МИНИСТЕРСТВО НАУКИ И ВЫСШЕГО ОБРАЗОВАНИЯ РОССИЙСКОЙ ФЕДЕРАЦИИ

федеральное государственное бюджетное образовательное учреждение

высшего образования

«РОССИЙСКИЙ ГОСУДАРСТВЕННЫЙ

ГИДРОМЕТЕОРОЛОГИЧЕСКИЙ УНИВЕРСИТЕТ»

Кафедра прикладной океанологии ЮНЕСКО-МОК и КУПЗ

ВЫПУСКНАЯ КВАЛИФИКАЦИОННАЯ РАБОТА

(Магистерская диссертация)

«Уровень Каспия как индикатор влагообмена системы океан-атмосфера-суша»

Исполнитель: Лактионов Павел Петрович, студент группы ПО-М18-1-7

Научный руководитель: д.г.н., профессор, Малинин Валерий Николаевич

«К защите допускаю»

Заведующий кафедрой

_____

(подпись)

_____

(ученая степень, ученое звание)

_____

(фамилия, имя, отчество)

«___»_____20__г.

Санкт–Петербург

2020

Введение

Актуальность работы заключается в том, что Каспийское море является уникальным водоемом, в водах которого встречается более 850 видов животных, более 450 видов растений. В недрах Каспия сосредоточены тысячи тонн полезных ископаемых, проходят границы пяти стран, добывается 90% всей осетровой икры.

Проблема выявления генезиса колебаний уровня Каспийского моря имеет не только огромное исследовательское, но и экономическое значение. Но, несмотря на это, на текущий момент времени исследователи не дали на этот вопрос достоверный и однозначный ответ. Нестеров Е.С пишет об этом так: «…данная область знаний еще не достигла того научного уровня, при котором ее основные положения и принципы были бы окончательно установлены и общепризнаны…». В этой работе мы выдвинем гипотезу о том, что уровень Каспийского моря является интегральным показателем тепло-влагообмена системы «Океан – Атмосфера - Суша»

Целью работы является выявление влияния тепло-влагообмена в Северной Атлантике на межгодовую изменчивость уровня Каспийского моря.

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

1. Обоснование влияния энергоактивных зон Северной Атлантики на горизонтальный перенос влаги с океана на материк.
2. Построение диагностических моделей деревьев решений суммарной теплоотдачи в энергоактивных зонах Северной Атлантики с переносом водяного пара на меридиональном разрезе 5° в. д.
3. Выделение реперных квадратов в ЭАЗО имеющих наилучшую связь со значениями зонального потока влаги на меридиональном разрезе 5° в. д.

4. Построение прогностических моделей суммарной теплоотдачи Ньюфаулендской и Норвежской энергоактивных зон с зональным переносом влаги методом деревьев решений.
5. Построение диагностических моделей суммарной теплоотдачи в энергоактивных зонах Северной Атлантики с приращениями уровня Каспийского моря.

Научная новизна работы состоит в том, что впервые в данной работе использовались данные суммарной теплоотдачи в атмосферу из энергоактивных зон Северной Атлантики для прогноза зонального переноса водяного пара на меридиональном разрезе 5° в. д.

Личный вклад автора заключается в том, что самостоятельно обработаны исходные данные зональных потоков влаги и суммарной теплоотдачи ЭАЗ Северной Атлантики, построены прогностические и диагностические модели деревьев решений зонального переноса влаги через меридиональный разрез 5° в.д., построена диагностическая модель приращений уровня Каспийского моря. Выполнен научный анализ полученных результатов. Это позволило подготовить к публикации статью «Долгосрочный прогноз уровня Каспийского моря» в сборнике «КИМО 2020.

Положения, выносимые на защиту:

1. Энергоактивные зоны океана в Северной Атлантике как важный фактор влияния на межгодовую изменчивость зонального переноса влаги с океана на материк.
2. Диагностические и прогностические модели деревьев решений с зональным переносом влаги на меридиональном разрезе 5° в.д.

3. Диагностические модели суммарной теплоотдачи в энергоактивных зонах Северной Атлантики с приращениями уровня Каспийского моря.

Практическая значимость работы состоит в разработке диагностических и прогностических моделей суммарной теплоотдачи энергоактивных зон со значениями зонального потока влаги, а также разработке диагностических моделей с приращениями уровня Каспийского моря.

Работа состоит из четырех разделов, введения и заключения.

В первом разделе представлен обзор влияния тепло-влагообмена Северной Атлантики на формирование осадков в бассейне Каспийского моря. Дается описание генетической концепции образования межгодовых колебаний увлажнения европейской части территорий России, предложенной Малининым В.Н. Обсуждается концепция энергоактивных зон Северной Атлантики.

Во втором разделе описываются исходные данные и метод деревьев решений.

В третьем разделе описывается выделение реперных квадратов в ЭАЗО, оценивается влияние суммарных потоков тепла в энергоактивных зонах Северной Атлантики на меридиональный перенос водяного пара на 5° в.д. на основе модели деревьев решений

В четвертом разделе выполнен прогноз значений зональных потоков влаги на меридиане 5° в. д. Строится диагностическая модель суммарной теплоотдачи ЭАЗ Северной Атлантики с приращениями уровня Каспийского моря.

В заключении сформулированы выводы, дан ответ на вопрос можно ли воспринимать уровень Каспия, как показатель влагообмена в системе “океан-атмосфера-суша”.

Содержание:

1. Связь между уровнем Каспийского моря и Северной Атлантикой.

1.1 Обзор уровня Каспия, как глобального индикатора системы «Океан – атмосфера – Суша». Прогноз приращений УКМ.

Осадки, выпадающие в бассейне реки Волги, почти полностью зависят от процессов теплообмена в Северной Атлантике. Это утверждение большое количество раз подтверждалось различными работами, к примеру [1,2,3,4,5]. В работе [1] рассматривалась генетическая концепция образования межгодовых колебаний увлажнения территории ЕЧР, предложенная В.Н. Малининым. Суть концепции изображена на схеме ниже (Рис. 1.1).

Рисунок 1.1. Алгоритм формирования межгодовых колебаний УКМ (по В.Н. Малинину [4]).

«…Если в результате крупномасштабного взаимодействия океана и атмосферы в СА в холодный период года усиливается зональный перенос водяного пара на континент, включая ЕЧР, то, соответственно, большее

количество осадков выпадает в бассейне р. Волги и других крупных рек ЕЧР, что усиливает увлажнение и приводит к повышению годового стока рек. Если зональный перенос влаги ослабевает и, соответственно, усиливается меридиональный перенос, то картина меняется на противоположную…» (Малинин, Гордеева, 2020).

Также, чем сильнее циклоническая активность в Северной Атлантике, тем больше воды испаряется с поверхности, и тем ниже становятся значения ТПО, в свою очередь увеличение циклонической активности приводит к росту скорости горизонтального переноса атмосферной влаги.

Эта концепция формирования межгодовых колебаний уровня Каспия нашла отражение в том числе и разработке долгосрочного прогноза уровня Каспийского моря. [4,5]

В работе [5] прогнозировалось приращение уровня Каспийского моря с использованием данных суммарной теплоотдачи ЭАЗО Северной Атлантики (Подробнее об ЭАЗО в разделе 1.3).

Значения приращения уровня прогнозировались на год вперед с помощью модели множественной линейной регрессии. Даже с 5 переменными R^2=0.7, с 21 переменной уже равен 0.99, что говорит о крайне высокой степени связи между значениями суммарной теплоотдачи ЭАЗО Северной Атлантики и УКМ (Таб. 1.1).

На рис 1.2 представлен график спрогнозированных и фактических значений. Модель с 21 предиктором практически полностью повторяет график хода приращений уровня, с 5 – имеет допустимые ошибки. В таблице 1.1 приводятся статистические характеристики, определяющие высокий уровень качества прогноза.

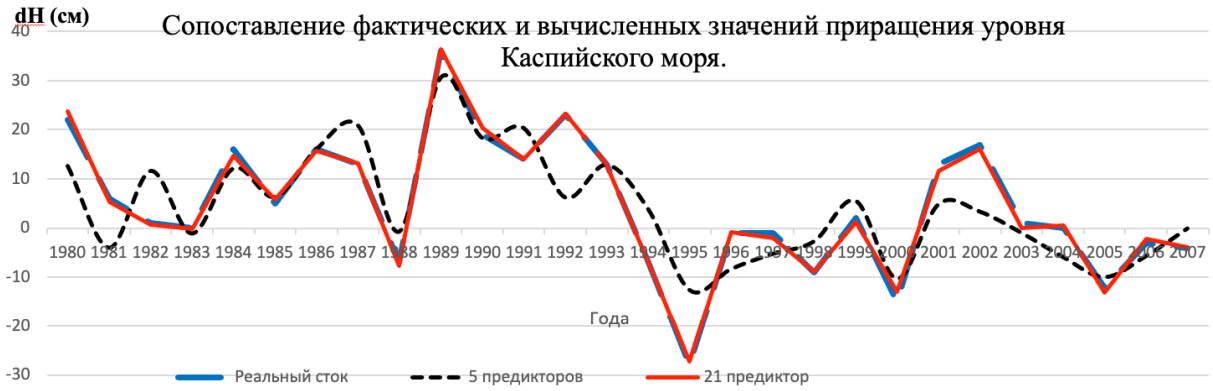


Рисунок 1.2. График фактических и спрогнозированных приращений УКМ (5, 21 предикторов модели) с заблаговременностью 1 год, на основе данных о суммарной теплоотдаче в ЭАЗО СА методом МЛР.

Таблица 1.1. Статистические параметры модели прогноза УКМ (5, 21 предикторов модели) с заблаговременностью 1 год, на основе данных о суммарной теплоотдаче в ЭАЗО СА методом МЛР.

Сдвиг	Число переменных	Характеристики		
		К-т детерминации	К-т Фишера	Стандартная ошибка, см/год
1 год	21	0,99	67,1	1,87
	5	0,70	9,44	21,08

Также стоит отметить важнейшую роль, через зонального переноса влаги через меридиональный разрезе 5° в.д, именно через него, как в холодный, так и в теплый период года вносится более половины всех осадков на ЕЧР[3]. На западную часть ЕЧР, оказывается большее влияние, чем на восточную. Средний вклад зонального потока влаги в дисперсию осадков ЕЧР составил R^2 = 0.55. Что подтверждает гипотезу о том, что зональный поток влияет на количество осадков в бассейне Волги, а следовательно и на уровень Каспийского моря.

1.1 Исторический очерк об изучении связи уровня Каспийского моря и Северной Атлантикой.

То как влияет Северная Атлантика на формирование атмосферных процессов, а, следовательно, и погоды, на ЕТР неоднократно поднималась в различной литературе [6-13]. Необходимо понимать, что от теплового состояния СА зависит погода на ЕЧР, да и на континенте в целом. Тенденция увлажнения [14], длившаяся на протяжении целого столетия (1891-1990 гг.) состоит в том, чтобы сухие и влажные периоды сменяли друг друга, тем самым формируя цикличность формирования осадков. В холодном полугодии длительность этих периодов будет существенно выше, нежели в теплом. К сухим периодам относятся временные отрезки с 1926 по 1935 г. и с 1941 по 1965 г. На протяжении пяти лет, в рассматриваемых периодах, сухие процессы встречались чаще дождливых на 26%. К наиболее выдающемуся периоду стоит отнести влажный период, длившийся около 10 лет (1981-1990 гг.). Тогда дождливые процессы происходили существенно чаще, нежели сухие, примерно на 20%. В период с 1986 по 1990 г. Повторяемость таких процессов пересекла критический рубеж – 38%. В теплом полугодии происходила периодичная, но не цикличная смена сухих и влажных периодов. Если до 1940 г. Происходила частая смена между сухими и влажными периодами, то начиная с 1945 до 1975 г. частая смена периодов прекратилась, сухие периоды стали гораздо чаще влажных. Был сделан вывод о связи между количеством выпадения осадков, а следовательно и массы стока рек с циклонической и антициклонической активностью в бассейне этих рек. Очевидна связь между скоростью выпадения осадков на ЕЧР, и тем насколько быстро набирает рост сток Волги. Повышение

уровня влажности водосборного бассейна Каспийского моря объясняется тем, что была активизирована западная форма циркуляции в 1970-е годы [15]. На территории Атлантико-Европейского сектора присутствовало неординарное продвижение циклонической циркуляции. Многие ученые доказали, что перемены в УКМ обусловлены сопряженностью с переменами циркуляционных эпох [16-21]. К их характерным чертам относятся зональные, меридиональные или смешанные атмосферные процессы. Однако нельзя говорить о какой-то единой цикличности «циркуляционных эпох», так как они происходят в хаотичном порядке. Если потребуется произвести длительный прогноз УКМ, то потребуется взять в расчет непосредственно текущую действующую циркуляционную эпоху. В работе [22] было обращено внимание, на то, что такая концентрация изменений присутствует в те периоды времени, где наблюдаются сильные «циркуляционные скачки». Подобную концентрацию изменений удалось установить и в иных гидрометеорологических характеристиках, включая осадки, давление, температуру воздуха и пр.

Л.С. Берг и В.Ю Визе [23, 24] сопоставили поведение УКМ за некоторый период с условиями плавания в арктических морях, и выяснили, что периоды с низким уровнем совпадают с легкими условиями плавания в Арктике, тяжелые условия – наоборот. Тогда же было выяснено, что, если в зимнее время над бассейном Волги господствует антициклон, он способствует сухому климату, соответственно и понижению стока рек. В Арктическом регионе наоборот устанавливается теплый период, что благоприятствует навигации. Эту особенность удалось подтвердить Н.А. Белинскому [22], продемонстрировавшему зависимость цикло-антициклонической деятельности и ледовитости арктических морей. А.А. Гирс [25], взяв за основу труды Г.Я. Вангенгейма, сделал однозначный вывод, что УКМ низкий, в годы, когда

преобладает восточная циркуляция атмосферы, и высок, когда главенствующую роль играет меридиональная циркуляция атмосферы. Западная форма приводит либо к постоянству значений уровня, либо к их росту. Также от форм циркуляции зависит и количество осадков.

Гирс, оперируя изложенной информацией, смог объяснить поведение УКМ на различных временных отрезках, чем подтвердил свои выводы о связи Северной Атлантики и уровня Каспия.

Еще двум исследователям, а именно И.С. Ким и С.П. Никулиной [26] удалось выяснить, что колебания морского уровня и атмосферная циркуляция, в различных её проявлениях, взаимосвязаны (в основе лежит классификация Вангенгейма-Гирса). Согласно прогнозу, выдвинутому исследователями в 1994 г., в ближайшие несколько лет уровень должен был прирастать со скоростью 8-9 см/год, после чего занять стабильную планку, несущественно видоизменяться. Начиная с 2025 г. уровень Каспийского моря должен вновь возрасти. А.Н. Кренке и В.В. Попова [27] произвели корреляцию между колебаниями УКМ и цикличностью типов атмосферной циркуляции, используя классификацию Дзердзеевского. Удалось установить, что подъем морского уровня и усиление меридиональной циркуляции взаимосвязаны. Падение морского уровня объясняется активизацией зональной циркуляции. Если преобладающим видом циркуляции выступает меридиональная, то будет больше дождей и снегопадов, в сравнении с годами, когда доминирует зональная циркуляция.

Н.С. Сидоренков и В.И. Швейкина [19] утверждают, что возрастание и снижение УКМ варьируется под влиянием климатических эпох, обусловленных «долгопериодным неравномерным вращением Земли». Их анализ также подтвердил связь, что когда морской уровень понижался, то присутствовало меньше атмосферных осадков. Была и совершенно противоположная ситуация.

Начиная с 1975 г. преимущественными были аномалии с отрицательной облачностью, с 1976 по 1983 г. – сугубо противоположные. Так авторы пояснили произошедший в 1978 г. прирост морского уровня. До 1976 г. атмосферные осадки в волжском бассейне были отрицательными, сток Волги уменьшился. После 1976 г. осадки начали превышать минимальный объем, испарение уменьшилось, сток начал увеличиваться - морской уровень начал повышаться.

1.3 Энергоактивные зоны Северной Атлантики.

В 1970 году Тимонов В.В. в одном из своих трудов, высказал идею о том, что в океане существуют очаги – зоны, в которых процесс обмена энергией, протекающий между поверхностью океана и слоями атмосферы, намного интенсивнее, чем в остальных. Данные зоны отображались на картах суммарной теплоотдачи в атмосферу за счет турбулентного потока тепла (P) и затрат тепла на испарение (LE). В данных исследованиях для района Северной Атлантики было выявлено 4 очага: Бермудский, Ньюфаундлендский, Западно-Гренландский и Норвежский». В каких-то случаях, изменение значений атмосферных показателей в описанных выше зонах влияет на незначительные колебания при энергетическом обмене между океаном и атмосферой. После выделения «активных» областей и создания данных карт стало возможно альтернативным способом формировать прогнозы изменения климата и изменение погодных условий, в том числе и долгосрочные, то есть с заблаговременностью от года.

Концепция энергоактивных зон океана была уточнена Г.И. Марчуком в 1976 г [28, 29, 30] в рамках советской наблюдательной части программы «Разрезы», это был проект, исследовавший роль океана в колебаниях климата, для составления климатических прогнозов. Стоит отметить, что в рамках этого проекта были получены и другие результаты, используемые до сих пор.

В работе [44] были рассчитаны статистические х-ки данных суммарной теплоотдачи. Согласно данным расчетам, на первом месте по мощности располагается Бермудская ЭАЗО. В центре зоны суммарная теплоотдача (P+LE) равняется 359.6 Вт/м^2. Коэффициент вариации не превышает 9%.

Бермудская ЭАЗО (Рис. 1.3) вытянута вдоль течения Гольфстрим и непосредственно этим и определена ее мощность, дело в том, что холодные воздушные массы с континента, попадая в зону теплых вод являются причиной большой разницы между температурой приводного слоя и ТПО. Непосредственно из-за этого формируется тот самый мощный поток тепла в атмосферу. Также играет свою роль и ее близость, как к тропической, так и субтропической зонам.

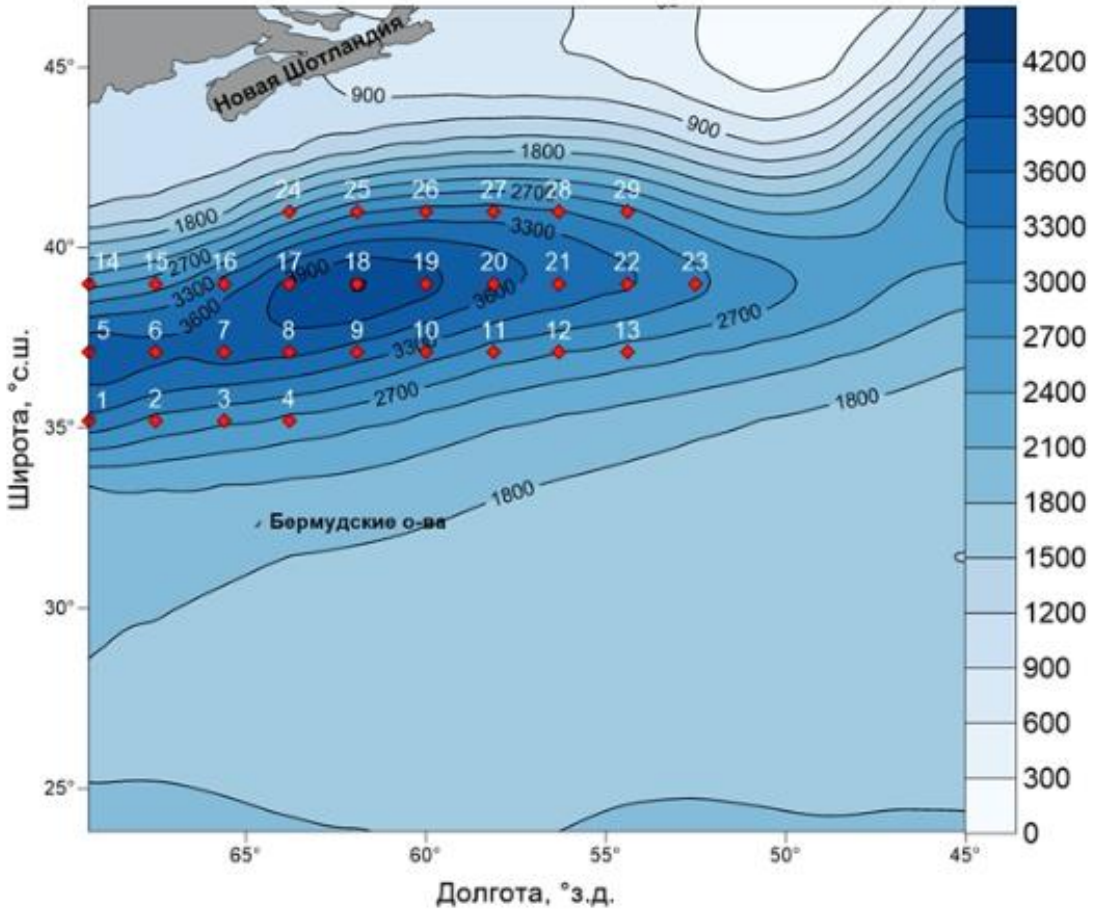


Рисунок 1.3. Карта распределения суммарной теплоотдачи (P+LE) в Бермудской ЭАЗО СА за 1980-2009 гг. в Вт/м^2. Красные точки на карте – станции Бермудской ЭАЗО.

На втором месте по мощности расположилась Ньюфаулендская ЭАЗО (Рис 1.4). В центре зоны суммарная теплоотдача (P+LE) равняется 249.6 Вт/м2. Коэффициент вариации не превышает 11%. Также, как и Бермудскую, эту зону определяет ее местоположение, она расположена между холодными и теплыми течениями, Лабрадорским и Северо-Атлантическим. В этом регионе сильно развита активность циклонов, регулярно меняются атмосферное состояние, сила ветра. Особенно выделяется осенне-зимний период.

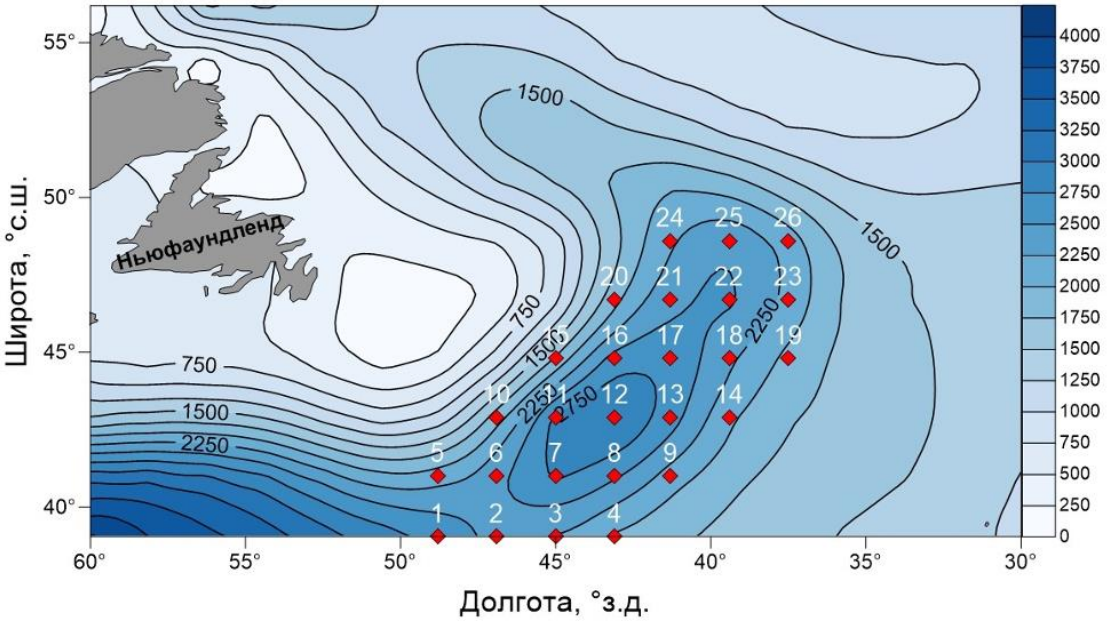


Рисунок 1.4. Распределение суммарной теплоотдачи (P+LE) в Ньюфаулендской ЭАЗО СА за 1980-2009 гг. в Вт/м^2. Красные точки на карте – станции Ньюфаулендской ЭАЗО.

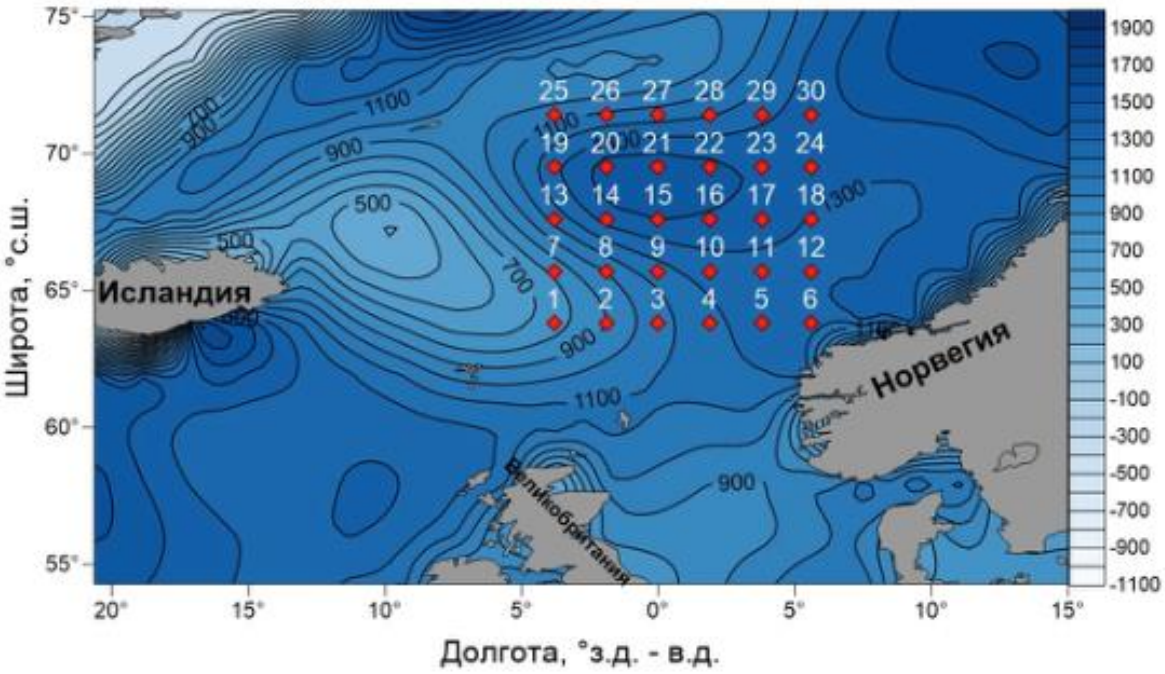


Рисунок 1.5. Распределение суммарной теплоотдачи (P+LE) в Норвежской ЭАЗО СА за 1980-2009 гг. в Вт/м^2. Красные точки на карте – станции Норвежской ЭАЗО.

Наиболее слабая из энергоактивных зон – Норвежская(Рис 1.5). Главным фактором формирования климата в данной зоне является Норвежское течение с его теплыми водами. В центре зоны суммарная теплоотдача (P+LE) равняется 126.1 Вт/м2. Коэффициент вариации не превышает 12.5%. Несмотря на менее выраженную интенсивность, чем у других зон, ее соседство с ЕЧР и Полярным бассейном делает ее крайне важной для изучения. Именно из нее на эти территории приходят практически все атлантические циклоны, которые в свою очередь приводят к мощным осадкам. В этой ЭАЗО в зимний период можно наблюдать мощные шторма, которые как мы знаем, приводят к высоким значениям теплообмена. Также важным климатическим фактором, влияющим на эту ЭАЗО, является Исландский центр действия атмосферы.

2. Исходные данные и методы расчетов.

2.1 Исходные данные

Исходными данными послужили данные о суммарной теплоотдаче, в Ньюфаулендской и Норвежской энергоактивных зонах Северной Атлантики. Для 59 станций (Рис 1.3, 1.4) был сформирован архив среднегодовых значений суммарной теплоотдачи (P+LE) за период 1980-2009 гг., данные брались из архива AMIP-II (Reanalysis-2». P – турбулентный поток тепла, LE – затраты тепла на испарение. Одним из недостатков данного архива, является тот факт, что данные в нем начали собираться лишь с 1979 года. Архив постоянно обновляется, данными из него для исследователей – ученых являются “open source”.

Из Ньюфаулендской ЭАЗО для прогноза использовалось 29 реперных квадратов, в Норвежской ЭАЗО - 30.

Данные о суммарной теплоотдаче использовались для создания прогностических и диагностических моделей зонального потока влаги на меридиане 5° в.д.

Данные зональных потоков влаги были рассчитаны из срочных данных, и усреднялись до среднегодовых.

При этом данные зонального потока влаги с помощью метода главных компонент были разделены на три области [3].

1) северную (от 5° в.д. 62.5 с.ш. до 5° в.д. 70 с.ш.);
2) центральную (от 5° в.д. 50 с.ш. до 5° в.д. 60 с.ш.);
3) южную (от 5° в.д. 40 с.ш. до 5° в.д. 47.5 с.ш.).

С каждой из этих областей строились полные модели деревьев решений с данными суммарной теплоотдачи Норвежской и Ньюфаулендской ЭАЗО. Для составления прогностической модели зонального потока влаги данные набор исходных данных делился на независимую и зависимую выборки. В зависимой выборке – на которой происходило обучение модели, использовались данные с 1981 по 2009 гг., независимая 2010-2014гг.

Также для подтверждения гипотезы о роли уровня Каспия, «как интегрального индикатора влагообмена в глобальной системе (океан-атмосфера-суша)», использовались данные о приращении УКМ, данные рассчитывались из декабрьских значений уровня на метеопосте в городе Баку. Данные по декабрьским значениям уровня брались из материалов [1,4,5].

2.2 Деревья решений [2].

Ключевым статистическим способом исследования был определен способ «decision trees», который дословно переводится как «дерево решений». Он имеет прямое отношение к технологии, называемой «data mining».

Сегодня нет дословного и утвержденного перевода словоформы (discovery-driven data mining). Буквальный перевод этой фразы – добыча информации. Если рассматривать практическую позицию, то «data mining» - интеллектуальная обработка экспериментальной информации, основной целью которой является нахождение объективных, а также неординарных закономерностей. Основная особенность таких закономерностей – это наличие конкретных выводов.

Data Mining включает в себя различные методики выявления закономерностей, в том числе и метод деревьев решений. Метод деревьев решений позволяет успешно решить, в том числе и гидрометеорологические задачи. Используется для составления прогнозов, или решения проблем классификации. Если у зависимой переменной будут дискретные значения, то используя способ дерева решений можно решить задачу классификации.

Основные преимущества деревьев решений:

1) Возможность дополнительно обрабатывать те значения, которые были упущены в процессе;

2) У моделей, сформированных деревьями решений, присутствует поразительная вычислительная точность, сравнимая с другими статистическими методами.
3) Помимо числовых переменных алгоритм деревьев решений допускает использование категориальных переменных.
4) В данном методе вычислений алгоритм позволяет не выделять при расчетах переменные с наиболее значимыми характеристиками, а классифицирует их независимо от вас.
5) Структура и визуализация данного метода легко поддается анализу и удобна при дальнейшем использовании.

«…В наиболее простом виде дерево решений — это способ представления классифицирующих правил в виде иерархической структуры. Ее основой, имеющей вид дерева, являются правила типа “если… то…. (if – then). Для принятия решения, к какому классу отнести некоторый объект или ситуацию, требуется ответить на вопросы, стоящие в узлах этого дерева, начиная с его корня. Корень — исходный вопрос, внутренний узел дерева, является узлом проверки определенного условия. Далее идет следующий вопрос и т.д., пока не будет, достигнут конечный узел дерева, являющийся узлом решения. Вопросы имеют вид значение параметра “А больше х?..” Если ответ “да”, то осуществляется переход к правому узлу следующего уровня, если “нет” — то к левому узлу; затем снова следует вопрос, связанный с соответствующим узлом…» (Гордеева, Малинин, 2016).

Алгоритм заканчивает формировать решения в узлах дерева, называемых «термальными вершинами».

«Statsoft 12», а именно в нем выполнялась основная расчетная часть, на выбор предоставляется три алгоритма ветвления дерева, каждый из которых

выбирается непосредственно в зависимости от ситуации. Дискриминантное одномерное ветвление для категориальных и порядковых предикторов, дискриминантное ветвление по линейным комбинациям и алгоритм «CART», на котором остановимся подробнее.

При использовании CART просчитываются всевозможные решения для одномерных ветвлений. Данный алгоритм можно применять при условии, что тип переменных разнороден, в том числе и для комбинированных видов. Данный способ базируется на том, что для нахождения оптимального решения, необходимо учесть всевозможные исходы, что делает данную процедуру достаточно долгой. Формирования структуры ветвей решений имеет следующей характер: от высшей точки дерева к его основанию, где наблюдается максимальное количество возможных вариаций.

Крайне важно использовать критерий расщепления, непосредственно он помогает разбить множество входящих данных на несколько частей. В алгоритме CART (создан в 1970е годы) используют индекс Gini.

Индекс определяется формулой:

$$Gini\ (T) = 1 - \sum_{j=1}^{n} p_j^2$$

Где, T – ведущий узел; p_j – вероятность класса j в узле T; n – количество классов. Индекс “Gini” помогает работает на принципе определения расстояний между распределениями классов.

С помощью алгоритма CART строится бинарное дерево решений. Это такое дерево, где у каждого узла, если его разбить, будет лишь по две дочерние ветви. Количество узлов зависит от количества исходного набора переменных.

На каждом разветвлении формируется так называемое «правило узла», в зависимости от его соблюдения идет разделение на правую, либо левую ветви.

В способе decision trees высший приоритет будет у классификации, имеющей меньшую цену. Здесь ценой выступает доля объектов, которые были классифицированы неверно. Объекты, неверно классифицированные – объекты, причисленные к одному классу, но относящиеся к иному. Лучший показатель будет у того прогноза, который выдает минимальное количество ошибок.

Вопрос о том, как остановить ветвление дерева имеет множество сложных аспектов. Логично, что чем больше ветвей у дерева, тем более точно оно классифицирует или прогнозирует данные, но проблема в том, что с ростом дерева падает качество точной классификации, так как количество переменных возрастает. По сути, хорошее дерево должно обладать двумя, казалось бы, взаимоисключающими условиями. Во-первых, условию простой интерпретации, во-вторых должно учитывать как можно больше значимых параметров. Именно из-за этого требуется использовать различные критерии оптимальности.

Суть алгоритма CART состоит в том, чтобы получить последовательность меньших в размере деревьев. Рассматриваются лишь исключительные деревья, имеющие наилучшие характеристики.

Для гидрометеорологических задач из множества всех алгоритмов предлагаемых методом деревьев решений на наш взгляд более всего подходит алгоритм CART. С помощью этого алгоритма можно построить полное дерево, а затем, путем его уменьшения выбирать оптимальное.

3. Оценка влияния суммарных потоков тепла в энергоактивных зонах Северной Атлантики на меридиональный перенос водяного пара на 5° в.д. на основе модели деревьев решений.

Как было сказано в главе 1, два основных фактора, формирующих перенос влаги в Европу – это испарение с акватории океана и циркуляция атмосферы. Наибольшее испарение имеет место в ЭАЗО СА, причем вклад испарения в суммарную теплоотдачу является определяющим. На рис. 3.1 представлено среднемноголетнее распределение годовых значений суммарной теплоотдачи (P+ LE) в трех ЭАЗО Северной Атлантики. В этой работе использовались данные для Ньюфаулендской и Норвежской ЭАЗО.

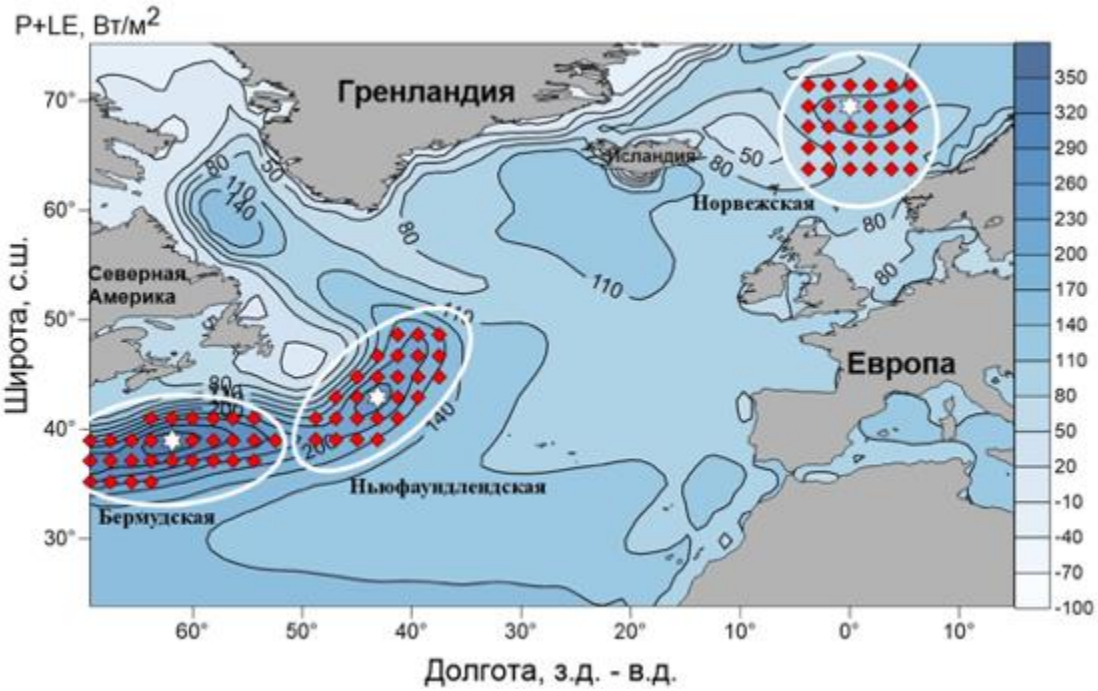


Рисунок 3.1. Среднемноголетнее распределение годовых значений суммарной теплоотдачи (P+LE) в Северной Атлантике за 1980-2017 гг. в Вт/м2.

Овалами обведены три ЭАЗО (Бермудская, Норвежская, Ньюфаудендская). Звездочками обозначены центры ЭАЗО с максимальной теплоотдачей.

На основе модели деревьев решений оценивалось влияние суммарных потоков тепла в энергоактивных зонах Северной Атлантики на меридиональный перенос водяного пара на 5° в.д. Первым делом была построена полная модель деревьев решений для самой южной точки этого меридионального разреза 5° в.д. 40 ° с.ш. Так как их полная модель содержала 10 деревьев, нам потребовалось решать задачу выбора оптимального дерева. Чтобы обнаружить оптимальное дерево алгоритм CART, рассчитывает для каждого дерева значения CVcost, CVкрит, а также стандартную ошибку CV, оптимальным принимается то дерево у которого значение CVcost оказывается меньше CVкрит. К сожалению, часто этот алгоритм не срабатывает.

Для полной модели станции 5° в.д. 40 ° с.ш. была выполнена кросс-проверка (Табл 3.1). По результатам мы можем видеть один из тех случаев, когда алгоритм оптимальности не сработал. Наилучшей моделью он предлагает считать тот случай, когда модель, состоит всего из одной вершины, то есть по сути представляет собой исходную выборку. Из-за этого нам приходится использовать другую тактику поиска наилучшей модели деревьев решений.

Таблица 3.1. Результат кросс-проверки полной модели деревьев решений для зонального потока влаги в точке с координатами 5° в.д. 40 ° с.ш.

Показатель	Число терми-нальных вершин (Terminal nodes)	Цена кросс-проверки (CV cost) Вт/м^2	Стандартная ошибка кросс-проверки (CV std. error) Вт/м^2	Цена проверки на обучающей выборке (Resubstituting cost) Вт/м^2	Сложность узла (Node complexity)
Дерево 1	10	220,278	48,87	8,82	0,00
Дерево 2	9	221,05	48,85	10,73	1,9
Дерево 3	8	207,35	48,35	13,49	2,75
Дерево 4	7	182,49	36,10	18,85	5,35
Дерево 5	6	176,13	36,55	24,78	5,93
Дерево 6	5	181,86	37,49	30,83	6,05
Дерево 7	4	173,39	39,96	38,84	8,00
Дерево 8	3	167,68	39,13	51,92	13,08
Дерево 9	2	163,48	39,21	71,37	19,44
Дерево 10	1	137,29	25,56	128,25	56,88

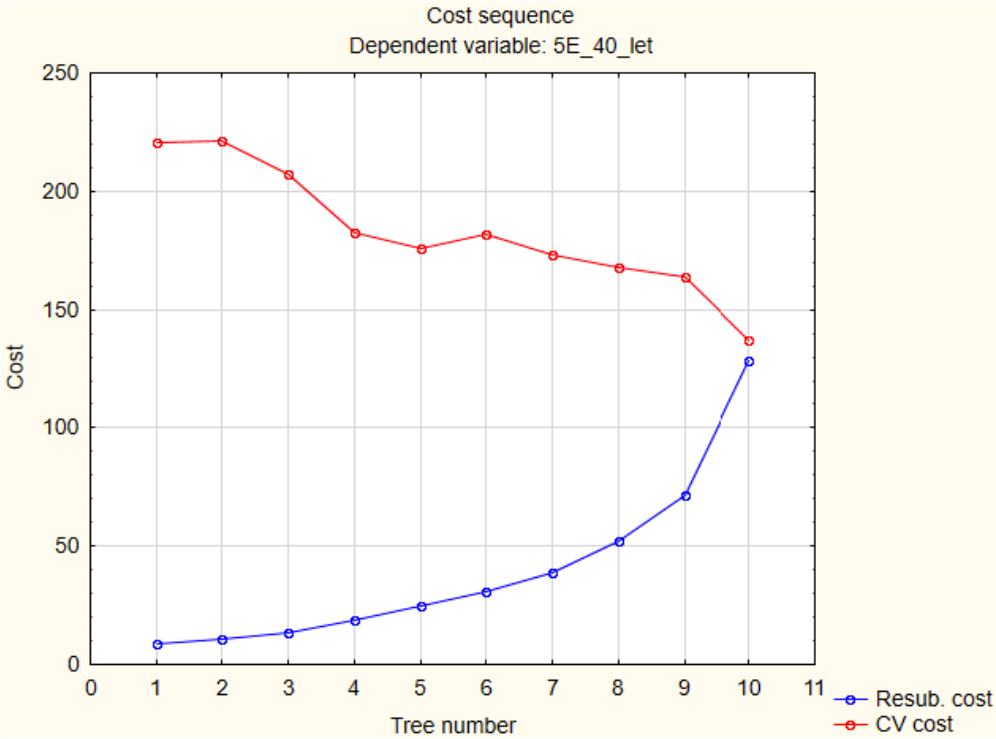


Рисунок 3.2. Распределение значений CV-cost и Resub-cost для модели полного дерева решений зонального потока влаги в точке 5° в.д. 40 ° с.ш.

Для этого мы оцениваем коэффициент детерминации, и вклад ошибки в СКО для каждого дерева. Как видно из (табл. 3.2) уже на 5 шаге R^2=0.75, а вклад ошибки в СКО менее 0.56 и равен 0.48. Это значение можно назвать непосредственно граничным, видно, что скорость уменьшения ошибки после этого рубежа, начинает сильно падать, и составляет всего 0.5 долей СКО за шаг, а к полному дереву снижается до 0.3 долей СКО за шаг. Дерево №6 содержит 5 терминальных и 4 нетерминальных вершин, и этого достаточно для поставленной задачи.

Таблица 3.2. Статистические оценки для всех моделей деревьев решений зонального потока влаги в точке с координатами 5° в.д. 40 ° с.ш. зависимости от годовых значений суммарной теплоотдачи в 56 квадратах Ньюфаулендской и Норвежской ЭАЗО за 1980-2014 гг.

№ Шага	Номер дерева	Число вершин терминальных/нетерминальных	Коэффициент детерминации модели дерева	Стандартная ошибка модели дерева в долях СКО
1	10	1/0	0,002	0,98
2	9	2/1	0,44	0,73
3	8	3/2	0,59	0,62
4	7	4/3	0,69	0,54
5	6	5/4	0,75	0,48
6	5	6/5	0,8	0,43
7	4	7/6	0,85	0,37
8	3	8/7	0,89	0,31
9	2	9/8	0,91	0,28
10	1	10/9	0,93	0,25

Выполним подобный этому анализ модели деревьев решений для самой северной точки меридионального разреза 5° в.д. Станции имеющей координаты 5° в.д. 70 с.ш. Полная модель содержала 12 деревьев, и также как в прошлый раз нам потребовалось решать задачу выбора оптимального дерева. К сожалению, критерий оптимальности, предложенный программой “Statistika 12” опять не сработал. Причина аналогична той, что была с самой южной точкой разреза, $CV\text{-}_{cost}$ был меньше $CV_{крит}$. лишь на последнем дереве, что можно увидеть из графика (Рис 3.3), это дерево является самым простым, по такому дереву невозможно найти те реперные квадраты значений суммарной

теплоотдачи ЭАЗО, оказывающих наибольшее влияние на зональный поток влаги. Следовательно, нам пришлось прибегнуть к тактике выбора наилучшего дерева по статистическим характеристикам (Таб. 3.3)

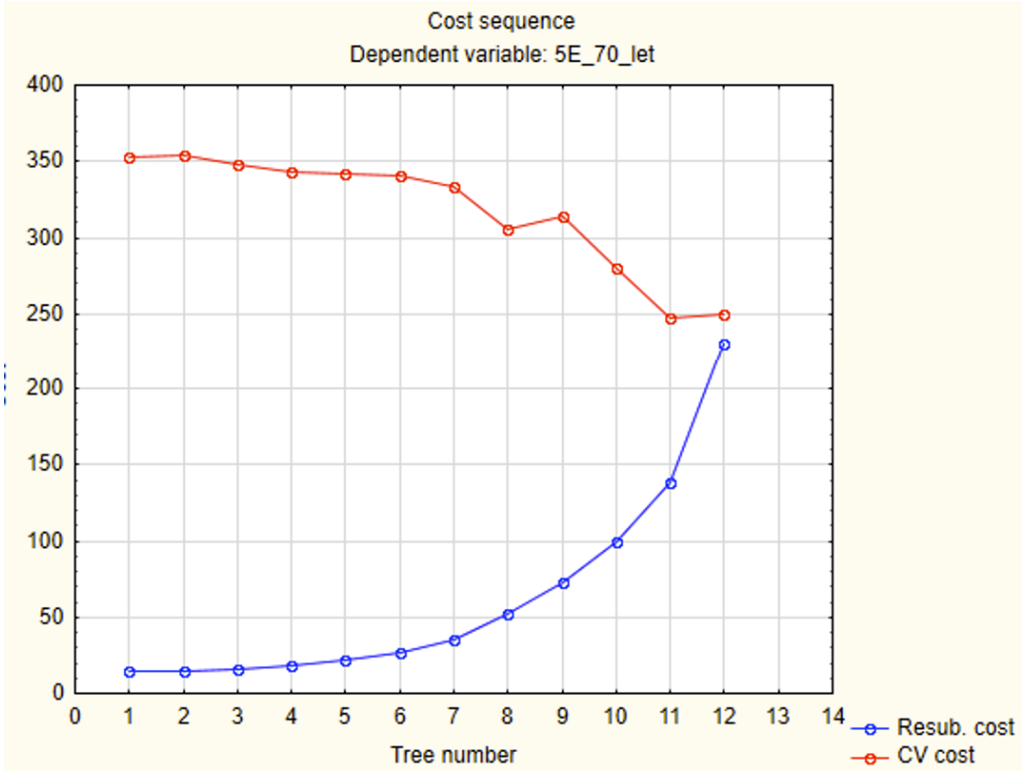


Рисунок 3.3. Распределение значений CV-cost и Resub-cost для модели полного дерева решений зонального потока влаги в точке 5° в.д 70 ° с.ш.

Таблица 3.3. Статистические оценки для всех моделей деревьев решений зонального потока влаги в точке с координатами 5° в.д. 70 ° с.ш. зависимости от годовых значений суммарной теплоотдачи в 56 квадратах Ньюфаулендской и Норвежской ЭАЗО за 1980-2014 гг.

№ Шага	Номер дерева	Число вершин терминальных/нетерминальных	Коэффициент детерминации модели дерева	Стандартная ошибка модели дерева в долях СКО
1	12	1/0	0,05	0,98
2	11	2/1	0,4	0,76
3	10	3/2	0,56	0,64
4	9	4/3	0,69	0,55
5	8	5/4	0,77	0,46
6	7	6/5	0,84	0,38
7	6	7/6	0,88	0,33
8	5	8/7	0,90	0,30
9	4	9/8	0,92	0,27
10	3	10/9	0,93	0,26
11	2	11/10	0,93	0,24
12	1	12/11	0,94	0,23

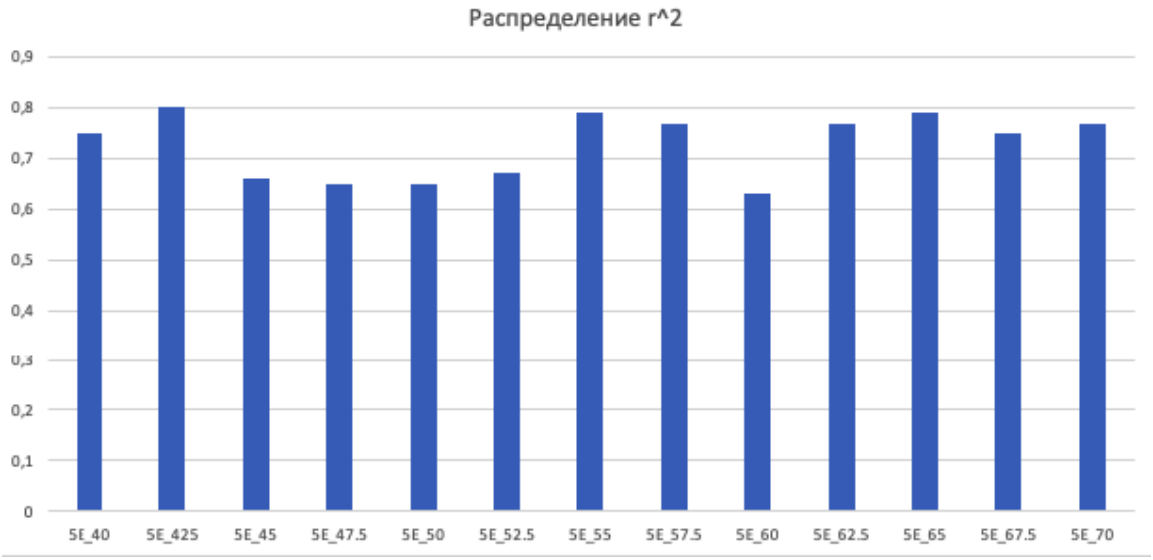
Дерево №1 имеет 12 терминальных вершин, непосредственно участвующих в ветвлении, и 11 нетерминальных вершин, которые в данном дереве в ветвлении не участвуют. Последнее дерево №12 имеет всего лишь 1 терминальную вершину, то есть представляет собой исходную выборку.

Как видно из таблицы уже на 5 шаге R^2=0.77, а вклад ошибки в СКО менее 0.56 и равен 0.46. Это значение также как и в прошлый раз следует назвать непосредственно граничным, и мы видим, что скорость уменьшения ошибки после этого рубежа, начинает сильно падать, и составляет всего 0.8 долей СКО за шаг, а к полному дереву снижается до 0.1 долей СКО за шаг. Дерево №8 содержит 5 терминальных и 4 нетерминальных вершин, и этого достаточно для поставленной задачи.

Подобным анализу этих двух точек, (5° в.д. 40 ° с.ш., 5° в.д. 70 ° с.ш.), также выполнялся анализ и для остальных 11 точек, то есть были построены полные модели деревьев решений. Всего было построено 125 моделей, в среднем 9-10 для каждой точки. Также как и в предыдущем случае для построения применялся алгоритм CART, полные модели деревьев имели по 8-12 терминальных вершин. Критерий оптимальности предложенный программой “Statistika” сожалению ни разу не сработал, поэтому в каждом случае наилучшее дерево выбиралось по статистическим характеристикам, а не по критерию оптимальности в программе “Statistika 12”. Также для наглядности и в отсутствии явно выраженного стандарта по отбору моделей, для сравнения качества моделей был выбран один шаг, шаг №5.

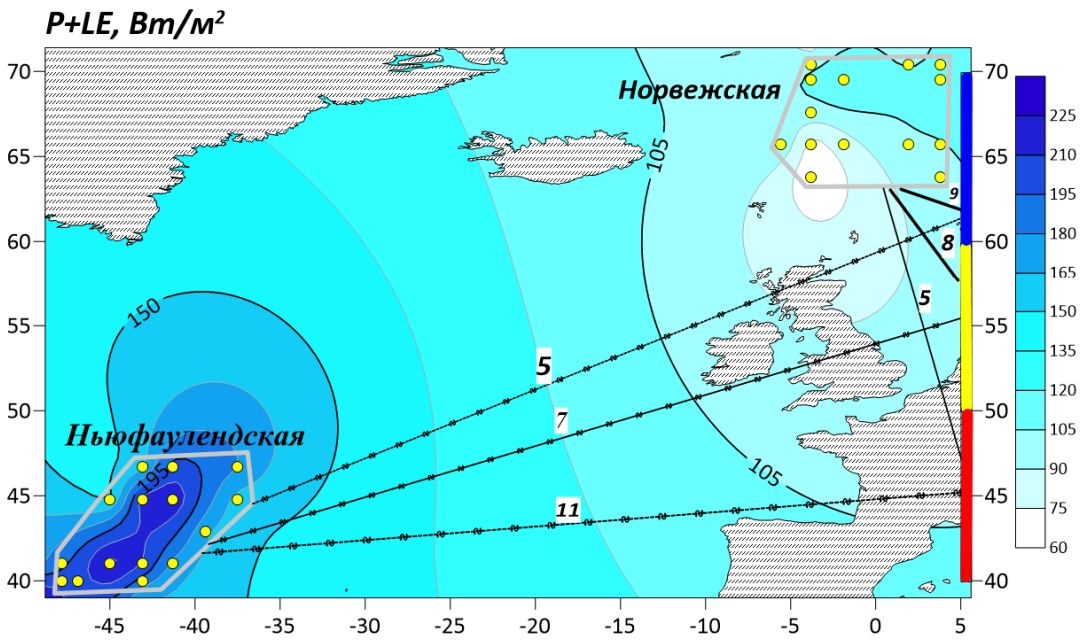
На Рис. 3.4 приведен график распределение коэффициента детерминации для разреза 5° в.д., на графике отчетливо видно равномерное влияние на каждую точку разреза. Среднее значение R2= 0.72, что говорит о высоком качестве моделей, и непосредственно доказывает наличие связи между суммарной теплоотдачей ЭАЗО и зональным потоком влаги на меридиональном разрезе.

Рисунок 3.4. Распределение коэффициента детерминации модели дерева для всех 13 точек разреза 5° в.д. на 5 шаге.



Следующий этап работы заключался в выборе станций суммарной теплоотдачи в Норвежской и Ньюфаулендской ЭАЗО, которые имеют наилучшую связь с данными зонального потока. Проанализировав, 5 шаг всех моделей из 59 станций осталось 29. На карте распределения значений суммарной теплоотдачи (Рис. 3.5) обведены обе ЭАЗО. В них помещены те 29 станций, имеющих наилучшую связь с меридиональным разрезом. Далее от этих областей проведены линии к северной (5E_625-5E_70), центральной (5E_50-5E_60) и южной (5E_40 – 5E_475) частям разреза. На северный участок, оказывает влияние 5 станций из Ньюфаудендской ЭАЗО, 9 из Норвежской, на центральную 7 и 8, на южную 11 и 5 соответственно. Эти значения подписаны над каждой из линий. Линии имеют различную толщину пропорциональную числу точек.

Рисунок – 3.5. Распределение влияния реперных квадратов среднегодовых значений суммарной теплоотдачи Норвежской и Ньюфаундлендской ЭАЗ на формирование значений зонального потока влаги, на южной (5E_40 – 5E_475), центральной (5E_50 – 5E_60) и северной (5E_625 – 5E_70) части разреза 5E цифрами с белым фоном на карте обозначено число реперных квадратов суммарной теплоотдачи.



4.Построение прогностических моделей зонального переноса влаги на меридиане 5° в.д. методом деревьев решений. Построение диагностической модели уровня Каспийского моря.

4.1 Прогноз зональных потоков.

В этом разделе решалась задача прогноза значений зонального потока влаги на меридиональном разрезе 5° в.д. Разрез был разделен на три части северную (5E_625-5E_70), центральную (5E_50-5E_60) и южную (5E_40 – 5E_475). Каждая часть разреза прогнозировалась отдельно, со сдвигом на 1 год. Предикторами выступили значения суммарной теплоотдачи с 29 станций ЭАЗО. Подробнее о выборе этих станций говорилось в главе №3. Методом моделирования были выбраны деревья решений. Алгоритм – CART.

На рис. 4.1, 4.2, 4.3 можно видеть, что минимальное значение CV-cost, как для северной части разреза, так и для центральной и южной соответствует последнему дереву, содержащему всего одну вершину. Это означает, что данные деревья решений следует считать оптимальным, однако они практически не позволяют осуществлять прогноз стока. Как и в предыдущей главе, критерий оптимальности не сработал.

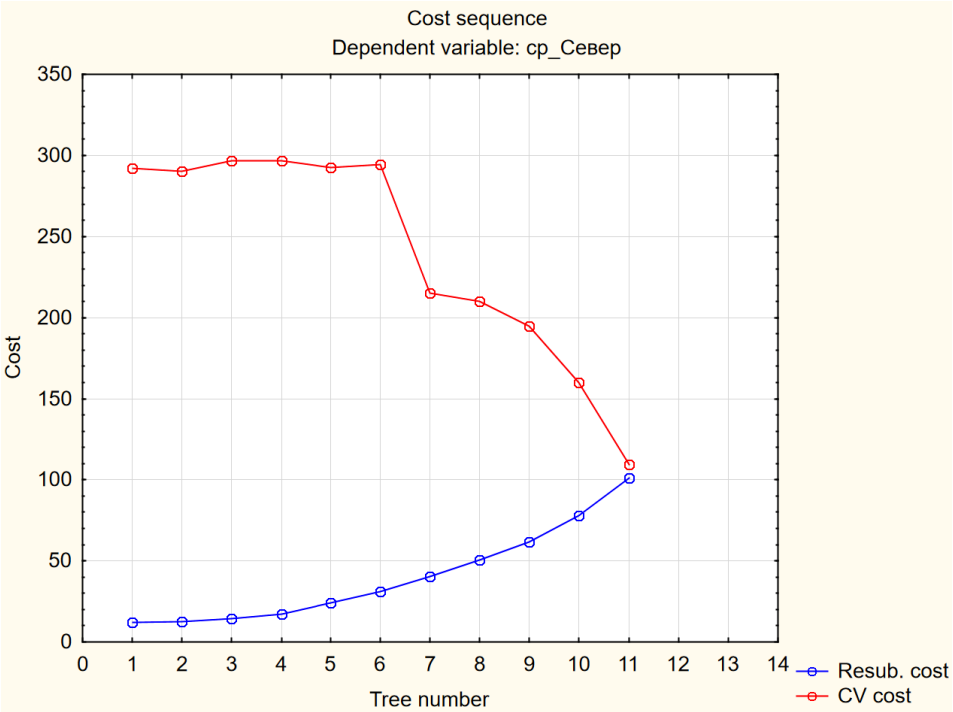


Рисунок 4.1. Распределение значений CV-cost и Resub-cost для прогностической модели полного дерева решений зонального потока влаги в северной части разреза 5° в.д.

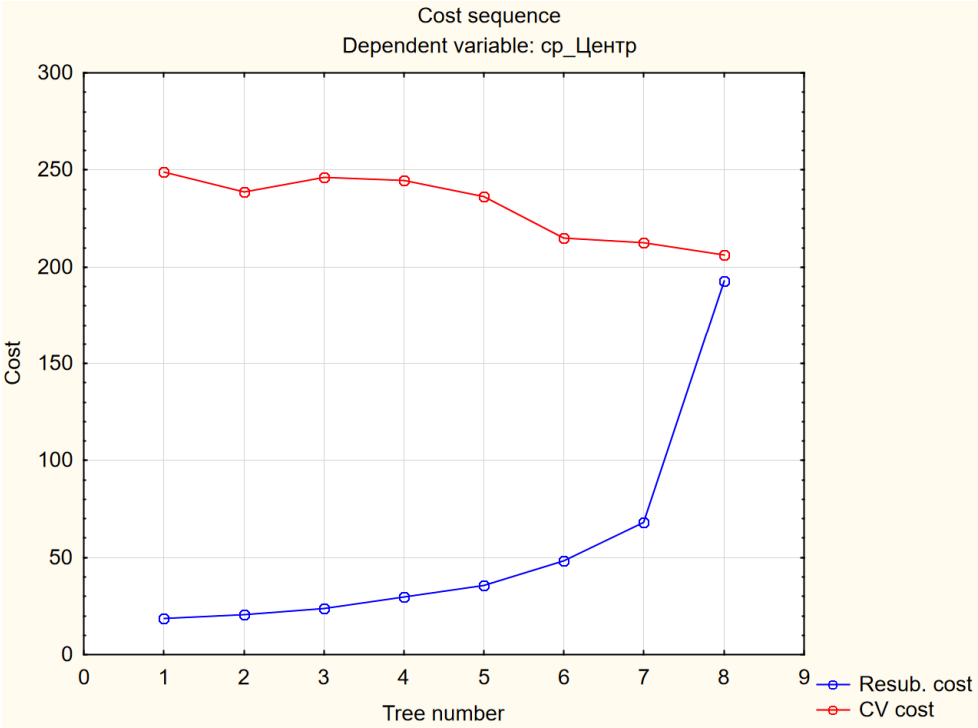


Рисунок 4.2. Распределение значений CV-cost и Resub-cost для прогностической модели полного дерева решений зонального потока влаги в центральной части разреза 5° в.д.

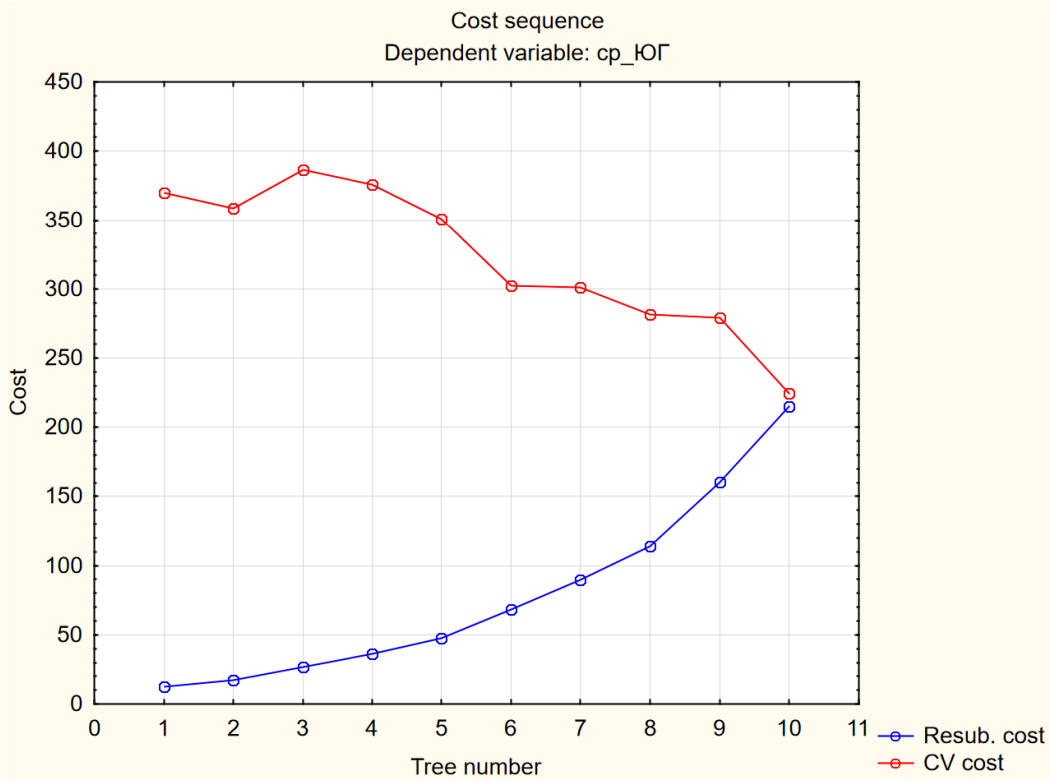


Рисунок 4.3. Распределение значений CV-cost и Resub-cost для прогностической модели полного дерева решений зонального потока влаги в южной части разреза 5° в.д

Так как критерий оптимальности предлагаемый программой “Statistika 12” не сработал мы использовали ту же тактику, что и в третьей главе. Вначале рассматривали результаты прогнозирования зонального потока влаги после первого ветвления, затем после второго и так далее, и так до последнего дерева. Было построено 11 деревьев решений для северной части меридионального разреза, 8 для центральной и 10 для южной.

Таблица 4.1 Статистические оценки зонального потока влаги для северной части меридионального разреза 5° в.д. по зависимой (1981–2009 гг.) и независимой (2010–2014 гг.) выборке для всех деревьев решений.

№ Шага	Номер дерева	Число вершин терминальных/нетерминальных	Стандартная ошибка модели дерева в долях СКО на зависимой выборке (1981-2009)	Коэффициент детерминации модели дерева	Стандартная ошибка модели дерева в долях СКО на независимой выборке (2010-2014)
1	11	1/0	0,98	0,02	1,12
2	10	2/1	0,88	0,19	0,85
3	9	3/2	0,781	0,37	0,80
4	8	4/3	0,71	0,47	0,68
5	7	5/4	0,62	0,60	0,70
6	6	6/5	0,57	0,66	0,43
7	5	7/6	0,50	0,73	0,49
8	4	8/7	0,42	0,81	0,48
9	3	9/8	0,381	0,85	0,77
10	2	10/9	0,34	0,87	0,58
11	1	11/10	0,34	0,88	0,88

Во всех трех случаях модель обучалась на зависимой выборке (1981-2009гг.), Из таблиц 4.1-4.3 следует, что чем больше дерево, тем больше коэффициент детерминации и меньше вклад ошибки в СКО. Для полных деревьев R^2 = 0.88, 0.87, 0.94 для северной, центральной и южной частей меридионального разреза соответственно. Стандартная ошибка моделей

деревьев для независимой выборки довольно сильно варьируется, можно сказать, что имеют случайный характер, но все таки можно выбрать наилучшие деревья.

Для северной части разреза оптимальным деревом будет дерево №6. (Рис. 4.4) Оно имеет 6 терминальных и 5 нетерминальных вершин. На первом ветвлении разделителем выступают значения суммарной теплоотдачи на станции Nuf_8, если значение менее 198,22 Вт/м^2, то нам требуется перейти ко второму ветвлению и смотреть значения на станции Nuf_14, если величина будет более 151,96 Вт/м^2 , то прогнозируемый, для этих случаев зональный поток будет равен 69,98 кг/м·с, когда менее 151,96 Вт/м^2 – 81,87 кг/м·с. При сценариях, когда значения на Nuf_8 более 198,22 Вт/м^2 , происходят различные уточнения еще на 3 ветвлениях, где решающую роль играют непосредственно значения суммарной теплоотдачи на станциях: Nuf_20 и Nuf_16. Другие переменные из исходной выборки не дают какого-либо вклада в описание дисперсии зонального потока влаги. Примечателен тот факт, что для качественного прогноза не понадобились значения суммарной теплоотдачи из Норвежской ЭАЗО, скорее всего это связано с близким расстоянием от Ньюфаулендской ЭАЗО до меридиана 5° в.д. Полученные значения зонального потока варьируют от 56 до 81 кг/м·с.

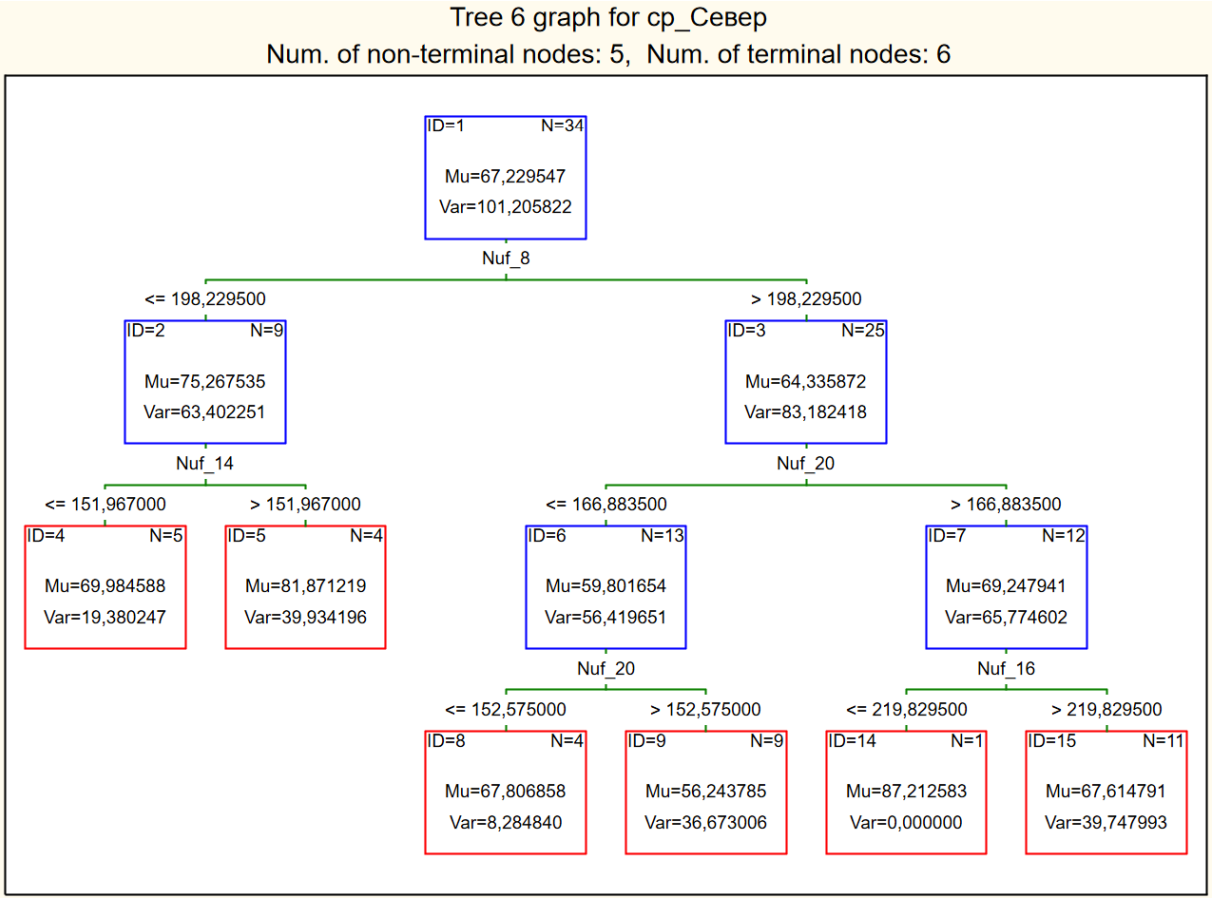


Рисунок 4.4 Дерево решений № 6 (шестой шаг ветвления), описывающее формирование зонального потока влаги для северной части меридионального разреза 5° в.д. в зависимости от значений суммарной теплоотдачи в некоторых квадратах Ньюфаулендской ЭАЗО за 1980-2013гг.

Таблица 4.2 Статистические оценки зонального потока влаги для центральной части меридионального разреза 5° в.д. по зависимой (1981–2009 гг.) и независимой (2010–2014 гг.) выборке для всех деревьев решений.

№ Шага	Номер дерева	Число вершин терминальных/нетерминальных	Стандартная ошибка модели дерева в долях СКО на зависимой выборке (1981-2009)	Коэффициент детерминации модели дерева	Стандартная ошибка модели дерева в долях СКО на независимой выборке (2010-2014)
1	8	4/3	0,98	0,02	1,3
2	7	5/4	0,59	0,64	0,73
3	6	6/5	0,49	0,74	0,62
4	5	7/6	0,43	0,80	0,49
5	4	8/7	0,39	0,84	0,45
6	3	9/8	0,37	0,85	0,74
7	2	10/9	0,36	0,86	0,34
8	1	11/10	0,34	0,87	0,55

Для центральной части разреза оптимальным деревом будет дерево №4. (Рис. 4.5) Оно имеет 8 терминальных и 7 нетерминальных вершин. На первом ветвлении разделителем выступают значения суммарной теплоотдачи на станции Norv_25, когда значение на станции менее 87,54 Вт/м^2, нам требуется обратиться ко второму ветвлению и обратить внимание на значение суммарной теплоотдачи на станции Norv_11, в случае когда оно будет более 95,02, мы прогнозируем зональный поток в зависимости от значений на станции Norv_1, когда менее, зональный поток равен 86,43 кг/м·с. При других сценариях мы рассматриваем значения на станциях Norv_13 и Nuf_4. В этих случаях

спрогнозированный зональный поток варьируется от 48 до 80 кг/м·с. 4 из 5 станций находятся в Норвежской ЭАЗО, столь малый процент значимых станций их Ньюфаулендской ЭАЗО связан с ее расстоянием до центральной части разреза.

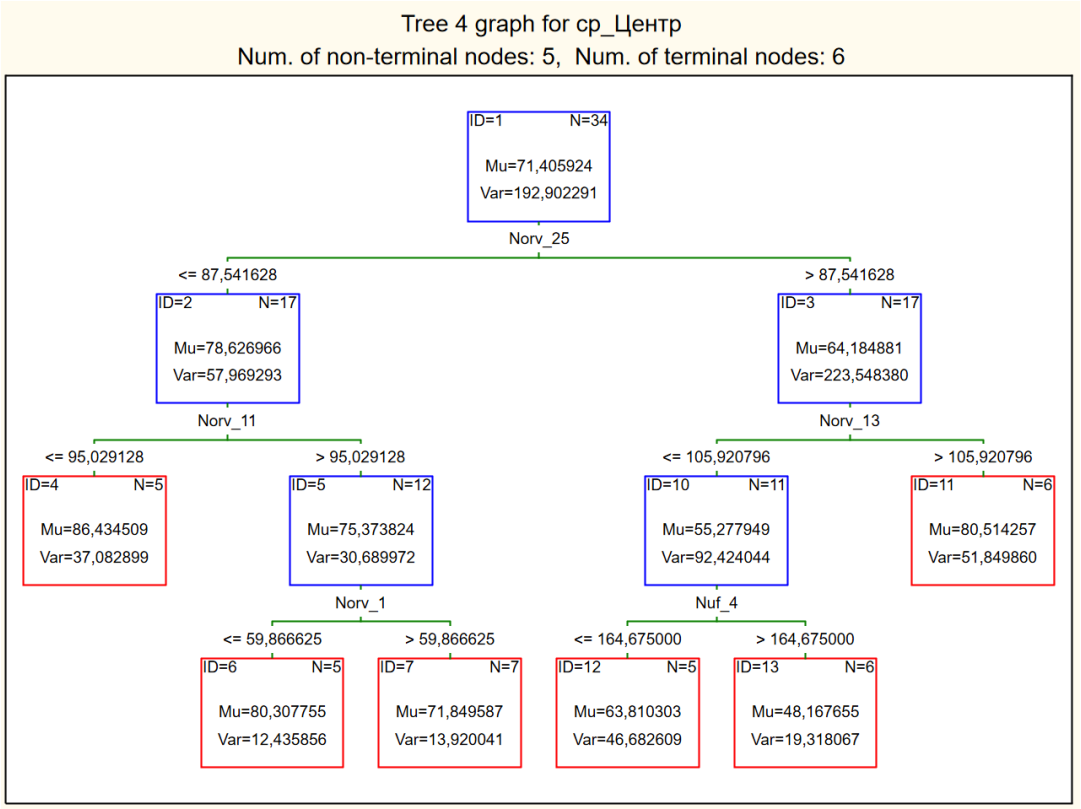


Рисунок 4.5 Дерево решений № 4 (пятый шаг ветвления), описывающее формирование зонального потока влаги для центральной части меридионального разреза 5° в.д. в зависимости от значений суммарной теплоотдачи в некоторых квадратах Ньюфаулендской и Норвежской ЭАЗО за 1980-2013гг

Для южной части разреза оптимальным деревом будет дерево №5. (Рис. 4.6) Оно имеет 7 терминальных и 6 нетерминальных вершин. На первом ветвлении разделителем выступают значения суммарной теплоотдачи на

станции Nuf_7, на этом ветвлении отсекаются экстремальные значения суммарной теплоотдачи, более 237 Вт/м^2, далее происходит уточнение этого сценария и рассматриваются значения на станции Nuf_15, в зависимости от значений происходит достаточно заметное разделение спрогнозированных значений, когда теплоотдача на станции менее 147,6 Вт/м^2 мы прогнозируем зональный поток равный 1,8 кг/м·с, более – 30,01. Если после первого ветвления, то есть отсечения экстремальных значений, мы рассматриваем более средние значения на станциях Norv_11, Norv_25, Nuf_16 – то мы можем спрогнозировать зональный поток равный от 16 до 58 кг/м·с.

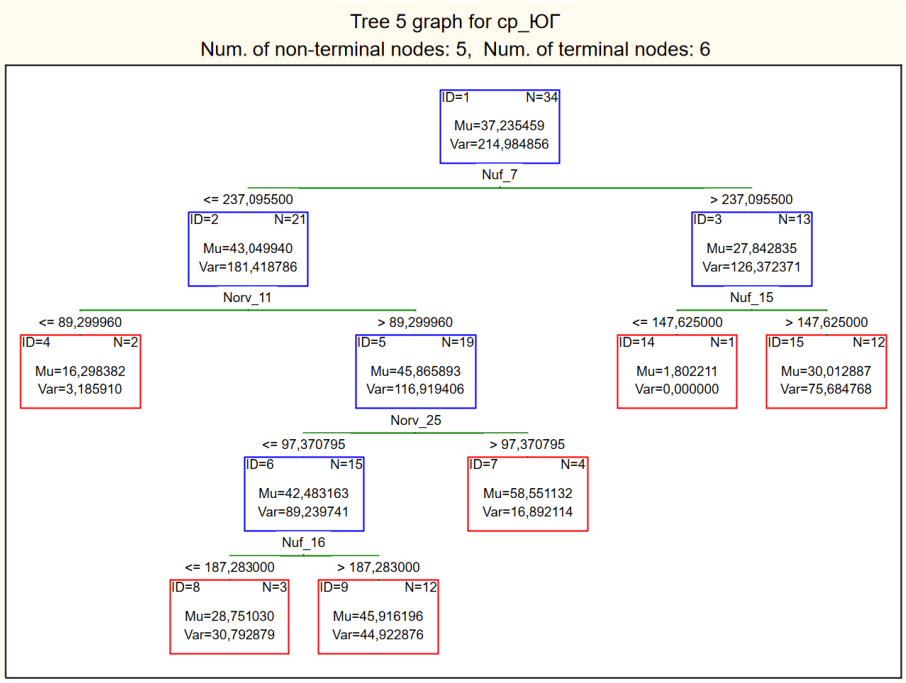


Рисунок 4.6 Дерево решений № 5 (шестой шаг ветвления), описывающее формирование зонального потока влаги для южной части меридионального разреза 5° в.д. в зависимости от значений суммарной теплоотдачи в некоторых квадратах Ньюфаулендской и Норвежской ЭАЗО за 1980-2013гг.

Таблица 4.3 Статистические оценки зонального потока влаги для южной части меридионального разреза 5° в.д. по зависимой (1981–2009 гг.) и независимой (2010–2014 гг.) выборке для всех деревьев решений.

№ Шага	Номер дерева	Число вершин терминальных/нетерминальных	Стандартная ошибка модели дерева в долях СКО на зависимой выборке (1981-2009)	Коэффициент детерминации модели дерева	Стандартная ошибка модели дерева в долях СКО на независимой выборке (2010-2014)
1	10	2/1	0,98	0,03	1,2
2	9	3/2	0,83	0,28	0,79
3	8	4/3	0,65	0,56	0,9
4	7	5/4	0,59	0,64	0,77
5	6	6/5	0,48	0,76	0,79
6	5	7/6	0,45	0,78	0,56
7	4	8/7	0,40	0,83	0,81
8	3	9/8	0,35	0,87	0,7
9	2	10/9	0,28	0,91	0,78
10	1	11/10	0,23	0,94	0,93

4.2 Проверка гипотезы о выявленной связи от Северной Атлантики до уровня Каспийского моря.

Несмотря на то, что хорошая корреляция спрогнозированных и фактических значений зонального потока во всех трех зонах, от северной до южной, отлично подтверждает гипотезу, о наличии причинно-следственных связей от Северной Атлантики до уровня Каспийского моря, высказанную в 1 главе. Было принято решение связать значения суммарной теплоотдачи в ЭАЗО именно с приращением уровня в Каспийском море с помощью доказавшеего свою состоятельность метода деревьев решений. С этой целью мы выполнили расчет деревьев решений для внутригодовых приращений уровня моря в Баку за 1981-2009 гг (n=29), матрица состояла из 29 станций суммарной теплоотдачи ЭАЗО, тех же, что были использованы для прогноза зонального потока, в первой части этого раздела. Моделирование выполнялось с помощью алгоритма CART.

На рис. 4.7 представлен график распределения Resubstitution cost и CV cost. Как и во всех остальных случаях, деревом с наименьшей ценой ошибки, оказывается последнее дерево, которое имеет всего 1 вершину, поэтому такая модель не является оптимальной для нас, хоть и выбрана автоматически алгоритмом CART, как наилучшая. Нам она не подходит, так как спрогнозировать какие-либо значения с помощью нее не выйдет.

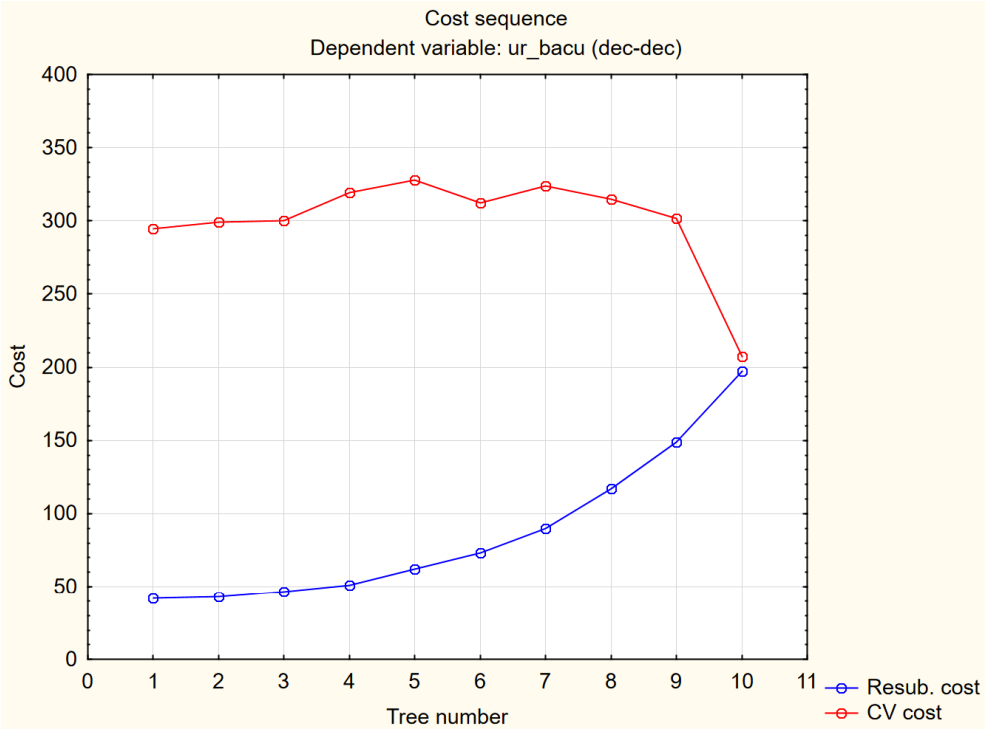


Рисунок 4.7. Распределение значений CV-cost и Resub-cost для диагностической модели полного дерева решений приращения уровня Каспийского моря со значениями суммарной теплоотдачи ЭАЗ Северной Атлантики.

Для определения дерева оптимального размера мы использовали ту же тактику, что и в предыдущих случаях. Вначале рассматривали результаты моделирования уровня Каспия после первого ветвления, затем после второго и так далее, и так до последнего дерева. На наш взгляд оптимальное дерево №5. (Рис. 4.8) Оно имеет 7 терминальных и 6 нетерминальных вершин.

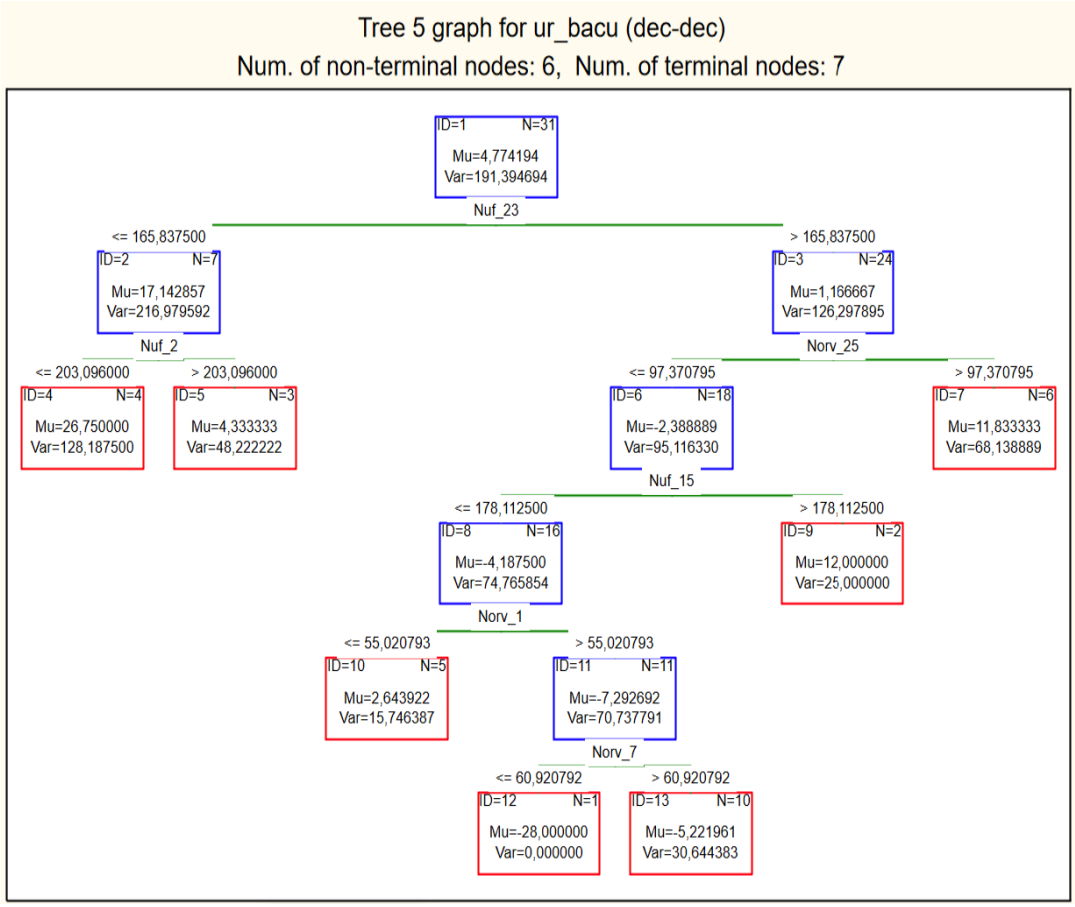


Рисунок 4.8. Дерево решений номер 5 (шестой шаг ветвления), характеризующее связь между значениями уровня Каспийского моря и годовыми значениями суммарной теплоотдачи в 29 квадратах Ньюфаулендской и Норвежской ЭАЗО за 1980-2009 гг.

На первом ветвлении разделителем выступают значения суммарной теплоотдачи на станции Nuf_23, если значение менее 165,83 Вт/м^2, то нам требуется перейти ко второму ветвлению и смотреть значения на станции Nuf_2, в случае когда величина более 203,09, прогнозируемое, для этих случаев приращение уровня будет равно 4,3 см, когда менее 203,09 – 26,75. При сценариях, когда значения на Nuf_23 более 165,83 Вт/м^2 , происходят различные уточнения еще на 5 ветвлениях, где решающую роль играют непосредственно значения суммарной теплоотдачи на станциях: Norv_25,

Nuf_15, Norv_1, Norv_7. Другие станции никак не описывают приращение уровня в этой модели.

Для оценки качества диагностической модели были рассчитаны коэффициенты детерминации и стандартная ошибка для каждого дерева (табл. 4.4). Чем больше дерево, тем больше R^2, и меньше вклад ошибки в СКО. Для последнего, наиболее полного дерева R^2 = 0,84, а стандартная ошибка = 0.38 в долях СКО.

Таблица 4.4 Статистические оценки для всех моделей деревьев решений зонального потока влаги в точке с координатами 5° в.д. 40 ° с.ш. зависимости от годовых значений суммарной теплоотдачи в 59 квадратах Ньюфаулендской и Норвежской ЭАЗО
за 1980-2014 гг.

№ Шага	Номер дерева	Число вершин терминальных/нетерминальных	Коэффициент детерминации модели дерева	Стандартная ошибка модели дерева в долях СКО
1	10	1/0	0,02	0,98
2	9	2/1	0,23	0,86
3	8	3/2	0,38	0,77
4	7	4/3	0,53	0,67
5	6	5/4	0,61	0,61
6	5	7/6	0,74	0,49
7	4	8/7	0,80	0,43
8	3	9/8	0,83	0,40
9	2	10/9	0,84	0,38
10	1	11/10	0,85	0,38

Заключение

В ходе работы были полностью решены поставленные задачи:

1) Теоретически и практически выявлено и обосновано влияние энергоактивных зон Северной Атлантики на горизонтальный перенос влаги с океана на континент. Энергоактивные зоны Северной Атлантики вносят серьезный вклад в крупномасштабное взаимодействие океана с атмосферой, они фактически полностью определяют климат в Северной Атлантике. Норвежская ЭАЗ генерирует практически все атлантические циклоны, которые в свою очередь приводят к мощным осадкам на материке, таким образом, и на европейской части России.

2) Построено 13 полных диагностических моделей деревьев решений суммарной теплоотдачи энергоактивных зон Северной Атлантики с переносом влаги на меридиональном разрезе 5° в. д. Из этих полных моделей деревьев, на основе анализа статистических х-к: R^2 и вклада ошибки в СКО, для оценки общего качества моделей и выделения энергоактивных зон имеющих наилучшую связь со значениями зонального потока влаги, выбран единый шаг – шаг №5, который обеспечивает высокий коэффициент детерминации модели и малую ошибку СКО. При этом средний коэффициент детерминации составляет 0.72, а средняя ошибка по 13 моделям составляет 0.44 в долях СКО. В среднем наиболее высокая точность модели свойственна северной части меридионального разреза, R^2 составляет 0.78.

3) На основе полученных оптимальных моделей было выделено 29 реперных квадратов, в том числе в Норвежской ЭАЗ – 15, в Ньюфаулендской - 14, дающих максимальный вклад в межгодовую изменчивость зонального переноса влаги на меридиональном разрезе 5° в. д. При этом на северный участок зонального разреза, оказывает влияние 5 станций из Ньюфаудендской

ЭАЗО, 9 из Норвежской, на центральную 7 и 8, на южную 11 и 5 соответственно.

4) С заблаговременностью 1 год были построены прогностические модели энергоактивных зон Северной Атлантики с зональным потоком влаги, разбитым на три участка с выбранными 29 станциями ЭАЗО. Во всех трех случаях модель обучалась на зависимой выборке (1981-2009гг.). Для полных деревьев R^2 = 0.88, 0.87, 0.94 для северной, центральной и южной частей меридионального разреза соответственно. Стандартная ошибка моделей деревьев для независимой выборки довольно сильно варьируется, можно сказать, что имеет случайный характер, но, несмотря на это выбрать наилучшие деревья представляется возможным, что и было сделано.

5) Также была построена диагностическая модель значений суммарной теплоотдачи энергоактивных зон Северной Атлантики с приращениями уровня Каспийского моря, для оценки качества диагностической модели были рассчитаны коэффициенты детерминации и стандартная ошибка для каждого дерева. Уже шестом шаге R^2 оказался равен 0.74, а стандартная ошибка = 0.49 в долях СКО, 6 терминальных вершин.

На наш взгляд полученные результаты подтверждают гипотезу о роли уровня Каспия, как индикатора влагообмена системы «океан-атмосфера-суша».

Список использованной литературы.

1. Гордеева С.М., Малинин В.Н. О предвычислении годового стока крупных рек Европейской территории России на основе метода деревьев решений (decision trees)// Ученые записки РГГМУ. 2018. No 50. С. 53—65.
2. Малинин В.Н. Статистические методы анализа гидрометеорологической информации. Учебник,- СПб.: изд. РГГМУ, 2008
3. В.Н. Малинин, С.М. Гордеева. Уровень Каспийского моря как индикатор крупномасштабного влагообмена в системе «океан – атмосфера – суша» // Труды КарНЦ РАН. No 4. Сер. Лимнология. 2020.
4. Малинин В.Н. Долгосрочное прогнозирование уровня Каспийского моря// Известия Ран, Серия географическая, 2009, № 6, с. 7-16
5. Лактионов П.П., Малинин В.Н. Долгосрочный прогноз уровня Каспийского моря // сборник материалов конференции КИМО-2020
6. Малинин В.Н. Проблема прогноза уровня Каспийского моря. – СПб: РГГМУ, 1994. – 160с
7. Нестеров Е.С. Североатлантическое колебание: атмосфера и океан. –М.: Триада лтд, 2013. – 143 с.
8. Нестеров Е.С. О фазах североатлантического колебания // Метеорология и гидрология. – 2003. – № 1. – С. 64–74.
9. Нестеров Е.С. О связи индексов колебаний циркуляции атмосферы и уровня Каспийского моря // Гидрометеорологические аспекты проблемы Каспийского моря и его бассейна. – СПб: Гидрометеоиздат, 2003. – С. 323–325.

10. Нестеров Е.С. Низкочастотная изменчивость циркуляции атмосферы и уровень Каспийского моря во второй половине XX века // Метеорология и гидрология. – 2001. – № 11. – С. 27–36.
11. Раткович Д.Я., Жданова И.С., Привальский В.Е. К проблеме уровенного режима Каспийского моря // Водные ресурсы. – 1973. – № 3. – С. 45–69.
12. Шулейкин В.В. Взаимодействие звеньев в системе океан-атмосфера-материк // Материалы V съезда ВГО. – 1970. – Вып. 17
13. Malinin, V.N. & Shmakova, V.Yu. (2018). VARIABILITY OF THE ENERGY-ACTIVE OCEAN ZONES IN NORTH ATLANTIC. Fundamental and Applied Climatology. 4. 55-70. 10.21513/2410-8758-2018-4-55-70.
14. Клименко Л.В. Об увлажнении центральной части ЕТР и о происхождении засух // Метеорология и гидрология. – 1996. – № 7. – С. 21–25.
15. Шнитников А.В. Закономерности колебания уровня Каспийского моря в связи с изменчивостью общей увлажненности материков Северного полушария // Сверхдолгосрочные прогнозы уровня Каспийского моря. – М.,1957. – С. 23–47.
16. Гирс А.А. Эпохальные преобразования форм атмосферной циркуляции и связанные с ними колебания уровня Каспийского моря // Изв. АН СССР. Сер. географ. – 1957. – № 1. – С. 102–108.
17. Дзердзеевский Б.Л. Циркуляционные механизмы в атмосфере Северного полушария в XX столетии. – М.: Межвед. Геофиз ком., 1968. – 176 с.
18. Клиге Р.К., Ковалевский В.С., Федорченко Е.А. Влияние глобальных климатических изменений на водные ресурсы Волжского бассейна // Глобальные изменения природной среды (климат и водный режим). – М.: Научный мир, 2000. – С. 220–236.

19. Сидоренков Н.С., Швейкина В.И. Изменение климатического режима бассейна Волги и Каспийского моря за последнее столетие // Водные ресурсы. – 1996. – Т. 23, № 4. – С. 401–406.
20. Технико-экономический доклад (ТЭД): Анализ изменений уровня Каспийского моря. – М.: 1992. – 592 с.
21. Тужилкин В.С., Косарев А.Н., Архипкин В.С., Никонова Р.Е. Многолетняя изменчивость гидрологического режима Каспийского моря в связи с вариациями климата // Вестник МГУ. Сер. 5. География. – 2011. – № 2. – С. 62–71.
22. Белинский Н.А. Использование некоторых особенностей атмосферных процессов для долгосрочных прогнозов. – Л.: Гидрометеоиздат, 1957. – 203 с.
23. Берг Л.С. Избранные труды. Т. III. – М.: АН СССР, 1960. – 455 с.
24. Визе В.Ю. Колебания солнечной активности и ледовитости арктических морей. – М.: Изд-во ГУСМП, 1945. – 6 с.
25. Гирс А.А. Многолетние колебания атмосферной циркуляции и долгосрочные гидрометеорологические прогнозы. – Л.: Гидрометеоиздат,1971. –280с.
26. Ким И.С., Никулина С.П. Изменение уровня Каспийского моря и циркуляция атмосферы // Метеорология и гидрология. – 1994. – № 7. – С. 76–80.
27. Кренке А.Н., Попова В.В. Колебания уровня Каспийского моря и глобальная циркуляция атмосферы // Материалы метеорологических исследований. – 1992. – № 15. – С. 116–122.

28. Алексеев Г.В., Николаев Ю.В., Романцов В.А. Норвежская энергоактивная зона//Итоги науки и техники. Атмосфера, океан, космос – программа «Разрезы». М., ВИНИТИ. 1985. –Т.5. – С.45-62.
29. Баранов Е.И. и др. Гидрометеорологическая характеристика Ньюфаундлендской энергоактивной зоны// Итоги науки и техники. Атмосфера, океан, космос – программа «Разрезы». М., ВИНИТИ. 1985. – Т.5. – С.63-93.
30. Бурков В.А. Гидрометеорологический обзор энергоактивной зоны Гольфстрима// Итоги науки и техники. Атмосфера, океан, космос – программа «Разрезы». М., ВИНИТИ. 1985. –Т.5. – С.4-44.