С.М. Гордеева, В.Н. Малинин

ТИПИЗАЦИЯ ПОЛЕЙ АНОМАЛИЙ ТЕМПЕРАТУРЫ ПОВЕРХНОСТИ ЮГО-ВОСТОЧНОЙ ЧАСТИ ТИХОГО ОКЕАНА

S.M. Gordeyeva, V.N. Malinin

TYPE DESIGNS OF ANOMALY FIELDS OF SURFACE TEMPERATURE OF PACIFIC OCEAN'S SOUTH-EASTERN PART

Предлагается решение задачи автоматизированного построения типизации гидрометеорологических полей большого объема на примере поля среднемесячных аномалий температуры поверхности океана (АТПО) юго-восточной части Тихого океана. При этом задача разделяется на три основных этапа: снижение размерности поля АТПО, его классификация с целью выделения типовых полей и распознавание реальных (фактических) полей АТПО. Обсуждается каждый из этапов, приводятся результаты типизации полей АТПО за период январь 1982 — апрель 2006 гг.

Ключевые слова: температура поверхности океана, классификация полей, межгодовая изменчивость, анализ полей, факторный анализ, кластерный анализ, статистические критерии, подбор аналогов.

Problem solution of automated type design constructing of large size hydrometeorological fields based on the example of the field of monthly-averaged ocean surface temperature anomalies (OSTA) of Pacific Ocean's south-eastern part is proposed. The problem is divided into three main stages-decreasing of OSTA field dimensionality, its classification for selection of typical fields and identification of real (actual) OSTA fields. Each of the stages is discussed and the results of the OSTA field type designs for the period January, 1982 – April, 2006 are given.

Key words: ocean surface temperature, classification of fields, interannual variability, analysis of fields, factor analysis, cluster analysis, statistical criteria, fitting of analogs.

Район юго-восточной части Тихого океана (ЮВТО) в течение многих лет являлся важнейшим объектом рыбопромысловых интересов России. До 1991 г. там добывалось примерно 1 млн т тихоокеанской ставриды, затем промысел был прекращен в связи с общим экономическим кризисом в стране. Из-за истощения рыбных запасов в районах традиционного промысла интерес к этому региону в настоящее время снова возрос. В связи с этим возникла необходимость адекватного океанологического обеспечения возможного широкомасштабного промысла. В рамки этой задачи входит исследование крупномасштабных процессов в океане и атмосфере на основе новых, полученных в последние годы, данных с использованием современных вычислительных технологий.

С другой стороны, состояние сырьевой базы рыболовства в промысловых районах океана обычно основывается на долгосрочном прогнозе океанологических процессов и прежде всего прогнозе температуры воды. Очевидно, что предварительное построение системы мониторинга ТПО и определяющих ее колебания факторов во многом может облегчить решение данной задачи. При этом весьма важным представляется комплексное изучение их пространствен-

но-временной изменчивости при различных масштабах осреднения, которое осуществляется с помощью различных методов многомерного анализа.

Вообще говоря, при анализе пространственно-временной изменчивости мы имеем две противоположные задачи: типизация временной изменчивости исходного поля и районирование акватории по характеру колебаний. Действительно, обе задачи могут быть осуществлены по одной и той же матрице исходных данных. Однако если в первом случае типизация временной изменчивости гидрометеорологической характеристики выполняется с учетом ее пространственных изменений, то во втором – районирование акватории осуществляется уже по характеру временных колебаний этой характеристики в отдельных узлах акватории. Очевидно, в общем случае задача типизации поля АТПО большого объема может быть разделена на три этапа:

- 1. Снижение размерности исходного поля АТПО методом главных компонент.
- 2. Классификация полей среднемесячных значений АТПО на основе главных компонент.
- 3. Идентификация реальных (фактических) полей в соответствии с полученной классификацией.

Рассмотрим каждый из указанных этапов. Район ЮВТО был принят в пределах от 30 до 50° ю.ш. и от 73 до 100° з.д., что соответствует Южному промысловому подрайону [Промыслово-океанологические исследования..., 2002], характерной особенностью которого является возможность осуществления там круглогодичного рыбного промысла. В качестве источника информации о ТПО был использован архив NOAA NCEP EMC CMB GLOBAL ReynSmithOIv2, который является развитием известного архива COADS [Reynolds, 2002]. Из архива были выбраны среднемесячные ряды ТПО с января 1982 по апрель 2006 г. в 540 точках сетки $1 \times 1^{\circ}$.

Из исходных рядов предварительно был исключен сезонный ход, определяющий около 90 % общей дисперсии. Для этого были рассчитаны климатические аномалии ТПО (АТПО) как отклонения фактической ТПО от ее климатического сезонного хода.

Вначале исходная матрица (540 точек на 292 месяца) подверглась разложению методом главных компонент. Несмотря на большой размер исходной матрицы, сходимость собственных чисел оказалась весьма высокой, ибо первые семь чисел описывают уже более 90 % дисперсии поля АТПО. Кроме того, следует отметить весьма быстрое убывание относительных величин собственных чисел до четвертого номера. Затем скорость убывания резко замедляется. Второй скачок в убывании собственных чисел начинается с седьмого номера. Так как первые семь чисел описывают более 90 % дисперсии исходного поля, то достаточно уверенно можно сделать вывод о том, что остальными числами разложения можно пренебречь. Кроме того, другим пороговым числом вполне возможно является 4, которое ограничивает дисперсию поля АТПО на уровне 84 %.

Чтобы принять решение, какое число компонент использовать в дальнейшем анализе, оценивалась точность восстановления поля АТПО в отдельных узлах с помощью четырех и семи главных компонент разложения. Для этого были рассчитаны средние квадратические ошибки восстановления полей АТПО на основе четырех и семи главных компонент (ГК). Естественно, точность восстановления АТПО по четырем ГК несколько хуже, однако это касается только зоны, расположенной вдоль побережья Южной Америки. При этом максимальные стандартные ошибки восстановления полей АТПО по четырем ГК составляют $0.3\,^{\circ}$ C, а по семи компонентам $-0.25\,^{\circ}$ C. Расхождения в стандартных ошибках восстановления АТПО указанными способами для подавляющего большинства точек не превышают $0.1\,^{\circ}$ C, т.е. находятся в пределах точности наблюдений.

Сопоставление полей ошибок с полем стандартных отклонений АТПО показало, что коэффициент «надежности», определяемый как отношение значений средних квадратических ошибок восстановления полей АТПО к стандартным отклонениям АТПО для всей акватории океана как по семи, так и по четырем ГК меньше 0,5. Это означает высокую точность восстановления полей АТПО обоими способами. Но поскольку меньшее число ГК упрощает содержательный анализ, то ограничимся в дальнейшем первыми четырьмя ГК.

Следующий этап — это выполнение классификации аномалий ТПО. Классификация аномалий ТПО в пространстве первых четырех компонент была выполнена иерархическим алгоритмом методом Уорда на основе евклидовой метрики.

На рис. 1 приводится дендрограмма классификации, из которого видно, что однозначное выделение числа классов, вообще говоря, вряд ли возможно, поскольку отсутствуют явно выраженные скачки в пороговых расстояниях между отдельными классами. Из общих соображений понятно, что число однородных состояний временных колебаний АТПО, с одной стороны, не может быть очень малым, а с другой — чрезмерное увеличение их приводит к излишней детализации, мелкомасштабности и размыванию границ. Очевидно, в рассматриваемом случае оптимальное число классов может составлять от 3 до 7.

Каждый класс, представленный на дендрограмме, это совокупность фактических полей аномалий ТПО, близких друг к другу по пространственной структуре. В связи с этим возникает необходимость построения некоторых типовых (средних) полей АТПО, объединяющих особенности пространственного распределения аномалий ТПО в своем классе. С этой целью предварительно осуществлялось осреднение значений, вошедших в рассматриваемый класс, для каждой из четырех компонент, а затем производилось восстановление среднего (типового) поля АТПО, соответствующего данному классу.

Рассмотрим теперь распределение типовых полей АТПО для отдельных классов, соответствующих дендрограмме на рис. 1. На первом шаге вся совокупность из 292 полей АТПО разделяется на два класса, из которых первый класс (A) представляет собой распределение отрицательных аномалий ТПО,

а второй (\mathbf{B}) – распределение положительных аномалий ТПО (рис. 2). Отметим, что пространственное распределение обоих типов аномалий ТПО (особенно первый класс) практически в точности повторяет пространственное распределение 1-й ГК.

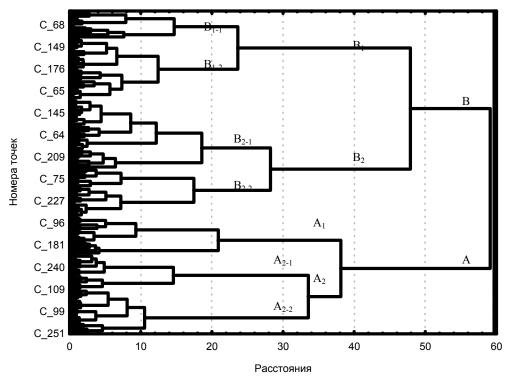


Рис. 1. Дендрограмма аномалий ТПО в пространстве четырех ГК, построенная методом Уорда (числами представлены варианты разбиения на классы).

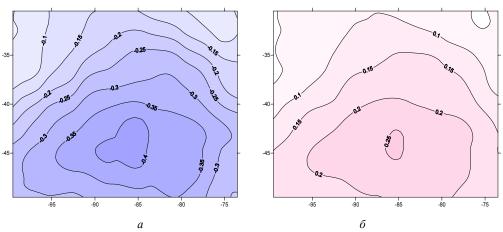


Рис. 2. Типовые поля ATПО, соответствующие классам $\mathbf{A}(a)$ и $\mathbf{B}(\delta)$.

Дальнейшее деление класса отрицательных аномалий A происходит на третьем шаге, когда первый класс разбивается на два класса A_1 и A_2 , а на следующем шаге класс A_2 делится ещё раз на классы A_{2-1} и A_{2-2} (рис. 3).

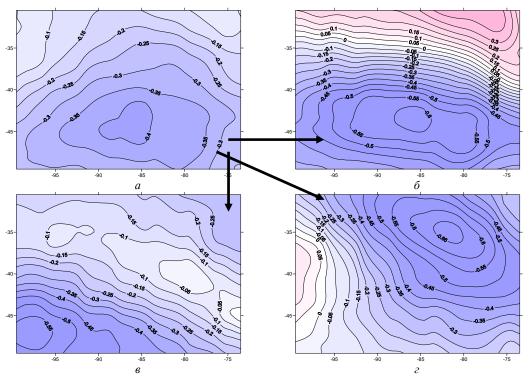


Рис. 3. Типовые поля АТПО, соответствующие классам A(a), $A_1(\delta)$, $A_{2-1}(\epsilon)$, $A_{2-2}(\epsilon)$.

Для класса A_1 , так же как и на первом шаге, пространственное распределение аномалий ТПО в общих чертах похоже на распределение 1-й ГК и отличается от нее интенсификацией отрицательных аномалий и появлением в северовосточной части ЮВТО слабых положительных аномалий ТПО. Промежуточный класс A_2 имеет характерное поле отрицательных аномалий ТПО, но только в отличие от класса A оно имеет иную конфигурацию: максимальные значения аномалий отмечаются в северо-восточной части ЮВТО. Дальнейшее деление этого класса на класс A_{2-1} лишь интенсифицирует отрицательные аномалии ТПО. Класс A_{2-2} , наоборот, имеет максимальные отрицательные аномалии ТПО в юго-западной части ЮВТО.

Типовое поле с положительными аномалиями ${\bf B}$ на втором шаге делится на классы B_1 и B_2 (рис. 4). Класс B_2 показывает, что значительные положительные аномалии ТПО сосредоточены в северо-западной части ЮВТО. Для класса B_1 характерна оппозиция аномалий ТПО, более высокие значения АТПО в северо-западной части региона и более низкие значения АТПО в его юго-западной части. Это практически точно соответствует распределению 2-й Γ K.

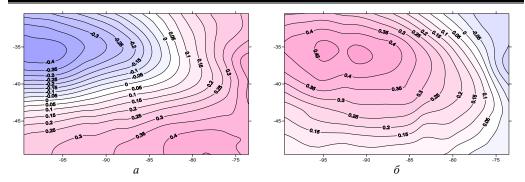


Рис. 4. Типовые поля АТПО, соответствующие классам $B_1(a)$, $B_2(\delta)$.

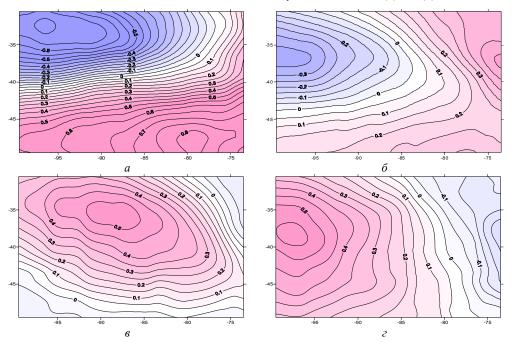


Рис. 5. Типовые поля АТПО, соответствующие классам $B_{1-1}(a)$, $B_{1-2}(\delta)$, $B_{2-1}(\epsilon)$, $B_{2-2}(\epsilon)$.

Далее класс B_2 делится на классы $B_{2\text{-}1}$ и $B_{2\text{-}2}$, а класс B_1 разделяется на классы $B_{1\text{-}1}$ и $B_{1\text{-}2}$ (рис. 5). Распределение аномалий для классов $B_{2\text{-}1}$ и B_2 весьма схоже, а в классе $B_{2\text{-}2}$ по сравнению с $B_{2\text{-}1}$ очаг положительных аномалий ТПО смещается к югу и захватывает практически всю западную часть региона. Аналогичным образом обстоит дело и при делении класса B_1 . Один из новых классов $B_{1\text{-}1}$ его почти полностью повторяет, в то время как для другого ($B_{1\text{-}2}$) характерен разворот оппозиции положительных и отрицательных аномалий ТПО против часовой стрелки с ослаблением самих аномалий.

Статистические характеристики для типовых полей приведены в табл. 1. Заметим, что среднее значение характеризует центр тяжести каждого класса, а стандартное отклонение представляет степень рассеяния в классе относительно центра. При условии нормального распределения исходных данных эти параметры являются исчерпывающими характеристиками статистического ряда.

Таблица 1
Статистические характеристики для типовых полей аномалий ТПО
при лелении их на семь классов

npii gettennii nx nu cemb muceob							
Характеристика	Класс 1 (A ₁)	Класс 2 (A ₂₋₁)	Класс 3 (A ₂₋₂)	Класс 4 (B ₁₋₁)	Класс 5 (B ₁₋₂)	Класс 6 (B ₂₋₁)	Класс 7 (B ₂₋₂)
Число случаев	36	32	39	24	56	62	43
Среднее значение, °С	-0,232	-0,232	-0,307	0,133	0,079	0,212	0,184
Ст.отклонение, °С	0,303	0,146	0,203	0,480	0,201	0,183	0,220
Размах колебаний, °С	1,009	0,609	0,817	1,543	0,835	0,661	0,856
Максимум, °С	0,381	0,019	0,150	0,826	0,464	0,532	0,570
Минимум,°С	-0,628	-0,591	-0,668	-0,717	-0,371	-0,129	-0,286

Нетрудно видеть, что для классов более холодного состояния АТПО наибольшие отрицательные аномалии ТПО в среднем характерны для класса A_{2-2} , в то время как для двух других классов они практически одинаковы. Что касается оценок стандартного отклонения, то они существенно различны для всех трех классов. При этом для поля A_1 стандартное отклонение почти в три раза выше по сравнению с полем A_{2-1} . По критерию Фишера расхождения между значениями дисперсии всех трех типовых полей с отрицательными аномалиями ТПО являются значимыми.

По сравнению с группой классов с отрицательными аномалиями различия между полями классов с положительными аномалиями ТПО существенно меньше.

Итак, из анализа классификации следует, что возможно два альтернативных варианта оптимального числа классов: первый – 5 классов, второй – 7 классов. Поскольку, исходя только из физических соображений, выбрать лучший из них не представляется возможным, то для оценки этих классификаций обратимся к статистическим критериям. Действительно, если по критериям Стьюдента и Фишера центры классов при их делении различаются значимо, то это означает, что классы хорошо различимы. Очевидно, дополнительно должен использоваться также и коэффициент корреляции. Если сравниваемые классы имеют высокую положительную корреляцию, то это значит, что классы повторяют друг друга. В табл. 2 приводятся оценки расхождений типовых полей аномалий ТПО на основе указанных статистических критериев, критические значения которых даны при уровне значимости $\alpha = 0,05$.

Как видно из табл. 2, классы A_1 и A_{2-1} , а также A_{2-1} и A_{2-2} значимо различаются по критерию Фишера, классы B_{2-1} и B_{1-1} — одновременно по обоим критериям, классы B_{2-1} и B_{2-2} по критериям Стьюдента и Фишера неразличимы, а классы B_{1-1} и B_{1-2} значимо различаются по критерию Фишера. Дополнительно

отметим, что между классами $B_{1\text{--}1}$ и $B_{1\text{--}2}$ отмечается высокая положительная корреляция, т.е. они в значительной степени повторяют друг друга.

Отсюда следует, что разделение исходной совокупности на семь классов вряд ли оправданно. Очевидно, целесообразно ограничиться только пятью классами, так как при этих условиях различия между центрами классов B_1 и B_2 по обоим критериям уже значимы. Правда, следует иметь в виду, что тогда число случаев появления полей с положительными АТПО резко увеличивается и значительно превышает число случаев для отрицательных аномалий ТПО.

Таблица 2 Оценка расхождений типовых полей аномалий ТПО, полученных метолом Уорла с помощью статистических критериев

полу тенных методом в орда е помощью статиети теских критериев							
Пары классов	Критерий Стьюдента		Критер	Vannanguug			
	$t_{\scriptscriptstyle m BHP}$	$t_{\rm kp}(\alpha=0.05)$	$F_{\text{выч}}$	$F_{\kappa p}(\alpha = 0.05)$	Корреляция		
А ₁ и А ₂₋₁	0,01	2,00	4,28	1,80	0,29		
А ₂₋₁ и А ₂₋₂	1,80	2,00	1,97	1,78	-0,63		
В ₂₋₁ и В ₁₋₁	3,23	1,97	2,51	1,41	-0,85		
В ₂₋₁ и В ₂₋₂	0,69	1,98	1,44	1,60	0,33		
В ₁₋₁ и В ₁₋₂	0,53	1,99	5,70	1,73	0,63		

Статистические характеристики для типовых полей аномалий ТПО при делении их на пять классов представлены в табл. 3.

Таблица 3 Статистические характеристики для типовых полей аномалий ТПО при делении их на пять классов

Характеристика	Класс 1 (A ₁)	Класс 2 (A ₂₋₁)	Класс 3 (A ₂₋₂)	Класс 4 (B ₁)	Класс 5 (B ₂)
Число случаев	36	32	39	80	105
Среднее значение, °С	-0,232	-0,232	-0,307	0,09	0,20
Ст.отклонение, °С	0,303	0,146	0,203	0,26	0,16
Размах колебаний, °С	1,009	0,609	0,817	0,88	0,64
Максимум, °С	0,381	0,019	0,150	0,45	0,46
Минимум, ^о С	-0,628	-0,591	-0,668	-0,43	-0,18

Итак, из результатов анализа полей АТПО с помощью иерархической классификации можно сделать вывод, что предпочтительней является вариант разбиения их на пять типовых ситуаций (классов), которые достаточно четко отличаются друг от друга и хорошо физически понятны.

Выделенные типы полей АТПО, так же, например, как формы атмосферной циркуляции Вангейгема-Гирса, есть некоторый идеальный «слепок» реальной временной изменчивости температуры воды за довольно продолжительный период времени. В этом легко убедиться, если обратиться к табл. 4, в которой даны результаты сравнения «гипотетических» типовых полей, полученных в результате иерархической классификации методом Уорда и последующего вос-

становления их по четырем ГК с результатами осреднения реальных (фактических) полей АТПО, соответствующих выделенным классам. Как видно из табл. 4, корреляция между этими полями близка к функциональной линейной связи. Средняя квадратическая разность восстановленных и осредненных полей АТПО меньше 0,04 °C, причем максимальная разность не превышает 0,1 °C.

Результаты сравнения типовых полей АТПО с осредненными фактическими полями АТПО, соответствующими выделенным классам

Таблица 4

Характеристика	A_1	B_1	B_2	A ₂₋₁	A ₂₋₂
Коэффициент корреляции	0,992	0,998	0,998	0,961	0,993
Среднее квадратическое расхождение, °С	0,04	0,02	0,01	0,04	0,02
Максимальная разность полей, °С	0,10	0,05	-0,05	0,10	-0,09

Естественно, что единичное реальное поле АТПО обладает локальными неоднородностями, мезомасштабной изменчивостью и далеко не во всех случаях точно копирует типовое поле, т.е. не является его точным «слепком». Именно поэтому и возникает задача распознавания реальных полей АТПО на принадлежность их типовым полям. По-видимому, решение данной задачи возможно несколькими способами:

- оценка среднего квадратического расхождения между фактическими и типовыми полями АТПО;
- оценка коэффициента корреляции между фактическими и типовыми полями АТПО;
 - дискриминантный анализ.

Все указанные способы были реализованы в данной работе. Однако наилучшие результаты были получены по первому способу. СКО-оценка каждого (*j*) месяца от типового (эталонного) поля рассчитывается как

$$\delta_{j} = \sqrt{\sum_{i=1}^{n} (x_{i,j(\phi \text{akt})} - x_{i(\text{TMII})})^{2}},$$

где n — число точек, равное 540; $x_{ij(\phi a \kappa r)}$ — фактическая аномалия ТПО в i-й точке поля для (j) месяца; $x_{i(\tau u n)}$ — «типовая» аномалия ТПО в i-й точке поля.

Заметим, что данная формула эквивалентна евклидовой метрике. Если рассчитать оценки среднего квадратического расхождения фактических полей и каждого из типовых полей аномалий ТПО, то получим матрицу $k \times 292$, где k – число типовых полей. Отнесение конкретного поля аномалий ТПО к типовому полю осуществляется по минимальному в строке значению δ_i .

Оправдываемость распознавания фактических полей АТПО определялась следующим образом. Если минимальная СКО-оценка приходится на класс, определенный путем классификации, то это считалось правильным распознаванием или, другими словами, оправдавшимся случаем. Если же для какого-либо месяца рассчитанная СКО-оценка оказывалась минимальной не для «своего»

типового поля, а для «чужого», то это считалось неоправдавшимся случаем. В табл. 5 даются оценки оправдываемости и СКО для пяти классов, полученных методом Уорда. При этом статистические характеристики СКО считались только по числу оправдавшихся случаев.

Таблица 5 Сравнение результатов оправдываемости распознавания аномалий ТПО

на основе СКО-оценок для иерархической классификации							
Характеристика	A_1	\mathbf{B}_{1}	B_2	A ₂₋₁	A ₂₋₂	В целом	
Всего случаев	36	80	105	32	39	292	
Число неоправдавшихся случаев	7	6	6	11	3	33	
Оправдываемость, %	80,6	92,5	94,3	65,6	92,3	88,7	
Максимальная величина СКО	0,71	0,93	0,79	0,49	0,46	0,93	
Минимальная величина СКО	0,20	0,16	0,21	0,21	0,16	0,16	

Общее число не оправдавшихся случаев распознавания составляет 33, что соответствует оправдываемости 88,7%. Наиболее «слабым звеном» классификации является класс A_{2-1} с отрицательными аномалиями ТПО (оправдываемость 65,6%).

Отметим, что точность распознавания аномалий ТПО для классификации из семи классов существенно хуже по сравнению с пятью классами. Так, общее число не оправдавшихся случаев распознавания полей АТПО составляет 74, что соответствует оправдываемости 74,7 %. Следовательно, число не оправдавшихся случаев практически в два раза больше по сравнению с классификацией из пяти классов. Существенно больше и значения СКО, причем как максимальные, так и минимальные. Таким образом, результаты, полученные выше на основе статистических критериев, полностью подтвердились.

Проанализируем теперь не оправдавшиеся случаи распознавания. В иерархической классификации, как уже отмечалось выше, самым «плохим» является класс A_{2-1} , в котором из 11 не оправдавшихся случаев 8 попадает на класс A_{2-2} , 2- на A_1 и 1- на B_1 . Причины этого заключаются в том, что различия между классами A_{2-1} и A_{2-2} не столь значительны, как между другими классами. Действительно, центры классов практически неразличимы по критерию Стьюдента. Очень важно также, что для всех не оправдавшихся случаев распознавания класс A_{2-1} имел вторую величину СКО. Кроме того, отметим, что из 11 неоправдавшихся случаев на лето (с декабря по март) приходится 5, т.е. почти половина всех ошибок. Это означает, что при независимом распознавании нужно выбирать две минимальные оценки СКО и потом визуально определять, какой из двух выбранных типовых классов лучше соответствует фактическому полю АТПО.

Выше были рассмотрены оценки СКО для акватории ЮВТО в целом. Однако представляет интерес анализ пространственного распределения ошибок распознавания фактических полей аномалий ТПО для отдельных классов. С этой целью выполнен расчет средних квадратических отклонений реальных полей от типовых (модельных) для пяти классов иерархической классификации в каждой точке акватории, пространственное распределение которых приводит-

ся на рис. 6, где диаметр точки пропорционален величине СКО: чем он больше, тем больше величина СКО.

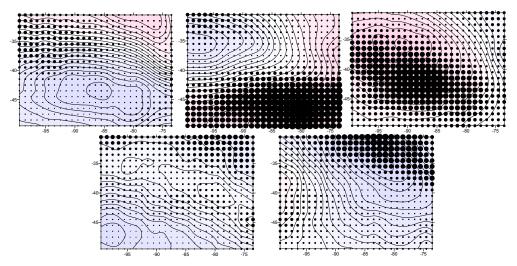


Рис. 6. Распределение средних квадратических отклонений фактических полей АТПО от типовых в узлах сеточной области ЮВТО для пяти классов иерархической классификации. Изолиниями представлены АТПО. Диаметр точек пропорционален величине СКО.

Как видно из рис. 6, значения СКО в каждом классе имеют ярко выраженное неравномерное распределение, причем очаги больших ошибок практически не повторяются. Так, в классе A_1 очаг максимальных ошибок сосредоточен в северо-западной части ЮВТО, в классе B_1 – в юго-восточном районе, в классе B_2 – в центральной части акватории, в классе A_{2-2} – в северо-восточном районе. И только в классе A_{2-1} значения СКО распределяются более или менее равномерно по акватории.

Итак, предложенное разделение полей АТПО на пять компактных однородных групп (типов) адекватно отражает крупномасштабные особенности изменчивости температуры поверхности юго-восточной части Тихого океана. Отметим, что применение комплекса методов многомерного статистического анализа позволяет довольно эффективно решать задачи, связанные с временной типизацией полей гидрометеорологических характеристик большого объема. В частности, рассмотренная выше процедура типизации может быть положена в основу создания каталога с однородными состояниями полей гидрометеорологических характеристик.

Литература

- 1. Промыслово-океанологические исследования в Атлантическом океане и южной части Тихого океана (по результатам исследования АтлантНИРО и запрыбпромразведки) / Под ред. В.Н. Яковлева. Калининград: изд. АтлантНИРО, 2002. 510 с.
- 2. Reynolds R.W. et al. An Improved in Situ and Satellite SST Analysis for Climate [текст] // J Climate, 2002, 15, p. 1609–1625.