

УДК 502/504

ОПТИМИЗАЦИЯ СХЕМ ОТБОРА ГЕОЭКОЛОГИЧЕСКИХ ПРОБ В ТРУДНОДОСТУПНЫХ АРКТИЧЕСКИХ РАЙОНАХ: СТАТИСТИЧЕСКИЙ ПОДХОД*

¹Мискевич И.В., ¹Котова Е.И., ¹Лохов А.С., ^{1,2}Нецветева О.П.

¹Институт океанологии им. П.П. Ширшова РАН, Москва, e-mail: subarct@gmail.com;

²Национальный парк «Русская Арктика», Архангельск

Для объективного анализа изучаемого процесса необходимо в кратчайшие сроки выполнить большой объем наблюдений, однако в арктических условиях это редко бывает возможным в основном из-за их труднодоступности. Также вопрос о минимально необходимом числе наблюдений может подниматься при высокой стоимости или трудоемкости обработки и анализа образцов. Авторами проведено исследование, призванное ответить на вопрос: какое минимальное количество наблюдений необходимо для получения достоверных результатов? В качестве минимального объема для геоэкологических наблюдений было выбрано предложенное Д. Тьюки число три. Проверка этого утверждения проведена на основании данных по содержанию взвешенных веществ в снеге и воде, а также кислородонасыщению вод (устье р. Кянды, Белое море), путем исследования различных статистических характеристик. Рассчитаны изменчивость показателей среднеарифметического значения, медианы и трехсреднего значения при различной длине выборок (от 2 до 20 значений) на основе этих данных. В результате расчетов четко прослеживается максимальная изменчивость среднеарифметического значения и минимальная изменчивость медианы по мере возрастания размера выборок. Показан возможный диапазон изменчивости Т-критерия Стьюдента для малых выборок. Даны рекомендации по отбору минимум трех проб различных компонентов природной среды, с учетом характера их изменчивости. Например, в случае исследования гидролого-гидрохимических характеристик приливных устьев рек, многие из которых являются периодическими, рекомендовано проводить отбор не менее трех наблюдений в каждую фазу приливо-отливных явлений. И учитывая изменчивость этих параметров для данного случая – не менее чем в 3 местах, в соответствии с особенностями функционирования маргинальных фильтров рек.

Ключевые слова: мониторинг, Арктика, оптимизация пробоотбора, статистика, геоэкология

**Исследования проведены в ходе выполнения государственного задания по теме «Современные и древние донные осадки и взвесь Мирового океана – геологическая летопись изменений среды и климата: рассеянное осадочное вещество и донные осадки морей России, Атлантического, Тихого и Северного Ледовитого океанов – литологические, геохимические и микропалеонтологические исследования; изучение загрязнений, палеообстановок и процессов в маргинальных фильтрах рек» (№ FMWE-2021-0006).*

ENVIRONMENTAL SAMPLING SCHEMES OPTIMIZATION IN HARD-TO-ACCESS ARCTIC AREAS: A STATISTICAL APPROACH

¹Miskevich I.V., ¹Kotova E.I., ¹Lokhov A.S., ^{1,2}Netsvetaeva O.P.

¹Shirshov Institute of Oceanology, Russian Academy of Sciences, Moscow,

e-mail: subarct@gmail.com

²Russian Arctic National Park, Arkhangelsk

The Arctic is a hard-to-reach region. This circumstance significantly increases the cost and complexity of any research. Thus, it is necessary to collect the enough minimum number of samples for objective analysis in the shortest possible time. Paper investigates a question – what is the minimum sampling number required to obtain reliable research? We chose the number three proposed by John Tukey as the minimum sample number in environmental research. John Tukey argued that statistics begin with three analyzed numbers. To test this hypothesis, we studied the statistical characteristics (arithmetic mean, median, Tukey's trimean) for selections with different sizes (from 2 to 20 values). We use three different selections of environmental data. This is suspended matter concentration in snow and water as well as water oxygen saturation in the Kyanda river mouth (White Sea). According to calculation, it conclude that with increase of the selections size arithmetic mean variability increase and the median variability decrease. Paper describes calculation of the Student's t-test possible variability range for arbitrary small samples. Paper gives recommendations to collect at least three samples of various components taking into account them variability. For example to investigate hydrological and hydrochemical characteristics of the tidal river mouths we recommend to collect at least three sample for each tidal phase. Because most of these characteristics are periodical. According to variability of these characteristics, samples need to be taken in at least three places with pay attention to river marginal filter specialties.

Keywords: monitoring, Arctic, sampling optimization, statistic, environmental science

При исследовании природных объектов на арктических территориях в условиях отсутствия развитой транспортной инфраструктуры продолжительные и масштабные наблюдения требуют очень больших финансовых

и технических затрат. С одной стороны, для объективного анализа изучаемого процесса должно быть выполнено большое количество наблюдений, с другой – время для проведения наблюдений должно быть очень коротким.

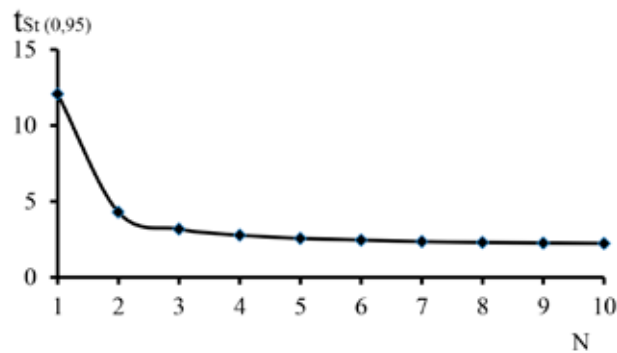


Рис. 1. Изменчивость критерия Стьюдента при доверительной вероятности в 95% для различной длины выборки данных

В первую очередь, это касается арктических островов и труднодоступных территорий, где для проведения исследований обычно необходимо применять вертолеты, ограниченные полетным временем. Определенные ограничения подобного рода возникают в морских экспедициях при высадке экспедиционного отряда на берег на шлюпке с борта морского судна для проведения серии наблюдений.

Наличие ограничений по объему возможных перевозимых грузов также накладывает ограничения на возможности объема пробоотбора. В данных ситуациях возникает вопрос определения оптимального количества необходимых наблюдений, а также достаточного объема проб, которые будут подлежать транспортировке.

В итоге должен быть найден определенный компромисс, который позволяет получать достоверную статистическую оценку изучаемого процесса при минимуме финансовых и технических затрат на проведение геоэкологических наблюдений. Он должен базироваться на определении минимального количества наблюдений, позволяющего достичь поставленных научных и (или) прикладных задач.

Известный американский математик Джон Тьюки утверждал, что статистика начинается с наличия трех анализируемых цифр [1]. Действительно, три значения, в отличие от двух значений, при проведении статистической обработки уже позволяют проводить как оценку масштаба положения (среднеарифметическое значение, медиана и т.п.), так и оценку масштаба изменчивости (стандартное отклонение, межквартильный размах и т.п.). При этом число «три» позволяет проводить такую обработку как с помощью стандартных параметров при

гипотезе сохранения нормального закона распределения данных, так и с использованием робастных статистик без учета его наличия [2].

В пользу указанного числа также говорит связь критерия Стьюдента с длиной используемой выборки данных. С трех цифр выборки начинается плавное понижение его величины (рис. 1), что снижает погрешность производимых статистических расчетов.

Исследования в данном направлении, проведенные в системе Росгидромета в конце прошлого века для гидролого-гидрохимических рядов наблюдений, позволили дать следующие рекомендации [2; 3]:

- при длине выборки пять наблюдений и меньше лучше использовать медиану;
- при длине выборки более пяти наблюдений целесообразно использовать трехсреднее значение.

Напомним, что медиана разделяет ранжированную выборку данных пополам и определяется по формуле (1):

$$C_M = \begin{cases} C_{(n+1)/2}, & \text{если } n \text{ нечетное,} \\ 0,5(C_{n/2} + C_{n/2+1}), & \text{если } n \text{ четное,} \end{cases} \quad (1)$$

где n – количество данных в выборке. Отметим, что в данном случае для расчетов использовалась её интегральная форма.

Трехсреднее значение находится по выражению (2):

$$C_{3cp} = 0,25(C_{0,25} + 2C_M + C_{0,75}), \quad (2)$$

где $C_{0,25}$ и $C_{0,75}$ – квартили, или, другими словами, медианы для половинок выборки.

Интерквартильный размах, который может выступать в качестве оценки изменчивости, определяется по формуле (3):

$$H = C_{0,75} - C_{0,25}. \quad (3)$$

Однако вышеуказанные рекомендации были получены на основе исследований выборок температуры воды и солености, которые, в отличие от многих геоэкологических параметров, обладают большей консервативностью в пределах внутрисезонных колебаний и не имеют прямой зависимости от техногенного влияния. Последнее может сформировать даже внутри малой выборки значения, отличающиеся друг от друга на 1–3 порядка.

Материалы и методы исследования

Для проверки выбора числа «три» в качестве минимального объема для геоэкологических наблюдений были исследованы статистические характеристики различных параметров для выборок данных различного размера. Рассматривались приведенные ниже параметры.

– Высота снежного покрова на метеостанции «Холмогоры» Приморского района Архангельской области. Побережье Белого моря находится в зоне активной циклонической деятельности, которая наиболее интенсивна осенью и зимой. Высота снежной толщи, помимо количества выпавших осадков, зависит от многих факторов, в частности от наличия оттепелей, которые могут приводить к уплотнению снежной толщи. В результате значения характеристик снежной толщи меняются в широком диапазоне год от года [4].

– Содержание взвешенных веществ в снежном покрове в Приморском районе Архангельской области в марте 2019 года. Вследствие многофакторности процесса формирования состава снежного покрова содержание веществ в снеге имеет значительную пространственную и временную неоднородность [5].

– Содержание взвешенных веществ в воде и степень насыщенности вод кислородом (кислородонасыщение) в зоне смешения речных и морских вод в устье р. Кянды в юго-восточной части Онежского залива Белого моря в августе 2016 года. Воды этого объекта обладают большой изменчивостью геоэкологических характеристик за счет короткопериодной приливной изменчивости фронтального раздела между пресными и солеными водами при наличии антропогенного влияния, т.к. водосбор устья р. Кянды располагается в зоне дренируемых сельскохозяйственных угодий (в период исследований здесь производилась заготовка сена и выпас скота) [6].

Результаты исследования и их обсуждение

Результаты статистических исследований показаны на рисунке 2 и в таблице.

Анализ полученных материалов показал, что данные по высоте снежного покрова распределены по нормальному закону, поэтому их статистическая интерпретация стандартным методом дает результаты не хуже, чем робастный подход. В рамках данной статьи далее они не рассматривались.

Характеристика размаха различных оценок масштаба положения для геоэкологических характеристик при изменении объема наблюдений

Показатель	N	C _{макс.} – C _{мин.}		
		C _{ср.}	C _М	C _{Зср.}
Содержание взвеси в снеге, мг/л	12	2,20	1,34	1,57
Содержание взвеси в водах устья р. Кянды	20	8,8	4,4	5,1
Кислородонасыщение вод устья р. Кянды	20	6,4	4,2	5,4

На рисунке 2 и в таблице для рассматриваемых показателей четко прослеживается максимальная изменчивость параметра C_{ср.} и минимальная – параметра C_М по мере возрастания количества наблюдений (N).

Таким образом, при планировании геоэкологических исследований в труднодоступных арктических районах для снижения финансовых издержек рекомендуется ориентироваться на отбор минимум трех проб различных компонентов природной среды.

При статической обработке полученных данных в качестве критерия оценки величины необходимо использовать медиану. Она, в отличие от среднеарифметического значения, минимизирует погрешности, возникающие при значительном сокращении количества отбираемых проб воды, снега (льда) и почвы (донных отложений).

В рамках рассматриваемой проблемы также можно дать рекомендации по минимальному количеству отбираемых проб при изучении маргинальных фильтров в устьях рек Северного Ледовитого океана. Напомним, что, согласно модели маргинального фильтра академика Лисицына А.П., в зоне устьевых вод по мере возрастания солености формируется три последовательные зоны: мутьевая, геохимическая и биологическая «пробки» [7].

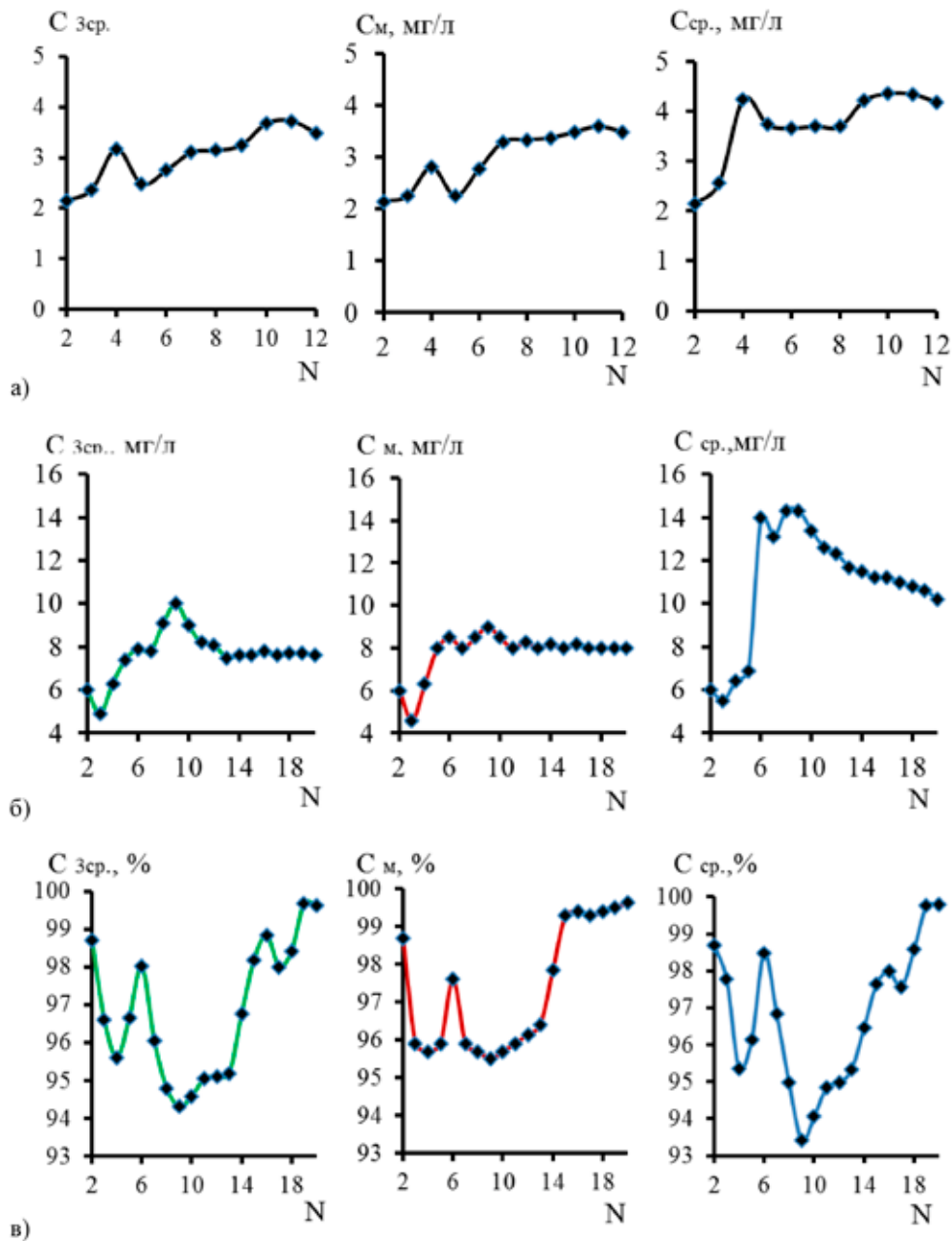


Рис. 2. Изменчивость статистик: $C_{cp.}$ – среднеарифметическое значение, C_m – медиана, $C_{3cp.}$ – трехсреднее значение, N – количество наблюдений) для:
 а) содержания взвеси в пробах снега; б) содержания взвеси в водах устья р. Кянды;
 в) кислородонасыщения вод устья р. Кянды

В первой из них фиксируются максимальные концентрации взвеси, во второй зоне отмечается наибольшая интенсивность геохимических процессов, а в третьей – повышенная биопродуктивность морских вод. Например, в Белом море первой зоне соответствует интервал солёности в 0,5–5‰, второй зоне – 5–20‰ и третьей зоне – солёность более 20‰ [8; 9].

Минимальное количество отбираемых проб при изучении маргинальных фильтров в устьях рек Северного Ледовитого океана составляет по три пробы в каждой из вышеперечисленных зон, тогда их общий минимум равняется 9.

Помимо этого, необходимо учитывать особенности конкретных геоэкологических показателей, например при исследо-

вании содержания взвесей в водах с явно выраженными приливно-отличными явлениями минимум составит 3 пробы в каждой зоне на каждую фазу прилива/отлива. Это число уже позволяет проводить необходимый корреляционный и регрессионный анализ полученных данных. Данная задача особенно актуальна для отбора воды и донных отложений на анализ дорогостоящих ингредиентов (ртуть, ХОП, ПХБ, диоксины и т.п.) [10].

Заключение

Анализ статистических характеристик для выборок геоэкологических данных различного размера показал, что при планировании геоэкологических исследований в труднодоступных арктических районах для снижения финансовых издержек рекомендуется ориентироваться на отбор минимум трех проб каждого из компонентов природной среды.

В качестве критерия оценки параметра необходимо использовать медиану, которая минимизирует погрешности, возникающие при значительном сокращении количества отбираемых проб.

При отборе проб следует учитывать особенности конкретных геоэкологических показателей и проводить три отбора в разных условиях, что кратко увеличит количество наблюдений.

Список литературы

1. Tukey J.W. *Exploratory Data Analysis* – Pearson, 2019. 720 p.
2. Микулинская С.М., Рожков В.А. Обработка малых выборок // Режимобразующие факторы, информационная база и методы ее анализа. Л.: Гидрометеоздат, 1989. С. 167–176.
3. Микулинская С.М., Михайлов Н.Н., Рожков В.А. и др. Метод анализа сезонной изменчивости неэквидистантных временных рядов гидролого-гидрохимических данных // Режимобразующие факторы, информационная база и методы ее анализа. Л.: Гидрометеоздат, 1989. С. 109–132.
4. Коршунова Н.Н., Давлетшин С.Г., Аржанова Н.М. Изменчивость характеристик снежного покрова на территории России // Фундаментальная и прикладная климатология. 2021. Т. 7. № 1. С. 80–100. DOI: 10.21513/2410-8758-2021-1-80-100.
5. Котова Е.И., Шевченко В.П. Влияние Дальнего атмосферного переноса на формирование ионного состава атмосферных осадков и снежного покрова прибрежной зоны Западного сектора Российской Арктики // Фундаментальные исследования. 2014. № 12–11. С. 2378–2382.
6. Мискевич И.В., Алабян А.М., Коробов В.Б., Демиденко Н.А., Попядухин А.А. Исследования короткопериодной изменчивости гидролого-гидрохимических характеристик устья реки Кянда в Онежском заливе Белого моря (28 июля–15 августа 2016 г.) // Океанология. 2018. Т. 58. № 3. С. 369–373.
7. Лисицын А.П. Маргинальный фильтр океанов // Океанология. 1994. Т. 4. № 5. С. 735–747.
8. Гордеев В.В., Филиппов А.С., Кравчишина М.Д., Новигатский А.Н., Покровский О.С., Шевченко В.П., Дара О.М. Особенности геохимии речного стока в Белое море // Система Белого моря. Т. II. Водная толща и взаимодействующие с ней атмосфера, криосфера, речной сток и биосфера. М.: Научный мир, 2012. С. 225–308.
9. Лещёв А.В. Особенности трансформации и переноса взвеси в устьевой области реки Северной Двины // Проблемы региональной экологии. 2013. № 3. С. 114–119.
10. Коробов В.Б., Шевченко В.П., Котова Е.И. Нерешенные задачи исследования устьевой области Северной Двины // Океанологические исследования. 2022. Т. 50. № 2. С. 125–138.