», «nctool» и прочие). Для дальнейшего анализа результатов методов КМ, GM и SOM использовались решения с наименьшими V или функцией правдоподобия по итогам 1000 прогонов алгоритма с различными инициализациями, для HW результат постоянный и не зависит от числа запусков алгоритма.

1.4.2. Поля погодных режимов в ЕАТ

Средние характерные поля аномалий z500 режимов, полученные путем усреднения всех суточных полей аномалий z500, попавших в соответствующие режимы, представлены на рисунке 1.7. Режимы, полученные методом SOM, не приведены на рисунке. Результаты кластеризации методом SOM оказались практически идентичны результатам, полученным по методу КМ – число суточных полей, попавших в другие, чем в методе КМ, режимы, составило 12 и 10 полей из 7491 и 7636 для зимних и летних режимов соответственно, то есть менее 0.2% от длины выборки (также см. сравнение внутрикластерных дисперсий в таблице 1.1), поэтому далее в данной главе результаты метода SOM не приведены, так как принимаются идентичными таковым по методу КМ.

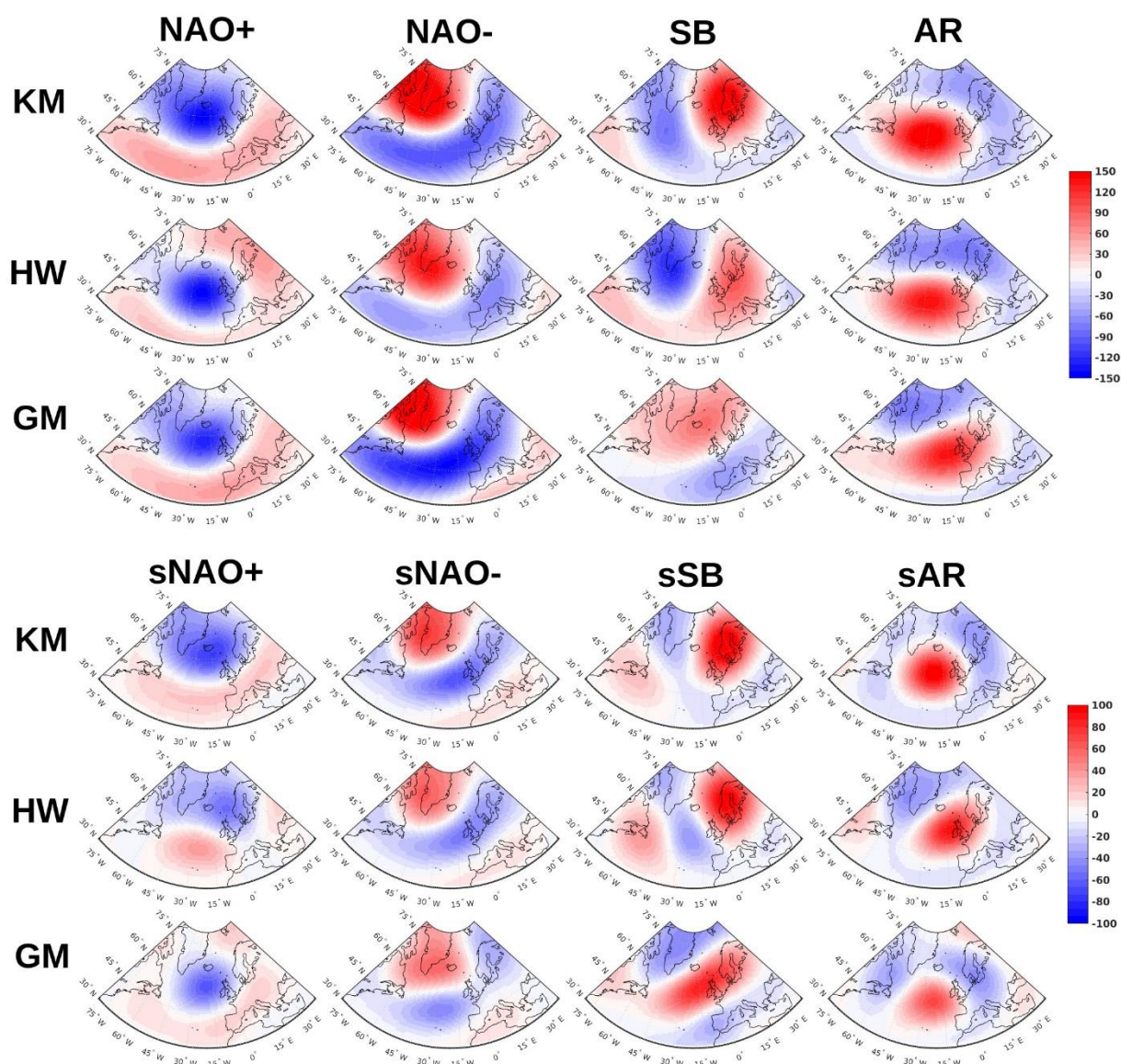


Рисунок 1.7. Средние поля аномалий z500 зимних и летних режимов в ЕАТ секторе за 1940–2022 гг., полученных с помощью различных методов кластеризации – k-means (KM), иерархическая кластеризация Уорда (HW), гауссова смесь (GM). Верхние три ряда – зимние режимы (декабрь, январь, февраль), нижние три ряда – летние (июнь, июль, август). Справа от полей приведены шкалы величины аномалий z500 в метрах для зимних и летних режимов.

Таблица 1.1. Сумма внутрикластерных расстояний зимних и летних погодных режимов, полученных разными методами кластеризации. Единица измерения чисел в таблицах – 10^4 м^2 .

Зимние погодные режимы						
Метод кластеризации	NAO+	NAO-	SB	AR	Сумма внутрикластерных дисперсий	Доля объясненной дисперсии

KM	1512	1374	1303	1323	5511	32.64%
SOM	1514	1374	1304	1319	5511	32.64%
HW	1267	2208	1248	1030	5753	29.68%
GM	1537	676	2122	1561	5896	27.94%
Летние погодные режимы						
Метод кластеризации	sNAO+	sNAO-	sSB	sAR	Сумма внутрикластерных дисперсий	Доля объясненной дисперсии
KM	462	603	494	476	2035	23.62%
SOM	461	604	494	477	2035	23.62%
HW	551	713	364	443	2070	22.29%
GM	746	631	424	338	2139	19.72%

Названия NAO+, NAO-, SB, AR даны для режимов, полученных методом KM, так как визуально они наиболее совпадают с традиционными зимними погодными режимами в ЕАТ, полученными этим методом в других исследованиях [23, 27, 66]. Не все режимы, полученные по методам HW и GM, воспроизводят вышеназванные режимы, однако они сопоставлены с режимами по методу KM таким образом, чтобы средняя пространственная корреляция режимов, названных аналогично, была максимальной. На рис. 1.7 и далее в главе для летних режимов используются аналогичные зимним режимам (NAO+, NAO-, SB, AR) названия с припиской «s» (summer).

Из рисунка видно, что летние поля режимов являются менее выраженными (поля на рисунках более бледные даже с учетом использования уменьшенной в 1.5 раза шкалы) по сравнению с зимними, что является следствием более низкой изменчивости полей z_{500} в летние месяцы по сравнению с зимними.

В таблице 1.1 приведено сравнение внутрикластерных дисперсий в зависимости от выбранного метода кластеризации. Видно, что KM и SOM произвели кластеры с близкими внутрикластерными расстояниями (разница в пользу k-means в 5-ом знаке после запятой, не приведено в таблице), при этом кластеризации HW и GM оказываются хуже в терминах суммарных внутрикластерных расстояний и доли объясненной дисперсии, то есть в среднем кластеры HW и GM менее плотные, несмотря на выраженность аномалий z_{500} в некоторых из них (например, NAO- в GM). Стоит отметить, что доля объяснённой дисперсии (которая считается как $1 - V_K/V_1$) зависит от метода предобработки данных и выбранного числа ЭОФ разложения

оригинальных полей, однако тенденция превосходства КМ при разном числе ЭОФ сохраняется.

В таблице 1.2 представлено сравнение пространственных корреляций средних полей режимов, полученных разными методами, а также сравнение пространственных корреляций зимних и летних полей по результатам кластеризации для одного и того же метода. По приведенным значениям в таблице видно, что результаты кластеризации GM и HW плохо коррелируют друг с другом (< 0.5 в среднем для обоих сезонов), но хорошо коррелируют с КМ (> 0.7 в среднем для обоих сезонов). Сравнение пространственных корреляций зимних и летних полей аналогичных режимов показывает, что наиболее похожие летние режимы по сравнению с зимними получаются при методе КМ, хуже всего при методе HW, что так же свидетельствует о том, что КМ является более предпочтительным методом при кластеризации воспроизводимых полей циркуляции атмосферы.

Таблица 1.2. Пространственные корреляции средних полей режимов, полученных разными методами для зимних и летних сезонов.

Корреляции режимов, полученных разными методами					
Методы кластеризации	(s)NAO+	(s)NAO-	(s)SB	(s)AR	В среднем
	Зимние погодные режимы				
КМ и HW	0.55	0.94	0.67	0.89	0.76
КМ и GM	0.97	0.92	0.20	0.71	0.70
HW и GM	0.67	0.74	-0.56	0.79	0.41
	Летние погодные режимы				
КМ и HW	0.79	0.99	0.92	0.63	0.83
КМ и GM	0.66	0.84	0.65	0.73	0.72
HW и GM	0.19	0.84	0.37	0.37	0.44
Корреляции аналогичных зимних и летних режимов, полученных одним методом					
Методы кластеризации	NAO+ и sNAO+	NAO- и sNAO-	SB и sSB	AR и sAR	В среднем
КМ	0.90	0.81	0.85	0.81	0.84
HW	-0.20	0.62	0.47	0.48	0.34
GM	0.83	0.83	0.25	0.51	0.60

1.5. Выводы

В главе приведен обзор четырех наиболее часто используемых методов кластеризации для выделения режимов крупномасштабной атмосферной циркуляции – k-means (KM), иерархической кластеризации Уорда (HW), модели Гауссовой смеси (GM) и Самоорганизующихся карт Кохонена (SOM), а также некоторых численных методов для определения оптимального количества режимов. По суточным данным реанализа ERA5 для высоты геопотенциальной поверхности на уровне 500 гПа (z500) с помощью вышеперечисленных методов идентифицированы погодные режимы в зимние и летние месяцы в ЕАТ и проведено сравнение их взаимное сопоставление.

Для региона ЕАТ по данным z500 реанализа ERA5 для периода 1940-2022 гг. в зимние месяцы показано, что минимальное оптимальное число режимов K равно 4. Для летних месяцев единственного статистически значимого K , использованные автором диссертации методы не дали. В целом, задача определения числа погодных режимов в ЕАТ является предметом дискуссии по сей день [21, 24]. В данной работе для обоих сезонов выделялись 4 погодных режима, как в большинстве работ по выделению зимних режимов в Евро-Атлантике (Cassou, 2008 [19]; Michelangeli et al., 1995 [43]; Fabiano et al., 2020 [27]; Dawson et al., 2012 [23]; Charlton-Perez et al., 2018 [66]) с помощью метода KM по данным о высоте геопотенциала. Летние режимы, выделенные методом KM, оказались визуально похожи на зимние, средний коэффициент пространственной корреляции характерных полей режимов оценен равным 0.84.

Режимы, полученные с помощью метода SOM, оказались практически неотличимы от режимов, выделенных методом KM. По результатам кластеризации из 7491 (7636) суточных полей z500 за зимний (летний) сезон для периода 1940-2022 гг. лишь 12 (10) суточных полей (менее 0.2%) было отнесено к другим, чем при использовании метода KM, режимам. Это не повлияло на основные статистические характеристики режимов, в связи с чем результаты, полученные методом SOM, можно считать идентичными результатам, полученным методом KM.

Режимы, полученные с помощью методов HW и GM, в среднем менее выражены (имеют более низкую долю объясненной дисперсии аномалий z500), отличаются визуально и имеют другие статистические характеристики – относительную повторяемость, характерную продолжительность и вероятности переходов. Сопоставление структуры режимов, проведенное

путём расчета коэффициентов пространственной корреляции их средних полей, показывает, что: а) режимы, полученные методами КМ и НВ и методами КМ и ГМ более похожи, чем режимы, полученные методами НВ и ГМ; б) летние режимы, полученные методом КМ, меньше отличаются от зимних, чем при использовании других методов – средний коэффициент пространственной корреляции между полями зимних и их аналогичных летних режимов составил 0.84 для метода КМ против 0.34 и 0.60 у методов НВ и ГМ соответственно. Согласно полученным результатам, метод КМ представляется более предпочтительным в задачах выделения погодных режимов, чем методы НВ и ГМ. Метод SOM требует настройки большего числа параметров, чем КМ, а его преимущество относительно других методов, состоящие во взаимной сортировке (самоорганизации) кластеров, при небольшом числе K не требуется.

Таким образом, режим k-means среди четырех протестированных методов выбран как оптимальный метод кластерного анализа для задачи идентификации погодных режимов по следующим признакам: 1) лучшее сходство полей режимов с другими методами, чем у остальных методов по сравнению друг с другом; 2) лучшее сходство полей режимов в различные сезоны; 3) минимальная сумма внутрикластерных расстояний; 4) простота относительно метода SOM, производящего идентичные погодные режимы.

ГЛАВА 2. ХАРАКТЕРИСТИКИ ПОГОДНЫХ РЕЖИМОВ В РЕГИОНАХ ЕВРО-АТЛАНТИКЕ И СЕВЕРНОЙ ЕВРАЗИИ В ЗИМНИЙ И ЛЕТНИЙ СЕЗОНЫ

В первой главе были представлены методы кластерного анализа, использующиеся для идентификации погодных режимов, по итогу сравнения которых метод k-means выбран как наиболее оптимальный. В данной главе будут представлены характеристики погодных режимов в Евро-Атлантике (ЕАТ), полученных с помощью кластеризации суточных полей z500 методом k-means. Так же для сопоставления будут приведены результаты, полученные с помощью других методов, результаты остальных методов для региона ЕАТ приведены скорее для справки, либо для подтверждения общих тенденций для полученных четырех погодных режимов (NAO+, NAO-, SB, AR), проявившихся в том числе по результатам кластеризации другими методами (GM и HW).

Помимо ЕАТ, аналогичным образом методом k-means проведена кластеризация z500 для идентификации погодных режимов в области Северной Евразии (СЕ) и представлены пространственные и временные характеристики для режимов в этой области. Несмотря на широкое применение режимного подхода не только в ЕАТ, но и в других областях внетропической части Северного полушария [16, 25, 26, 27], погодные режимы в СЕ, непосредственно влияющие на погоду на территории России, до сих пор не идентифицировались отдельно для данной области.

2.1. Виды временных характеристик погодных режимов

По результатам кластеризации каким-либо методом, каждое суточное поле, характеризующее атмосферную циркуляцию, присваивается одному из K режимов. Простейшей статистической характеристикой режима можно считать его частоту или относительную повторяемость, то есть долю суточных полей, отнесенных данному режиму. Другими статистическими характеристиками режимов являются их характерная продолжительность, наиболее вероятные и статистически значимые [67] переходы между режимами, которые можно выявить, посчитав матрицу переходов и сравнив её со сгенерированными матрицами, посчитанными путем случайного перемешивания режимных событий.

В некоторых исследованиях используют переходные дни [68], которые не относят к какому-либо из режимов. Их наличие меняет подход к изучению переходов между режимами, добавляя системе «переходное» состояние и позволяя рассматривать события, когда режимы переходят не только из одного в другой, но и возвращаются «сами в себя» через переходное состояние. В данной работе не будут использоваться переходные дни.

Так как у каждого из режимов по итогу кластеризации оказывается различная относительная повторяемость, то, при прочих равных, вероятность перехода в режимы с высокой относительной повторяемостью будет выше по сравнению с режимами с низкой относительной повторяемостью, если бы переходы происходили чисто случайно. Чтобы сравнить, насколько полученные вероятности переходов между режимами являются статистически значимыми, то есть отличными от чисто случайных, существует метод, предложенный в [67]. Суть метода состоит в следующем: генерируются N случайных временных рядов чисел $k =$ от 1 до K с длиной, равной длине оригинального временного ряда режимных событий. Ряды генерируются таким образом, чтобы общее количество режимных событий каждого k -го из K режимов сгенерированного временного ряда равнялось количеству режимных событий оригинального временного ряда. Далее, для каждой из сгенерированных выборок считается своя матрица переходов $TM(i, j)_{n, \text{ген}}$ где $n=1:N$, которая затем сравнивается с оригинальной матрицей переходов $TM(i, j)$. Переход между режимами $i \Rightarrow j$ будет считаться статистически значимо частым на уровне значимости P , если в сгенерированных матрицах переходов такой переход окажется равновероятен или более вероятен, то есть $TM(i, j)_{n, \text{ген}} \geq TM(i, j)$, не более, чем в $N \cdot (1 - P)$ случаев. В работе [67] $N=10000$, при $P=0.95$ (уровень значимости 5%) число случаев, когда $TM(i, j)_{n, \text{ген}} \geq TM(i, j)$, должно быть меньше 500.

Как правило, наиболее населенные режимы являются в то же время самыми продолжительными и наиболее вероятными для переходов из других режимов, как бы стягивая переходы в свою сторону в виду большей населенности, но при этом могут не быть значимо частыми. Стоит отметить, что характерная (средняя или медианная) продолжительность режимов, описывающих низкочастотную изменчивость атмосферной циркуляции, существенно зависит от способа предобработки полей циркуляции, а именно применения или неприменения временной фильтрации данных для отсеивания синоптической изменчивости и соответствующего выбора периода отсечки фильтра.

В различных исследованиях с выделением четырех погодных режимов в Евро-Атлантическом секторе – положительной и отрицательной фаз Североатлантического колебания (NAO+ / NAO-), Скандинавского блокинга (SB) и Атлантического антициклона (AR) – их относительная населенность различается в зависимости от выбранного способа предобработки, границ временного периода и сектора, но для последних десятилетий прослеживается общая картина – самым населенным ($\approx 30\%$ дней) и продолжительным является режим NAO+, средними по населенности оказываются режимы SB и AR ($\approx 23\text{--}27\%$ дней), наименее населенным NAO- ($\approx 20\%$ дней) (Cassou, 2008 [19]; Dawson et al., 2012 [23]; Charlton-Perez et al., 2018 [66]).

2.2. Временные характеристики погодных режимов в ЕАТ

Пространственные характеристики, то есть структуры полей погодных режимов в ЕАТ, можно увидеть на рисунке 1.7 по методу КМ. Среди четырех режимов NAO+ можно считать стандартным типом течения, при котором теплые западные ветра из Атлантики проникают на территорию Европы. Остальные четыре режима – NAO-, SB и AR соответствуют антициклоническим образованиям и препятствуют типичному движению западных ветров в сторону континентальной Европы. Летние режимы sNAO+, sNAO-, sSB, sAR, полученные по методу КМ, по пространственной структуре в целом аналогичны зимним с разницей в том, что абсолютная величина аномалий $z500$ для них в 1.5–2 раза ниже, чем зимой.

1) Относительная повторяемость режимов в ЕАТ

Для полученных разными методами погодных режимов в ЕАТ (см. раздел 1.4.2 диссертации) анализировались и сравнивались относительная и сезонная повторяемости, средняя продолжительность и матрицы переходов. Относительная повторяемость режимов представлена в таблице 2.1, где также для справки приведена повторяемость зимних погодных режимов в ЕАТ, полученных другими авторами методом КМ за разные временные периоды и с немного отличающимися способами предобработки данных.

Таблица 2.1. Относительная повторяемость зимних и летних погодных режимов, полученных разными методами. Повторяемость традиционных зимних погодных режимов в ЕАТ-секторе, полученных в других работах методом КМ, выделена курсивом. Режимы, для которых тренды сезонной повторяемости за 1940-2022 гг. значимые на p -уровне = 0.05, отмечены жирным шрифтом и соответствующим знаком тренда.

Метод кластеризации	(s)NAO+	(s)NAO-	(s)SB	(s)AR
Зимние погодные режимы				
КМ	30.2% +	21.4% -	23.7%	24.7%
Cassou, 2008	30%	20%	27%	23%
Dawson et al., 2012	29.6%	20.4%	27.6%	22.4%
Charlton-Perez et al., 2018	29.7%	20.0%	28.6%	21.8%
HW	23.6%	31.5% -	23.1%	21.8%
GM	30.0% +	12.3%	30.4% -	27.4%
Летние (s) погодные режимы				
КМ	25.5%	28.1%	23.4%	23.0%
HW	26.9%	33.4%	17.4%	22.4%
GM	37.8%	25.4% +	19.1%	17.8%

По таблице 2.1 видно, что полученная в работе повторяемость режимов NAO+ и NAO- в пределах 1.5% от относительной повторяемости согласуется с таковой в других работах (Cassou, 2008 [19]; Dawson et al., 2012 [23]; Charlton-Perez et al., 2018 [66]). Полученные автором повторяемости режимов SB и AR сильнее отличаются от их повторяемостей в других работах, что, скорее всего, является следствием большей чувствительности этих режимов к выбору временного интервала (1940 – 2022 гг. в настоящей работе) и способу предобработки данных: в работе выбрано 13 ЭОФ в зимние месяцы, вычитание сезонного цикла сглаживалось 5-дневным скользящим средним и применялся 10-дневный low-pass фильтр, см. раздел 1.4.1), а в работах вышеуказанных авторов эти цифры могут различаться.

При сопоставлении результата в настоящей работе с работами других автором прослеживается общая тенденция – самый частый режим – NAO+, средние по частоте SB и AR, а NAO- – самый редкий. При кластеризации z500 другими методами данная тенденция не прослеживается – в HW режим NAO- оказался наиболее продолжительным с частотой в 31.5%, в GM NAO- –

наименее продолжительный, но со слишком низкой относительной повторяемостью 12.3%.

Несмотря на хорошую пространственную корреляцию средних полей летних режимов по сравнению с зимними при методе КМ (средняя корреляция 0.84, см. таблицу 1.2), их относительная повторяемость отличается – самым частым оказался режим sNAO- с повторяемостью 28.1% против 21.4% у NAO- зимой, а повторяемость sNAO+ составила 25.5% против 30.2% у NAO+ зимой. В целом, относительная населенность летних режимов получилась более равномерной по сравнению с зимними, а создание равномерно населенных кластеров является одним из свойств метода КМ при плохой кластеризуемости данных (доля объясненной кластеризацией дисперсии 23.62% летом против 32.64% зимой, см. таблицу 1.1).

Помимо относительной повторяемости режимов, анализировались временные ряды сезонной повторяемости (количество суток за сезон, когда наблюдался данный режим) и их линейные тренды. Для сезонной повторяемости зимних режимов, полученных разными методами, получились разные по величине и значимости тренды, но наблюдается общая картина – положительный тренд повторяемости NAO+, который оказался значимым (на α -уровне 5%) для режимов, полученных по методам КМ и GM, и отрицательный тренд повторяемости NAO-, который оказался значимым для режимов, полученных по методам КМ и HW.

Временные ряды и линейные тренды режимов NAO+ и NAO-, полученных методом КМ, представлены на рисунке 2.1. Режимы NAO+ и NAO- ассоциированы с волнами тепла (NAO+) и холода (NAO-) соответственно над территориями Европы и Западной России (Cattiaux et al., 2010 [29]; Бардин и др., 2019 [69]), поэтому их многолетние тренды сезонной повторяемости могут частично объяснять общий тренд потепления в данных регионах в последние десятилетия.



Рисунок 2.1. Временные ряды (линии с точками), скользящие средние с окном осреднения в 10 лет (черные жирные кривые) и линейные аппроксимации (пунктирные линии) сезонной повторяемости режимов NAO+ и NAO-, полученных методов КМ. Линейный тренд NAO+: $+0.22$ дней за сезон / год, p -уровень = 0.004; линейный тренд NAO-: -0.15 дней за сезон / год, p -уровень = 0.050.

Зимние режимы SB и AR, полученные по различным методам, имеют незначимые разнонаправленные тренды сезонной повторяемости, кроме значимого отрицательного тренда SB при методе GM, который по своей пространственной структуре слабо похож на свои аналоги в методах HW и KM (см. рисунок 1.7 и таблицу 1.2), то есть, вообще говоря, не представляет собой скандинавский блокинг. Для летних режимов значимые тренды сезонной повторяемости отсутствуют практически для всех режимов, полученных разными методами, кроме положительного тренда для режима sNAO-, полученного по методу GM.

2) Характерная продолжительность режимов в ЕАТ

В таблице 2.2 представлены средние продолжительности режимов. Самым продолжительным оказался режим NAO-, причем как для зимних месяцев (> 10 дней по методам КМ и НВ), так и для летних (sNAO-, >9 дней по методам КМ и НВ). Такая повышенная продолжительность отмечается на фоне самой низкой (по методу КМ) среди четырех режимов относительной повторяемости в зимние месяцы. Наименее продолжительными оказались режимы SB в зимние месяцы и sAR в летние.

Таблица 2.2. Статистические характеристики продолжительности режимов, полученных разными методами. Столбцы, обозначенные символом «w» – средние взвешенные (на относительную повторяемость) значения по всем четырем режимам.

Метод кластерно- го анализа	Зимние режимы					Летние режимы				
	NAO+	NAO-	SB	AR	w	sNAO+	sNAO-	sSB	sAR	w
Средняя продолжительность, суток										
КМ	9.5	10.4	7.8	8.2	9.0	7.2	9.1	7.3	6.7	7.6
НВ	8.7	11.6	7.6	8.8	9.4	8.3	11.2	7.8	7.9	9.1
ГМ	8.6	8.6	7.4	8.5	8.2	8.5	7.1	7.0	5.9	7.4
Медианная продолжительность, суток										
КМ	7.0	8.0	6.0	6.0	6.7	6.0	7.0	6.0	5.0	6.1
НВ	7.0	9.0	6.0	7.0	7.4	6.0	8.0	6.0	6.0	6.7
ГМ	7.0	7.5	6.0	6.0	6.5	6.0	5.0	5.0	4.0	5.2
Среднеквадратичное отклонение, суток										
КМ	8.6	8.2	6.0	7.4	7.6	5.9	7.4	6.0	5.2	6.2
НВ	6.9	9.4	6.3	6.5	7.5	6.4	9.5	5.7	6.0	7.2
ГМ	7.3	7.1	5.9	7.8	7.0	7.8	5.9	6.5	4.7	6.5

Сравнение средних и медианных значений продолжительности режимов, полученных разными методами, показывает, что наиболее продолжительные режимы получаются при кластеризации методом НВ, а наименее продолжительные при кластеризации ГМ. Для всех режимов отмечается высокая величина стандартного отклонения их продолжительности, которая близка к средней продолжительности и превышает медианную, что говорит о том, что среди событий режимов часто

встречаются как и очень короткие, продолжительностью не более трех суток, которые можно условно назвать переходными, так и довольно продолжительные, которые могут занимать до двух месяцев (средняя максимальная продолжительность по всем режимам и методам для зимних месяцев составляет 45 суток, для летних 40 суток). На рисунке 2.2 представлены гистограммы распределения продолжительностей для режимов, полученных по методу КМ. Из рисунка 2.2 и таблицы 2.2 (строки КМ) видно, что летние режимы оказываются менее продолжительными, чем зимние, также по рисунку заметно отличие в распределении продолжительностей режимов NAO- и sNAO- относительно остальных зимних и летних режимов в меньшем количестве событий низкой продолжительности и как следствие большей средней продолжительности. Самые длительные события по рисунку 2.2. отмечаются в диапазоне от 49 до 54 суток для зимних режимов NAO+ и AR, а для летних в диапазоне от 40 до 45 суток за исследуемый период 1940-2022 гг. дважды наблюдались события sSB и sNAO-.

Стоит упомянуть, что характерная продолжительность режимов зависит в том числе от способа обработки данных, а именно от применения или неприменения фильтрации по времени и выбора периода отсечки. В данной работе применялся фильтр Баттерворта низких частот с периодом отсечки в 10 суток, из-за которого общая дисперсия суточных полей аномалий $z500$ снижается примерно в 1,5 раза, а характерная продолжительность режимов увеличивается.

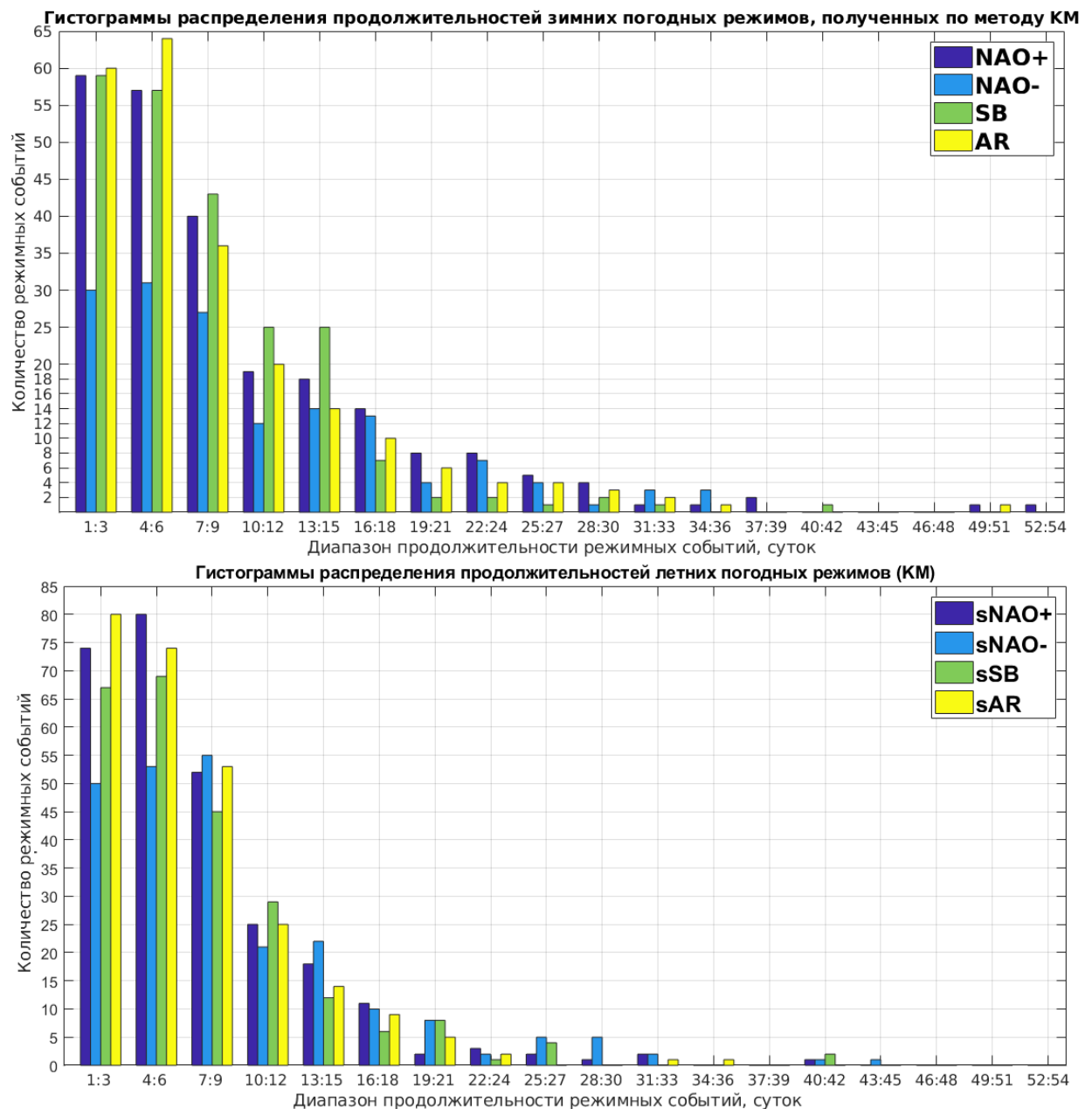


Рисунок 2.2. Гистограммы распределения продолжительностей зимних и летних погодных режимов, полученных по методу КМ.

4) Переходы между режимами в ЕАТ

Были рассчитаны матрицы переходов между режимами, а также статистически значимые переходы (см. раздел 2.1) с целью выяснить, какие переходы более или менее вероятны, что может иметь потенциальное применение в задачах предсказуемости. Вероятностные матрицы переходов режимов, а также статистически значимые из них, полученные для каждого из методов, представлены в таблице 2.3. Вероятностные матрицы получены

путем деления числа переходов из данного режима в другие на суммарное число переходов из данного режима (то есть сумма вдоль каждой строки в табл. 5 равна 1 без учета особенностей округления).

Таблица 2.3. Матрицы вероятностей переходов зимних и летних режимов в ЕАТ, полученных разными методами. Первый столбец – «Название режима» – указывает, из какого режима осуществляется переход ($X \rightarrow \dots$), последние четыре столбца – в какой режим осуществляется переход ($\dots \rightarrow Y$). Жирным шрифтом (курсивом) обозначены статистически значимые частые (редкие) переходы на уровне 5%.

Название режима	Метод кластеризации	Зимние режимы			
		NAO+	NAO-	SB	AR
NAO+	KM	-	<i>0.11</i>	0.48	0.41
	HW	-	0.36	0.48	<i>0.16</i>
	GM	-	0.12	<i>0.39</i>	0.49
NAO-	KM	0.33	-	0.33	0.34
	HW	0.35	-	<i>0.28</i>	0.37
	GM	0.52	-	0.48	<i>0.00</i>
SB	KM	<i>0.33</i>	0.33	-	0.35
	HW	0.42	<i>0.23</i>	-	0.35
	GM	<i>0.37</i>	0.25	-	0.37
AR	KM	0.51	0.22	<i>0.27</i>	-
	HW	<i>0.20</i>	0.36	0.44	-
	GM	0.41	<i>0.00</i>	0.59	-
Название режима	Метод кластеризации	Летние режимы			
		sNAO+	sNAO-	sSB	sAR
sNAO+	KM	-	<i>0.21</i>	0.45	0.34
	HW	-	<i>0.35</i>	0.27	0.38
	GM	-	<i>0.37</i>	0.26	0.37
sNAO-	KM	0.45	-	<i>0.20</i>	0.35
	HW	0.52	-	0.23	<i>0.26</i>
	GM	0.58	-	0.21	<i>0.21</i>
sSB	KM	<i>0.25</i>	0.35	-	0.40
	HW	<i>0.16</i>	0.45	-	0.39
	GM	<i>0.37</i>	0.37	-	0.26
sAR	KM	0.41	0.33	<i>0.27</i>	-
	HW	0.47	<i>0.27</i>	0.26	-
	GM	0.49	<i>0.25</i>	0.27	-

Статистические значимые переходы получены по методике, предложенной в (Vautard et al., 1990 [67]), то есть это такие переходы, которые происходят в 95% случаев чаще или реже, чем в матрицах, случайно сгенерированных путем перемешивания режимных событий с сохранением общего числа событий каждого режима. Так как переходы рассматривались между режимными событиями разной продолжительности, а не между отдельными сутками, то переходы режимов в самих себя (например, NAO+ $\rightarrow$ NAO+) при таком подходе отсутствуют.

Из таблицы 2.3 видно, что вероятность перехода из режима NAO+ в «противоположный» режим NAO- довольно низкая – 0.11 для режимов по методу KM и 0.12 по GM, однако 0.36 по HW. При этом вероятность перехода из NAO- в NAO+ в 3-4 раза выше, чем из NAO+ в NAO-, что может объясняться как свойствами атмосферной циркуляции в Евро-Атлантике, так и разницей в повторяемости этих режимов, так как более населенные режимы «перетягивают» переходы в свою сторону, однако NAO+ является более населенным только по методам KM и GM, из-за чего, вероятно, такой разницы в вероятности переходов не наблюдается для этих режимов, полученных методом HW (0.36 для перехода из NAO+ в NAO- и 0.35 для перехода из NAO- в NAO+), при этом режим NAO+ в HW заметно отличается от NAO+ по методам KM и GM (см. таблицу 1.2).

Аналогичный, но менее выраженный результат получился для летних режимов sNAO+ и sNAO- – вероятность перехода из sNAO- в sNAO+ в 1.5 – 2 раза выше (для каждого из методов), чем из sNAO+ в sNAO-, что, при схожести летних и зимних режимов NAO разных фаз может свидетельствовать о том, что разница в вероятности переходов между этими режимами объясняется не только разницей в их относительной повторяемости, так как в летние месяцы повторяемость sNAO- выше, чем sNAO+ для всех методов, кроме GM, о чем так же свидетельствует тот факт, что для всех методов переход из sNAO- в sNAO+ оказался статистически значимо частым, а переход из sNAO+ в sNAO- статистически значимо редким.

Интересным результатом оказалась нулевая вероятность перехода между режимами NAO- и AR при кластеризации методом GM, то есть за исследуемый период между полученными режимами такие переходы не наблюдались ни разу, однако стоит отметить, что режим AR для метода GM отличен от классического режима AR, полученного с помощью KM (см. рисунок 1.7 и таблицу 1.2). При других методах кластеризации между этими режимами наблюдаются либо незначимые, либо статистически значимо

частые (NAO- $\rightarrow$ AR в HW) вероятности переходов. Также для метода GM, в отличие от остальных, наблюдаются цепочки статистически значимых частых переходов, проходящие через каждый из режимов: «NAO+ $\rightarrow$ AR $\rightarrow$ SB $\rightarrow$ NAO- $\rightarrow$ NAO+ $\rightarrow$...» и, аналогично, «sNAO+ $\rightarrow$ sAR $\rightarrow$ sSB $\rightarrow$ sNAO- $\rightarrow$ sNAO+ $\rightarrow$...», то есть системы режимов, полученных методом GM для зимних и летних месяцев, со статистически значимо высокой вероятностью поочередно проходят через все свои состояния, несмотря на то, что летние режимы при методе GM относительно плохо воспроизвелись по сравнению с зимними (средняя корреляция полей аналогичных режимов 0,6), что может являться как простым совпадением, так и свойством метода GM.

Для других методов не наблюдается замкнутых цепочек статистически значимых переходов, проходящих через все режимы, так как не для каждого из режимов, полученных по методам KM и HW существуют статистически значимые переходы, например, их нет для режимов NAO- и sAR по методу KM и нет у режимов AR и sNAO+ по методу HW.

2.3. Идентификация погодных режимов в Северной Евразии

В этом разделе будут исследованы погодные режимы и их свойства в регионе, охватывающем Северную Евразию (СЕ), который менее изучен с точки зрения режимного подхода. Основное различие между режимами в СЕ и более известными режимами в ЕАТ заключается в том, что последние влияют на погоду в основном над Европой, в то время как здесь применяется аналогичный подход к более обширной континентальной территории Евразии. Погодные режимы в Северной Евразии идентифицированы путем применения метода кластерного анализа k-means к предварительно обработанным суточным полям z500 на основе данных ERA5, которые приблизительно описывают особенности атмосферной циркуляции в середине тропосферы. Поля z500 из реанализа ERA5 для идентификации режимов в Северной Евразии использовались отдельно как для зимнего, так и для летнего сезонов, чтобы учесть возможные различия в характеристиках погодных режимов в противоположные сезоны. Проанализированы основные статистические свойства полученных погодных режимов, такие как повторяемость и продолжительность, а также особенности переходов между режимами.

2.3.1. Предобработка данных

В качестве региона Северной Евразии выбран сектор от 0° до 180° восточной долготы и от 40° до 80° северной широты. В качестве переменной для идентификации режимов в СЕ использовались среднесуточные поля $z500$ с 1940 по 2022 год по данным из реанализа ERA5 [64]. Для зимы и лета выбраны месяцы декабрь–январь–февраль без 29 февраля и июнь–июль–август соответственно. Хотя в некоторых исследованиях погодных режимов используются расширенные определения вышеупомянутых сезонов, которые включают ноябрь/март и май/сентябрь, например [61], в данном случае рассматриваются сезоны продолжительностью 3 месяца, поскольку климат Северной Евразии континентальный, и сезонный цикл здесь сильнее, чем в Евро-Атлантике.

Перед кластеризацией полей $z500$ к ним применяется стандартная процедура предобработки, аналогично тому как это описано в разделе 1.4.1 настоящей диссертации. Во-первых, сезонный рассчитывается сезонный цикл путем усреднения полей $z500$ за одни и те же календарные даты для всех лет набора данных. Этот сезонный цикл сглаживается путем применения среднего значения за 5 дней (как в [23]), затем посчитанный сглаженный сезонный цикл вычитается из оригинального временного ряда $z500$ для получения полей аномалий $z500$. Во-вторых, к полям $z500$ применяется временная фильтрация нижних частот с периодом отсечки 10 дней, чтобы удалить из данных высокочастотную синоптическую изменчивость, для этого используется фильтр Баттерворта [65] первого порядка.

После получения фильтрованных полей аномалий $z500$ (здесь и далее по тексту фильтрованные поля аномалий $z500$ будут обозначаться « $dz500$ »), временной ряд $dz500$ разделяется на зимний и летний месяцы, и к зимнему и летнему временным рядам $dz500$ по отдельности применяется разложение на эмпирические ортогональные функции (ЭОФ). ЭОФ-разложение применяется к зимним и летним полям $dz500$, взвешенным на $\sqrt{\cos(lat)}$, чтобы учесть разницу площадей в узлах сетки $dz500$ на разных широтах, вносящих вклад в общую дисперсию полей $dz500$. Для анализа погодных режимов в СЕ было решено оставить 8 ведущих ЭОФ для зимних полей $dz500$ и 12 ЭОФ для летних полей, так чтобы каждый набор из зимних и летних ведущих ЭОФ объяснял $> 80\%$ дисперсии $dz500$ для своего сезона (8 ведущих зимних ЭОФ объясняют 82% дисперсии для зимних полей $dz500$ и 12 летних ЭОФ объясняют 81% дисперсии для летних полей $dz500$). В литературе используется разное количество ведущих ЭОФ для кластеризации погодных режимов, от 4 [27] до 20 [61]. Обычно утверждается, что большее количество

ЭОФ не сильно изменяет результаты кластеризации по сравнению с меньшим количеством [24]. Для расчета ЭОФ и главных компонент (РС) использовалась матрица ковариаций, в результате чего полученные ЭОФ и РС отсортированы по доле объясненной дисперсии полей dz500. После получения временных рядов главных компонент зимних и летних полей dz500, к ним, то есть к РС их ЭОФ-разложения, применяется алгоритм кластеризации k-means. Перед применением k-means, значения главных компонент не нормируются, чтобы каждая из оставленных РС вносила вклад в квадрат евклидова расстояния (формула 1.8) между объектами в пространстве главных компонент (представляющими суточные поля dz500) пропорционально дисперсии, которую они объясняют.

2.3.2. Выбор числа режимов

Алгоритм k-means требует, чтобы исследователь заранее знал количество K кластеров, которые он хотел бы получить. Выбор оптимального количества погодных режимов в исследуемой области является сложной задачей, и несколько исследований погодных режимов были сосредоточены именно на этой проблеме [21, 24]. Наиболее распространенное число режимов для наиболее изученного региона, Евро-Атлантики, обычно составляет четыре [23, 43, 56], но в некоторых недавних исследованиях выбирается шесть [24] или более [9] режимов. Существуют также исследования погодных режимов в Северной Америке, где выбирается четыре режима [25, 70].

Чтобы оценить оптимальное число режимов в Северной Евразии, было использовано два подхода: 1) индексы классифицируемости и воспроизводимости, описанные в [43] и разделе 1.3 настоящей диссертации, и 2) критерий отношения дисперсии (VR, Variance Ratio) Калински–Харабаша [59], который также использовался, например, в [60]. Основное различие между этими двумя подходами заключается в том, что индексы классифицируемости и воспроизводимости используют в качестве метрики для выбора оптимального K пространственную корреляцию, которая учитывает различия в пространственной структуре различных режимов, но не в их амплитуде (т. е. величине аномалий dz500). Критерий VR рассчитывается с использованием квадрата Евклидова расстояния, которое учитывает как форму (пространственную структуру), так и амплитуду полей dz500 в различных режимах. Для определения K значения VR для оригинальной выборки данных сравниваются со значениями VR, рассчитанными для

сгенерированной выборки на основе красного (броуновского) шума, имеющего одинаковое среднее, СКО и автокорреляцию lag-1, как и у оригинальной выборки. Результаты двух подходов для потенциальных значений $K = 2:10$ для обоих сезонов представлены на рисунке 2.3.

Индексы классифицируемости и воспроизводимости для зимних и летних режимов в СЕ, полученные на основе кластеризации временных рядов главных компонент полей dz500, показывают, что оптимальное число погодных режимов для зимнего сезона составляет 4. Для лета наблюдаются высокие значения индекса воспроизводимости при $K = 3$ и незначимый пик индекса классифицируемости при $K = 7$, при этом значения индексов резко убывают при $K > 4$. Уровни значимости индексов VR (рисунки 2.3 в, г) получены путем сравнения актуальных значений VR со значениями, посчитанными на основе 500 случайно сгенерированных выборок на основе красного (броуновского) шума с одинаковой длиной и СКО, как и у оригинальной выборки. Индекс VR также указывает, что для зимы оптимальное $K = 4$, при этом ни для какого K не получено статистически значимых значений VR в летний сезон, для которого, судя по рисунку 2.3 г, случайно сгенерированный красный шум кластеризуется лучше, чем оригинальная выборка данных. С учетом результатов тестов на оптимальное K , и принимая во внимание что $K = 4$ так же является наиболее частым выбором для идентификации погодных режимов в Евро-Атлантическом и Североамериканском регионах, автор выбрал $K = 4$ как для зимнего, так и для летнего сезонов. Использование $K = 4$ для летнего сезона полезно возможностью сопоставления характеристик погодных режимов в разные сезоны.

Для получения финальных результатов кластеризации с выбранным $K=4$ была использована модификация k-means simulated annealing (метод «имитации отжига») с 1000 случайных инициализаций кластеров для каждого сезона. Для итогового анализа были отобраны лучшие решения с минимальным внутрикластерным расстоянием V (формула 1.10).

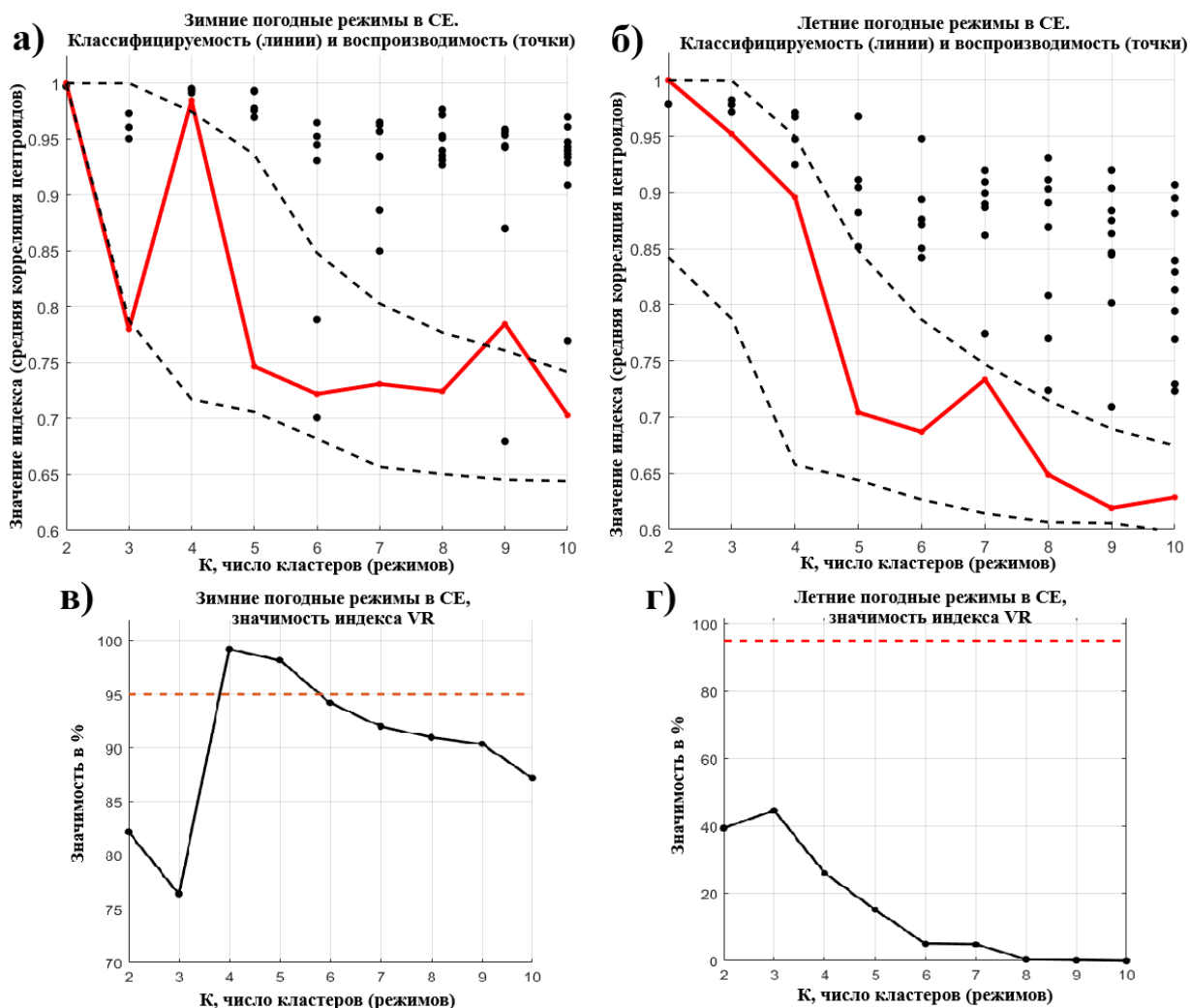


Рисунок 2.3. Индексы классифицируемости и воспроизводимости для кластеризации k-means зимних (а) и летних (б) погодных режимов в СЕ. Красные линии представляют значения индекса классифицируемости оригинальной выборки, черные пунктирные линии представляют 90-й и 10-й процентиля индексов классифицируемости сгенерированных выборок. Черными точками обозначены значения индекса воспроизводимости для каждого $k=1:K$ из кластеров для выбранного K (чем выше точки, тем лучше); На рисунках (в) и (г) представлены значения индекса VR для оригинальной выборки относительно ранжированных индексов VR, полученных по сгенерированным выборкам. Красные пунктирные линии означают 95-ый перцентиль значимости. Значимость указывает на долю случайно сгенерированных выборок, значения VR которых оказались ниже, чем значение для оригинальной выборки.

2.3.3. Анализ временных характеристик режимов в Северной Евразии

В результате кластеризации главных компонент ЭОФ-разложения полей $dz500$ методом k-means мы получаем временной ряд $idx(t)$ со значениями от 1 до 4, указывающими на номера кластеров, к которым были отнесены суточные поля $dz500(lon, lat, t)$. Стоит отметить, что главные компоненты и ЭОФ-разложение использовались только для кластеризации и, после получения временного ряда номеров сменяющих друг друга во времени кластеров $idx(t)$, автор работал с полями $dz500(lon, lat, t)$ (например, для отрисовки полей режимов, как на рисунке 1.7), не разложенными на ЭОФ и не взвешенными по пространству на $\sqrt{\cos(lat)}$. Алгоритм k-means относится к категории hard clustering, то есть каждое суточное поле $dz500(lon, lat, t)$ приписывается одному и только одному кластеру, которое ассоциируется со своим собственным погодным режимом, поэтому в результате кластеризации мы имеем по 4 погодных режима для зимнего и летнего сезона. Переменная t в $idx(t)$ включает все дни одного и того же сезона (зимнего или летнего) за весь исследуемый период 1940 – 2022 гг. Временной ряд $idx(t)$ для зимних режимов в СЕ имеет длину 7470, то есть 83 зимы по 90 дней, включая 82 полных зимы и январь и февраль 1940 года и декабрь 2022 года, а временной ряд $idx(t)$ для летних погодных режимов имеет длину 7636, что есть 83 полных летних сезона по 92 суток в каждом. В качестве зимы года $YYYY$ (например, 1950) автор имеет в виду последовательно идущие декабрь предыдущего года $YYYY-1$ (1949) и январь и февраль года $YYYY$ (1950). Два неполных сезона: 1) январь и декабрь 1940 года и 2) декабрь 2022 были исключены при расчете временных рядов сезонной повторяемости, гистограмм продолжительности и матриц переходов, то есть за исследуемый период 1940-2022 гг. имелось лишь 82 полных зимы и 83 полных лета.

Для изучения предсказуемости режимов в СЕ, анализировались их переходы в форме стохастической матрицы TM . Для её подсчета временные ряды зимних и летних $idx(t)$ разделяются по годам на $idx(t_s)$ и отдельно для каждого года s ($s = 1940, 1941, \dots, 2022$) считается матрица TM_s путем подсчета всех переходов $idx(t_s) \Rightarrow idx(t_s + 1)$, где t_s и $t_s + 1$ представляют сутки внутри одного и того же зимнего (или летнего) сезона одного и того же года (например, зимы 1950). Такой подсчет происходит для всех лет исследуемого периода 1940 – 2022 гг. Так, за 82 полных зимних периода набирается 7298 переходов (89 переходов за сезон), а для 83 летних периодов набирается 7553 переходов (по 91 переход за сезон). Например, если $idx(t_s) = i$, а $idx(t_s + 1) = j$,

1) = j , то соответствующий матричный элемент $TM_s(i, j)$ получает +1 к общему числу таких переходов. Далее, все TM_s за все года s складываются в общую матрицу $TM(i, j)$, и все элементы каждой её строки i делятся на сумму строки, чтобы каждый отдельный матричный элемент представлял собой вероятность, а сумма вероятностей перехода из режима i в любой другой j (от 1 до 4) равнялась 1.

Матричные элементы $TM(i, j)$ представляют вероятности перехода из режима i в режим j на следующие сутки, так что их размерность это сутки⁻¹, а диагональные элементы $TM(i, i)$ указывают на вероятность системы остаться в одном и том же режиме на следующей день. Тогда, вероятность выйти из режима i в любой другой режим $j \neq i$ на следующий день равна $1 - TM(i, i)$, а характерную продолжительность режима в сутках можно оценить по формуле:

$$Pr(i) = (1 - TM(i, i))^{-1} = \frac{1}{\sum_{j \neq i} TM(i, j)} \quad (2.1)$$

Помимо матриц переходов, анализировались продолжительности погодных режимов в СЕ. Для этого наборы из 82 зимних и 83 летних временных рядов $idx(t_s)$ трансформировались в ряды *режимных событий* ***idxe(s)***, указывающих на текущий номер погодного режима, и ряды продолжительностей режимов ***pers(s)***, указывающих на количество дней, проведенных системой в данном режиме. Например, $idx(t_s) = \langle \dots 3, 3, 3, 3, 3, 1, 1, 4, 4, 4, 2, 2, 2, 2 \dots \rangle$ трансформируется в ряд режимных событий ***idxe(s) = \langle \dots 3, 1, 4, 2 \dots \rangle*** и ряд продолжительностей ***pers(s) = \langle \dots, 5, 2, 3, 4, \dots \rangle***, где s указывает на сезон года, например, лето 1960. Стоит отметить, что сумма значений $pers(s)$ для любого зимнего сезона s равна 90, а для летнего 92. Существенная разница между $idx(t_s)$ и ***idxe(s)*** в том, что соседние числа во втором случае всегда разные, что отражает тот факт, что один погодный режим всегда сменяется другим. Использование ***idxe(s)*** и ***pers(s)*** позволяет посчитать общее и сезонное число режимных событий для анализа их распределения по продолжительностям.

2.4. Пространственные структуры погодных режимов в Северной Евразии

Результаты кластеризации временных рядов главных компонент полей dz_{500} при $K = 4$ представлены на рисунке 2.4. Усредненные поля dz_{500} для зимних и летних погодных режимов на рисунке 2.4 представлены в порядке убывания их относительной повторяемости.

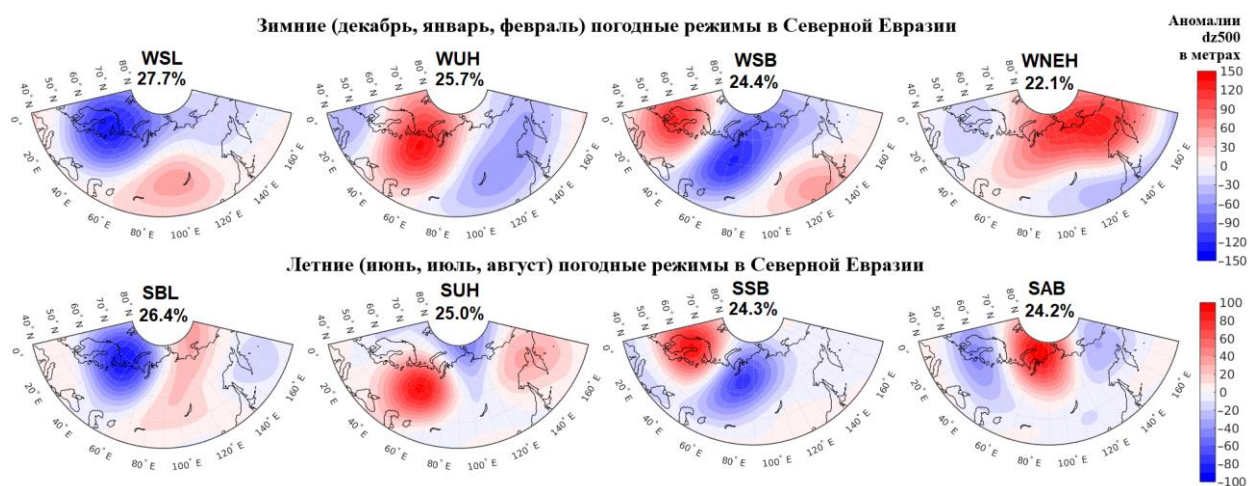


Рисунок 2.4. Средние поля dz_{500} четырех зимних и летних погодных режимов в Северной Евразии. На верхней строке изображены зимние погодные режимы, на нижней – летние. Подписи к режимам указывают на их названия и относительную повторяемость в процентах за весь исследованный период 1940 – 2022 гг. Шкалы аномалий dz_{500} , соответствующих погодным режимам, представлены справа.

Самый частый зимний погодный режим (27.7% суток) назван WSL (Winter Scandinavian Low, Зимний Скандинавский Минимум), потому что его главной особенностью является отрицательная аномалия dz_{500} над Скандинавией, такая структура аномалий dz_{500} должна продлевать и усиливать западный перенос воздушных масс из Европы на территорию России. Второй по частоте зимний погодный режим (25.7% суток) назван WUH (Winter Ural High, Зимний Уральский Максимум), выделяющийся положительной аномалией dz_{500} над Уральскими горами. Третий по частоте режим (24.4% суток) назван WSB (Winter Scandinavian Blocking, Зимний Скандинавский Блокинг), его особенностью является положительная аномалия dz_{500} над Скандинавией, сопровождаемая отрицательной аномалией на востоке от Уральских гор. Режим с положительной аномалией dz_{500} над Скандинавией так же часто идентифицируется в исследованиях погодных режимов в Евро-Атлантике, например, в работах [19, 23, 39] или см.

режим SB на рисунке 1.7 среди режимов, идентифицированных автором настоящей диссертации для региона ЕАТ. Наименее редкий зимний режим (22.1% суток) назван WNEH (Winter North Eastern High, Зимний Северо-Восточный Максимум), так как он характеризуется положительной аномалией dz500 над Северо-Восточной частью Сибири.

Наиболее частым летним погодным режимом (26.4% суток) является SBL (Summer Barents Low, Летний Баренцев Минимум), его особенностью является отрицательная аномалия высоты геопотенциальной поверхности над Баренцевым морем. Хотя SBL выглядит похожим на WSL и должен оказывать аналогичное действие, усиливая западный перенос на европейской территории России, центр отрицательной аномалии dz500 в нём смещен восточнее в сторону Баренцева моря по сравнению с WSL, у которого центр аномалии находится над Кольским полуостровом. Второй и третий по частоте летние погодные режимы (25.0% и 24.3%) названы SUH (Summer Ural High, Летний Уральский Максимум) и SSB (Summer Scandinavian Blocking, Летний Скандинавский Блокинг) соответственно. Эти режимы имеют структуру, похожую на структуру их зимних аналогов, WUH и WSB соответственно, поэтому имеют аналогичное антициклоническое действие, препятствуя западному переносу на территории России. Четвертый летний режим (24.2% суток) назван SAB (Summer Arctic Blocking, Летний Арктический Блокинг), его основной структурой является положительная аномалия высоты геопотенциальной поверхности над Карским морем, сопровождаемая более слабыми отрицательными аномалиями к юго-западу и юго-востоку. Режим SAB имеет пространственную структуру, похожую на «омега-блокирование».

Средняя абсолютная величина аномалий dz500 летних погодных режимов в 1.5-2 раза ниже (см. шкалы на рисунке 2.4), чем у их зимних аналогов (то же самое наблюдается для режимов в ЕАТ, см. рисунок 1.6), что по большей части объясняется меньшей изменчивостью полей z500 в летние месяцы. Средняя по площади абсолютная величина аномалий dz500 в регионе СЕ за 1940-2022 гг. в зимние месяцы составляет 79 метров, 66 метров в среднем за год и 51 метр в летние месяцы.

2.5. Временные характеристики погодных режимов в Северной Евразии

При сравнении относительной повторяемости погодных режимов в Северной Евразии, можно заметить, что кластеризация k-means

идентифицировала режимы с примерно равной относительной повторяемостью, с большей разницей в повторяемости у зимних режимов (27.7% у WSL против 22.1% у WNEH), чем у летних (26.4% у SBL против 24.2% у SAB). Это различие в зимних и летних повторяемостях так же свидетельствует о том, что режимы «лучше» кластеризуются в зимние месяцы, чем в летние, так как k-means склонен к созданию равных по величине кластеров в тех случаях, когда данные распределены равномерно и не имеют отличимых областей высокой плотности, по сути кластеров, в фазовом пространстве кластеризуемого набора данных (в данном случае, в пространстве главных компонент ЭОФ-разложения полей dz500).

Сезонная повторяемость погодных режимов в СЕ была проверена на наличие статистически значимых трендов с помощью теста Манн-Кендалла [71], результаты представлены в таблице 2.4. На рисунке 2.5. представлены временные ряды сезонной повторяемости режимов в СЕ для обоих исследованных сезонов.

Таблица 2.4. Статистические свойства и линейные тренды сезонной повторяемости погодных режимов в СЕ. Статистически значимые на уровне 0.05 тренды и их характеристики выделены полужирным шрифтом.

Статистика и линейные тренды сезонной повторяемости погодных режимов в СЕ								
Название режима	WSL	WUH	WSB	WNEH	SBL	SUH	SSB	SAB
Среднее значение (суток/сезон)	25.0	23.1	22.2	19.7	24.3	23.0	22.4	22.3
Коэффициент вариации, СКО/среднее	0.494	0.590	0.611	0.692	0.470	0.644	0.560	0.549
Линейный тренд, суток в год	-0.040	+0.004	-0.004	+0.041	-0.016	+0.243	-0.118	-0.109
Коэффициент детерминации, R^2	0.006	< 0.001	< 0.001	0.005	0.001	0.156	0.052	0.046
Пройден ли тест Манн-Кендалла значимости тренда	Нет	Нет	Нет	Нет	Нет	Да	Да	Да
Значимость тренда, p-уровень	0.428	0.904	0.826	0.767	0.937	0.002	0.047	0.044

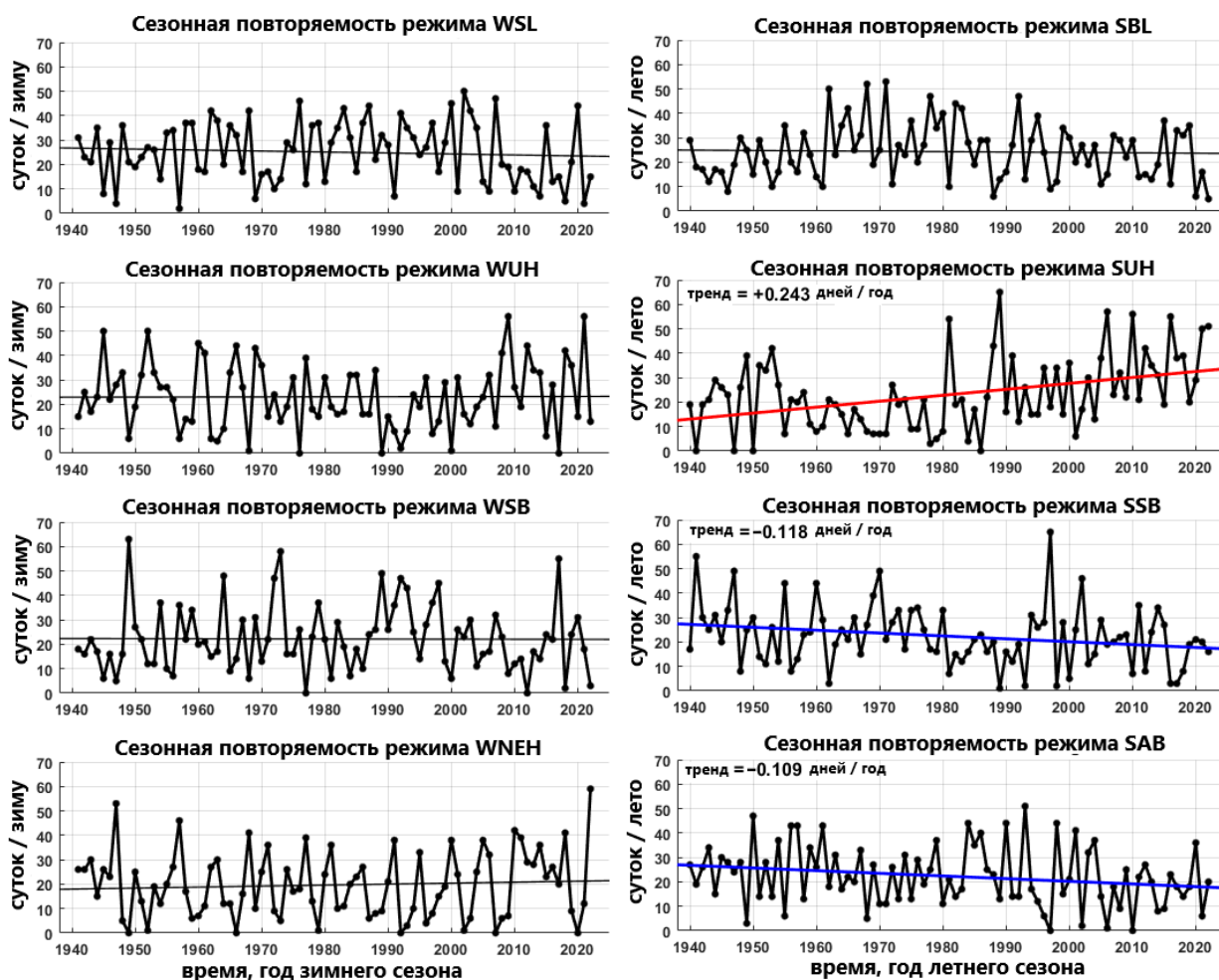


Рисунок 2.5. Временные ряды сезонной повторяемости погодных режимов в СЕ (в сутках за сезон). Статистически значимые на уровне 5% линейные тренды согласно тесту Манн-Кендалла показаны цветными линиями, красным положительный тренд (SUN) и синим отрицательные тренды (SSB and SAB); незначимые линейные тренды показаны черными линиями.

Как показано в таблице 2.4 и на рисунке 2.5, для зимы нет статистически значимых трендов сезонной повторяемости режимов, в то время как для трёх из четырех летних погодных режимов наблюдаются тренды, статистически значимые на уровне 5%, несмотря на высокую межгодовую изменчивость сезонной повторяемости (см. коэффициент вариации в таблице 2.4), характерную как для зимних, так и для летних погодных режимов. Так как сумма сезонной повторяемости всех режимов для каждого сезона постоянная (90 для зимы и 92 для лета), положительный тренд SUN (Уральского максимума) компенсируется отрицательными трендами режимов SSB (Скандинавский блокинг) и SAB (Арктический блокинг). Режим SUN соответствует блокирующей активности над европейской территорией России, связанной с волнами жары, например, рекордной жарой 2010 года

[72], сопровождавшейся блокингом. Для 2010 года наблюдается локальный пик режима уральского максимума SUN, при этом абсолютный максимум приходится на лето 1989 года.

Так же были проанализированы характерные продолжительности погодных режимов в СЕ путем анализа функций $idxe(s)$ и $pers(e)$, которые описаны в части 2.3.3. Средние значения продолжительностей режимов в СЕ зимой составила 8.3, 8.2, 8.2, 8.2 суток для режимов WSL, WUH, WSB, WNEH соответственно, в то время как для летних режимов SBL, SUN, SSB, SAB продолжительности составили 7.8, 8.3, 7.8, 8.5 соответственно. Ниже на рисунке 2.6 представлены гистограммы распределения событий погодных режимов $idxe(s)$ по их продолжительности $pers(s)$.

Из рисунка 2.6 можно отметить различие в распределении продолжительностей между зимними и летними погодными режимами, например, различие в числе режимных событий короткой продолжительности – зимой примерно одинаковое число режимных событий продолжительностью 1-3 суток и 4-6 суток (по ~50 событий для каждого режима), в то время как летом событий продолжительностью 4-6 суток больше, чем коротких событий продолжительностью 1-3 суток. Как для зимы, так и для лета можно отметить наличие редких режимных событий с очень большой продолжительностью (>30 суток), причем зимой такие события наблюдаются преимущественно для режима скандинавского блокирования WSB, а летом для режима уральского максимума SUN.

Несмотря на применение фильтрации нижних частот перед кластеризацией полей $z500$, никаких дополнительных ограничений на продолжительность режимов не применялось, поэтому среди множества режимных событий присутствуют продолжительности в том числе менее 10 суток, такие события наблюдаются и зимой, и летом. При этом режимные события продолжительностью 10 суток и более занимают в среднем 60% времени каждого сезона за период 1940-2022 гг., в то время как события продолжительностью <10 суток занимают оставшиеся, в среднем, 40% времени, однако этот процент варьируется – 33.6% общей продолжительности для режима SAB до 47.1% общей продолжительности для режима SBL (см. рисунок 2.6 г, колонки режима SBL для диапазонов продолжительности 4-6 суток и 7-9 суток выше, чем у остальных летних режимов). Короткие режимные события продолжительностью 1-3 суток, по сути переходные между более длительными событиями, занимают около 6% общего числа суток как для зимних, так и для летних сезонов.

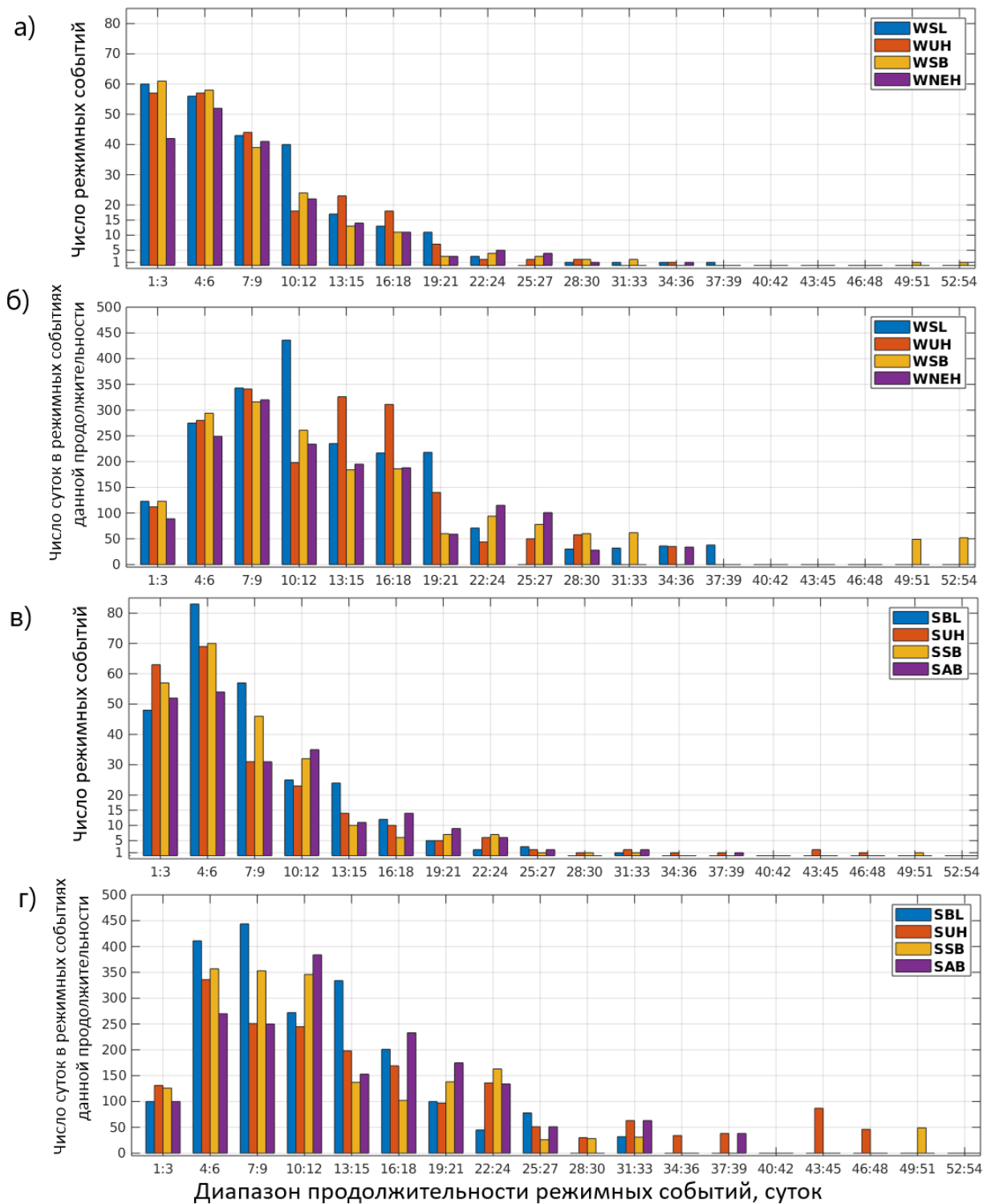


Рисунок 2.6. Гистограммы распределения режимных событий по их продолжительности для зимы (а, б) и лета (в, г) за исследуемый период 1940-2022 гг. (а), (в) – общее число режимных событий в данном диапазоне продолжительности (ось X). (б), (г) – общее число суток, проведенное системой в режимных событиях в данном диапазоне продолжительности.

Матрицы вероятностей переходов для зимних и летних погодных режимов в СЕ представлены в таблице 2.5. Используя уравнение (2.1) (см. раздел 2.3.3), можно получить оценки продолжительности режимов – 9.0, 9.0, 8.9 и 8.9 для зимних WSL, WUH, WSB и WNEH соответственно, и 8.5, 9.0, 8.3 и 9.2 для летних SBL, SUH, SSB и SAB, соответственно. Эти оценки примерно на 1 суток продолжительности больше, чем фактические средние значения продолжительностей, полученные на основе анализа рядов $idxe(s)$ и $pers(e)$ и приведенные на 3 абзаца выше. Можно отметить, что характерные продолжительности между летними режимами варьируются сильнее, чем между зимними, это можно отметить как по рисунку 2.6, так и по средним продолжительностям, полученным как по фактическим временным рядам, так и по оценкам по формуле (2.1).

Таблица 2.5. Матрицы суточных вероятностей переходов между режимами для зимнего и летнего сезонов. Значения, указанные в скобках для недиагональных элементов, представляют собой вероятности перехода к другому режиму, когда диагональные вероятности не учитываются, т.е. вероятность выбора данного режима при условии смены режима на следующие сутки. Статистически значимые частые переходы выделены жирным шрифтом.

Матрица переходов для зимних погодных режимов в СЕ				
	... => WSL	... => WUH	... => WSB	... => WNEH
WSL => ...	88.9%	3.0% (27.4%)	4.4% (39.8%)	3.6% (32.7%)
WUH => ...	2.8% (25.1%)	88.9%	4.3% (38.6%)	4.0% (36.2%)
WSB => ...	6.6% (58.7%)	3.1% (27.4%)	88.8%	1.6% (13.9%)
WNEH => ...	3.3% (29.4%)	5.8% (51.1%)	2.2% (19.4%)	88.7%
Матрица переходов для летних погодных режимов в СЕ				
	... => SBL	... => SUH	... => SSB	... => SAB
SBL => ...	88.2%	3.6% (30.3%)	2.2% (18.8%)	6.0% (50.9%)
SUH => ...	2.6% (23.9%)	88.9%	5.5% (49.8%)	2.9% (26.3)
SSB => ...	7.4% (61.7%)	3.1% (25.7%)	87.9%	1.5% (12.6%)
SAB => ...	2.8% (26.1%)	4.5% (41.2%)	3.5% (32.7%)	89.1%

Анализируя временные ряды режимных событий $idxe(s)$ (см. раздел 2.3.3), можно получить матрицу переходов, которая выглядит так же, как если считать её по суточным переходам $idx(t_s)$ за тем лишь исключением, что диагональные элементы в случае $idxe(s)$ равны нулю, так как один режим обязательно сменяется другим. Для исследования того, являются ли переходы между режимами чисто случайными или переходят друг в друга в некотором определенном порядке, актуальные ряды режимных событий $idxe(s)$ сравнивались со случайно сгенерированными рядами режимных событий,

полученных случайным перемешиваем оригинального ряда. Это перемешивание производилось, следуя следующим правилам: 1) длина сгенерированного ряда равна длине оригинального; 2) количество событий каждого из режимов в сгенерированном ряду равно количеству событий в оригинальном ряду; 3) каждое режимное событие в сгенерированном ряду (например, WSL) обязательно должно сменяться другим режимом (WUH, WSB, WNEH, но не опять WSL), а не тем же самым. Используя сгенерированные ряды режимных событий, по ним можно, аналогично оригинальному ряду, считать матрицы переходов. Те переходы между режимами, которые в оригинальном временном ряду режимных событий происходили чаще, чем в 95% сгенерированных матрицах перехода, называются статистически значимыми частыми переходами [67] на уровне 5%.

На рисунке 2.7 представлена схема таких статистически значимых частых переходов для погодных режимов в Северной Евразии. Стоит отметить, что значимо частые переходы не обязательно являются наиболее вероятными, так как каждый режим имеет разную относительную повторяемость режимных событий. Из рисунка 2.7 можно отметить интересную последовательность, характерную как для зимы, так и для лета – статистически значимые частые переходы образуют цепь переходов, поочередно проходящую через все режимы, «WSL => WNEH => WUH => WSB => ...» для зимы и «SBL => SAB => SUH => SSB => ...» для лета. Таким образом, система погодных режимов в Северной Евразии стремится пройти через все свои возможные состояния в определенном порядке. Для зимы так же отмечается цепь статистически значимых частых переходов из двух режимов – «WUH <=> WNEH».

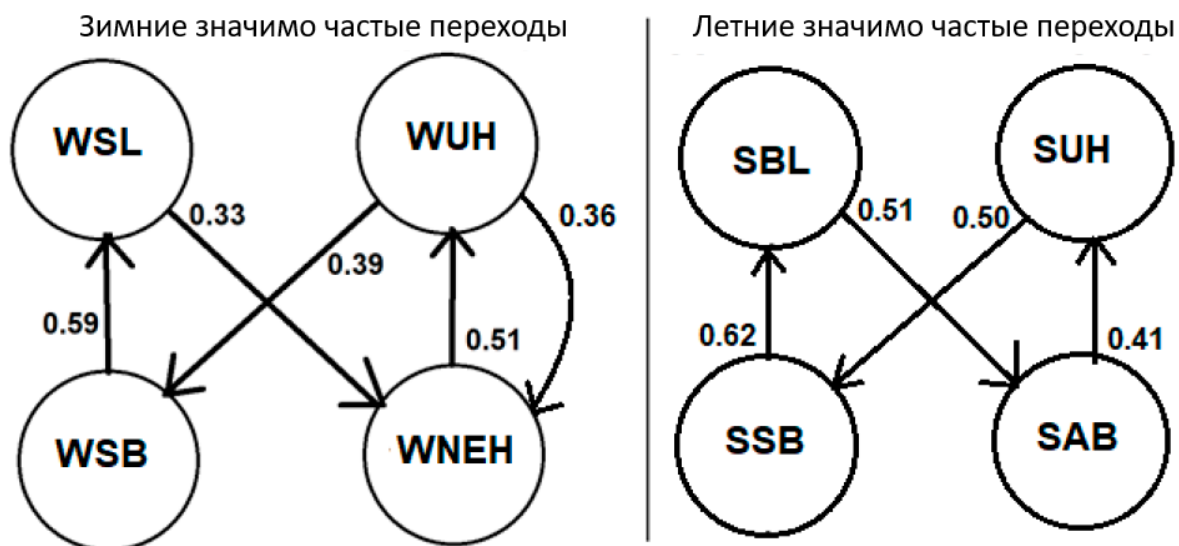


Рисунок 2.7. Схематическое изображение статистически значимых частых переходов (на уровне значимости 5%) между погодными режимами в Северной Евразии. Цифрами возле стрелок обозначены вероятности этих переходов (те же цифры, что в таблице 2.5 в скобках).

2.6. Выводы

В Главе 2 представлены характеристики четырех погодных режимов в Евро-Атлантике – NAO+, NAO-, SB, AR за 1940-2022 гг., такие как полная и сезонная повторяемости, продолжительности и вероятностей матрицы переходов. Пространственные характеристики этих режимов, то есть средние поля $z500$, приведены в Главе 1.

Для режимов в ЕАТ, полученных автором методом k-means, показано качественное совпадение с повторяемостью четырех режимов в ЕАТ, полученных другими авторами – наибольшая относительная повторяемость режима «NAO+» (30%) и наименьшая повторяемость режима «NAO-» (21%) среди четырех режимов. Показаны временные ряды и тренды повторяемости режимов «NAO+» и «NAO-». Показано, что, несмотря на наименьшую относительную повторяемость «NAO-», его средняя и медианная продолжительности, то есть число суток, которое система пребывает в данном режиме до перехода в другой режим, является наибольшей среди четырех режимов. Для режимов в ЕАТ посчитаны матрицы вероятностей перехода и показано что существуют значимо частые переходы, то есть переходы из одного режима в ЕАТ в другой с большей вероятностью происходят по определенной траектории, что важно для их предсказуемости.

Для области Северной Евразии за 1940-2022 гг. впервые получены погодные режимы по методике, аналогично той, которая часто используется в литературе для Евро-Атлантики. Показано, что оптимальное число режимов в зимние месяцы равно четырем. Аналогично Евро-Атлантике, в зимние месяцы поля $z500$ в Северной Евразии зимой кластеризуются лучше, чем летом, а средние поля режимов (центроиды кластеров) получаются более выраженными по величине аномалий.

Зимой в Северной Евразии выделено четыре режима – скандинавский минимум WSL, уральский максимум WUH, скандинавский блокинг WSB и северо-восточный максимум WNEH. Для лета так же выделено четыре погодных режима – летний баренцевский минимум SBL, уральский максимум SUH, скандинавский блокинг SBL, арктический блокинг SAB.

Из четырех зимних и летних погодных режимов, зимние режимы не имеют долгосрочных трендов сезонной повторяемости за исследованный период 1940-2022 гг., а для трёх из четырех летних режимов обнаружены значимые (p -уровень = 5%) линейные тренды – положительный тренд режима уральского максимума SUH (+0.24 суток/сезон в год) и отрицательные тренды повторяемости режимов скандинавского и арктического блокинга (SSB и SAB).

Проведен анализ распределения продолжительностей режимных событий в Северной Евразии. Средняя продолжительность погодных режимов в Северной Евразии составила 8-9 суток, с большей изменчивостью от режима к режиму в летний сезон, при этом показано, что большая часть времени (60%) за исследованный период 1940-2022 гг. приходится на режимные события продолжительностью от 10 суток и более, при этом отдельные эпизоды погодных режимов имеют продолжительность от 30 вплоть до 52 суток.

Анализ переходов между режимами выявил наличие статистически значимо частых последовательностей переходов, формирующих замкнутые цепочки, проходящие через всю систему из четырех режимов для зимы и лета. Показано, что в зимний период система режимов стремится сменяться в порядке «WSL => WNEH => WUH => WSB => ...», а летом в порядке «SBL => SAB => SUH => SSB => ...».

ГЛАВА 3. ВОСПРОИЗВЕДЕНИЕ ЕВРО-АТЛАНТИЧЕСКИХ ПОГОДНЫХ РЕЖИМОВ В МОДЕЛЯХ КЛИМАТА

Режимный подход может использоваться в качестве индикатора способности климатических моделей и моделей прогноза погоды воспроизводить особенности крупномасштабных атмосферных течений [73]. Например, в работе [23] рассматривается влияние пространственного разрешения на воспроизведение четырех традиционных погодных режимов в Евро-Атлантике (NAO+, NAO-, Scandinavian Blocking, Atlantic Ridge). Показано, что при использовании более высокого разрешения модель среднесрочного прогноза погоды IFS ECMWF лучше воспроизводит погодные режимы в Евро-Атлантическом регионе относительно данных реанализа, чем при использовании более грубого разрешения, типичного для климатических моделей.

В отличие от зимнего сезона, для летних месяцев режимный подход при исследовании атмосферной циркуляции в Евро-Атлантическом регионе применяется реже, работ по определению погодных летних погодных режимов методами кластерного анализа существенно меньше (хотя они есть, например, [62]), вероятно, в связи с меньшими суточными аномалиями полей давления/геопотенциала и, как следствие, худшей их кластеризуемостью, то есть меньшей долей дисперсии полей $z500$ и SLP, объясняемой кластеризацией. В связи с этим, важно проверить воспроизведение в моделях не только традиционных зимних погодных режимов в Евро-Атлантике, но и их летних аналогов.

В данной главе будет приведено количественное сравнение успешности 8 климатических моделей, участвовавших в проекте CMIP6, включая модель ИВМ РАН INM-CM5-0 [74], воспроизводить в экспериментах historical характеристики традиционных зимних погодных режимов в Евро-Атлантике (NAO+, NAO-, SB, AR), получаемых кластеризацией полей $z500$ методом k-means, а также их летних аналогов (sNAO+, sNAO-, sSB, sAR), а именно средние поля $z500$ этих режимов, их повторяемости, продолжительности и матрицы вероятностей переходов.

3.1. Данные и методы

Для воспроизведения погодных режимов в Евро-Атлантике использовались данные z500 из реанализа ERA5 и по результатам экспериментов historical (r1i1p1f1) восьми климатических моделей, участвовавших в проекте CMIP6 – 1) INM-CM5-0 (ИВМ РАН, Россия), 2) CESM2 (NCAR, США), 3) GFDL-CM4 (NOAA, США), 4) EC-Earth3 (Европа), 5) MIROC6 (JAMSTEC, Япония), 6) NorESM2-LM (NCC, Норвегия), 7) CanESM5 (CCCMA, Канада), 8) ACCESS-CM2 (CSIRO, Австралия). Эти 8 моделей были выбраны по признаку наилучшей доступности их суточных данных z500 по результатам экспериментов historical.

Для возможности сопоставления результатов кластеризации модельных данных и данных реанализа использовался одинаковый временной интервал – зимние (декабрь, январь, февраль без 29-го февраля) и летние (июнь, июль, август) месяцы с 1950 по 2014 гг. – и одна и та же пространственная область – регион Евро-Атлантики, от 80 гр. з.д. до 40 гр. в.д. и от 30 гр. с.ш. до 80 гр. с.ш. Поля z500 реанализа ERA5 и всех моделей для удобства были предварительно интерполированы на одинаковую прямоугольную сетку с шагом в 1 градус. Перед кластеризацией из оригинальных полей z500 были получены поля аномалий, к полям аномалий z500 применена временная низкочастотная фильтрация, далее отфильтрованные поля аномалий были разложены на эмпирические ортогональные функции (ЭОФ).

Для получения полей аномалий вычитался сезонный ход z500, который рассчитывался путем усреднения полей z500 за одинаковые календарные даты разных лет и последующего сглаживания этого хода скользящим средним с окном в 5 суток. После вычитания сезонного хода к полям аномалий применялась временная низкочастотная фильтрация (low-pass) с периодом отсечения в 10 суток для удаления низкочастотной синоптической изменчивости полей z500, не относящейся к длительным режимным событиям, для низкочастотной фильтрации применялся фильтр Баттерворта. Для пропорционального учета вклада дисперсии z500 на ячейках сетки разных широт, поля аномалий z500 (далее dz500) нормировались на квадратный корень из косинуса широты. Далее, уже нормированные поля dz500 раскладывались на ЭОФ. Зимние и летние dz500 раскладывались отдельно на свои собственные сезонные ЭОФ, а не на общие круглогодичные.

Описанная выше процедура предобработки применялась как к полям z500 реанализа, так и к модельным полям z500, к каждой модели по

отдельности. Таким образом, модельные погодные режимы выделялись в собственном фазовом пространстве модельных полей dz500. В отличие от других работ, где модельные поля dz500 проецировались на ЭОФ полей dz500 реанализа, чтобы получить так называемые псевдо-главные компоненты (как, например, в [27, 75]), в данной работе модельные поля dz500 раскладывались на собственные модельные ЭОФ. Это делалось для того, чтобы, в случае наличия в модели собственных, отличных от реанализа, максимумов плотности распределения dz500 в пространстве модельных ЭОФ, идентифицировать режимы, отличные от режимов реанализа, что указывало бы на худшее воспроизведение реальной динамики атмосферы в модели относительно реанализа.

Непосредственно алгоритму кластеризации k-means подвергались главные компоненты разложения полей dz500 на 10 первых ЭОФ. В различных работах используется разное число первых ЭОФ разложения полей z500 для кластеризации, как правило, в диапазоне от 4 до 20. В работе [24] показано, что при использовании от 10 ЭОФ и более результаты кластеризации практически перестают отличаться от результатов кластеризации оригинальных неразложенных полей dz500. 10 первых ЭОФ разложения по данным реанализа объясняют 90.8% дисперсии в зимние месяцы и 82.4% дисперсии в летние. Для моделей доли объясненной дисперсии z500 ведущими ЭОФ примерно совпадают таковыми по реанализу – в моделях 10 первых ЭОФ объясняют от 89.7% (MIROC6) до 92.7% (NorESM2-LM) в зимний сезон и от 80.2% (MIROC6) до 88.5% (NorESM2-LM) в летний.

Автор решил придерживаться того же числа кластеров для выделения погодных режимов в Евро-Атлантике, которое было использовано для их идентификации в Главе 1 за период 1940-2022 гг., $K = 4$, так как это наименьшее число режимов, являющееся оптимальным за данный период для ЕАТ на основе совокупности методов, включая индексы классифицируемости воспроизводимости и Калински-Харабаша. Для более короткого периода, используемого в данной главе, 1950 – 2014 гг., в зимний период так же выделяется $K = 4$ как наименьшее оптимальное число режимов. В летний период для возможности сопоставления результатов двух сезонов выбрано такое же $K = 4$, как и зимой. Другими исследователями погодных режимов в ЕАТ $K = 4$ так же использовалось как оптимальное число режимов для Евро-Атлантики в летнее время [62, 76].

3.2. Воспроизведение полей погодных режимов

Полученные по результатам кластеризации средние поля $dz500$ зимних и летних погодных режимов (далее, ПР) в Евро-Атлантике по данным реанализа ERA5 и данным экспериментов historical некоторых моделей представлены на рисунке 3.1.

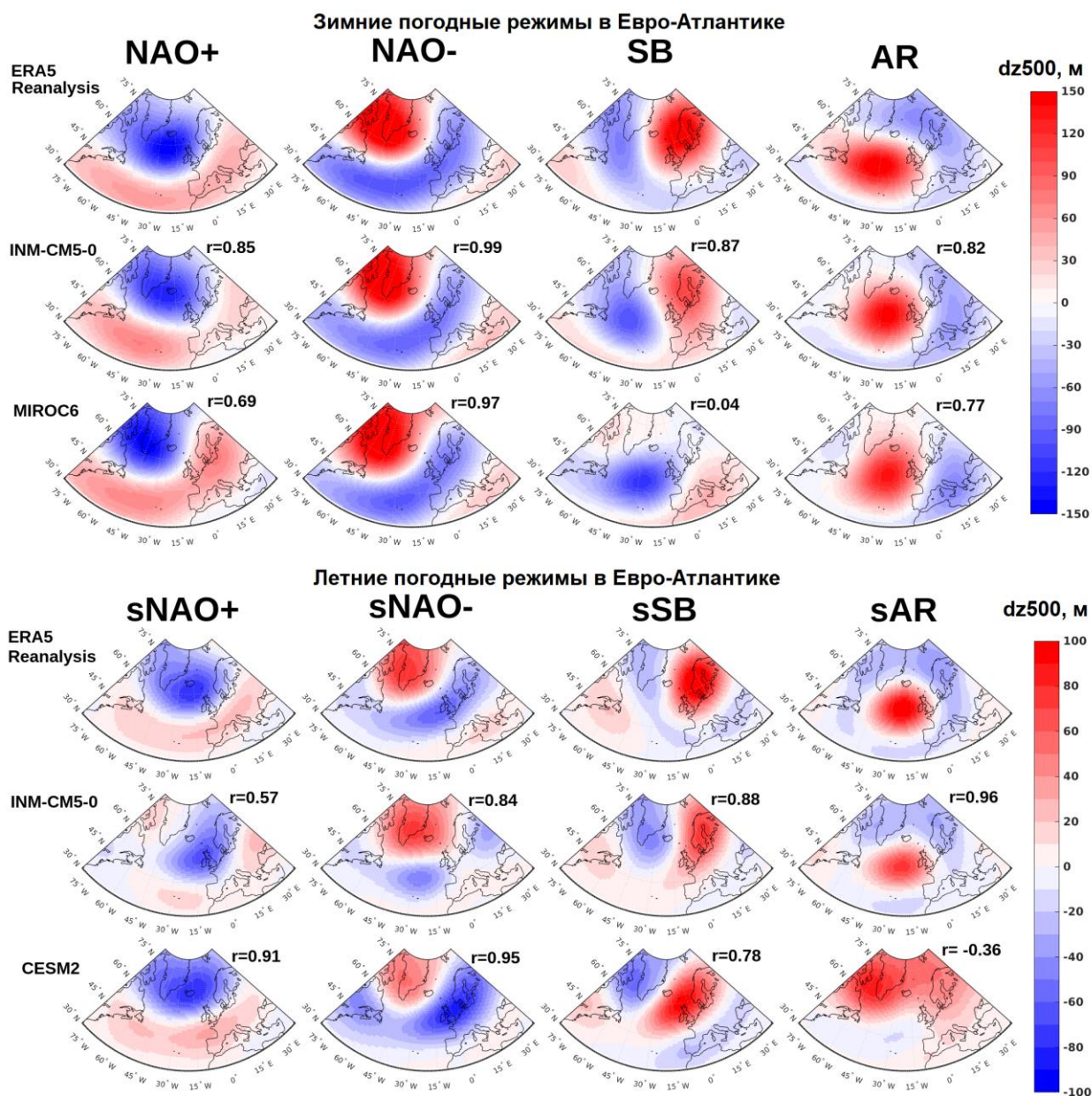


Рисунок 3.1. Средние поля аномалий $z500$ (в метрах) зимних и летних ПР в Евро-Атлантике за 1950 – 2014 гг. по данным ERA5 и моделей INM-CM5-0, MIROC6, CESM2. Символом « r » обозначены коэффициенты пространственной корреляции поля данного модельного ПР с аналогичным полем ПР реанализа.

Средние поля dz500 зимних ПР, полученные по данным реанализа ERA5, визуально совпадают с полями ПР, полученными в других работах (напр., [19, 23, 27]) и названы аналогичным образом – NAO+, NAO-, SB (Scandinavian Blocking), AR (Atlantic Ridge). Летние ПР названы аналогично с припиской «s» (summer) – sNAO+, sNAO-, sSB, sAR.

Пространственные структуры или паттерны (здесь и далее паттернами называются средние поля dz500 погодных режимов) модельных ПР сопоставлялись с таковыми по реанализу путем расчета пространственных корреляций по формуле 1.5, но вместо $z500_w(i, j, t)$ использовались центроиды кластеров $\langle z500_w(i, j, t) \rangle_{t \in k}$. Перед расчетом корреляций паттерны взвешивались на корень из косинуса широты для учета разности площадей ячеек сетки на разных широтах. Модельные режимы, для паттернов которых попарная корреляция с паттернами какого-либо из режимов NAO+, NAO-, SB, AR по данным реанализа оказалась максимальной, сочтены модельными аналогами этих режимов и названы так же, как режимы реанализа. Аналогичное сопоставление проведено для летних режимов.

Не для всех моделей воспроизвелись характерные паттерны зимних и летних ПР, то есть в матрице 4x4 попарных корреляций не для каждого i-го паттерна модельного ПР нашелся единственный уникальный j-ый паттерн ПР реанализа, то есть для некоторых j из четырех ПР реанализа (NAO+, NAO-, SB, AR) максимальная корреляция наблюдалась для одних и тех же i-ых ПР модели. Такой результат получился для зимних ПР модели MIROC6 (не воспроизвелся режим SB) и для летних ПР модели CESM2 (не воспроизвелся режим sAR). Эти модели, не сумевшие воспроизвести все из четырех паттернов ПР реанализа представлены на рисунке 3.1 (наряду с моделью INM-CM5-0 для сопоставления, которая смогла воспроизвести основные паттерны как зимних, так и летних ПР реанализа). Для зимних ПР MIROC6 и летних ПР CESM2 сопоставление с режимами реанализа производилось вручную – предпочтение отдавалось поиску «лучшей» пары для зимних ПР NAO+ и NAO- и летних sNAO+ и sNAO- из реанализа (лучшей, то есть пары с максимальной пространственной корреляцией).

Пространственные корреляции паттернов зимних модельных ПР с паттернами ПР реанализа представлены в таблице 3.1, в которой приведены только максимальные парные корреляции для каждого из четырех ПР реанализа, то есть максимальные значения для каждого j-го столбца в матрицах 4x4 (i-ая строка – паттерн ПР модели, j-ый столбец – паттерн ПР

реанализа) попарных корреляций паттернов ПР модель-реанализ. В таблице 3.2 представлено аналогичное сопоставление, но для летних ПР.

Таблица 3.1. Лучшие (максимальные из четырех) значения коэффициентов пространственных корреляций паттернов зимних модельных ПР для каждого из зимних ПР реанализа. Самые низкие значения пространственной корреляции для каждого ПР среди моделей отмечены жирным.

Название модели	Название режима				Среднее значение пространственной корреляции
	NAO+	NAO-	SB	AR	
INM-CM5-0	0.85	0.99	0.87	0.82	0.88
CESM2	0.97	0.96	0.91	0.95	0.95
GFDL-CM4	0.93	0.98	0.86	0.95	0.93
EC-Earth3	0.94	0.97	0.95	0.96	0.95
MIROC6	0.69	0.97	0.04	0.77	0.62
NorESM2-LM	0.87	0.98	0.90	0.90	0.91
CanESM5	0.69	0.80	0.81	0.92	0.80
ACCESS-CM2	0.68	0.91	0.87	0.68	0.78
Среднее по всем моделям	0.83	0.95	0.78	0.87	0.85

Таблица 3.2. То же, что в таблице 3.1, но для летних погодных режимов.

Название модели	Название режима				Среднее значение пространственной корреляции
	sNAO+	sNAO-	sSB	sAR	
INM-CM5-0	0.57	0.84	0.88	0.96	0.81
CESM2	0.91	0.95	0.78	-0.36	0.57
GFDL-CM4	0.85	0.93	0.88	0.81	0.87
EC-Earth3	0.94	0.96	0.96	0.67	0.88
MIROC6	0.84	0.88	0.53	0.65	0.73

NorESM2-LM	0.83	0.96	0.82	0.57	0.80
CanESM5	0.90	0.92	0.79	0.69	0.82
ACCESS-CM2	0.95	0.91	0.93	0.78	0.90
Среднее по всем моделям	0.85	0.92	0.82	0.60	0.80

Так как пространственная корреляция не всегда является лучшей метрикой определения схожести двух полей и не учитывает интенсивность характерных паттернов ПР, то есть величину аномалий $dz500$ в паттернах режимов (условно, если значения модельных $dz500$ в каждой точке поделить на 2 а значения $dz500$ реанализа оставить теми же самыми, то корреляция не изменится), паттерны модельных ПР сравнивались с паттернами ПР реанализа еще и путем подсчета сумм квадратов расстояний (СКР), которые можно посчитать по формуле 1.4, вместо $z500_w(i, j, t)$ используя центроиды $\langle z500_w(i, j, t) \rangle_{t \in k}$ моделей и реанализа. В отличие от коэффициента корреляции, чем ближе значение СКР к нулю, тем больше два паттерна похожи друг на друга. Аналогично пространственной корреляции, при расчете СКР паттерны ПР так же нормировались пропорционально корню из косинуса широты ячеек для уравнивания вклада ячеек, имеющих разную площадь (в формуле 1.4 это уже учтено). Результаты представлены в таблицах 3.3 и 3.4 для зимы и лета соответственно. Перед расчетом паттерны нормировались так, что значения в таблицах следует интерпретировать как средний по полю квадрат разницы $dz500$ между паттерном ПР модели и паттерном ПР реанализа.

Таблица 3.3 СКР зимних паттернов ПР реанализа с аналогичными паттернами зимних ПР моделей, значения указаны m^2 (среднее по всем ячейкам полей $dz500$). Худшие (т.е. самые высокие) парные значения СКР для каждого режима реанализа отмечены жирным.

Название модели	Название режима				Среднее значение квадрата расстояния
	NAO+	NAO-	SB	AR	
INM-CM5-0	762	148	831	926	667
CESM2	240	445	650	283	404
GFDL-CM4	358	208	843	352	440
EC-Earth3	438	444	343	247	368
MIROC6	2010	386	4800	1160	2090

NorESM2-LM	646	351	650	642	572
CanESM5	1530	2400	1160	443	1380
ACCESS-CM2	1910	1030	889	2480	1580
Среднее значение по всем моделям	987	677	1270	817	938

Таблица 3.4. То же, что в таблице 3.3, но для летних погодных режимов.

Название модели	Название режима				Среднее значение квадрата расстояния
	sNAO+	sNAO-	sSB	sAR	
INM-CM5-0	432	295	232	90	262
CESM2	125	291	423	2420	815
GFDL-CM4	169	136	211	318	208
EC-Earth3	147	114	86	513	215
MIROC6	187	220	675	542	406
NorESM2-LM	225	134	321	982	416
CanESM5	134	156	371	539	300
ACCESS-CM2	56	206	135	337	183
Среднее значение по всем моделям	184	194	307	718	351

В результате сравнительного анализа паттернов модельных ПР с ПР реанализа можно сделать следующие выводы: 1) большинство из исследованных моделей в исторических экспериментах хорошо воспроизводят как летние, так и зимние паттерны ПР реанализа; 2) зимние паттерны традиционных евро-атлантических погодных режимов (NAO+, NAO-, SB, AR) в моделях воспроизводятся лучше, чем их летние аналоги (sNAO+, sNAO-, sSB, sAR), средняя пространственная корреляция по паттернам зимних ПР составила 0.85, а по летним 0.8 (см. таблицы 3.1, 3.2).

3.3. Воспроизведение временных характеристик погодных режимов

Помимо паттернов ПР, производилось сравнение относительных повторяемостей и продолжительностей модельных ПР с ПР реанализа. Сравнивались повторяемости и продолжительности тех модельных ПР с ПР реанализа, которые при сопоставлении паттернов оказались наиболее похожи по метрике пространственной корреляции. Схожесть величин относительных

повторяемостей и средних продолжительностей режимов оценивалась по формуле:

$$\Delta F_{\text{ПР модель}} = \frac{|F_{\text{ПР модель}} - F_{\text{ПР ERA5}}|}{F_{\text{ПР ERA5}}} * 100\%, \quad (3.1)$$

где $F_{\text{ПР}}$ – одна из статистических характеристик ПР, то есть либо относительная повторяемость одного из четырех ПР (как доля всего исследуемого периода), либо его средняя продолжительность (в сутках). Таким образом, чем меньше величина $\Delta F_{\text{ПР модель}}$, тем лучше модель воспроизводит характеристику данного режима. В таблицах 3.5 и 3.6 приведено сравнение повторяемостей модельных ПР с ПР реанализа для зимних и летних ПР соответственно.

Таблица 3.5. Разницы относительных повторяемостей зимних погодных режимов по данным моделей с режимами по данным реанализа, величины указаны в %, рассчитаны по формуле 3.1. Фактическая относительная повторяемость модельных ПР и ПР реанализа указана в скобках следом за разницей, в % от всей длины исследуемого периода. Самые высокие различия для каждого режима отмечены жирным шрифтом. Для ПР SB модели MIROC6 стоит «-», т.к. данный режим не воспроизвелся в модели.

Название модели и реанализа	Название режима				Средняя разница, %
	NAO+	NAO-	SB	AR	
ERA5	(29.6)	(22.2)	(23.7)	(24.5)	
INM-CM5-0	2.9 (30.5)	5.7 (20.9)	11.1 (26.3)	9.0 (22.3)	7.2
CESM2	8.1 (32.1)	18.4 (18.1)	6.0 (25.1)	1.0 (24.8)	8.4
GFDL-CM4	9.4 (32.4)	0.2 (22.1)	1.8 (23.2)	9.5 (22.2)	5.2
EC-Earth3	1.7 (29.1)	6.9 (20.6)	7.5 (25.4)	1.1 (24.8)	4.3
MIROC6	4.8 (28.2)	12.7 (19.4)	-	8.6 (26.7)	8.7
NorESM2-LM	10.0 (32.6)	9.9 (20.0)	8.5 (21.6)	5.0 (25.8)	8.4
CanESM5	4.3 (28.4)	15.0 (18.8)	2.2 (24.2)	16.6 (28.6)	9.6
ACCESS-CM2	4.8 (28.2)	18.1 (26.2)	15.5 (27.3)	25.5 (18.3)	16.0

Средняя разница и повторяемость по ансамблю	5.8 (30.2)	10.9 (20.8)	7.5 (24.7)	9.5 (24.2)	8.4
--	------------	-------------	------------	------------	-----

Таблица 3.6. То же, что в таблице 3.5, но для летних погодных режимов. Для ПР sAR модели CESM2 стоит «-», т.к. данный режим не воспроизвелся в модели.

Название модели и реанализа	Название режима				Средняя разница, %
	sNAO+	sNAO-	sSB	sAR	
ERA5	(24.5)	(28.8)	(23.3)	(23.3)	
INM-CM5-0	10.5 (27.1)	19.7 (23.1)	1.8 (23.7)	11.5 (26.0)	10.9
CESM2	19.6 (29.4)	14.2 (24.7)	7.7 (25.1)	-	13.8
GFDL-CM4	27.1 (31.2)	16.8 (24.0)	4.3 (22.3)	3.4 (22.5)	12.9
EC-Earth3	1.8 (25.0)	13.8 (24.8)	2.6 (23.9)	12.6 (26.2)	7.7
MIROC6	2.7 (25.2)	12.0 (25.4)	14.2 (26.6)	2.2 (22.8)	7.8
NorESM2-LM	42.2 (34.9)	27.1 (21.0)	4.4 (24.3)	15.3 (19.7)	22.2
CanESM5	0.8 (24.7)	12.7 (25.2)	11.5 (26.0)	3.3 (24.1)	7.1
ACCESS-CM2	22.6 (30.1)	19.1 (23.3)	3.4 (24.1)	3.6 (22.5)	12.2
Средняя разница и повторяемость по ансамблю	15.9 (28.5)	16.9 (23.9)	6.2 (24.5)	7.4 (23.4)	11.6

При анализе относительных повторяемостей ПР можно отметить, что за исследованный период в зимние месяцы режим NAO+ по данным реанализа оказался самым частым с повторяемостью в 29.6%, а самым редким режимом оказался «противоположный» режим NAO- с повторяемостью в 22.2%. Тот факт, что NAO+ - самый продолжительный режим, а NAO- – самый редкий, успешно воспроизвели почти все из исследованных моделей, включая INM-

CM5-0, кроме модели ACCESS-CM2, которая показала худшее воспроизведение повторяемостей для зимних месяцев со средней ошибкой воспроизведения повторяемости режимов в 16.0%. Лучший результат в воспроизведении относительных повторяемостей режимов оказался у модели EC-Earth3 со средней по всем четырем режимам разницей в 4.3%, средняя разница у INM-CM5-0 составила 7.2%.

Относительные повторяемости летних аналогов традиционных зимних ПР, полученных по реанализу ERA5, слабо различаются и составляют 24.5%, 28.8%, 23.3%, 23.3% для режимов sNAO+, sNAO-, sSB, sAR соответственно. Стоит отметить, что в отличие от зимнего режима NAO-, являющегося самым редким в зимние месяцы, летний режим sNAO-, наоборот, самый частый. Аналогично сравнению паттернов зимних и летних модельных ПР с ПР реанализа, летние относительные повторяемости модельных режимов хуже воспроизводят относительные повторяемости ПР, полученных по реанализу. Так, средняя по всем моделям и режимам разница относительных повторяемостей в зимние месяцы составила 8.4%, а для летних режимов 11.6%. Худший результат в воспроизведении относительных повторяемостей летних ПР оказался у модели NorESM2-LM со средней разницей повторяемостей относительно реанализа в 22.2%, режим sNAO+ у которой оказался перенаселенным с повторяемостью в 34.9%, что на 42.2% превышает повторяемость sNAO+ в реанализе, составляющую 24.5%. Лучше всего в воспроизведении относительных повторяемостей показала себя канадская модель CanESM5 со средней разницей по всем режимам относительно реанализа в 7.1%.

В таблицах 3.7 и 3.8 приводится сопоставление средних продолжительностей режимов. Анализ продолжительностей зимних и летних ПР, полученных по реанализу, показывает, что самым продолжительными оказываются режимы NAO- и sNAO- (средняя продолжительность составила 9.5 суток как для зимы, так и для лета), несмотря на то, что в зимние месяцы NAO- является самым редким режимом, а в летние месяцы самым частым, такой результат может указывать на то, что не существует прямой связи между повторяемостью и продолжительностью режима. Также стоит отметить, что зимние режимы в целом оказались более продолжительными, чем летние (как по реанализу, так и по модельным данным), что может указывать на большую «режимность» атмосферной циркуляции в зимнее время, чем летнее, то есть на большую устойчивость паттернов атмосферных течений в данный сезон.

Таблица 3.7. Разницы средних продолжительностей зимних погодных режимов по данным моделей с режимами по данным реанализа, величины указаны в %, рассчитаны по формуле 3.1. Фактическая средняя продолжительность (количество суток) модельных ПР и ПР реанализа указана в скобках следом за разницей. Самые высокие разницы для каждого режима отмечены жирным шрифтом. Для ПР SB по данным модели MIROC6 стоит «-», т.к. данный режим не воспроизвелся в модели.

Название модели и реанализа	Название режима				Средняя разница, %
	NAO+	NAO-	SB	AR	
ERA5	(8.6)	(9.5)	(7.4)	(7.8)	
INM-CM5-0	2.9 (8.9)	4.2 (9.9)	8.7 (8.0)	5.4 (7.3)	5.3
CESM2	7.1 (9.2)	3.6 (9.1)	3.8 (7.7)	6.1 (7.3)	5.1
GFDL-CM4	8.9 (9.4)	7.6 (8.7)	5.8 (7.0)	5.9 (7.3)	7.0
EC-Earth3	6.8 (8.0)	16.1 (7.9)	1.9 (7.3)	13.0 (6.8)	9.5
MIROC6	13.8 (9.8)	12.0 (8.3)	-	3.1 (8.0)	9.6
NorESM2-LM	10.6 (9.5)	9.3 (10.3)	5.5 (7.0)	5.6 (8.2)	7.7
CanESM5	2.4 (8.4)	3.9 (9.8)	3.9 (7.7)	0.4 (7.8)	2.6
ACCESS-CM2	9.9 (9.5)	18.1 (11.2)	8.6 (8.0)	9.9 (7.0)	11.6
Средняя разница и продолжительность по всем моделям	7.8 (9.1)	9.3 (9.4)	5.5 (7.4)	6.2 (7.5)	7.3

Таблица 3.8. То же, что в таблице 3.7, но для летних погодных режимов. Для ПР sAR по данным модели CESM2 стоит «-», т.к. данный режим не воспроизвелся в модели.

Название модели и реанализа	Название режима				Средняя разница, %
	sNAO+	sNAO-	sSB	sAR	
ERA5 Reanalysis	(6.9)	(9.5)	(7.1)	(6.8)	

INM-CM5-0	17.1 (8.1)	6.4 (8.9)	12.7 (8.0)	5.9 (7.2)	10.5
CESM2	28.0 (8.8)	8.2 (8.7)	17.8 (8.3)	-	18.0
GFDL-CM4	12.3 (7.7)	9.4 (8.6)	3.6 (7.3)	12.5 (7.6)	9.4
EC-Earth3	15.3 (7.9)	11.8 (8.3)	9.8 (7.8)	13.1 (7.7)	12.5
MIROC6	18.2 (8.1)	7.9 (8.7)	8.7 (7.7)	9.5 (7.4)	11.1
NorESM2-LM	51.4 (10.4)	6.6 (8.8)	10.0 (7.8)	15.0 (7.8)	20.7
CanESM5	14.2 (7.9)	3.1 (9.2)	19.4 (8.5)	12.0 (7.6)	12.2
ACCESS-CM2	6.6 (7.3)	14.4 (8.1)	3.9 (7.4)	3.1 (6.6)	7.0
Средняя разница и продолжительность по всем моделям	20.4 (8.3)	8.5 (8.7)	10.7 (7.8)	10.2 (7.4)	12.7

При сопоставлении продолжительностей модельных ПР с ПР реанализа выявлено, что практически все модели удачно воспроизводят тот факт, что режимы NAO- и sNAO- является самыми продолжительными. Для 4 из 8 моделей самый продолжительный режим NAO-, у оставшихся 4 из 8 NAO- на втором месте по продолжительности после NAO+. Для летнего sNAO- аналогичный результат – в 6 моделях режим sNAO- самый продолжительный, в оставшихся двух самым продолжительным оказался sNAO+.

Сравнение разниц продолжительностей модельных ПР и ПР реанализа показывает, что худшие результаты в зимние месяцы продемонстрировала модель ACCESS-CM2, а в летние - модель NorESM2-LM, что идентично результату при сравнении относительных повторяемостей модельных ПР с ПР реанализа. Таким образом, можно предположить, что правильное воспроизведение повторяемости связано с воспроизведением продолжительности режимов, несмотря на то, что непосредственная связь между двумя этими характеристиками режимов может отсутствовать, то есть большая повторяемость режима не приводит к большей продолжительности, как можно заметить, сравнивая повторяемость и продолжительность режимов NAO+, NAO- и sNAO+, sNAO- по данным реанализа.

В целом, при воспроизведении продолжительностей ПР в моделях относительно реанализа наблюдается результат, аналогичный полученному

при сравнении паттернов и относительных повторяемостей – средние продолжительности режимов лучше воспроизводятся для традиционных зимних ПР (NAO+, NAO-, SB, AR), чем для их летних аналогов, средняя по всем моделям и режимам разница продолжительностей ПР составила 7.3% зимой против 12.7% летом.

3.3.1. Матрицы вероятностей переходов погодных режимов

В данном разделе проведен анализ матриц вероятностей переходов между режимами как по данным реанализа, так и по модельным данным, а также подсчитаны статистически значимые частые переходы для этих матриц. Матрицы вероятностей переходов между ПР, полученными по данным реанализа, представлены в таблице 3.9, на которой отмечены статистически значимые частые переходы на уровне 5%, посчитанные по методике, описанной в разделе 2.1. Статистически значимые частые переходы не обязаны быть наиболее вероятными, и наоборот, наиболее вероятные не обязаны являться статистически значимо частыми, так как при подсчете статистической значимости учитывается разница в количестве режимных событий, которое отличается у каждого из режимов, при этом для переходов из одного режима ($i \Rightarrow \dots$) может оказаться несколько режимов (очевидно, не больше $K-2$, то есть не больше двух для $K=4$), переход в которые будет статистически значимо частым.

Матрица вероятностей переходов между погодными режимами являются тонкой характеристикой, чувствительной к основным параметрам погодных режимов, таким как средние поля режимов и их относительные повторяемости, в связи с этим матрицы вероятностей переходов между ПР, полученными различными моделями, довольно существенно различаются. Так, при сопоставлении значимых переходов между режимами в зимние месяцы из 8 моделей, они полностью совпали со значимыми переходами для ПР по данным ERA5 лишь у одной модели (GFDL-CM4), в летние месяцы при взаимном сопоставлении значимых переходов, то есть логических матриц 4×4 , где для значимого перехода $i \Rightarrow j$ соответствующее значение матрицы равно 1, а для незначимого 0, ни для одной из моделей все 16 значений матрицы не совпали полностью с результатами для переходов между режимами по данным ERA5.

Таблица 3.9. Матрица вероятностей переходов зимних и летних ПР в Евро-Атлантике, полученных по данным реанализа ERA5 за 1950 – 2014 гг. Жирным шрифтом отмечены статистически значимо частые переходы на уровне значимости 5%. Строка указывает на то, из какого режима осуществляется переход, столбец – в какой режим осуществляется переход.

Зимние погодные режимы в Евро-Атлантике по ERA5 за 1950–2014 гг.				
	=> NAO+	=> NAO-	=> SB	=> sAR
NAO+ =>	-	0.14	0.44	0.42
NAO- =>	0.36	-	0.32	0.32
SB =>	0.30	0.35	-	0.35
AR =>	0.47	0.20	0.33	-
Летние погодные режимы				
	=> sNAO+	=> sNAO-	=> sSB	=> sAR
sNAO+ =>	-	0.19	0.46	0.34
sNAO- =>	0.46	-	0.21	0.34
sSB =>	0.25	0.36	-	0.39
sAR =>	0.41	0.31	0.28	-

В таблице 3.10 приведены средние по ансамблю матрицы вероятностей переходов зимних и летних ПР, полученных путем усреднения матриц вероятностей переходов для всех модельных ПР (кроме MIROC6 в зимние месяцы и CESM2 в летние, которые не воспроизвели все 4 поля в соответствующие сезоны). Несмотря на то, что отдельные матрицы переходов, полученные по ПР из разных моделей, сильно различаются между собой, между матрицами переходов, полученных по реанализу и усредненными матрицами переходов, полученными по ансамблю моделей, существует сходство. Так, статистически значимо частые переходы между зимними ПР (NAO+ => SB, SB => NAO-, AR => NAO+) являются таковыми как в реанализе, так и для 3 из 7 моделей. При этом совпадают некоторые другие особенности переходов, так, вероятность перехода из NAO+ в NAO- достаточно низкая и составляет 0.14 (с учетом округления) как по данным реанализа, так и в среднем по моделям, но вероятность обратного перехода из NAO- в NAO+ в разы выше и составляет 0.36 по данным реанализа и 0.29 в среднем по моделям. Аналогичная ситуация наблюдается для летних переходов между sNAO+ и sNAO- – вероятность перейти из sNAO+ в sNAO- составляет 0.19 по реанализу и 0.14 в среднем по моделям, однако вероятность обратного перехода из sNAO- в sNAO+ составляет 0.46 по ПР реанализа и 0.49 в среднем

по моделям, причем переход sNAO- в sNAO+ оказался статистически значимо частым как по данным реанализа, так и для 6 из 7 моделей.

Таблица 3.10. Усредненная по ансамблю моделей матрица вероятностей переходов между зимними и летними ПР, полученными по данным экспериментов historical. Для зимних месяцев не учитывались результаты модели MIROC6, а для летних результатов модели CESM2, так как эти модели не воспроизвели средние поля каждого из 4 режимов в соответствующие сезоны. В скобках после вероятностей указано количество моделей из ансамбля, для которых данный переход оказался статистически значимым частым на уровне 5%. Жирным шрифтом отмечены переходы, для которых у хотя бы 3 из 7 моделей они оказались статистически значимыми. Строка указывает на то, из какого режима осуществляется переход, столбец – в какой режим осуществляется переход.

Зимние погодные режимы в Евро-Атлантике по моделям за 1950 – 2014 гг.				
	=> NAO+	=> NAO-	=> SB	=> sAR
NAO+ =>	-	0.14 (0/7)	0.43 (3/7)	0.43 (1/7)
NAO- =>	0.29 (0/7)	-	0.43 (2/7)	0.29 (1/7)
SB =>	0.35 (1/7)	0.27 (3/7)	- (0/7)	0.38 (2/7)
AR =>	0.47 (3/7)	0.23 (2/7)	0.30 (0/7)	- (0/7)
Летние погодные режимы в Евро-Атлантике по моделям за 1950 – 2014 гг.				
	=> sNAO+	=> sNAO-	=> sSB	=> sAR
sNAO+ =>	-	0.24 (0/7)	0.40 (4/7)	0.35 (1/7)
sNAO- =>	0.49 (6/7)	-	0.24 (1/7)	0.27 (0/7)
sSB =>	0.35 (0/7)	0.24 (0/7)	-	0.41 (4/7)
sAR =>	0.32 (0/7)	0.34 (5/7)	0.34 (2/7)	-

3.4. Выводы

В данной главе исследовано воспроизведение характеристик зимних и летних режимов в Евро-Атлантике в моделях Земной системы, участвовавших в проекте CMIP6. Режимы анализировались за тот же временной период в той же пространственной области, что и по данным реанализа, для этого использовались модельные данные экспериментов historical. Характеристики модельных режимов сравнивались с характеристиками режимов, полученных по реанализу, включая паттерны погодных режимов, их относительные повторяемости, средние продолжительности и матрицы вероятностей переходов.

В результате сравнения воспроизводимости основных характеристик погодных режимов в моделях выявлено, что практически все модели воспроизводят средние паттерны погодных режимов в Евро-Атлантике в зимние и летние месяцы при сравнении их с паттернами погодных режимов реанализа. Воспроизведение паттернов ПР в моделях оценивалось с помощью двух метрик – пространственной корреляции и суммы квадратов расстояний. Из восьми протестированных моделей, не смогли воспроизвести все 4 режима только японская модель MIROC6 в зимние месяцы и американская CESM2 в летние. Получено, что в среднем модели лучше воспроизводят основные характеристики погодных режимов в зимние месяцы, чем в летние.

Большинство исследованных моделей воспроизводят так же и временные характеристики зимних и летних погодных режимов в ЕАТ, такие как повторяемости и продолжительности, при этом ошибка воспроизведения временных характеристик погодных режимов в ЕАТ зимой меньше, чем летом. Большинство исследованных моделей воспроизводят тот факт, что в зимние месяцы повторяемость режима NAO+ является наибольшей, а режима NAO- наименьшей, при этом модели так же воспроизводят тот факт, что средняя продолжительности зимнего режима NAO- и его летнего аналога sNAO- являются наибольшими по сравнению с другими ПР. Некоторые особенности матриц вероятностей переходов между режимами, такие как последовательности статистически значимых частых переходов, воспроизводятся в среднем по ансамблю, но не отдельными моделями.

ГЛАВА 4. ПРЕДСКАЗУЕМОСТЬ ПОГОДНЫХ РЕЖИМОВ ВЕДУЩИМИ МОДАМИ КЛИМАТИЧЕСКОЙ ИЗМЕНЧИВОСТИ И ВЛИЯНИЕ РЕЖИМОВ НА ЭКСТРЕМАЛЬНЫЕ ПОГОДНЫЕ ЯВЛЕНИЯ

В Главе 2 в рисунках 2.2 и 2.6 показаны гистограммы распределения числа режимных событий в ЕАТ и СЕ зависимости от их продолжительности, из которых видно, что существуют редкие эпизоды, когда погодные режимы могут длиться вплоть до 40-50 суток (NAO+, AR, sSB в Евро-Атлантике и WSB, SSB, SUN в Северной Евразии), занимая до половины времени всего сезона и определяя погоду в течение такого длительного времени. В связи с этим важно исследовать предсказуемость погодных режимов. Вопрос предсказуемости ПР для региона Евро-Атлантики исследуется давно. В данной главе представлены результаты анализа предсказуемости Северо-Евразийских режимов, полученных автором.

Помимо исследования предсказуемости ПР в Северной Евразии, в этой главе будет представлено влияние Евро-Атлантических и Северо-Евразийских погодных режимов на погодные аномалии в Северном полушарии и рассмотрен частный случай влияния ПР на экстремальные осадки на южном берегу Крымского полуострова.

4.1. Связь погодных режимов в Северной Евразии с $Nino3.4$ и колебаниями площади морского льда

Известно, что характеристики погодных режимов, включая их повторяемость и интенсивность, связаны как с внутрисезонными модами Земной климатической системы (например, известна связь погодных режимов в Евро-Атлантике с колебанием Маддена-Джулиана [19]), так и с более долгопериодными модами климатической изменчивости [77], например, с колебаниями температуры поверхности в экваториальной части Тихого океана, явление есть Эль-Ниньо [78], или концентрацией морского льда в Северном полушарии [60]. В последние годы было много исследований связи циркуляции атмосферы в средних широтах, например, блокинговой активности, с явлением Эль-Ниньо в зимние [79] и летние месяцы [80], а также циркуляции атмосферы в Евро-Атлантике с морским льдом [81]. Некоторые новые исследования показывают интересные, но противоречивые результаты, например, в статье 2024 года показано [82], что мгновенные и годовые

отклики паттернов NAO на колебание Эль-Ниньо противоположны, при этом предсказуемость явления Эль-Ниньо в различные сезоны [83] до сих пор активно исследуется [84], для чего разрабатываются новые методы [85, 86]. В связи с этим представляет интерес исследование связи погодных режимов в Северной Евразии, представленных автором диссертации в Главе 2, с колебанием Эль-Ниньо и площадью морского льда в Северном полушарии.

В разделе 4.1 будет представлен анализ связи повторяемости погодных режимов в Северной Евразии с колебаниями температуры поверхности океана в экваториальной части Тихого Океана с использованием индекса Nino3.4 [87], а также с колебаниями площади морского льда в Северном полушарии. Площадь морского льда испытывает долгосрочный отрицательный тренд в период регулярных спутниковых наблюдений, начавшийся в 1979 году [88]. В рамках данной диссертации для исследования предсказуемости режимов в Северной Евразии, оценивалась связь их повторяемости с индексом Nino3.4 и площадью морского льда в предшествующие режимам сезоны. Для зимних погодных режимов в Северной Евразии будет исследоваться их повторяемость с осенними Nino3.4 и морским льдом, для летних режимов – с весенними Nino3.4 и морским льдом.

Для расчета Nino3.4 и аномалий площади морского льда использовались данные реанализа ERA5: среднемесячные данные по температуре морской поверхности (sea surface temperature, SST) и концентрации морского льда (sea-ice concentration, SIC). Nino 3.4 рассчитан как детрендрованный временной ряд средней SST в области 170 з.д. – 120 з.д., 5 ю.ш. – 5 с.ш, а для расчета площади морского льда в Северном полушарии использованы все значения SIC севернее 40 градусов с.ш. Для анализа связи Nino3.4 и морского льда с ПР, были отобраны временные ряды средних за 3 месяца осенних и весенних значений Nino3.4 и суммарной площади морского льда в Северном полушарии. Для Nino3.4 использовался временной период, совпадающий с периодом исследования ПР в Северной Евразии – 1940 – 2022 гг., а для морского льда только значения с 1979 по 2022 гг., то есть с года начала спутниковых наблюдений.

После расчета временных рядов, все значения Nino3.4 и морского льда были разделены на 3 перцентильные группы (для удобства, далее будем называть их фазами) – 1) значения ниже 25-го перцентиля принимались как отрицательная фаза (или «Ла-Нинья» для Nino3.4); 2) значения между 25-ым и 75-ым перцентилем принимались как нейтральная фаза; 3) значения выше 75-го перцентиля как положительная фаза («Эль-Ниньо» для Nino3.4). Величины

относительной повторяемости режимов в Северной Евразии в последующие сезоны, перед которыми наблюдались значения, попавшие в различные группы перцентилей аномалий Nino3.4 или площади морского льда, сравнивались между собой. Результаты сравнения представлены на рисунке 4.1. Для получения квартилей и «усов» на графиках формата «ящик с усами» из рисунка 4.1 использовался бутстрэппинг временных рядов сезонных повторяемостей ПР в различных фазах Nino 3.4 и площади морского льда, генерировалось по 10000 случайных выборок.

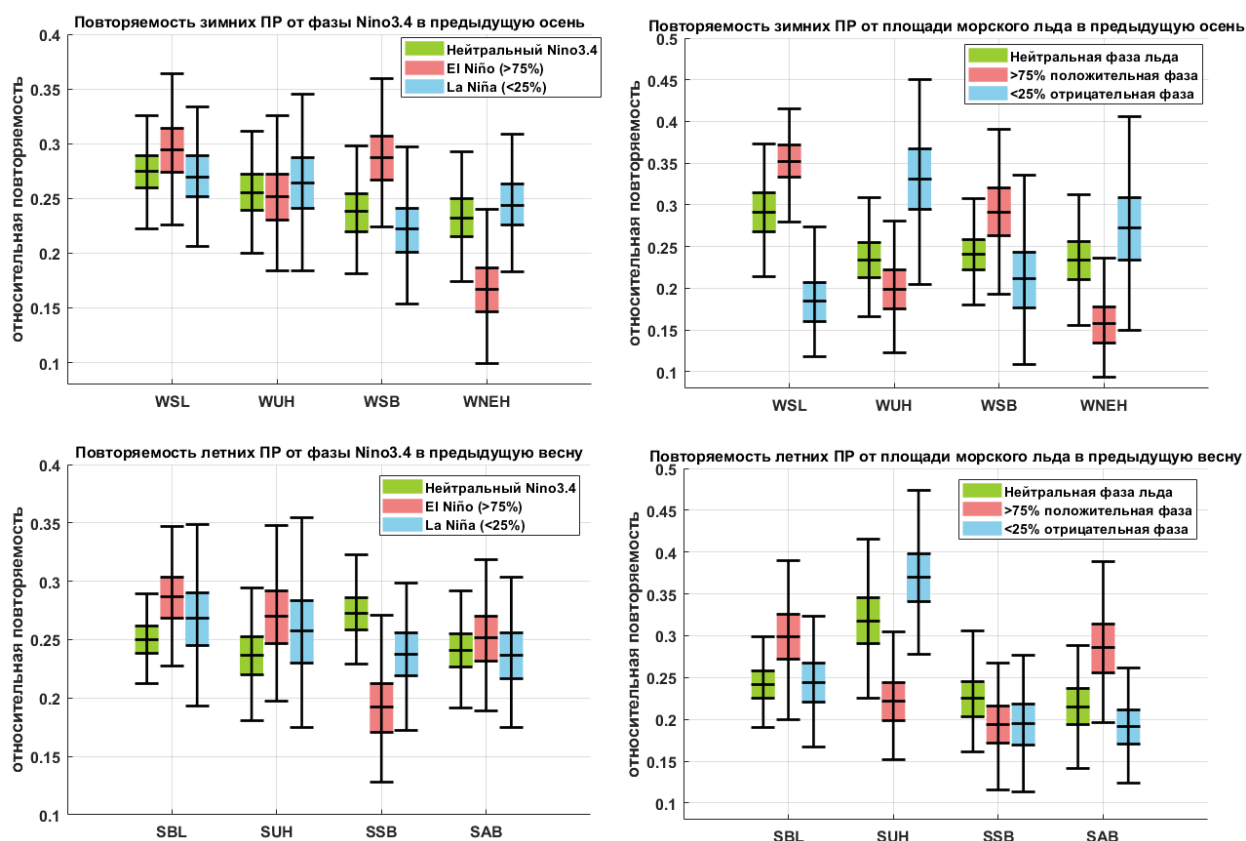


Рисунок 4.1. Средняя относительная повторяемость ПР в СЕ как доля всего сезона во время нейтральных (зеленые блоки), положительных (красные блоки) и отрицательных (синие блоки) фаз Nino3.4 и аномалий площади морского льда в Северном полушарии в предшествующие режимам сезоны. Линии в середине блоков указывают на фактические средние значения повторяемостей. Верхние и нижние границы блоков указывают на 3-ий и 1-ый квартили (75-ый и 25-ый перцентили) распределения, а верхние и нижние линии «усов» указывают на 99-й и 1-й перцентили. Границы блоков и «усы» получены с помощью метода бутстрэп.

Дополнительно к рисунку 4.1, результаты сравнения повторяемости ПР в СЕ в различных фазах индекса Nino 3.4 и площади морского льда приведены в таблице 4.1, где с использованием критерия Стьюдента приведено сравнение

повторяемостей ПР для противоположных фаз Nino 3.4 и площади морского льда, которые представлены красными и синими блоками на рисунке 4.1.

Таблица 4.1. Результаты двухвыборочного t-критерия Стьюдента для выборок повторяемостей погодных режимов при положительных и отрицательных фазах индекса Nino 3.4 и площади морского льда в Северном полушарии в предшествующие сезоны. Нулевая гипотеза H_0 – повторяемости при разных фазах принадлежат одной выборке. Значения «1» - H_0 может быть отвергнута на уровне значимости $\alpha = 10\%$, «0» - H_0 не может быть отвергнута при $\alpha = 10\%$. р-уровень – вероятность получить наблюдаемый или более экстремальный результат при условии верности гипотезы H_0 . Жирным шрифтом отмечены результаты, при которых выборки различны на заданном уровне значимости $\alpha = 10\%$, т. е. фазы Nino 3.4 или льда значимо связаны с повторяемостями ПР.

Повторяемости ПР в СЕ при противоположных фазах «Эль-Ниньо» (за 1940 – 2022 гг.) в предшествующие сезоны				
	WSL	WUH	WSB	WNEH
$H_0 (\alpha = 0.10)$	0	0	0	1
р-уровень	0.565	0.798	0.139	0.074
	SBL	SUH	SSB	SAB
$H_0 (\alpha = 0.10)$	0	0	0	0
р-уровень	0.692	0.788	0.295	0.744
Повторяемости ПР в СЕ при противоположных фазах площади морского льда в предшествующие сезон (за 1979 – 2022 гг.)				
	WSL	WUH	WSB	WNEH
$H_0 (\alpha = 0.10)$	1	1	0	0
р-уровень	0.002	0.058	0.259	0.103
	SBL	SUH	SSB	SAB
$H_0 (\alpha = 0.10)$	0	1	0	0
р-уровень	0.312	0.017	0.954	0.104

Хотя изменчивость повторяемости для каждого ПР велика при любой фазе Nino 3.4, как видно по рисунку 4.1 и коэффициентам вариации в таблице 2.4, можно увидеть заметное изменение в средней повторяемости для некоторых ПР в СЕ в зависимости от фазы Nino3.4. В годы с осенним «Эль-Ниньо» (положительной фазой Nino3.4) наблюдается повышенная повторяемость WSB (р-уровень = 0.139) и пониженная повторяемость WNEH (р-уровень = 0.074) в последующую зиму. Весенний «Эль-Ниньо» слабо связан с повторяемостями летних ПР в СЕ, наиболее заметное (видно по рисунку 4.1), но все же незначимое (р-уровень = 0.295) влияние весенний «Эль-Ниньо» оказывает на повторяемость режима летнего скандинавского блокинга SSB.

Из рисунка 4.1 и таблицы 4.1 видно, что флуктуации площади морского льда в Северном полушарии оказывают большее влияние на повторяемость ПР в СЕ в последующие сезоны, чем ЭНЮК. В годы с отрицательными аномалиями площади морского льда (когда льда мало) в осенние месяцы, наблюдается пониженная повторяемость WSL (p-уровень = 0.002) и повышенная повторяемость WUH (p-уровень = 0.058) и WNEH (p-уровень = 0.103) в последующие зимы. При отрицательных аномалиях площади морского льда весной, летом наблюдается повышенная повторяемость режима уральского максимума SUH (p-уровень = 0.017) и снижается повторяемость режима арктического блокирования SAB (p-уровень = 0.104). Примечательно, что отрицательные аномалии площади морского льда в Северном полушарии связаны с повышенными аномалиями режимов уральского максимума в последующие сезоны и зимой (WUH) и летом (SUH).

4.2. Связь погодных режимов в Северной Евразии с экстремальными аномалиями температур и осадков

В разделе выше была показана связь повторяемости погодных режимов в Северном полушарии с явлением Эль-Ниньо, которое в свою очередь связано с частотой погодных аномалий в Северной Евразии [89, 90], однако представляют интерес выявить связь непосредственно между погодными режимами в Северной Евразии и параметрами атмосферы в области их действия. Существуют работы в которых исследовалась связь погодных режимов в других регионах с такими параметрами, как температура [29], осадки и скорости ветра [30]. Особый интерес при этом представляет влияние режимов на распределение экстремальных погодных явлений [28], таких как очень жаркие или холодные дни, или дни с сильными осадками.

В рамках диссертации было исследовано влияние полученных автором зимних и летних погодных режимов в Северной Евразии на частоту редких аномалий среднесуточных температур и осадков – таких, которые происходят с вероятностью 5%. Например, для температур, это аномалии, которые превосходят 95-ый перцентиль (например, жаркие дни) или ниже 5-го перцентиля (холодные дни), а для осадков это такие дни, когда дневная сумма осадков превосходит 95-ый перцентиль, то есть дни с сильными ливнями или снегопадами.

Данные по температуре и осадкам были взяты из реанализа ERA5 для той же области Северной Евразии, для которой идентифицированы погодные режимы – от 0 до 180 в.д. и от 40 с.ш. до 80 с.ш., за тот же временной период 1940–2022 гг. В качестве температуры использовалась переменная T2M – температура в двух метрах над землей, для осадков TP – полная сумма осадков. Данные по T2M осреднялись до среднесуточных, по TP – суммировались. Чтобы проанализировать распределение редких аномалий температур и осадков независимо от сезона, из них вычитался сглаженный сезонный ход. Сезонный ход T2M и TP рассчитывался аналогично тому, как это делалось для z500, то есть осреднялись значения за одинаковые календарные даты разных лет, а затем сглаживались 5-дневным скользящим средним. Далее, для каждой переменной были рассчитаны пороги 95-го (или 5-го) перцентиля в каждом узле сетки, и для каждого погодного режима рассчитывалось количество раз, когда данный порог был превышен во время данного ПР. Количество превышений при данном режиме соотносилось с ожидаемым среднестатистическим числом превышений, равном $N_k * 0,05$, где N_k – общее количество суток, проведенное системой в данном k-ом ($k=[1, 2, 3, 4]$) режиме за все зимние или летние сезоны исследуемого периода 1940–2022 гг.

Отношение фактического количества наблюдаемых аномалий выше порога 95-го (или ниже 5-го) перцентиля T2M и TP к ожидаемому $N_k * 0,05$ в зависимости от режима представлено на рисунке 4.2. На рисунке приведены относительные вероятности аномалий выше 95-го перцентиля для температуры и осадков во время летних ПР и аномалий ниже 5-го перцентиля для температур для зимних ПР, так как именно преимущественно такие события можно рассматривать как опасные (т. е. жаркие и ливневые дни летом, холодные дни зимой). Результаты для T2M ниже 5-го перцентиля летом (холодные летние дни) и выше 95-го перцентиля для T2M и TP зимой не приведены (теплые зимние дни и снегопады). Как видно из рисунка 4.2, пространственное распределение вероятностей аномалий T2M и TP по форме сообразно с паттернами z500 погодных режимов, что особенно заметно для распределения вероятностей летних аномалий T2M > 95%.

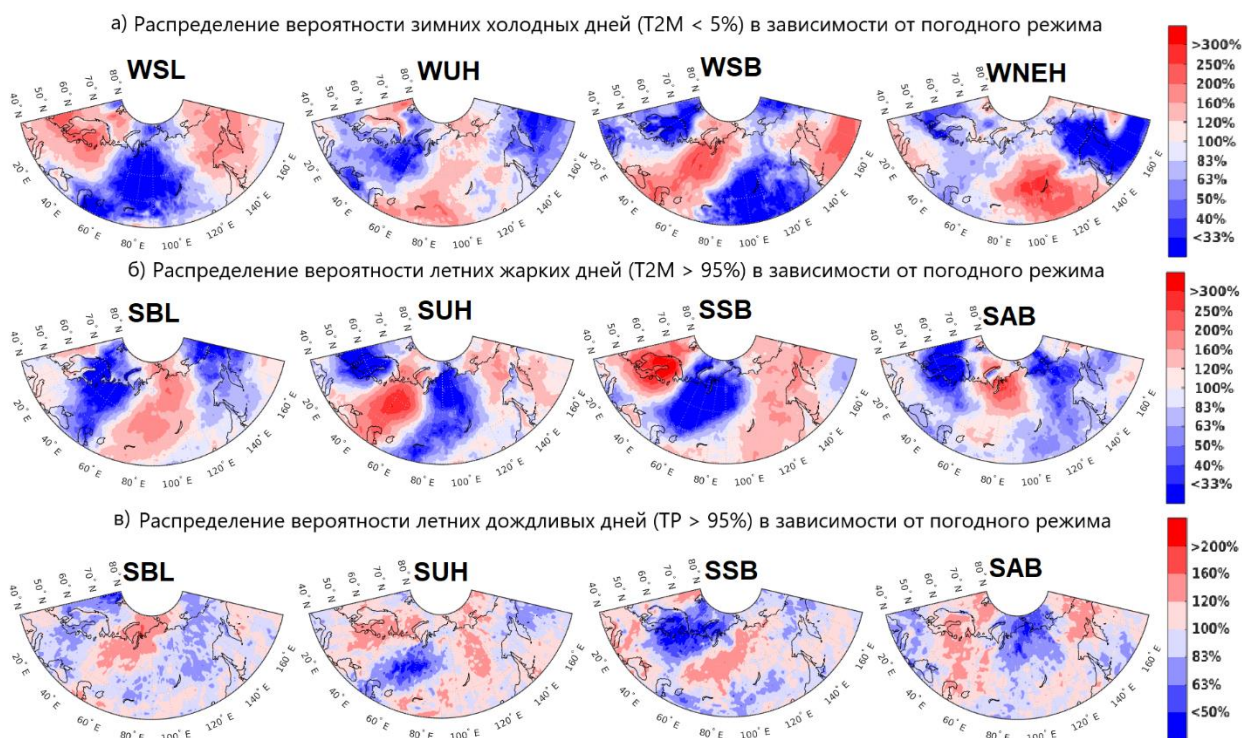


Рисунок 4.2. (а) пространственное распределение относительной частоты событий с аномалиями среднесуточной $T2M$ ниже 5-го перцентиля (холодные зимние дни) для зимних ПР в СЕ; (б) пространственное распределение относительной частоты событий с аномалиями среднесуточной $T2M$ выше 95-го перцентиля (жаркие летние дни) для летних ПР в СЕ; (в) как (б), но для TP , суточных сумм осадков (дождливые летние дни). Шкалы справа от карт указывают на наблюдаемое количество превышений частоты $T2M$ или TP для разных режимов за 1940-2022 гг. в % относительно среднестатистического количества таких дней ($N_k * 0,05$).

Поля на рисунке 4.2 можно интерпретировать следующим образом – во время зимнего режима WSL, например, вероятность аномально холодных дней ниже 5-го перцентиля в Средней Сибири составляет $< 33\%$ от среднестатистической вероятности (0.05), т.е. $p < 0.014$. Можно отметить, что во время летнего режима уральского максимума, SUH, вероятность аномально жарких дней на Европейской территории России (ЕТР) увеличивается примерно в 2.5-3 раза. Режим уральского максимума SUH ассоциируется с блокинговой активностью над ЕТР, что показано на рисунке 4.3, его повторяемость имела локальный максимум в 2010 году, когда в ЕТР наблюдалась рекордная жара [72]. Режим SUH, ассоциированный с повышенной вероятностью жарких дней в ЕТР, так же имел локальный пик повторяемости в 2016 году (см. рисунок 2.5), когда над ЕТР так же наблюдалась волна жары [91].

На рисунке 4.3 представлено сопоставление относительных повторяемостей летних ПР в СЕ за всё время и в дни, когда наблюдалось атмосферное блокирование по двум индексам – Леженаса-Окланда (ЛО) [92] и Тибальди-Молтени (ТМ) [93, 94]. Из рисунка видно, что примерно в 90% случаев по обоим индексам блокинг над ЕТР наблюдается либо во время режима уральского максимума SUN, либо во время режима Скандинавского блокирования SSB. При этом для узкого диапазона долгот ЕТР – от 40 в.д. до 60 в.д. – две трети случаев приходятся именно на режим уральского максимума SUN, для которого наблюдается долгосрочный тренд роста повторяемости (рис. 2.5) и который связан с повышенной вероятностью жарких дней над ЕТР (рис. 4.2). Рост повторяемости режима SUN, связанного с блокингами в ЕТР по индексам ЛО и ТМ, качественно согласуется с ростом повторяемостей блокинговой активности в Северном полушарии, наблюдаемым в последние десятилетия [95].

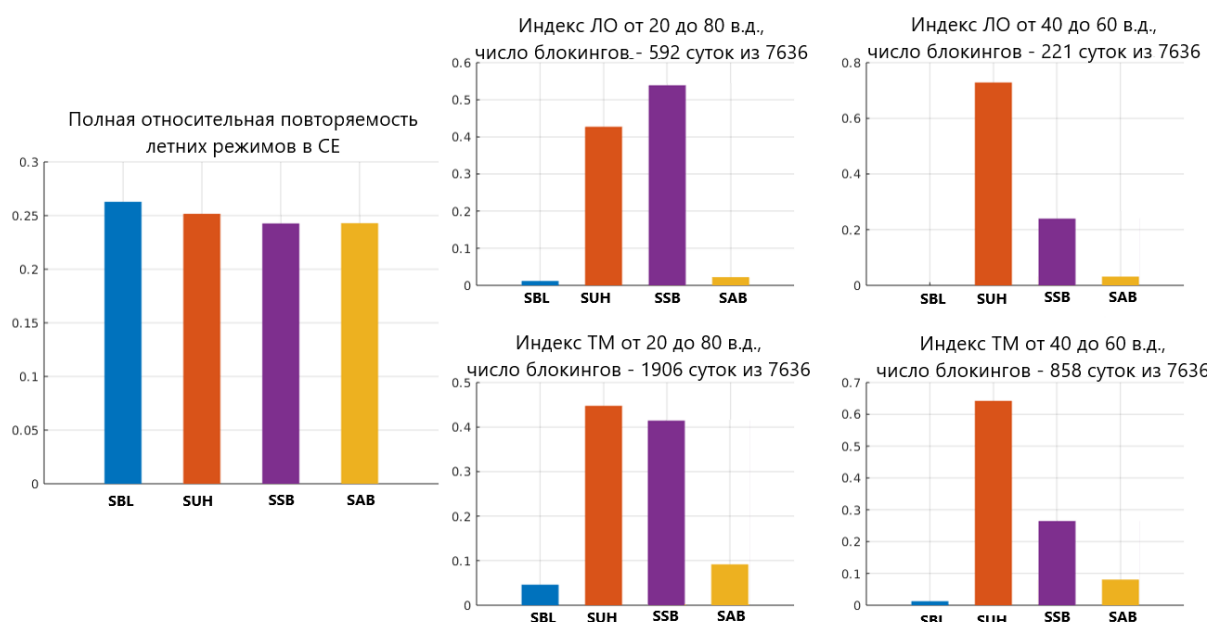


Рисунок 4.3. Относительная повторяемость летних ПР в СЕ за весь период 1940-2022 гг. и относительная повторяемость ПР в дни, когда на какой-либо из долгот на Европейской территории России в широком (от 20 до 80 в.д.) или узком диапазоне долгот (от 40 в.д. до 60 в.д.) наблюдались атмосферные блокирования по критериям Леженаса-Окланда (ЛО) или Тибальди-Мольтени (ТМ).

4.3. Современные погодные режимы в Евро-Атлантике и их влияние на экстремальные аномалии температур, осадков и ветра

В разделах 1.4.2, 2.2 и 3.2 представлены четыре традиционных погодных режима в Евро-Атлантике («NAO+», «NAO-», «SB», «AR») по данным реанализа ERA5 за временной период 1940 – 2022 гг., однако за такой длительный период наблюдаются тренды как самих погодных режимов (см. рисунок 2.1) в Евро-Атлантике, так и средних температур [96] и, вероятно, осадков и ветра, поэтому стоит исследовать влияние ПР в Евро-Атлантике на погодные аномалии в ЕАТ и прилегающих областях за более узкий современный период времени. В данном разделе будут представлены современные зимние погодные режимы в Евро-Атлантике за 1979-2021 гг. и их влияние на распределение температур на двух метрах над поверхностью (T2M), суточных сумм осадков (TP) и скорости ветра в 10 метрах над поверхностью (WS10) во внетропической зоне Северного полушария. Хотя для Евро-Атлантики за период с 1979 года существуют современные сеточные наборы данных более высокого разрешения, например [97], данные по z500 использовались из ERA5, так как для исследования крупномасштабных структур атмосферной циркуляции разрешения ERA5 с шагом 0.25° более, чем достаточно. Данные по осадкам, температуре и ветру использовались так же из реанализа ERA5.

Зимние (декабрь, январь, февраль) ПР в ЕАТ за 1979 – 2021 были получены аналогично процедуре, описанной в 1.4.1 и 2.3.1, но по более упрощенной схеме предобработки без фильтра низких частот – из оригинальных суточных полей z500 вычитался сглаженный сезонный ход, сглаженный 5-дневным скользящим окном для получения полей аномалий dz500. Поля dz500 раскладывались на 10 первых ЭОФ, которые объясняют 83% дисперсии зимних dz500 за 1979 – 2021 гг. К временным рядам главных компонент первых 10 ЭОФ применялся метод кластеризации k-means с $K = 4$ для воспроизведения четырех традиционных зимних ПР, полученных в других главах за более длительный период с 1940 по 2022 гг.

Из 3881 зимних дней за 1979 – 2021 гг. полная относительная повторяемость кластеров NAO+, NAO-, SB и AR составила 1250 (32.2%), 758 (19.5%), 1071 (27.6%) и 802 (20.7%) дней соответственно. Можно отметить факт того, что NAO+ самый частый режим, а NAO- самый редкий, такой результат сохраняется как на длительном промежутке с 1940 года (см. таблицу 2.1), так и на более коротком с 1979 года.

Поля зимних ПР в ЕАТ за 1979 – 2021 гг. представлены на рисунке 4.4 в полярной проекции. Эти режимы визуально совпадают с режимами, представленными на рисунках 1.7 (по КМ) и 3.1 и названы аналогично – NAO+, NAO-, SB, AR. Помимо самих средних полей ПР, на рисунке представлены так же тренды этих средних полей, посчитанные как тренды временного ряда, состоящего из средних полей режимов за каждую полную зиму исследуемого периода, то есть за 1979 – 2021 гг. для каждого из 4 режимов имелся ряд из 42 средних полей, полученных путем усреднения всех суточных dz_{500} , отнесенных к данному режиму за данный зимний сезон. Из рисунка видно, что для всех режимов во внетропической части Северного полушария наблюдается в целом положительный тренд (красные цвета на рисунке 4.4б), который может быть связан с общим климатическим трендом z_{500} в связи с термическим расширением тропосферы [98], при этом области значимых трендов (p -уровень ≤ 0.05) сосредоточены преимущественно на южных границах внетропической зоны Северного полушария.

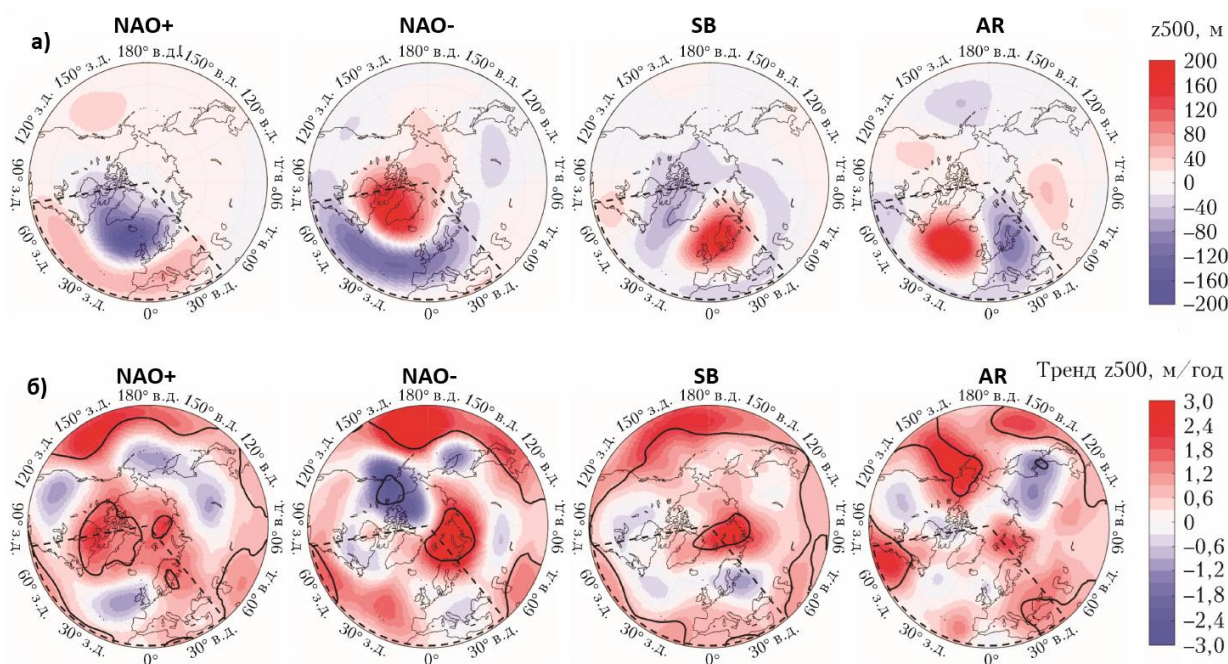


Рисунок 4.4 Средние поля (а) (в метрах) и тренды (б) полей аномалий z_{500} (в метрах/год) зимних ПР в ЕАТ за 1979-2021 гг. Черная жирная линия в (б) – области значимых на уровне трендов на уровне значимости 5%; пунктирная линия – границы ЕАТ сектора.

Непосредственно в зоне ЕАТ сектора, для которого считались режимы, области значимых положительных трендов также сосредоточены в основном на границах этого сектора. Можно отметить крупные области: 1) вокруг моря Баффина для режима NAO+; 2) на юго-западной и юго-восточной границе

ЕАТ сектора для режима AR. При этом значимые тренды для каждого из режимов находятся в удалении от абсолютных максимумов аномалий $z500$ средних полей этих режимов (рисунок 1, верхний ряд), которые и характеризуют эти режимы, что свидетельствует о том, что пространственная структура режимов в ЕАТ сохраняется в течение исследуемого периода.

Несмотря на небольшие области значимых трендов ПР в ЕАТ за 1979-2021 гг., их сезонная повторяемость за этот период не имеет статистически значимых трендов, что продемонстрировано на рисунке 4.5, несмотря на наличие трендов сезонной повторяемости у режимов в ЕАТ за 1940-2022 гг., продемонстрированных на рисунке 2.1, что объясняется как различием во временных интервалах, так и упрощенной методикой определения погодных режимов за укороченный современный период 1979-2021 гг.

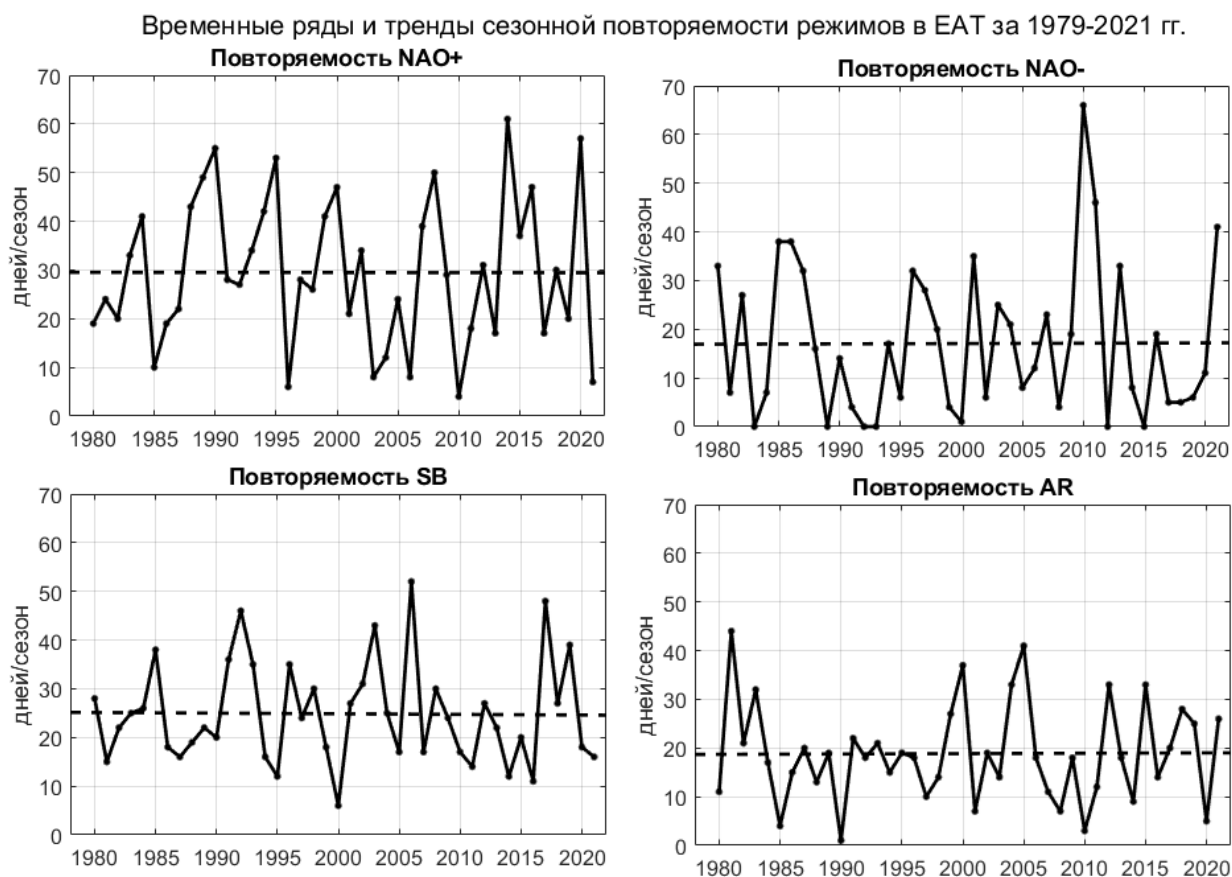


Рисунок 4.5. Временные ряды сезонной повторяемости погодных режимов в ЕАТ (сплошные линия с точками) за 1979 – 2021 гг., и их линейные тренды (пунктирная линия).

Стоит отметить, что значения в одни и те же года между режимами за 1940-2022 гг. и 1979-2021 гг. хоть и должны коррелировать, но не совпадают полностью, так как для получения современных ПР в ЕАТ использовалась

более упрощенная методика без применения фильтра низких частот, однако общие характеристики и особенности режимов при этом сохраняются, а временные ряды повторяемости NAO+/NAO- соответствуют индексу NAO [99] с корреляцией 0.89, что показано на рисунке 4.6



Рисунок 4.6 Сопоставление временных рядов индекса NAO [99] (левая шкала, серый цвет), осредненного за зимние месяцы, и разности повторяемости режимов NAO+ и NAO- (правая шкала, черный цвет). Коэффициент корреляции временных рядов 0.89.

Распределение экстремальных аномалий в зависимости от современных ПР в ЕАТ исследовалось аналогичному тому, как это делалось для ПР в СЕ в разделе 4.2 – рассчитывалось наблюдаемое количество экстремальных аномалий выше 95-го или ниже 5-го перцентиля во время погодных режимов по отношению к среднестатистической частоте таких событий (0.05 или раз в 20 суток), но с разницей в том, что для ПР в СЕ аномалии исследовались непосредственно в области определения режимов, а для современных ПР в ЕАТ во всей внетропической зоне Северного полушария (выше 30 градусов с.ш.) Для этого использовались температура в двух метрах над поверхностью T2M, суточная сумма осадков TP и среднесуточной скорости ветра в 10 метрах над поверхностью WS10. Аномалии T2M, TP и WS10 и их пороги 5-го и 95-го перцентиля рассчитывались после вычитания сезонного хода, сглаженного 5-дневным скользящим средним.

Повторяемость экстремальных событий (ниже 5-го и выше 95-го процентиля функции распределения вероятности), наблюдающихся во время полученных режимов, представлена на рисунке 4.7. Во время режима NAO+ характерна пониженная более чем в 3 раза частота аномально холодных и повышенная частота аномально теплых дней над европейским континентом. Противоположная картина наблюдается над Северной Атлантикой южнее Гренландии: для NAO+ наблюдаются повышенная в 1.2-2 раза по сравнению с ожидаемой частота сильных осадков на севере Европы и повышенная до 2.5 раз частота аномально сильного ветра на северо-западе от Европы.

Для режима NAO- наблюдается обратная режиму NAO+ смещенная на север картина как для температуры, так и для осадков и скорости ветра. Например, частота аномально холодных дней над севером Европы повышена по сравнению с ожидаемой до 3 раз и более. На юге Европы и севере Африки частота аномально теплых дней повышена до 2-2.5 раз. В центре Европы наблюдается нейтральная зона, в которой частота экстремальных аномалий температуры ($T2M > 95\%$) равна ожидаемой. На севере Европы и России начинается зона с пониженной в 3 раза и более частотой $T2M > 95\%$. К югу от Гренландии наблюдается повышение частоты аномально теплых дней в 3 раза и более и аналогичное понижение холодных дней. Отмечается повышенная в 2-3 раза частота сильных осадков на западе Пиренейского полуострова и повышенная до 3 раз частота сильного ветра на широте Азорских островов вдоль всего Атлантического океана.

Несмотря на то что для режимов NAO+ и NAO- наблюдаются противоположные картинки частоты аномалий $T2M$, TP и $WS10$, пространственные распределения повторяемости этих аномалий не являются полностью симметричными. Пространственное распределение повторяемости аномалий $T2M$, TP и $WS10$ при режиме NAO- в европейском регионе смещена на северо-запад относительно режима NAO+, т.е. в сторону максимума положительной аномалии $z500$ у режима NAO-. Визуально для указанных режимов наиболее симметричными друг другу оказались поля повторяемости экстремального ветра ($WS10 > 95\%$ на рис. 3).

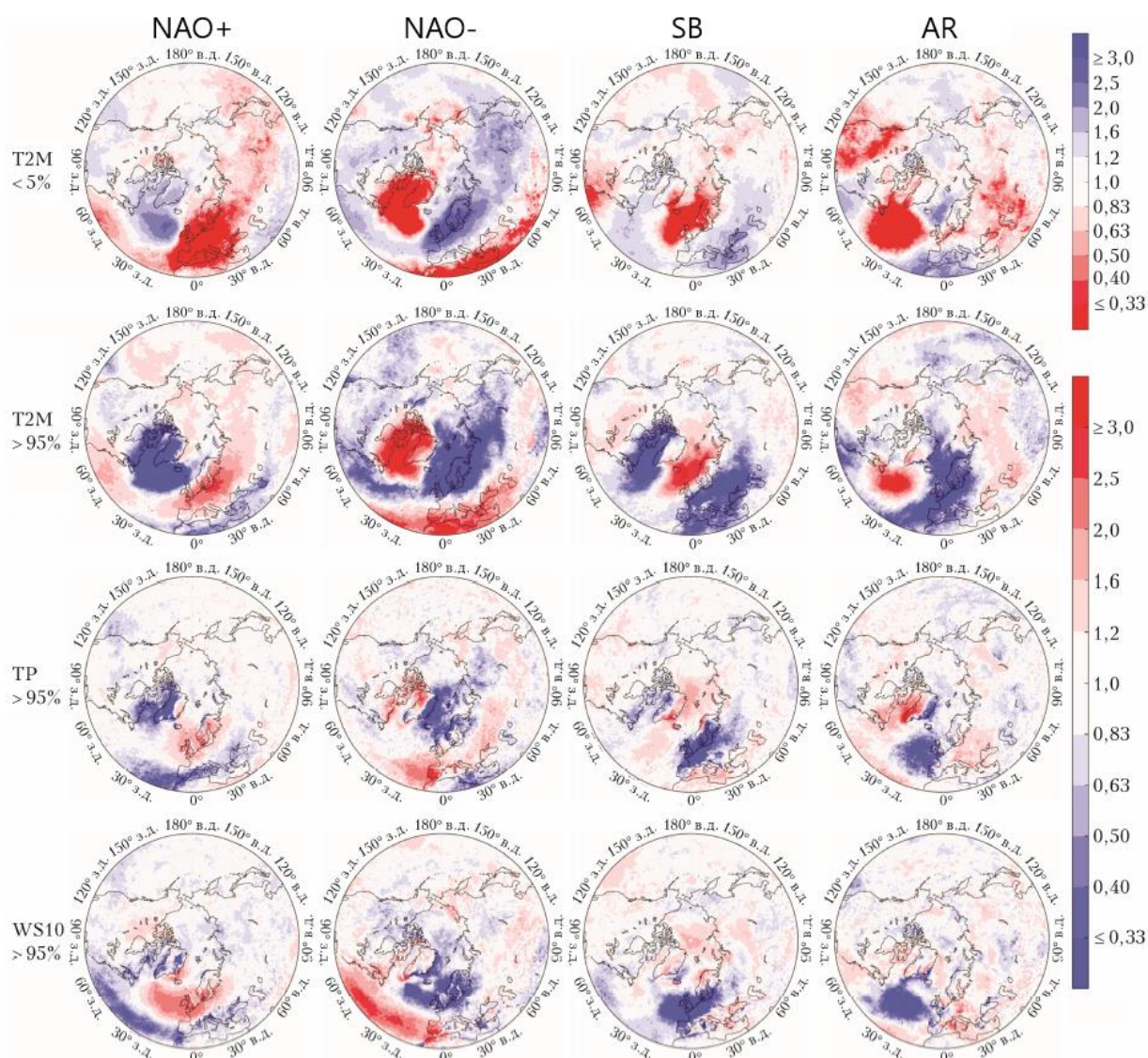


Рисунок 4.7. Связанная с зимними погодными режимами в ЕАТ (за 1979-2021 гг.) повторяемость экстремальных аномалий T2M, TP, WS10. Верхний ряд – повторяемость T2M ниже 5-го перцентиля, то есть холодные дни. Со 2-го по 4-ый ряд: повторяемость аномалий выше 95-го перцентиля, то есть повторяемость аномально теплых зимних дней, снегопадов и сильных ветров. На шкалах справа отношение фактически наблюдаемой частоты экстремальных событий при данном погодном режиме к среднестатистической (0.05 или раз в 20 суток).

Во время режима скандинавского блокинга SB наблюдается увеличение частоты экстремально теплых и уменьшение частоты экстремально холодных дней непосредственно над Скандинавией и Северным морем. Над континентальной Европой, кроме Скандинавии, и на юго-западе от Гренландии наблюдается пониженная до 3 раз и более частота аномально теплых дней, а частота аномально холодных дней немного повышена. На северо-западе Европы при режиме SB отмечается уменьшение частоты аномально сильных осадков и ветра.

Для территории России повторяемость аномально теплых и аномально холодных дней при режимах NAO+, NAO- и SB качественно согласуется с результатами работы [69], где исследуется влияние фазы NAO и Скандинавской моды на продолжительность аномальных волн тепла и холода на территории России, рассчитанных по данным станционных наблюдений. При погодном режиме NAO+ аналогично положительной фазе индекса NAO происходит рост (снижение) числа аномально теплых (холодных) дней над большей частью территории России (с уменьшением влияния на восток), а во время режимов NAO- и SB, аналогично отрицательной фазе NAO и положительной фазе Скандинавской моды, отмечается увеличение (уменьшение) числа аномально холодных (теплых) дней.

Для режима AR наблюдается повышенная (пониженная) частота аномально теплых (холодных) дней над зоной положительной аномалии высоты геопотенциала этого режима (см. рис. 4.4a). Над этой же областью отмечается пониженная частота дней с аномально сильным ветром, а немного восточнее – область с пониженной частотой сильных осадков. На западе Гренландии во время AR выделяется область с повышенной частотой сильных осадков.

В целом для всех режимов отмечается ослабление их влияния на повторяемость экстремальных событий в удаленных от области определения (80 з.д. – 40 в.д.) режимов регионах, однако для режимов NAO- и AR влияние распространяется и на достаточно удаленные области. Для AR заметно кратное снижение, т. е. в несколько раз меньше ожидаемого, повторяемости экстремально холодных дней (рис. 4.7, $T2M < 5\%$) вплоть до западного побережья Северной Америки, а для режима NAO- наблюдается кратное снижение повторяемости аномально теплых дней вплоть до Восточной Сибири.

Между полями повторяемости экстремальных положительных и отрицательных аномалий T2M для всех режимов наблюдается некоторая асимметрия как по пространству, так и по величине повторяемости (т.е. повышенная в 2 раза частота аномалий $T2M > 95\%$ в некоторой точке для конкретного рассматриваемого режима не означает пониженную в 2 раза частоту $T2M < 5\%$), что может объясняться несимметричностью кумулятивной функции распределения аномалий приземной температуры в зимние месяцы [100].

4.4. Связь интенсивности осадков в Крыму с летними погодными режимами на европейском континенте

В разделах 4.2 и 4.3 показано влияние крупномасштабных погодных режимов в Евро-Атлантике и Северной Евразии на распределение экстремальных аномалий температур, осадков и ветра в области определения режимов и прилегающих областях. В данном разделе представлены результаты исследования частного случая влияния режимов циркуляции атмосферы на распределение экстремальных осадков на южном берегу Крыма (ЮБК), для которого в последние годы наблюдалось несколько случаев экстремальных осадков со значительными негативными последствиями [101, 102].

Ранее метод погодных режимов использовался для исследований частных случаев экстремальных погодных явлений, например, аномально холодной зимы 2010 года в Европе [29], при этом для экстремальных осадков в Крыму использовались другие подходы с применением крупномасштабной атмосферной циркуляции, например, метод элементарных циркуляционных механизмов по типизации Дзердзеевского [103], ранее применявшийся и для других, соседствующих с Крымом, регионов [104], однако интерес представляет связь осадков в Крыму с погодными режимами, полученными с использованием современных методов кластерного анализа.

В данном разделе методом k-means simulated annealing получены режимы циркуляции атмосферы в Европейском регионе для летних месяцев, исследована повторяемость этих режимов и обнаружена связь между сильными осадками на ЮБК и полученными погодными режимами. Показано, что некоторые режимы способствуют значительно более сильным осадкам, чем другие, при этом для повторяемости некоторых из них существуют значимые тренды.

4.4.1. Идентификация погодных режимов в Европе

Для получения режимов циркуляции использовались данные реанализа ERA-5: среднее давление на уровне моря (Mean Sea Level Pressure, MSLP) и геопотенциал на высоте 500 гПа (z500) за 1979 – 2021 гг. (43 года), исследовались летние месяцы: июнь, июль, август. Был выбран следующий сектор: 30° с.ш. – 80° с.ш., 40° з.д. – 60° в.д. Выбранный сектор отличается от

традиционного сектора для ЕАТ (80 з.д. – 40 в.д.) и выбранного автором диссертации сектора Северной Евразии (0 в.д. – 180 в.д.), т.к. для Крыма был выбран более узкий сектор, влияющий на циркуляции непосредственно в Европейском континенте, включая европейскую территорию России.

Для анализа использовались данные по осадкам из эмпирического архива E-OBS [105], имеющие разрешение $0,1^\circ$ градуса. Для этих данных проведено осреднение вдоль узлов Южного побережья Крыма, примыкающего к Крымским горам, эта область представлена на рисунке 4.8. Также анализировался временной ряд осадков со станции в Феодосии (индекс ВМО 33976).

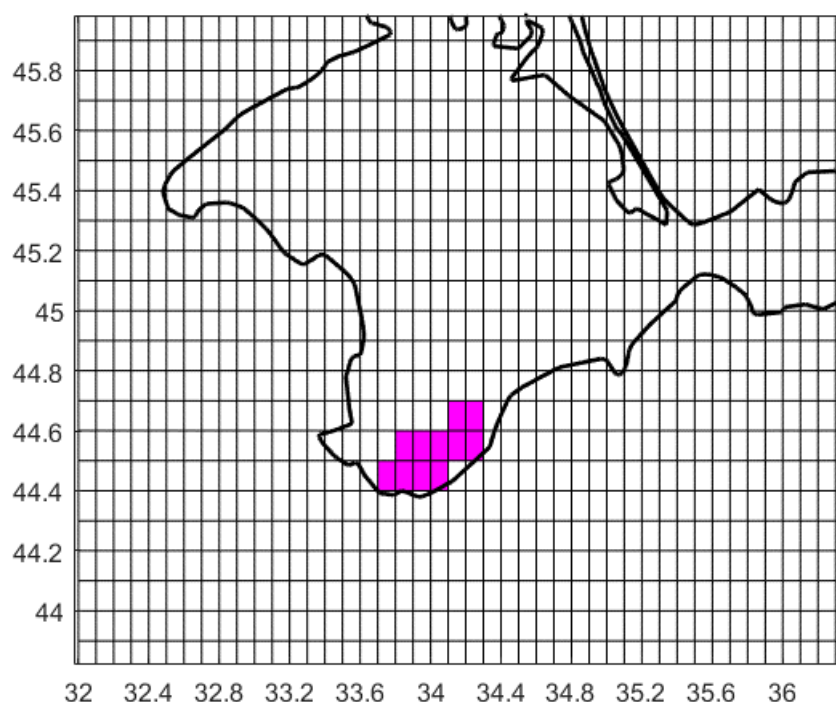


Рисунок 4.8. Ячейки сетки E-OBS, которые использовались для анализа осадков на южном берегу Крыма. Значения вдоль этих ячеек для последующего анализа осадков были осреднены в один временной ряд. По осям X и Y восточная долгота и северная широта в градусах.

Из практических соображений, по исходным данным MSLP и z500 с частотой 1/час и пространственным разрешением $0,25^\circ$ было проведено осреднение по времени до суточных полей и по пространству до шага в 1° . Далее, перед процедурой кластеризации была проведена следующая предобработка: из среднесуточных полей получены аномалии MSLP и z500 путём вычитания сезонного цикла, сглаженного скользящим средним с окном в 5 суток. Сезонный цикл получен осреднением суточных полей,

соответствующих одним и тем же календарным датам (по 43 поля на каждую дату). Из оригинальных суточных полей вычитался полученный сезонный цикл. С целью уменьшения размерности и сглаживания данных, из полей аномалий были получены временные ряды главных компонент эмпирических ортогональных функций (ЭОФ). Перед получением главных компонент поля аномалий были взвешены на корень из площади, чтобы уравновесить вклад дисперсии узлов сетки на разных широтах. По взвешенным летним полям аномалий были посчитаны временные ряды главных компонент, которые впоследствии кластеризовались. Для кластеризации оставлено по 50 первых главных компонент для MSLP и z500. Для MSLP они объясняют 98.06% дисперсии, для z500 98.96%, то есть практически всю изменчивость соответствующих полей.

Суточные поля MSLP и z500 кластеризовались вместе как одна точка в кластеризуемом наборе данных, чтобы учесть одновременно циркуляцию как на нижней границе, так и в свободной тропосфере. Аналогичная методика с одновременной кластеризацией нескольких переменных применялась, например, в [57, 106]. Чтобы уравновесить вклад главных компонент MSLP и z500 в процедуре кластеризации, временные ряды каждой из первых 50 главных компонент полей MSLP и z500 были нормированы на стандартное отклонение первой главной компоненты. Уже к нормированным временным рядам главных компонент применялся метод кластеризации k-means с оптимизацией simulated annealing [46].

Для определения оптимального числа кластеров использовался индекс классифицируемости [43], описанный в разделе 1.3. Результаты представлены на рисунке 4.9. Получено, что статистически значимым числом погодных режимов в Европе при кластеризации одновременно MSLP и z500 является $K=4$, либо $K \geq 6$, однако наибольшее различие в качестве кластеризации между оригинальной выборкой данных и случайно сгенерированными выборками получена при $K=7$, которое и выбрано для дальнейшего анализа связи погодных режимов с осадками в Крыму. Выбор большего K по сравнению с регионами EAT и CE в более узком регионе Европы потенциально может выявить более специфические типы крупномасштабной атмосферной циркуляции, для которых характерна повышенная частота экстремальных осадков в Крыму по сравнению с остальными. При меньших числах ($K = 4$) в одни и те же режимы циркуляции будут сливаться фактически разные поля MSLP и z500, для которых характерна различная частота экстремальных осадков в Крыму.

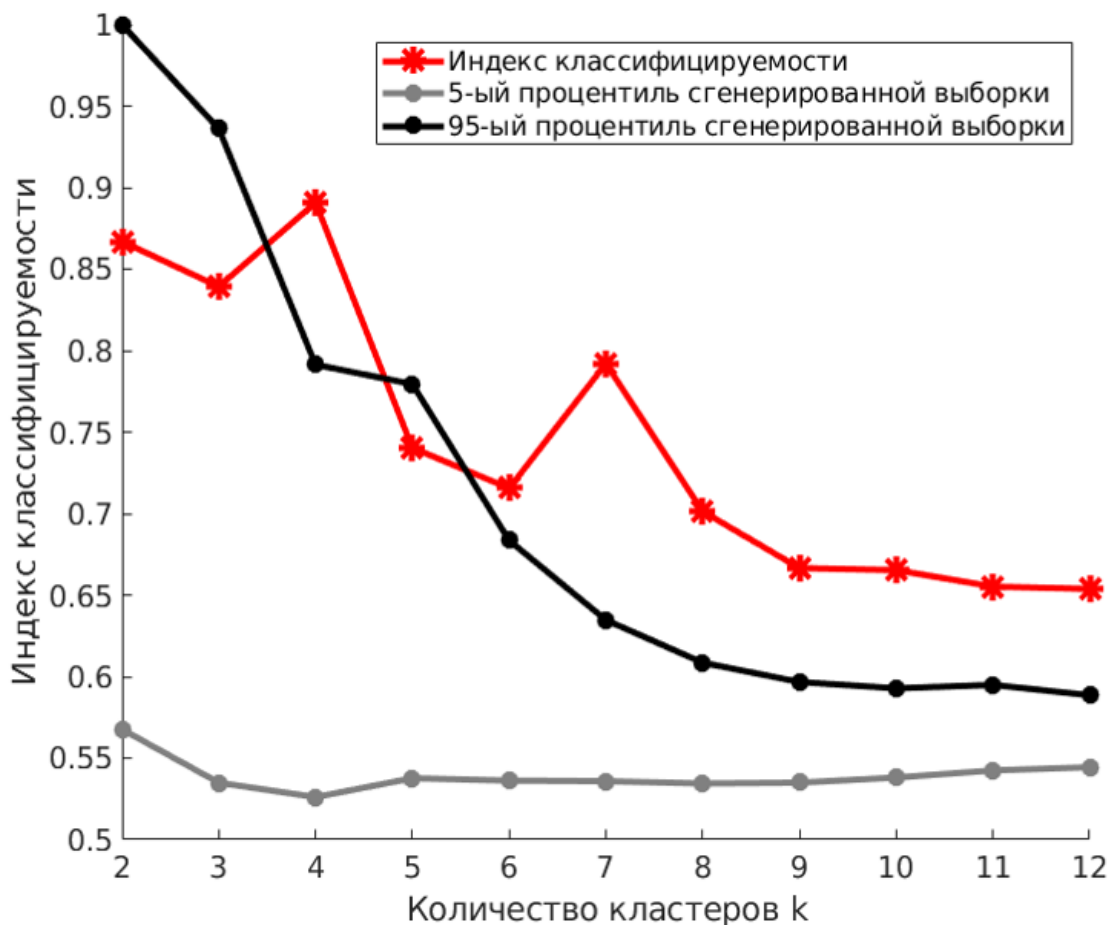


Рисунок 4.9. Значения индекса классифицируемости для летних погодных режимов в Европе от 40 з.д. до 60 в.д. при одновременной кластеризации полей MSLP и z500 методом k-means.

4.4.2. Анализ связи режимов в Европе с осадками в ЮБК

По итогу кластеризации каждые летние сутки из 43 лет (3956 суток) за 1979 – 2021 гг. отнесены к одному из 7 кластеров. На рисунке 4.10 представлены средние поля аномалий z500 по полученным кластерам. Средние поля аномалий MSLP полученных кластеров в целом похожи на поля z500, однако центры аномалий в них смещены на несколько градусов долготы на Запад относительно z500. Наибольшие осадки на ЮБК наблюдались за 17-18 июня 2021 года. По результатам кластеризации дни с 14 по 19 июня 2021 отнесены к режиму #5 (на рисунке 4.10д), 20 июня 2021 отнесен к режиму #1 (рисунок 4.10а). Предполагается, что антициклоническая и блокинговая активность к северу от Крыма способствует сильным осадкам в Крыму. Так, в середине июня 2021, когда наблюдались рекордные осадки на ЮБК [107], по

индексу Тибальди-Мольтени на долготах Крыма стоял блокинг, что проиллюстрировано на рисунке 4.10и.

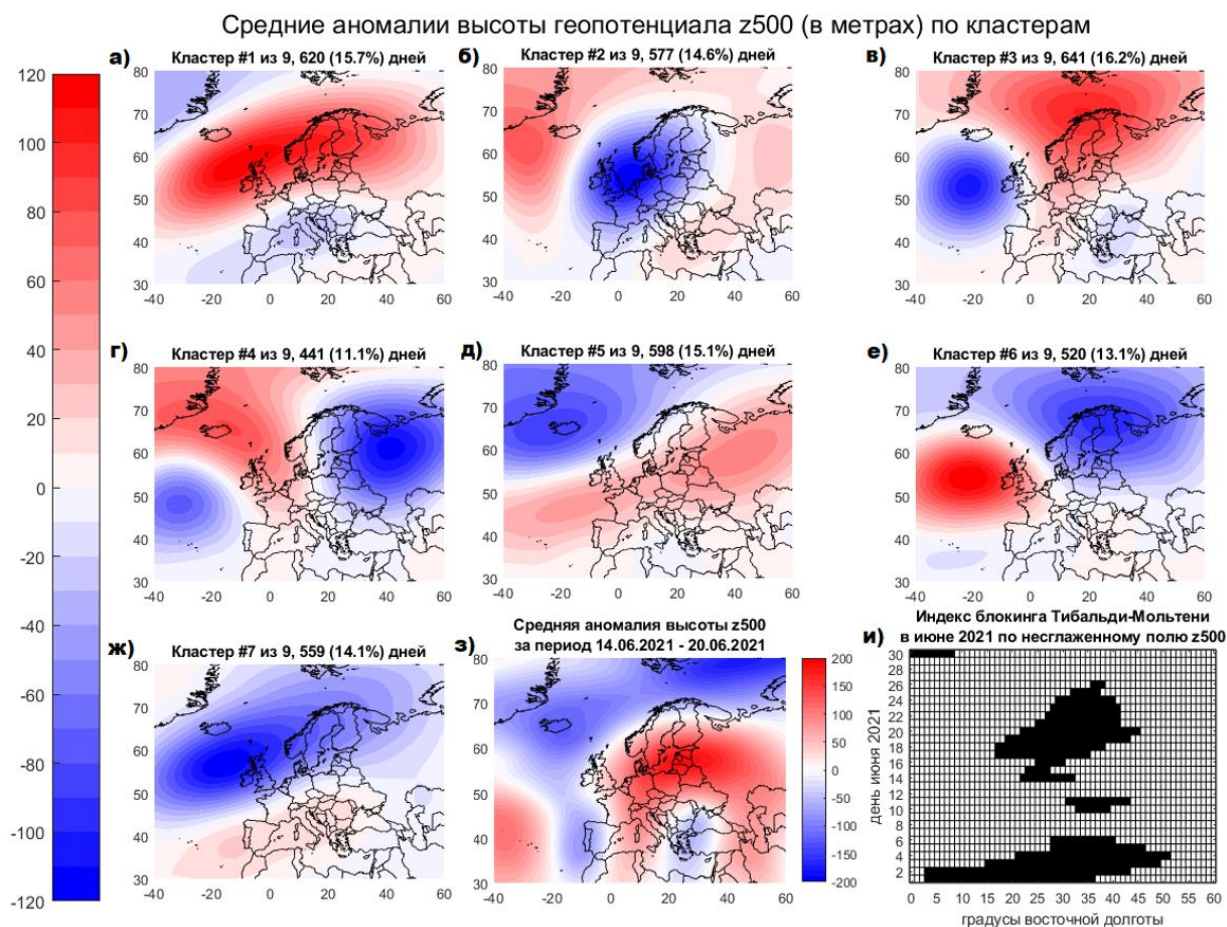


Рисунок 4.10 а-ж: Средние поля аномалий геопотенциала для 7 полученных летних режимов циркуляции. Шкала слева выражена в метрах аномалий высот z500; (з) Среднее за 7 суток поле аномалий высоты геопотенциала z500 в июне 2021-го года, когда наблюдались рекордные для ЮБК осадки (128 мм за 17 июня и 94 мм за 18 июня); (и) Наличие атмосферного блокирования для долгот вокруг Крыма в июне 2021 по индексу Тибальди-Мольтени (закрашенные черные ячейки – блокинг присутствует).

Для полученных погодных режимов в Европе были проанализированы осадки на ЮБК по базе данных E-OBS и станционным наблюдениям осадков на станции в г. Феодосия, для которой имелся наиболее полный временной ряд осадков за исследуемый период 1979-2021 гг. Результаты представлены на рисунке 4.11. Из рисунка видно, что средняя интенсивность осадков на ЮБК не сильно меняется в зависимости от номера кластера (режима циркуляции), однако частота экстремальных осадков, когда суточная сумма осадков превышала 95-ый процентиль интенсивности, меняется в несколько раз в зависимости от номера кластера. По совокупности результатов E-OBS и

станционных наблюдений в Феодосии можно отметить, что наибольшая частота экстремальных осадков приходится на погодные режимы под номерами 1, 3 и 5 (а, в, д на рис. 4.10), эти режимы антициклонические.

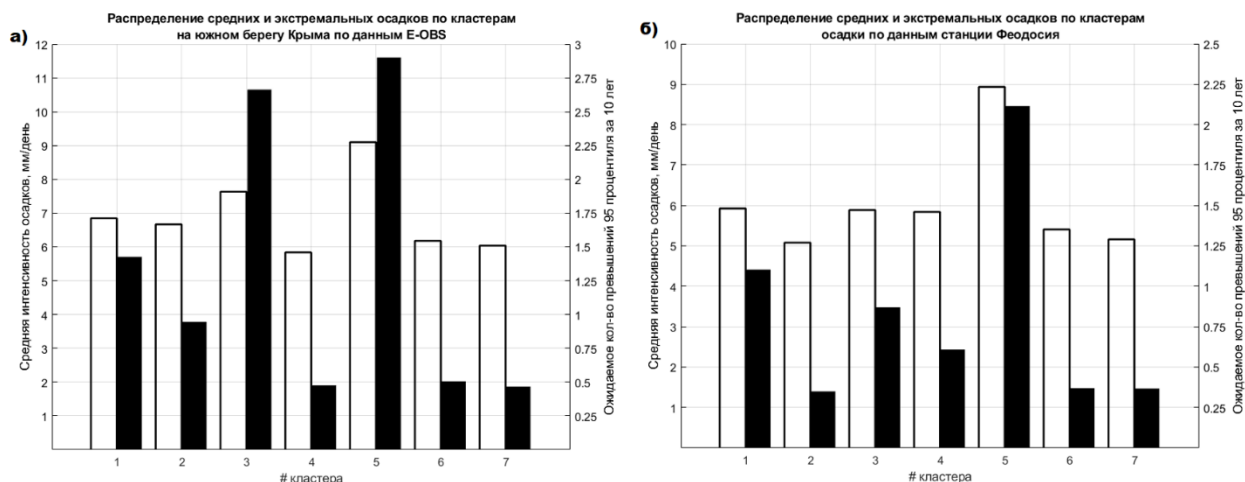


Рисунок 4.11 а): распределение средней интенсивности осадков (левая шкала, белые столбцы) и частоты экстремальных осадков выше 95 перцентиля интенсивности (правая шкала, черные столбцы) в зависимости от номера погодного режима (из рисунка 4.10) по данным E-OBS; б) то же самое, но по данным наблюдений на станции в Феодосии.

Совокупная сезонная повторяемость летних режимов в Европе, для которых наблюдается повышенная частота экстремальных осадков на Южном берегу Крыма, представлена на рисунке 4.12. Из рисунка видно, что для совокупной сезонной повторяемости режимов под номерами 1, 3 и 5 (из рисунка 4.10) наблюдается положительный линейный тренд, значимый на уровне 7% (p -уровень = 0.0677), что может частично объяснять рост числа событий с экстремальными осадками в Крыму за последнее десятилетие, при этом модельные прогнозы для второй половины XXI века не показывают статистически значимых изменений в числе экстремальных осадков для территории Крыма [108]. По отдельности для режимов 1, 3 и 5 не отмечено линейных трендов на уровне значимости менее 10%, что объясняется в том числе высокой изменчивостью их повторяемости от года к году.

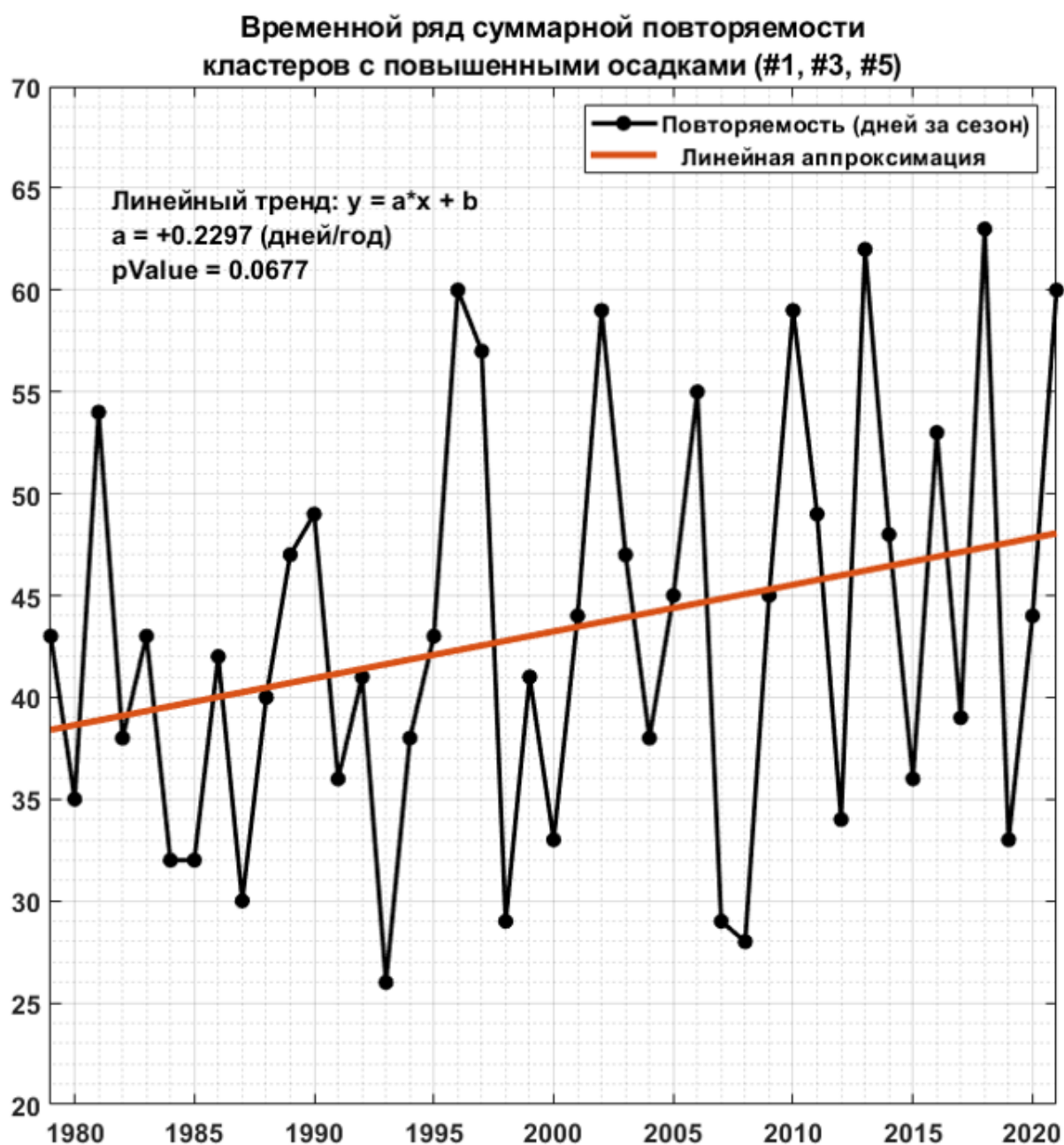


Рисунок 4.12. Временной ряд совокупной сезонной повторяемости режимов циркуляции #1, #3, #5 (из рисунка 4.10), для которых наблюдается повышенная частота экстремальных осадков на южном берегу Крыма.

4.5 Выводы

В четвертой главе настоящей диссертации представлен анализ изменения повторяемости погодных режимов в Северной Евразии в зависимости от индекса явления Эль-Ниньо, Nino3.4, а также от аномалий площади морского льда в Северном полушарии. Показано, что повторяемость некоторых погодных режимов в Северной Евразии в зимние и летние месяцы зависит от индекса Nino 3.4 и аномалий площади морского льда в предшествующие погодным режимам сезоны. Так, положительная фаза Nino3.4 осенью связана с повышенной повторяемостью WSB и сниженной повторяемостью WNEH зимой, а отрицательные аномалии площади морского льда осенью и весной связаны с повышенной повторяемостью режимов уральского максимума в последующие зимние (WUH) и летние (SUN) сезоны.

В разделах 4.2 и 4.3 показано влияние погодных режимов в Северной Евразии и Евро-Атлантике на частоту экстремальных погодных аномалий в Северном полушарии. Показано влияние каждого из полученных автором летних и зимних погодных режимов в СЕ на вероятности экстремальных погодных аномалий в зоне определения режимов, а для четырех зимних погодных режимов в Евро-Атлантике показано влияние на погодные аномалии во всей внетропической зоне Северного полушария. Так, летний режим уральского максимума SUN в Северной Евразии связан с повышенной частотой аномально жарких дней на Европейской территории России, а режим NAO- в Евро-Атлантике связан с кратным ростом вероятности аномально холодных дней на Севере Европы, включая ЕТР. Например, аномально холодная зима в Европе 2010 года совпадает с абсолютным максимумом повторяемости режима NAO- в тот год.

В разделе 4.4 проанализирован частный случай влияния крупномасштабных режимов атмосферной циркуляции на осадки на Южном берегу Крыма. Показано, что во время определенных антициклонических режимов кратно увеличивают вероятность сильных осадков, превышающих 95-й перцентиль интенсивности. Для этих режимов за исследуемый период 1971-2021 гг. наблюдается значимый на уровне 7% положительный тренд их совокупной сезонной повторяемости, что может привести к росту числа экстремальных осадков на ЮБК в будущем.

ЗАКЛЮЧЕНИЕ

В настоящей диссертационной работе проведено исследование крупномасштабной атмосферной циркуляции в различных областях Северного полушария с использованием погодных режимов (ПР), полученных методами кластерного анализа. Основные результаты работы состоят в следующем:

1. Проведен сравнительный анализ методов кластеризации, используемых для идентификации ПР в Евро-Атлантическом регионе. Получено, что среди методов: иерархической кластеризации Уорда (HW), модели гауссовой смеси (GM), k-means (KM) и самоорганизующихся карт Кохонена (SOM), KM является оптимальным. Оптимальность KM основывается на его простоте и способности выделять ПР в фазовом пространстве полей $z500$ с наименьшими суммами внутрикластерных расстояний. Расчет пространственных корреляций средних полей ПР, полученных разными методами, показал, что KM производит наиболее усредненный результат по сравнению с остальными методами.

2. В регионах Евро-Атлантики (ЕАТ) и Северной Евразии (СЕ) методом KM идентифицированы ПР, представлены и проанализированы их характеристики, включая повторяемости, продолжительности и матрицы переходов. Получено, что в зимний период оптимальное число режимов $K = 4$ как для ЕАТ, так и для СЕ. За период 1940–2022 гг. зимой в ЕАТ наблюдается положительный тренд повторяемости NAO^+ и отрицательный тренд повторяемости NAO^- . В Северной Евразии за 1940–2022 гг. повторяемость зимних ПР стабильна, а летом у трёх из четырех ПР статистически значимо на уровне 5% меняется повторяемость, наиболее значимый тренд у режима уральского максимума SUN – его повторяемость растёт на 0.243 суток за год. Для ПР в СЕ получены замкнутые цепочки статистически значимых по частоте переходов, указывающих на неслучайную последовательность смены погодных режимов.

3. Проведен анализ способности климатических моделей воспроизводить характеристики ПР в ЕАТ в экспериментах *historical*. Получено, что ПР, идентифицированные методом k-means в собственном фазовом пространстве модельных полей $z500$, качественно совпадают с полями по данным реанализа ERA5. Средняя пространственная корреляция зимних модельных ПР с ПР реанализа составила 0.85 в зимние месяцы и 0.80

в летние. Повторяемости и продолжительности ПР в моделях зимой воспроизводятся лучше, чем летом.

4. Выявлена связь повторяемости зимних и летних ПР в СЕ с индексом явления Эль-Ниньо, Nino3.4, и площадью морского льда в Северном полушарии в предшествующие сезоны. Явление Эль-Ниньо в осенние месяцы связан с ростом повторяемости режима WSB в последующие зимы. Отрицательные аномалии площади морского льда в Северном полушарии в осенний и весенний период связаны с ростом повторяемости режимов уральского максимума в последующие зимний (WUH) и летний (SUN) сезоны.

5. Показано влияние ПР в ЕАТ и СЕ на вероятности экстремальных аномалий температур, осадков и ветра. Летний режим SUN связан с ростом вероятности жарких дней на европейской территории России 2-3 раза, а зимних режим NAO- в ЕАТ связан с кратным ростом аномально холодных дней на территории Европы. Показана связь антициклонических погодных режимов в Европе с экстремальными осадками на южном берегу Крыма, и выявлен рост совокупной повторяемости этих режимов в период 1979–2021 гг.

СПИСОК ЛИТЕРАТУРЫ

1. Barnston A. G., Livezey R. E. Classification, seasonality and persistence of low-frequency atmospheric circulation patterns // *Monthly weather review*. 1987. V. 115. N. 6. P. 1083-1126.
2. Wallace J. M., Gutzler D. S. Teleconnections in the geopotential height field during the Northern Hemisphere winter // *Monthly weather review*. 1981. V. 109. N. 4. P. 784-812.
3. Ghil M., Robertson A. W. "Waves" vs. "particles" in the atmosphere's phase space: A pathway to long-range forecasting? // *Proceedings of the National Academy of Sciences*. 2002. V. 99. N. suppl_1. P. 2493-2500.
4. Cheng X., Wallace J. M. Cluster analysis of the Northern Hemisphere wintertime 500-hPa height field: Spatial patterns // *Journal of atmospheric sciences*. 1993. V. 50. N. 16. P. 2674-2696.
5. Vautard R. Multiple weather regimes over the North Atlantic: Analysis of precursors and successors // *Monthly weather review*. 1990. V. 118. N. 10. P. 2056-2081.
6. Baur, F., P. Hess & H. Nagel. Kalendar der Grosswetterlagen Europas 1881-1939 // Bad Homburg, Germany. 1944.
7. Дзердзеевский Б. Л., Курганская В. М., Витвицкая З. М. (1946). Типизация циркуляционных механизмов в северном полушарии и характеристика синоптических сезонов // Тр. науч.-исслед. учреждений Гл. упр. гидрометеорол. службы при Совете Министров СССР. Сер. 2. Синоптическая метеорология. Вып. 21. М.–Л.: Гидрометиздат, 1946. 80 с. 8.
8. Вангенгейм Г. Я. Основы макроциркуляционного метода долгосрочных метеорологических прогнозов для Арктики // *Труды ААНИИ*. 1952. Т. 34. С. 314.
9. Huth R. et al. Classifications of atmospheric circulation patterns: recent advances and applications // *Annals of the New York Academy of Sciences*. 2008. V. 1146. N. 1. P. 105-152.
10. James P. M. An objective classification method for Hess and Brezowsky Grosswetterlagen over Europe // *Theoretical and Applied Climatology*. 2007. V. 88. P. 17-42.
11. Lund I. A. Map-pattern classification by statistical methods // *Journal of Applied Meteorology* (1962-1982). 1963. P. 56-65.
12. Jones P. D., Hulme M., Briffa K. R. A comparison of Lamb circulation types with an objective classification scheme // *International Journal of Climatology*. 1993. V. 13. N. 6. P. 655-663.

13. Gritsun A. Statistical characteristics, circulation regimes and unstable periodic orbits of a barotropic atmospheric model // Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences. 2013. V. 371. N. 1991. P. 20120336.
14. Kondrashov D. et al. Predicting weather regime transitions in Northern Hemisphere datasets // Climate Dynamics. 2007. V. 29. P. 535-551.
15. Kimoto M., Ghil M. Multiple flow regimes in the Northern Hemisphere winter. Part I: Methodology and hemispheric regimes // Journal of Atmospheric Sciences. 1993. V. 50. N. 16. P. 2625-2644.
16. Robertson A. W., Ghil M. Large-scale weather regimes and local climate over the western United States // Journal of Climate. 1999. V. 12. N. 6. P. 1796-1813.
17. Dole R. M., Gordon N. D. Persistent anomalies of the extratropical Northern Hemisphere wintertime circulation: Geographical distribution and regional persistence characteristics // Monthly Weather Review. 1983. V. 111. N. 8. P. 1567-1586.
18. Deligiorgi D., Philippopoulos K., Kouroupetroglou G. An assessment of self-organizing maps and k-means clustering approaches for atmospheric circulation classification // Recent Advances in Environmental Science and Geoscience. 2014. V. 17.
19. Cassou C. Intraseasonal interaction between the Madden–Julian oscillation and the North Atlantic Oscillation // Nature. 2008. V. 455. N. 7212. P. 523-527.
20. Santos J. A., Corte-Real J., Leite S. M. Weather regimes and their connection to the winter rainfall in Portugal // International Journal of Climatology: A Journal of the Royal Meteorological Society. 2005. V. 25. N. 1. P. 33-50.
21. Christiansen B. Atmospheric circulation regimes: Can cluster analysis provide the number? // Journal of Climate. 2007. V. 20. N. 10. P. 2229-2250.
22. Visbeck M. H. et al. The North Atlantic Oscillation: past, present, and future // Proceedings of the National Academy of Sciences. 2001. V. 98. N. 23. P. 12876-12877.
23. Dawson A., Palmer T. N., Corti S. Simulating regime structures in weather and climate prediction models // Geophysical Research Letters. 2012. V. 39. N. 21.
24. Falkena S. K. J. et al. Revisiting the identification of wintertime atmospheric circulation regimes in the Euro-Atlantic sector // Quarterly Journal of the Royal Meteorological Society. 2020. V. 146. N. 731. P. 2801-2814.
25. Lee S. H., Furtado J. C., Charlton-Perez A. J. Wintertime North American weather regimes and the Arctic stratospheric polar vortex // Geophysical Research Letters. 2019. V. 46. N. 24. P. 14892-14900.

26. Yang S., Yuan H., Dong L. Offshore wind resource assessment by characterizing weather regimes based on self-organizing map // *Environmental Research Letters*. 2022. V. 17. N. 12. P. 124009.
27. Fabiano F. et al. A regime view of future atmospheric circulation changes in northern mid-latitudes // *Weather and Climate Dynamics Discussions*. 2020. V. 2020. P. 1-24.
28. Yiou P., Nogaj M. Extreme climatic events and weather regimes over the North Atlantic: when and where? // *Geophysical Research Letters*. 2004. V. 31. N. 7.
29. Cattiaux J. et al. Winter 2010 in Europe: A cold extreme in a warming climate // *Geophysical Research Letters*. 2010. V. 37. N. 20.
30. Garrido-Perez J. M. et al. Impact of weather regimes on wind power variability in western Europe // *Applied Energy*. 2020. V. 264. P. 114731.
31. Babanov B. A., Semenov V. A. Weather Regimes in Northern Eurasia: Statistics, Predictability and Associated Weather Anomalies // *Atmosphere*. 2024. V. 15. N. 11. P. 1392.
32. Lorenz E. N. Deterministic nonperiodic flow // *Journal of atmospheric sciences*. 1963. V. 20. N. 2. P. 130-141.
33. Greene C. A. et al. The climate data toolbox for MATLAB // *Geochemistry, Geophysics, Geosystems*. 2019. V. 20. N. 7. P. 3774-3781.
34. Hess, P. & H. Brezowsky. 1952. Katalog der Großwetterlagen Europas. Ber. Dt. Wetterd. in der US-Zone 33. Bad Kissingen, Germany.
35. Lamb, H.H. 1972. British Isles Weather Types and a Register of Daily Sequence of Circulation Patterns, 1861-1971. *Geophysical Memoir 116*, HMSO, London (UK), 85pp.
36. Гирс А. А. (1974). Макроциркуляционный метод долгосрочных метеорологических прогнозов. Л.: Гидрометеиздат. 485 с.
37. Дзердзеевский Б.Л.. Циркуляционные механизмы в атмосфере северного полушария в XX столетии // *Материалы метеорологических исследований*. – М.: Изд. ИГ АН СССР и Междувед. Геофиз. Комитета при Президиуме АН СССР, 1968. – 240 с.
38. Mukhin D. et al. Revealing recurrent regimes of mid-latitude atmospheric variability using novel machine learning method // *Chaos: An Interdisciplinary Journal of Nonlinear Science*. 2022. V. 32. N. 11.
39. Бабанов Б.А., Семенов В.А., Мохов И.И. Сравнение различных методов кластеризации для определения погодных режимов в Евро-Атлантическом секторе в зимний и летний сезоны // *Известия Российской академии наук. Физика атмосферы и океана*. 2023. Т. 59. №. 6. С. 686-706.

40. Roux M. A comparative study of divisive and agglomerative hierarchical clustering algorithms // J. Classification. 2018. V. 35. P. 345–366.
41. Схематическое изображение агломеративной кластеризации из сети Интернет [Электронный ресурс]. URL: <https://quantdare.com/wp-content/uploads/2016/06/AggloDivHierarClustering-800x389.png>
42. Ward Jr J. H. Hierarchical grouping to optimize an objective function // Journal of the American statistical association. 1963. V. 58. N. 301. P. 236-244.
43. Michelangeli P. A., Vautard R., Legras B. Weather regimes: Recurrence and quasi stationarity // Journal of the atmospheric sciences. 1995. V. 52. N. 8. P. 1237-1256.
44. Arthur D., Vassilvitskii S. K-means++ the advantages of careful seeding // Proceedings of the eighteenth annual ACM-SIAM symposium on Discrete algorithms. 2007. P. 1027–1035.
45. Selim S. Z., Alsultan K. A simulated annealing algorithm for the clustering problem // Pattern recognition. 1991. V. 24. N 10. P. 1003-1008.
46. Philipp A. et al. Long-term variability of daily North Atlantic–European pressure patterns since 1850 classified by simulated annealing clustering // Journal of Climate. 2007. V. 20. N. 16. P. 4065-4095.
47. Molteni F., Tibaldi S., Palmer T. N. Regimes in the wintertime circulation over northern extratropics. I: Observational evidence // Quarterly Journal of the Royal Meteorological Society. 1990. V. 116. N 491. P. 31-67.
48. Khan K., Rehman S. U., Aziz K., Fong S., and Sarasvady S. DBSCAN: Past, present and future // The fifth international conference on the applications of digital information and web technologies (ICADIWT 2014). IEEE, 2014. P. 232-238.
49. Banfield J. D., Raftery A. E. Model-based Gaussian and non-Gaussian clustering // Biometrics. 1993. P. 803-821.
50. Kearns M., Mansour Y., Ng A. Y. An information-theoretic analysis of hard and soft assignment methods for clustering // Learning in graphical models. 1998. P. 495-520.
51. Bilmes J. A. A gentle tutorial of the EM algorithm and its application to parameter estimation for Gaussian mixture and hidden Markov models // International computer science institute. 1998. V. 4. N 510. P. 126.
52. Yang M. S., Lai C. Y., Lin C. Y. A robust EM clustering algorithm for Gaussian mixture models // Pattern Recognition. 2012. V. 45. N 11. P. 3950-3961.
53. Smyth P., Ide K., Ghil M. Multiple regimes in northern hemisphere height fields via mixture model clustering // Journal of the Atmospheric Sciences. 1999. V. 56. N 21. P. 3704-3723.

54. Hannachi A. Low-frequency variability in a GCM: Three-dimensional flow regimes and their dynamics // *Journal of climate*. 1997. V. 10. N 6. P. 1357-1379.
55. Kohonen T. Self-organizing maps. Springer Science & Business Media. 2012. V. 30.
56. Polo I., Ullmann A., Roucou P., and Fontaine B. Weather regimes in the Euro-Atlantic and Mediterranean sector, and relationship with West African rainfall over the 1989–2008 period from a self-organizing maps approach // *Journal of Climate*. 2011. V. 24. N 13. P. 3423-3432.
57. Loikith P. C., Lintner B. R., Sweeney A. Characterizing large-scale meteorological patterns and associated temperature and precipitation extremes over the northwestern United States using self-organizing maps // *Journal of Climate*. 2017. V. 30. N 8. P. 2829-2847.
58. Shi C., Wei B., Wei S. et al. A quantitative discriminant method of elbow point for the optimal number of clusters in clustering algorithm // *EURASIP Journal on Wireless Communications and Networking*. 2021. V. 2021. N 1. P. 1-16.
59. Caliński T., Harabasz J. A dendrite method for cluster analysis // *Communications in Statistics-theory and Methods*. 1974. V. 3. N. 1. P. 1-27.
60. Crasemann B. et al. Can preferred atmospheric circulation patterns over the North-Atlantic-Eurasian region be associated with arctic sea ice loss? // *Polar Science*. 2017. V. 14. P. 9-20.
61. Matsueda M., Palmer T. N. Estimates of flow-dependent predictability of wintertime Euro-Atlantic weather regimes in medium-range forecasts // *Quarterly Journal of the Royal Meteorological Society*. 2018. V. 144. N. 713. P. 1012-1027.
62. Guemas V., Salas-Mélia D., Kageyama M., Giordani H., Voldoire A., and Sanchez-Gomez E. Summer interactions between weather regimes and surface ocean in the North-Atlantic region // *Climate dynamics*. 2010. V. 34. P. 527-546.
63. Folland C. K., Knight J., Linderholm H. W., Fereday D., Ineson S., and Hurrell J. W. The summer North Atlantic Oscillation: past, present, and future // *Journal of Climate*. 2009. V. 22. N 5. P. 1082-1103.
64. Hersbach H. et al. The ERA5 global reanalysis // *Quarterly Journal of the Royal Meteorological Society*. 2020. V. 146. N 730. P. 1999-2049.
65. Selesnick I. W., Burrus C. S. Generalized digital Butterworth filter design // *IEEE Transactions on signal processing*. 1998. V. 46. N 6. P. 1688-1694.
66. Charlton-Perez A. J., Ferranti L., Lee R. W. The influence of the stratospheric state on North Atlantic weather regimes // *Quarterly Journal of the Royal Meteorological Society*. 2018. V. 144. N 713. P. 1140-1151.

67. Vautard R., Mo K. C., Ghil M. Statistical significance test for transition matrices of atmospheric Markov chains // *Journal of Atmospheric Sciences*. 1990. V. 47. N 15. P. 1926-1931.
68. Kondrashov D., Ide K., Ghil M. Weather regimes and preferred transition paths in a three-level quasigeostrophic model // *Journal of the atmospheric sciences*. 2004. V. 61. N 5. P. 568-587.
69. Бардин М.Ю., Платова V.B. Долгопериодные вариации показателей экстремальности температурного режима на территории России и их связь с изменениями крупномасштабной атмосферной циркуляции и глобальным потеплением // *Метеорол. и гидрол.* 2019. №. 12. С. 5–19.
70. Liu Y. et al. How do North American weather regimes drive wind energy at the sub-seasonal to seasonal timescales? // *npj Climate and Atmospheric Science*. 2023. V. 6. N. 1. P. 100.
71. Mann H. B. Nonparametric tests against trend // *Econometrica: Journal of the econometric society*. 1945. P. 245-259.
72. Dole R. et al. Was there a basis for anticipating the 2010 Russian heat wave? // *Geophysical research letters*. 2011. V. 38. N. 6.
73. Delgado-Torres C. et al. Representation and Annual to Decadal Predictability of Euro-Atlantic Weather Regimes in the CMIP6 Version of the EC-Earth Coupled Climate Model // *Journal of Geophysical Research: Atmospheres*. 2022. V. 127. N 14. e2022JD036673.
74. Volodin E., Gritsun A. Simulation of observed climate changes in 1850–2014 with climate model INM-CM5 // *Earth System Dynamics*. 2018. V. 9. N. 4. P. 1235-1242.
75. Fabiano F. et al. Euro-Atlantic weather Regimes in the PRIMAVERA coupled climate simulations: impact of resolution and mean state biases on model performance // *Climate Dynamics*. 2020. V. 54. P. 5031-5048.
76. Cassou C., Terray L., Phillips A. S. Tropical Atlantic influence on European heat waves // *Journal of climate*. 2005. V. 18. N. 15. P. 2805-2811.
77. Mukhin D. et al. Principal nonlinear dynamical modes of climate variability // *Scientific reports*. 2015. V. 5. N. 1. P. 15510.
78. López-Parages J. et al. ENSO influence on the North Atlantic European climate: A non-linear and non-stationary approach // *Climate Dynamics*. 2016. V. 47. P. 2071-2084.
79. Мохов, И. И., Тимажев А. В. Повторяемость зимних атмосферных блокирований в Северном полушарии при различных фазах Эль-Ниньо, Тихоокеанской десятилетней и Атлантической мультидесятилетней

осцилляций // Известия Российской академии наук. Физика атмосферы и океана. 2024. Т. 60. №. 5. С. 653-666.

80. Мохов И. И., Тимажев А. В. Повторяемость летних атмосферных блокирований в Северном полушарии в разных фазах явлений Эль-Ниньо, Тихоокеанской десятилетней и Атлантической мультидесятилетней осцилляций // Известия Российской академии наук. Физика атмосферы и океана. 2022. Т. 58. №. 3. С. 1-11.

81. Barnes E. A., Screen J. A. The impact of Arctic warming on the midlatitude jet-stream: Can it? Has it? Will it? // Wiley Interdisciplinary Reviews: Climate Change. 2015. V. 6. N. 3. P. 277-286.

82. Scaife A. A. et al. ENSO affects the North Atlantic Oscillation 1 year later // Science. 2024. V. 386. N. 6717. P. 82-86.

83. Mukhin D. et al. An atmospheric signal lowering the spring predictability barrier in statistical ENSO forecasts // Geophysical Research Letters. 2021. V. 48. N. 6. e2020GL091287.

84. Seleznev A., Mukhin D. Improving statistical prediction and revealing nonlinearity of ENSO using observations of ocean heat content in the tropical Pacific // Climate Dynamics. 2023. V. 60. N. 1. P. 1-15.

85. Gavrillov A. et al. Linear dynamical modes as new variables for data-driven ENSO forecast // Climate Dynamics. 2019. V. 52. P. 2199-2216.

86. Gavrillov A., Kravtsov S., Mukhin D. Analysis of 20th century surface air temperature using linear dynamical modes // Chaos: An Interdisciplinary Journal of Nonlinear Science. 2020. V. 30. N. 12.

87. Trenberth K. E. The definition of el nino // Bulletin of the American Meteorological Society. 1997. V. 78. N. 12. P. 2771-2778.

88. Cavalieri D. J., Parkinson C. L. Arctic sea ice variability and trends, 1979–2010 // The Cryosphere. 2012. V. 6. N. 4. P. 881-889.

89. Mokhov I. I., Timazhev A. V. Seasonal temperature extremes in the North Eurasian regions depending on ENSO phase transitions // Atmosphere. 2022. V. 13. N. 2. P. 249.

90. Мохов И. И. Аномальные зимы в регионах Северной Евразии в разных фазах явлений Эль-Ниньо // Доклады Российской академии наук. Науки о Земле. 2020. Т. 493. №. 2. С. 93-98.

91. Bardin M. Y., Platova T. V., Samokhina O. F. Extreme Heat Waves and Extreme Summer Seasons in European Russia // *Russian Meteorology and Hydrology*. 2024. V. 49. N. 6. P. 467-481.
92. Lejenäs H., Økland H. Characteristics of Northern Hemisphere blocking as determined from a long time series of observational data // *Tellus A*. 1983. V. 35. N. 5. P. 350-362.
93. Tibaldi S., Molteni F. On the operational predictability of blocking // *Tellus A*. 1990. V. 42. N. 3. P. 343-365.
94. Lucarini V., Gritsun A. A new mathematical framework for atmospheric blocking events // *Climate Dynamics*. 2020. V. 54. N. 1. P. 575-598.
95. Lupo A. R. et al. Changes in global blocking character in recent decades // *Atmosphere*. 2019. V. 10. N. 2. P. 92.
96. Krauskopf T., Huth R. Temperature trends in Europe: comparison of different data sources // *Theoretical and Applied Climatology*. 2020. V. 139. N. 3. P. 1305-1316.
97. Гавриков А. В., Гулев Р. К. Высокора разрешающее климатическое моделирование атмосферы над Северной Атлантикой для задач исследования атмосферы и океана // *Океанология*. 2020. Т. 60. №. 6. С. 831-834.
98. Christidis N., Stott P. A. Changes in the geopotential height at 500 hPa under the influence of external climatic forcings // *Geophysical Research Letters*. 2015. V. 42. N. 24. P. 10798-10806.
99. Индекс NAO из базы данных Climate Prediction Center NCEP [Электронный ресурс]. URL: https://ftp.cpc.ncep.noaa.gov/wd52dg/data/indices/nao_index.tim.
100. Шукуров К. А., Семенов В. А. Характеристики зимних аномалий приземной температуры в Москве в 1970-2016 гг. при сокращении площади морских льдов в Баренцевом море // *Известия Российской академии наук. Физика атмосферы и океана*. 2018. Т. 54. №. 1. С. 13-27.
101. Ефимов В. В., Анисимов А. Е., Комаровская О. И. Характеристики экстремальных атмосферных осадков в Крыму // *Экологическая безопасность прибрежной и шельфовой зон моря*. 2022. № 2. С. 6–18.
102. Evstigneev V. P. et al. Severe precipitation phenomena in Crimea in relation to atmospheric circulation // *Atmosphere*. 2022. V. 13. N. 10. P. 1712.
103. Кононова Н. К. и др. Связь экстремальных осадков в Крыму с макроциркуляционными процессами // *Сложные системы*. 2020. №. 3. С. 4-29.
104. Кононова Н. К. Циркуляция атмосферы как фактор стихийных бедствий на Северном Кавказе в XXI веке // *Геополитика и экогеодинамика регионов*. 2012. Т. 8. №. 1-2. С. 72-103.

105. Cornes R. C. et al. An ensemble version of the E-OBS temperature and precipitation data sets // *Journal of Geophysical Research: Atmospheres*. 2018. V. 123. N. 17. P. 9391-9409.
106. Lennard C., Hegerl G. Relating changes in synoptic circulation to the surface rainfall response using self-organising maps // *Climate Dynamics*. 2015. V. 44. P. 861-879.
107. Гавриков А. В. и др. Экстремальные осадки в июне 2021 г. над Черным морем в контексте их долгопериодной климатической изменчивости // *Океанология*. 2022. Т. 62. №. 3. С. 357-364.
108. Semenov V. A., Aleshina M. A. Scenario-based forecasts of changes in the temperature and hydrological regime of Crimea in the XXI century by data of CMIP6 climate models // *Water Resources*. 2022. V. 49. N. 4. P. 661-670.