

УДК 656.073.7

DOI 10.26118/2782-4586.2024.93.44.014

ТУАРШЕВА ОЛЬГА АЛЕКСАНДРОВНА

**Государственный университет морского и речного флота имени
адмирала С.О. Макарова**

**Гистограммы как метод статистических технологий для
исследования потоков событий на флоте**

Аннотация. В статье представлена одна из статистических технологий, используемых для исследования потоков событий, в т.ч. в судоходстве и стивидорной деятельности, на основе гистограмм. Рассмотрена классификация этих потоков по содержанию. Приведены параметры, характеризующие рассматриваемые потоки событий, обоснован их приоритет по их значению с точки зрения безопасности функционирования и жизнеобеспечения. На примере распределения времени обслуживания судов на шлюзе показана неопределенность отображения его характера с помощью гистограмм. При этом рассмотрены три способа разбиения диапазона времени обслуживания судов на интервалы. Сравнение полученных гистограмм показало заметное их отличие друг от друга, несмотря на то, что все они отображают один и тот же поток событий в наблюдаемом диапазоне времени. Если каждый исследователь имеет возможность выбирать тот закон распределения, который кажется ему наиболее предпочтительным по его субъективным представлениям, то в этом случае нарушается объективность адекватного отображения распределения потоков событий в виде гистограмм. Таким образом, уже на этапе первичной обработки данных при поиске законов распределения наблюдаемых потоков событий на основе гистограмм традиционная статистика не дает однозначного ответа по решению выявленных проблем. В условиях высокой интенсивности обменных процессов постиндустриального общества на уровне предельных ограничений эти проблемы - актуальны и дают заказ для науки.

Ключевые слова: судоходство, стивидорная деятельность, флот, судно, шлюз, канал, статистическая технология, управление, ритм функционирования транспортных систем, режим движения транспортных потоков, интервал между событиями, частота событий, распределение событий.

TUARSHEVA OLGA ALEXANDROVNA

Admiral Makarov State University of Maritime and Inland Shipping

Histograms as a method of studying fleet flows

Annotation. The article presents one of the statistical technologies used to study the flows of events, including in shipping and stevedoring activities, based on histograms. The classification of these streams by content is considered. The parameters characterizing the considered streams of events are given. The priority of flows in terms of their importance from the point of view of safety of functioning and life support is justified. Using a specific example of the distribution of ship maintenance time at the gateway, the uncertainty of displaying its nature using histograms is shown. At the same time, three ways of dividing the time range of ship maintenance into intervals are considered. A comparison of the obtained histograms showed their noticeable difference from each other, despite the fact that they all display the same stream of events in the observed time range. If each researcher has the opportunity to choose the distribution law that seems to him the most preferable according to his subjective ideas, then in this case the objectivity of an adequate representation of the distribution of the observed streams of events in the form of histograms is violated. Thus, already at the stage of primary data processing, when searching for the laws of distribution of observed event flows based on histograms, traditional statistics do not provide an unambiguous answer to solve the identified problems. In the conditions of an intensive path of development of a post-industrial society with a high intensity of metabolic processes at the level of extreme restrictions, these problems are relevant and give an order for science.

Keywords: shipping, stevedoring, fleet, ship, gateway, channel, statistical technology, management, rhythm of functioning of transport systems, mode of movement of traffic flows, interval between events, frequency of events, distribution of events.

В транспортных системах наблюдаются потоки событий различного содержания. Одни из них характеризуют ритмы систем обслуживания транспортных потоков (в портах, на шлюзах и др.) другие – ритмы движения этих потоков (по водным магистралям). Третьи – потоки отказов разных компонент транспортных систем (причалов в порту, шлюзов), четвертые – потоки соблюдения или несоблюдения установленного регламента работы транспортных систем (например, нарушения нормативного времени стоянок флота и др.) или установленного режима обслуживания транспортных потоков. Пятые характеризуют степень соблюдения или несоблюдения уровня управляемости, согласованности или организованности режимов функционирования транспортных систем (портов, шлюзов) и режимов движения транспортных потоков (флота и смежных видов транспорта).

Эти потоки событий в транспортных системах характеризуются разными параметрами, значения которых в той или иной шкале, в той или иной метрике могут иметь частично организованный и частично неорганизованный характер распределения в определенном диапазоне – от минимума до максимума.

В режимах функционирования транспортных систем определяющими являются потоки событий, характеризующие ритм их работ, потому что они имеют первостепенное значение с точки зрения безопасности функционирования и жизнеобеспечения.

Следовательно, на первое место выступают потоки отказов и потоки нарушения нормативных режимов работы по управляемости, согласованности и организованности.

Приоритет первых потоков (потоков отказов) определяется их массовостью, регулярностью и непрерывной наблюдаемостью. А именно: наблюдая их, диспетчер (порта, шлюза), управляющий режимом их обслуживания, приобретает опыт интуитивного распознавания опасного и неопасного состояния ненаблюдаемых ритмов и обретает навыки по вводу их в безопасную нормативную область. Вырабатывается искусство восприятия, анализа, диагностики и управления качеством обслуживания транспортных потоков в пределах безопасных границ реализуемых ритмов движения и функционирования. Искусство и уровень мастерства диспетчерской службы определяются, в первую очередь, именно этими особенностями.

По этой причине потоки событий, характеризующие ритмы функционирования транспортных систем и режимов движения транспортных потоков, исследуются в первую очередь.

Эти потоки описывают два взаимобратных показателя:

- интервал между событиями (например, интервал прибытия судов в порт) и
- частота событий в принятую единицу времени (например, число судов, обслуженных в порту за единицу времени).

Наиболее известными и авторитетными публикациями, связанными с исследованием, являются работы [1 - 10] и другие. В работе [7] обращается внимание на проблему стационарности ограниченной совокупности наблюдений. В состоянии системы, характеризующие каждое наблюдение, система периодически возвращается с периодом возврата. Выдвинутая авторами гипотеза о том, что при подобных периодах возврата совокупности наблюдений таких периодов принадлежат одной неограниченной генеральной совокупности, уточняет решение статистических задач, но не избавляет от принадлежности событий к гипотетической генеральной совокупности. В работе [8] авторы впервые отказываются от гипотезы генеральной совокупности, но не ставят вопрос о поиске собственных спектров наблюдений совокупности данных. Подход, применяемый в

непараметрической статистике с представлением данных в номинальных шкалах, шкалах порядка, интервальных шкалах и в виде серий [9], вскрывает специфические свойства организации совокупности наблюдаемых данных, но остается в рамках парадигмы генеральной совокупности. В решении проблемы стратификации данных теория статистического анализа достигла значительных успехов: в технологии контрольных карт Шухарта и кумулянт при оценке качества наблюдаемых данных [1, 2, 3], в задачах кластеризации по различным критериям сходства и различия данных при их сравнении с разными кластерами [4, 6], в теории случайных процессов [5], и в других направлениях. Однако, оставаясь в пределах парадигмы генеральной совокупности с параметрическими законами распределения, традиционная теория статистических решений исследовала статистическую организованность данных с точностью до квадратичных форм и статистических моментов второго порядка, что является очень серьезным ограничением. Оценки с точностью до величин второго порядка для качественных характеристик транспортных систем оказываются нечувствительными и недостаточными. Требуется более глубокий анализ причинной связанности данных.

Особое место в череде исследований занимает работа [10], где развивается инвариантная статистика, определяется внутренняя организованность и степень неорганизованности произвольных потоков событий без декларации их законов распределения.

Автор статьи тоже внесла свой посильный вклад в развитие статистической теории применительно к транспортным потокам в сфере судоходства и стивидорной деятельности [11 - 15].

Наиболее распространенный, первичный и простой метод исследования характера распределения потоков событий в статистической традиции, в том числе в транспортных системах, – это гистограммы. В них полный диапазон изменения характеристик потока событий – от минимума до максимума – графически разбивается на ограниченное число интервалов, по которым

распределяются конкретные реализации событий; по ним подсчитывается число произошедших событий в каждом из выделенных интервалов. Это число часто называют «частотой попадания» событий в данный интервал, а все распределения частот по всем интервалам – распределением или спектром частот. Это определение не следует путать с частотой наступления событий в единицу времени.

Такое распределение частот хорошо представляется графически в виде гистограммы.

Однако сам по себе способ отображения распределения событий в виде гистограмм имеет принципиальную неопределенность. Получаемые оценки для качественных характеристик транспортных систем с использованием названной статистической технологии оказываются нечувствительными и недостаточными. Изначально не ясно: на сколько интервалов следует разбивать диапазон изменения характеристик наблюдаемого потока событий; должны ли быть интервалы распределения одинаковыми; как определить – принадлежит ли выброс наблюдаемому потоку событий или он в нем случаен; волна на гистограмме – это случайный всплеск или закономерность.

Рассмотрим особенности такого отображения потока событий на конкретном примере.

В табл. 1 зафиксирован поток событий, каждое из которых характеризуется временем обслуживания судов τ , следующих в одном направлении, на шлюзе канала за период с 01 – 04 число месяца.

Таблица 1

Обслуживание судов на шлюзе

τ	Время обслуживания судов на шлюзе, мин								
27	30	30	24	21	24	30	33	30	24
33	15	33	30	24	42	30	21	24	36
42	36	30	30	33	36	24	33	24	30
15	30	21	54	48	54	30	36	39	33
51	33	45	48	27	30	78	46	34	29
44						N = 51		$\bar{\tau} = 33$	

За период с 01 по 04 число месяца на шлюзе было обслужено $N = 51$ судно со средним временем обслуживания:

$$\bar{\tau} = \frac{1}{N} \sum_{k=1}^N \bar{\tau}_k = 33 \text{ мин.}$$

Наблюдаемый диапазон времени обслуживания судов составляет:

$$S_{\tau} = (15 \dots 78) = 78 - 15 = 63 \text{ (мин).}$$

Разобьем наблюдаемый диапазон времени обслуживания судов в шлюзе по трем вариантам, например, на 3, 6 и 10 интервалов.

В табл. 2, 3 и 4 приведены частоты распределения наблюдаемых потоков событий. В первой строке таблиц показаны пределы интервалов, во второй строке – их порядковый номер и в третьей – число событий, происходящих в каждом интервале.

Таблица 2

Распределение частот судопотока при разделении на 3 интервала

15...35	35...55	55..78
1	2	3
36	14	1

Таблица 3

Распределение частот судопотока при разделении на 6 интервалов

15...25	25...35	35...45	45...55	55...65	65...78
1	2	3	4	5	6
16	24	7	7	0	1

Таблица 4

Распределение частот судопотока при разделении на 10 интервалов

15...21	21...27	27...33	33...39	39...45	45...51	51...57	57...63	63...69	69...78
1	2	3	4	5	6	7	8	9	10
2	12	15	11	3	4	3	0	0	1

На рис. 1, 2, 3 построены гистограммы распределения частот распределения наблюдаемых событий, соответствующие данным табл. 2, 3, 4.

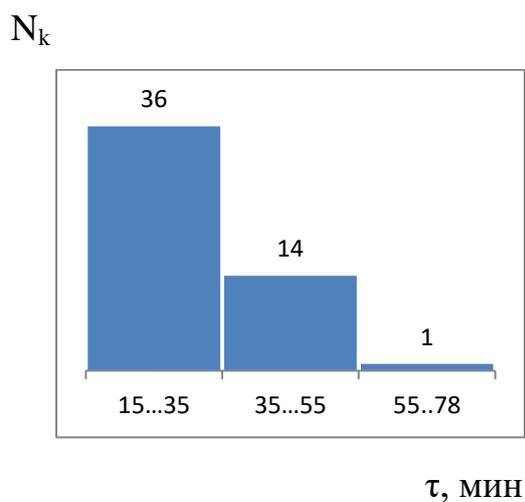
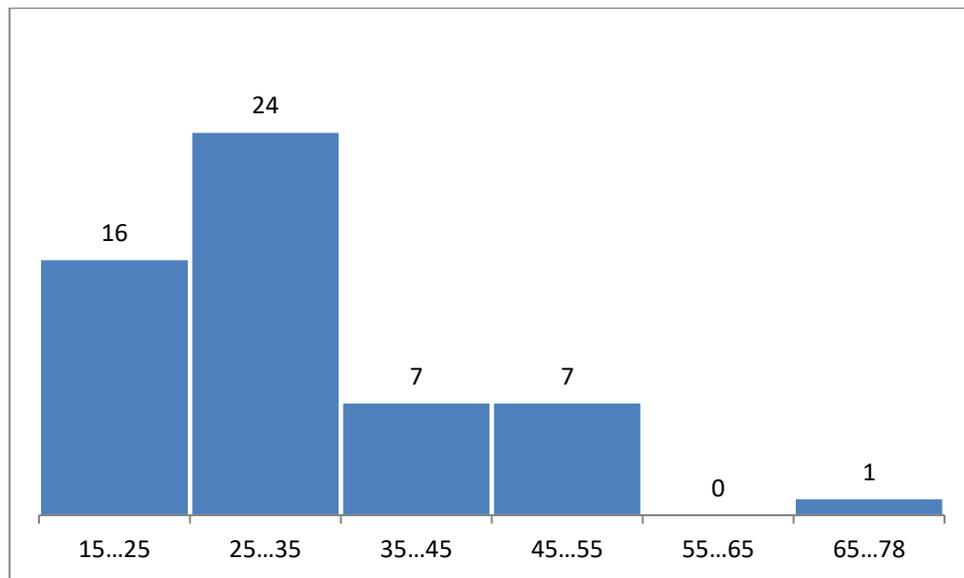


Рис. 1. Распределение частот времени обслуживания судопотока на шлюзе ($k = 3$)

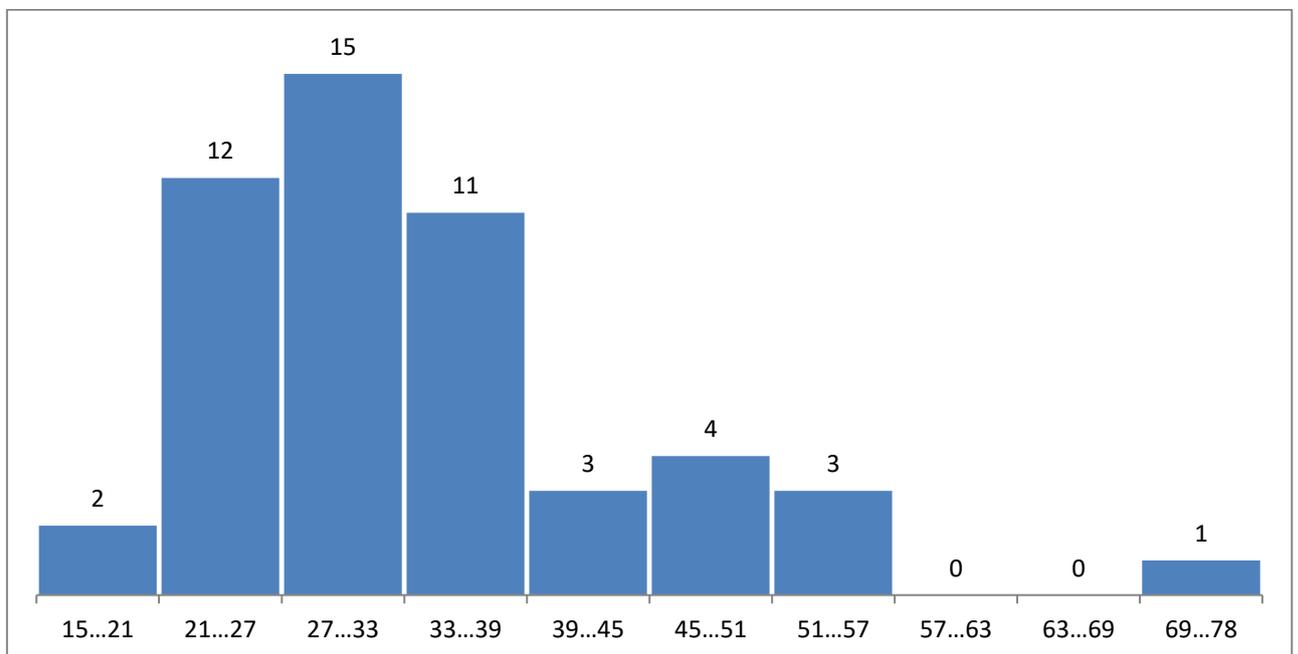
N_k



τ , МИН

Рис. 2. Распределение частот времени обслуживания судопотока на шлюзе ($k = 6$)

N_k



τ , МИН

Рис. 3. Распределение частот времени обслуживания судопотока на шлюзе ($k = 10$)

Сравнение гистограмм, приведенных на рис. 1 – 3, позволяет сделать следующие выводы:

1. Гистограммы для одного и того же потока событий с разным числом интервалов в наблюдаемом диапазоне заметно отличаются друг от друга. Создается впечатление, что они отражают разные потоки событий, но все гистограммы отображают один и тот же поток событий, но непохоже.

Первая гистограмма (рис. 1) изображает монотонно убывающий поток событий. Вторая (рис. 2) - выпуклый поток событий с максимумом на интервале 25...35 мин., провалом на интервале 55...65 мин. и выбросом на интервале 65...78 мин. Третья гистограмма (рис. 3) имеет более гладкую выпуклость с максимумом на интервале 27...33 мин., более глубокий провал на интервале 57...69 мин., выброс на интервале 69...78 мин. и волну на интервале 39...51 мин.

Изначально не ясно: на сколько интервалов следует разбивать диапазон изменения характеристик наблюдаемого потока событий, и какая гистограмма отображает поток более точно.

Традиционная статистика аргументированного ответа по решению этой задачи не дает. Но по существующей рекомендации каждое событие либо попадает, либо не попадает в свой интервал. Если таких интервалов k , то общее разнообразие, равное числу наблюдаемых событий N , равно:

$$2^{k-1} = N,$$

и, логарифмируя это выражение, получим:

$$k = 1 + \frac{\ln N}{\ln 2} \approx 1 + 3,3 \ln N.$$

В рассмотренном случае при $N = 51$ получаем:

$$k = 1 + \frac{\ln 51}{\ln 2} = 6,67; [k] = 6.$$

В соответствии с этим предпочтение следует отдать гистограмме с числом интервалов $k = 6$ (рис. 2).

Но в статистической традиции время обслуживания судов (в порту, на шлюзе) всегда аппроксимируется экспоненциальным законом распределения. Для случая такой аппроксимации лучше подходит гистограмма с числом интервалов $k = 3$ (рис. 1), тогда как по существующей рекомендации предпочтительна гистограмма рис. 2.

Понятно, что однозначного предпочтения какой-либо из диаграмм нет. Такая неопределенность допускает большую свободу в выборе законов распределения наблюдаемых потоков событий, то есть каждый исследователь получает возможность выбирать тот закон, который кажется ему наиболее предпочтительным по его субъективным представлениям. В такой ситуации нарушается объективность адекватного отображения распределения наблюдаемых потоков событий в виде гистограмм.

2. Другая проблема связана с выбором интервалов распределения. Как правило, на гистограммах интервалы одинаковы и равны:

$$\tau = \frac{S}{k}; S = \tau_{\max} - \tau_{\min},$$

где k – число интервалов.

Требуется выяснить – верно ли это. Следует подтвердить либо опровергнуть необходимость равенства интервалов между собой.

Если наблюдаемый поток событий имеет свою непроявленную внутреннюю организованность, свои ритмы, сгущения и разрежения, то эти ритмы не обязательно должны быть эквидистантными. Расстояние между ними может быть различным, диктоваться оно будет внутренними объективными законами равновесной самоорганизации потока событий в

сложившихся граничных условиях внешней среды, а не субъективным выбором по воле исследователя.

Статистическая традиция аргументированно не подтверждает и не опровергает необходимость равенства интервалов между собой.

Этот вопрос по сути связан с проблемой стратификации данных, с разбиением их на однородные группы событий, совокупности, классы, страты, кластеры, каждый из которых имеет специфический ритм сгущений и разряжений, ритм плотности потока событий. При этом существует проблема установления границ между смежными стратами и отнесения конкретного наблюдения к той или иной страте в случае, если оно находится в области границы между ними.

3. Третья проблема связана с появлением выбросов в наблюдаемом потоке событий и определением их принадлежности наблюдаемому потоку событий или их случайности.

Изначально не ясно: как определить – принадлежит ли выброс наблюдаемому потоку событий или он в нем случаен.

Выбросы в наблюдаемом потоке событий на гистограмме могут проявляться или не проявляться - в зависимости от числа интервалов их разбиения на выбранном типе гистограммы.

В рассмотренном выше примере на гистограмме рис. 1 ($k = 3$) выброс не проявлен; на гистограмме рис. 2 ($k = 6$) он появился; на гистограмме рис. 3 ($k = 10$) – явно выделился.

В первом случае этот выброс нужно включить в обработку. Во втором случае – исключить из обработки как ошибочное маловероятное наблюдение.

На вопрос о принадлежности выброса на гистограмме наблюдаемому потоку событий (когда выброс нужно включить в обработку) или о его случайности (когда выброс нужно исключить из обработки как ошибочное маловероятное наблюдение) традиционная статистика также не дает аргументированного ответа. При этом существующая рекомендация гласит, что, если выброс отстоит от среднего значения на расстояние, большее, чем

три среднеквадратичных отклонения, вероятность его наблюдения очень мала (меньше 1%). Такой выброс маловероятен, его рекомендуется исключить из рассмотрения - как случайное наблюдение, не принадлежащее наблюдаемому потоку событий.

Но нормальный закон распределения допускает и бесконечно большие выбросы с нулевой (очень редкой) вероятностью.

Если наблюдаемый выброс все же проявлен в исследуемом потоке событий и если это инструментальное измерение, то возможно, что выброс – ошибка. Но если рассматривается конкретное судно с большим временем шлюзования, то его никак нельзя отнести к разряду ошибок.

Следовательно, проблему оценивания выбросов следует решать осторожно, опираясь, по возможности, на объективные закономерности, но не на субъективную волю исследователя.

4. Возникает еще одна проблема: определение случайна или закономерна появившаяся на гистограмме волна.

На гистограмме рис. 3 ($k = 10$) появилась волна на интервале 39...57 мин. Изначально не ясно: как определить – волна на гистограмме – это случайный всплеск или закономерность.

В рамках статистической традиции при появлении на гистограмме волны обычно поступают так: если наблюдаемый поток событий однороден и не имеет закономерной внутренней дифференциации на зоны сгущения или разряжения, то наблюдаемая волна, скорее всего, носит случайный характер и при обработке наблюдений ее можно сгладить.

Когда же наблюдаемая волна есть следствие внутренней организованности потока событий, которая возникла в режиме самоорганизации равновесного состояния в сложившихся граничных условиях окружающей среды, могла возникнуть внутренняя дифференциация ритмов сгущения и разряжения событий, которая проявляется в виде волн плотности распределения событий. Эти волны несут информацию о внутренней организации потока событий. При их сглаживании при обработке

теряется эта информация. А она, как правило, является диагностическим признаком, характеризующим состояние организованности наблюдаемого потока событий.

Таким образом, исследуя один из методов изучения распределений транспортных потоков, определили, что уже на этапе первичной обработки данных при поиске законов распределения наблюдаемых потоков событий, в т.ч. в судоходстве и стивидорной деятельности, на основе гистограмм возникает ряд проблем фундаментального характера, связанного с внутренней самоорганизацией наблюдаемого потока событий при поиске равновесного состояния в сложившихся граничных условиях внешней среды, на которые традиционная статистика не дает однозначного ответа.

Наблюдаемый поток событий имеет свои ритмы, сгущения и разряжения.

Традиционная статистика не обращала внимания на внутреннюю самоорганизацию наблюдаемых событий.

Решать названные проблемы следует на основе объективных законов, а не на основе субъективных представлений исследователя.

При интенсивном пути развития постиндустриального общества в условиях высокой интенсивности обменных процессов на уровне предельных ограничений эти проблемы становятся актуальными и дают заказ для науки в области статистического анализа.

Требуется парадигма: наблюденные данные несут в себе всю информацию об их внутренней организованности. Необходимо проявить эту внутреннюю, скрытую от наблюдателя организованность данных и использовать ее для анализа источников, генерирующих эти данные, и управления ими.

Список источников

1. Балакришнан А.В. Теория фильтрации Калмана. - М.: Мир, 1988. - 166 с.
2. Бокс Дж., Дженкинс Г. Анализ временных рядов. Прогноз и управление/ Вып. 1 – М.: Мир, 1974. – 604 с.

3. Бриллинджер Д. Временные ряды. Обработка данных и теория/ Под ред. А.Н. Колмогорова. Пер с англ. - М.: Мир, 1980. - 536 с.
4. Джонстон Дж. Эконометрические методы. – М.: Статистика, 1980. - 443 с.
5. Карлин С. Основы теории случайных процессов. - М.: Мир, 1971. - 536 с.
6. Классификация и кластер/ Под ред. Дж. Вэн Райзина. - М.: Мир, 1980. - 390 с.
7. Крейн М., Лемуан О. Введение в регенеративный метод анализа моделей. – М.: Наука, 1982. - 104 с.
8. Методы анализа данных: Подход, основанный на методе динамических сгущений / Колл. авт. под рук. Э. Дидэ. Пер. с франц. – М.: Финансы и статистика, 1985. - 357 с.
9. Рунион Р. Справочник по непараметрической статистике. Современный подход. – М.: Финансы и статистика, 1982. - 198 с.
10. Трифанов В.Н. Методические основы синтеза динамических сетей. Алгебраическое равновесие и статистика/ Препринт № 148 под ред. В.М. Пономарева// АН СССР. – Л.: Ленинградский институт информатики и автоматизации, 1991. - с. 59-61.
11. Туаршева О.А. Альтернативный вариант статистического анализа потоков событий // Транспортное дело России. – 2018. - № 6 (139). - с. 307 – 308.
12. Туаршева О.А. Структура режимов транспортных коммуникационных сред // Транспортное дело России. – 2019. - № 5 (144). - с. 157- 159.
13. Туаршева О.А. Теория очередей на водном транспорте для потоков общей природы // Транспортное дело России. – 2023. - № 4 (167). - с. – 151 - 155.
14. Туаршева О.А. Управление очередью судов и регулирование очередности их обслуживания в портах и на шлюзах каналов // Транспортное дело России. – 2023. - № 4 (167). - с. – 166 - 170.

15. Туаршева О.А. Решение проблем управления в судоходстве и стивидорной деятельности на основе статистического анализа // Транспортное дело России. – 2023. - № 5 (168). - с. – 193 – 197.

Информация об авторе

ТУАРШЕВА ОЛЬГА АЛЕКСАНДРОВНА, кандидат технических наук, доцент, профессор кафедры «Экономики и технологий бизнес-менеджмента на водном транспорте», ФГБОУ ВО «ГУМРФ им. адмирала С. О. Макарова», г. Санкт-Петербург, Россия

Information about the author

TUARSHEVA OLGA ALEKSANDROVNA, Candidate of Technical Sciences, Associate Professor, Position - Professor of the Department, Admiral Makarov State University of Maritime and Inland Shipping Department of Economics and Technology of Business Management in Water Transport, St. Petersburg, Russia