

УДК 543.3

И.В. Мискевич, К.Г. Боголицын

Мискевич Игорь Владимирович родился в 1953 г., окончил в 1975 г. Ленинградское высшее инженерное морское училище, кандидат географических наук, старший научный сотрудник Института экологических проблем Севера УрО РАН. Имеет более 50 печатных трудов в области экологии водных объектов.



Боголицын Константин Григорьевич родился в 1949 г., окончил в 1971 г. Архангельский лесотехнический институт, доктор химических наук, профессор, заведующий кафедрой теоретической и прикладной химии Архангельского государственного технического университета, директор Научно-исследовательского института химии и химической технологии при АГТУ, академик МАНЭБ и РАИН, заслуженный деятель науки РФ. Имеет более 250 научных трудов в области разработки физико-химических основ процессов переработки древесины.

**НЕКОТОРЫЕ ОСОБЕННОСТИ СТАТИСТИЧЕСКОГО АНАЛИЗА ПАРАМЕТРОВ КАЧЕСТВА ВОД ПРИЛИВНЫХ УСТЬЕВ РЕК**

Показано, что при статистическом анализе характеристик качества вод приливных устьев рек целесообразно применять робастные параметры, при этом вместо среднеарифметического значения рекомендуется использовать медиану.

гидрохимическая характеристика, устья рек приливных морей, антропогенные воздействия.

Гидрохимические характеристики вод устьев рек приливных морей имеют большую пространственно-временную изменчивость, обусловленную приливами и их взаимодействием с другими природными явлениями. При этом колебания концентраций многих ингредиентов внутри устьевой системы могут на порядок превышать их соответствующие изменения в чисто речных и морских водах. Для многих устьевых областей рек типична значительная концентрация промышленных, транспортных и сельскохозяйственных объектов, сопровождающаяся четко выраженным антропогенным воздействием на качество эстуарных вод. Подобная ситуация приводит к появлению весьма сложных механизмов, формирующих химический состав вод приливных устьев рек в современных условиях.

Современные технологии комплексной переработки растительного сырья, реализуемые на предприятиях химико-лесного комплекса, относятся

к наиболее водоемким. Так, на производство 1 т целлюлозы потребляется 120 ... 170 м² воды, а доля загрязненных сточных вод целлюлозно-бумажных и деревообрабатывающих предприятий в общем промышленном сбросе составляет 19,4 %. Качество готовой продукции в значительной степени зависит от качества воды природного водоема, определяемого компонентным составом. Особенности Архангельской области обусловили концентрацию лесохимических предприятий в устьях наиболее крупных рек Белого моря – в дельте рек Северная Двина, Мезень, Онега, Кемь, Ковда. Поэтому характеристика особенностей статистического анализа параметров качества приливных устьев рек имеет особую актуальность для предприятий, интенсивно эксплуатирующих водные ресурсы.

Рациональное использование поверхностных вод и их охрана в условиях хорошо выраженного техногенного воздействия требует большого объема сведений о качестве этих вод. Данная задача может быть успешно решена при создании соответствующих информационных систем, опирающихся на мониторинг водной среды [1]. С другой стороны, прикладной эффект от подобной информации может быть получен лишь при применении показателей, адекватно отражающих состояние водного объекта. В противном случае предприятия-водопользователи будут иметь большие издержки в финансовом и технологическом отношениях, сопровождаемые возрастанием негативного влияния на устьевые экосистемы.

7*

Для реализации практических задач в этой области чаще всего используют не первичные гидрохимические данные, а результаты их статистического анализа. При этом обычно применяют среднеарифметическое (среднее) значение и экстремумы для всего массива имеющихся величин. Реже используют среднеквадратичное (стандартное) отклонение и повторяемость концентраций исследуемых ингредиентов выше уровня ПДК. Другие статистики, как правило, применяют только при проведении специализированных исследований.

Известно, что среднее значение имеет максимальную эффективность для нормального и гамма-распределения [2]. Но данный параметр, а также стандартное отклонение теряют свою эффективность для асимметричных распределений с «утяжеленными хвостами» и для малых выборок с наличием даже одного выделяющегося значения («выброса») [3, 4, 10]. Например, из четырех сезонных наблюдений для трех сезонов отмечено содержание загрязняющего вещества 0,1 ПДК, для одного сезона – 10 ПДК, т.е. среднее значение за год составит 2,6 ПДК. Получаем формальный вывод о значительном загрязнении водной среды по исследуемому гидрохимическому показателю. Если подобную цифру использовать для разработки нормативов ПДС, то при их реализации потребуется очищать сточные воды по рассматриваемому поллютанту до уровня ПДК, что является весьма сложным и дорогим мероприятием.

В данной ситуации величина в 10 ПДК может быть результатом аналитической ошибки или реальной величиной, но имеющей очень малую по-

вторяемость (не более 5 ... 10 %). Разумеется, для получения достоверной информации необходимо увеличить количество наблюдений, но это требует времени и значительных финансовых затрат на дополнительные анализы. Приведенный пример ярко иллюстрирует пагубность формального отношения к статистической обработке гидрохимических наблюдений, хотя, как будет показано ниже, и для вышеупомянутых четырех значений можно получить достаточно корректный результат.

В зоне смешения речных и морских вод индикатором химического состава водной среды может служить соленость (минерализация), которая определяет степень распреснения соленых вод. В устьях рек неприливногo моря квазистационарное пространственное распределение солености вдоль водотока (водоема) в меженные периоды можно аппроксимировать экспоненциальной кривой, значения которой соответствуют набору средних величин для совокупности выделенных створов внутри устьевой системы [5]. Сравнительно небольшие пространственные колебания солености в пределах меженного периода формируют для вышеупомянутых створов массив ее значений, которые в большинстве случаев подчиняются нормальному статистическому распределению.

В приливных устьях рек ситуация может кардинально измениться. Хотя вдольэстуарный профиль солености, осредненный для меженного периода, также аппроксимируется экспоненциальной кривой, для большинства устьевых створов экстремумы за счет приливо-отливных явлений достигают отметок, характерных для чисто речных и морских вод. Таким образом, специфика гидродинамики приливных устьев рек обуславливает наличие для створов, расположенных в начале и в конце зоны смешения речных и морских вод, сильно выраженную асимметрию в статистическом распределении солености. В ее распределении появляется «хвост», в первом случае тяготеющий к высокой солености, во втором – к низкой солености. В середине зоны смешения речных и морских вод распределение солености будет в наибольшей степени отвечать нормальному закону. Аналогичная картина будет прослеживаться для основных ионов (натрий, калий, кальций, магний, хлориды, сульфат и иногда гидрокарбонаты) и для большинства биогенных и органических веществ, но для них меняется знак асимметрии, так как высокие концентрации этих соединений типичны для речных вод.

В условиях антропогенного влияния для многих гидрохимических показателей законы распределения могут быть сильно трансформированы, но наиболее характерной чертой для них является асимметричность и наличие «выбросных» точек, особенно, при нестабильной работе очистных сооружений или отсутствии таковых. Следовательно, как природные, так и техногенные факторы создают предпосылки для потери эффективности стандартных статистик (среднее значение и дисперсия) при обработке результатов наблюдений за качеством вод приливных устьев рек.

Для корректного статистического анализа данных при отклонении распределения от нормального закона или неизвестном законе распределения, а также при коротких выборках рекомендуется применять так называемые

мые робастные, или устойчивые (помехоустойчивые), подходы [2 – 4, 10]. Согласно Хьюберу [11], их можно подразделить на подходы, базирующиеся на применении L -, R - и M -оценок. Первые представляют из себя линейные комбинации порядковых статистик (медиана, усеченное среднее, винзоризованное среднее и т.п.). R -оценки определяют с помощью ранговых статистик, M -оценки – аналогично методу максимального правдоподобия с помощью различных весовых функций.

Принято считать, что наиболее удачно в робастной статистике использование M -оценок, так как другие подходы не допускают удобных обобщений в многопараметрических ситуациях [6], правда их расчет требует итерационных процедур и сложен для варианта «ручного» вычисления. В однопараметрических задачах, к которым обычно относятся статистические исследования гидрохимических характеристик для выделенного створа (точки), весьма привлекательным является применение L -оценок из-за простоты их вычисления и смысловой сопоставимости со стандартными статистиками. Анализ литературных источников показывает, что для изучения состояния различных природных сред и гидрохимических показателей в качестве оценки центра распределения, в частности, рекомендуется использовать медиану и трехсреднее значение Тьюки (центральное среднее), а в качестве оценки масштаба распределения – интерквартильный размах и медиану абсолютных отклонений [2, 7 – 9, 11].

Медиана делит ранжированную выборку данных пополам, ее определяют по формуле

$$C_i = \begin{cases} C_{(n+1)/2}, & \text{если } n - \text{нечётное;} \\ 0,5 (C_{n/2} + C_{n/2+1}), & \text{если } n - \text{чётное,} \end{cases} \quad (1)$$

где n – число данных в выборке.

Трехсреднее значение находят по выражению

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

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

Интерквартильный размах определяют по формуле

$$H = C_{0,75} - C_{0,25}.$$

Медиана абсолютных отклонений (МАО) является медианой для выборки, сформированной из разностей между C_M и C_i .

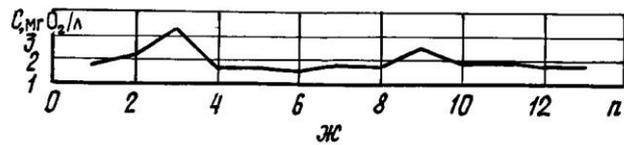
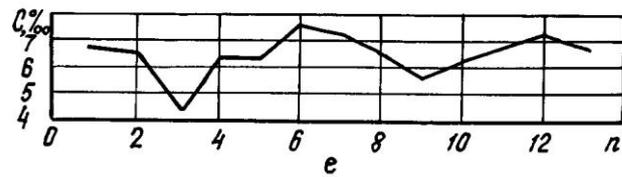
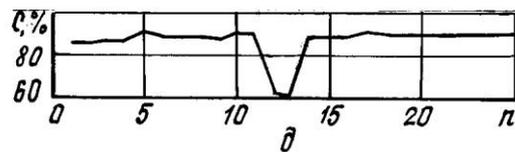
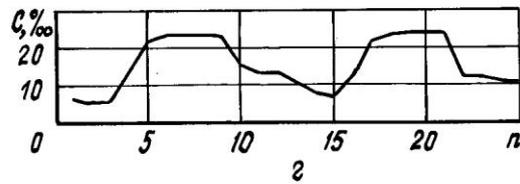
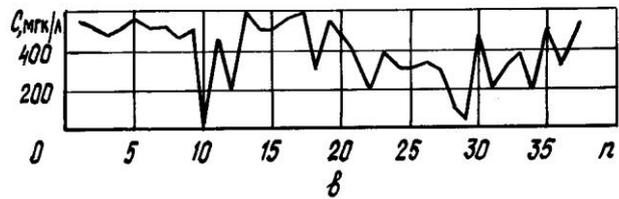
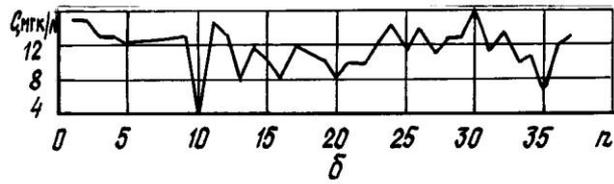
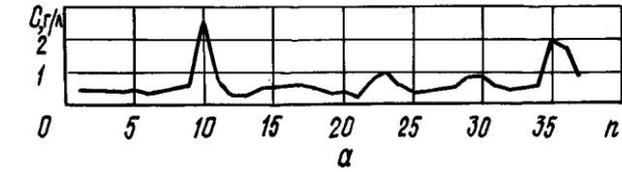
Для нормального распределения среднее значение, медиана и трехсреднее значение совпадают. Это позволяет пользователю, слабо разбирающемуся в статистике, рассматривать вышеприведенные робастные параметры как улучшенную версию среднеарифметического значения и устраняет сложности при интерпретации и использовании полученных результатов в практических задачах. Действительно, медиана для малой выборки из генеральной совокупности с наличием «выбросов» в большинстве случаев будет больше соответствовать среднему значению для этой совокупности, чем среднее значение для малой выборки. Специальные исследования показали, что при $n \leq 5$ лучше использовать медиану, а при $n > 5$ – трехсреднее значение [10]. Если взять ранее упоминавшийся теоретический пример для четы-

рех наблюдений, то получим, что $C_m = 0,1$ ПДК, т.е. уровень загрязнения за год сильно меняется по сравнению с расчетами, когда используют среднее значение. Стандартное отклонение σ связано с робастными оценками изменчивости данных следующими соотношениями:

$$H = 1,34\sigma; \text{MAO} = C_m + 0,67\sigma.$$

Для иллюстрации эффективности использования робастных статистик на реальных примерах выполним статистический анализ результатов гидрохимических наблюдений, сделанных в трех устьевых областях рек Белого моря: в протоке Кузнечиха дельты р. Северная Двина (29-й лесозавод), в устье р. Кемь (причал морпорта), в кутовой части Унской губы (мыс Красный). Эти точки были выбраны с учетом необходимости отразить воздействие приливных явлений.

Исследования протоки Кузнечиха проводили в течение трех суток (8–11 апреля 1987 г.) с дискретностью 2 ч. Они включали в себя определение хлоридов, аммонийного и нитритного азота. Рассматриваемый участок в зимнюю межень располагается в начале зоны смешения речных и морских вод и испытывает влияние сточных вод Соломбальского ЦБК, 29-го лесозавода и других промышленных объектов. Здесь расположен водозабор 29-го лесозавода и планируется разместить водозабор для строящегося метанолевого завода. Общее число наблюдений в протоке Кузнечиха по каждому ингредиенту составило 37.



Короткопериодные изменения содержания C хлоридов (a), нитритного (b) и аммонийного (c) азота в протоке Кузнечиха, солёности (e) и насыщенности кислородом (d) в устье р. Кемь, солёности (e) и БПК₅ ($ж$) в кутовой части Унской губы

На представленном рисунке *a* для хлоридов отчетливо наблюдается влияние морских вод, связанное с приливами. Отметим, что в полную воду приливного цикла в этой точке отбора фиксировались значения, которые на порядок превышали основной массив наблюдений. Подобной изменчивости отвечает статистическое распределение с явно выраженной правосторонней асимметрией.

Для нитритного и аммонийного азота прослеживается левосторонняя асимметрия, обусловленная подходом на рассматриваемый участок в фазу прилива более чистых морских вод (см. рисунок *b, в*).

Поверхностные воды устья р. Кемь исследовали в течение суток (11 октября 1988 г.) с дискретностью 1 ч на предмет определения солености и содержания кислорода. Точка отбора проб располагалась в середине зоны смешения речных и морских вод, примерно на 100 м ниже по течению от Кемского лесозавода. Для статистического анализа были выбраны не абсолютные значения содержания кислорода, а насыщенность им воды. Этот показатель прямо не связан с колебаниями температуры и солености морских вод. Как видно из рисунка *г, д*, в течение суток было зафиксировано кратковременное аномальное понижение этого показателя, очевидно связанное с выносом обедненных кислородом вод из сортировочного бассейна Кемского лесозавода на максимуме скорости отливного течения. В этой ситуации в статистическом распределении насыщенности вод кислородом прослеживается левосторонняя асимметрия. Для солености признаки явной асимметричности отсутствуют.

Наблюдения в кутовой части Унской губы около мыса Красный проводили на поверхностном горизонте в течение суток с дискретностью 2 ч. Точка отбора проб воды располагалась в начале зоны смешения речных и морских вод, где процессы распреснения носят сложный характер и определяются стоком ряда небольших рек. При этом определяли (см. рисунок *е, ж*) соленость и биохимическое потребление кислорода (БПК₅). Слабо выраженную асимметрию в статистическом распределении изучаемых показателей генерирует сток рек Карбасовка и Бабья, что хорошо прослеживается на малой воде приливного цикла.

Для анализа изменчивости оценок центра распределения исследуемых данных из исходного объема наблюдений создадим ряд выборок, соответствующих различной дискретности отбора проб в течение трех суток для протоки Кузнечиха и в течение суток для других водных объектов. При этом обязательно будем захватывать первое и последнее (или предпоследнее) наблюдения. Таким образом, для р. Кузнечиха помимо исходной выборки в 37 значений получим еще 9 промежуточных выборок, имеющих объем в 3, 4, 5, 6, 7, 8, 10, 13, 19 значений, которым соответствует дискретность отбора проб воды от 36 до 4 ч. Для устья р. Кемь будем иметь исходную выборку в 25 значений и 6 промежуточных выборок с объемами 3, 4, 5, 7, 9, 13 значений, для Унской губы – исходную выборку в 13 значений и 4 промежуточные выборки с объемами 3, 4, 5, 7 значений.

Таблица 1

Статистические характеристики содержания хлоридов, нитритного и аммонийного азота в протоке Кузнечиха в зависимости от дискретности отбора проб воды в течение 3-х суток (апрель 1987 г.)

Число наблюдений, n	Хлориды, г/л			Азот, мкг/л					
				нитритный			аммонийный		
	C_c	C_m	C_T	C_c	C_m	C_T	C_c	C_m	C_T
3	0,52	0,42	0,46	12,6	12,6	12,6	533	5540	538
4	0,45	0,38	0,42	11,6	11,8	11,7	490	530	508
5	0,92	0,53	0,57	10,9	12,4	11,9	346	510	415
6	0,78	0,59	0,62	12,0	12,0	11,9	355	395	385
7	0,44	0,40	0,41	11,5	11,1	11,4	451	510	489
8	0,61	0,49	0,45	12,0	12,0	12,1	414	440	432
10	0,48	0,46	0,46	11,6	12,2	11,6	436	490	472
13	0,64	0,53	0,50	10,6	10,8	10,9	364	480	420
19	0,58	0,48	0,49	11,4	12,0	11,8	428	460	450
37	0,62	0,48	0,48	11,5	12,0	11,9	389	450	425
	Стандартное отклонение σ								
	0,151	0,065	0,065	0,564	0,551	0,444	60,4	44,9	47,5
	Коэффициент вариации K_v								
	25,6	13,6	13,4	4,9	4,6	3,8	14,3	9,3	10,5
	Размах значений P								
	0,46	0,21	0,21	2,0	1,8	1,7	187	145	153

Критерием эффективности используемого показателя будем считать его устойчивость к изменениям числа наблюдений в период, охватываемый тремя сутками. Другими словами, чем меньше он будет зависеть от параметра n , тем корректнее будет его использование как статистики в прикладном отношении. Эту задачу можно решить, используя такие оценки, как размах значений (разница между максимумом и минимумом), среднеквадратическое отклонение и коэффициент вариации. Результаты соответствующих вычислений представлены в табл. 1–3 (C_c – среднее значение, K_v – коэффициент вариации, P – размах значений, остальные обозначения взяты из уравнений (1), (2)).

Анализ полученных статистик указывает на предпочтительность использования медианы и трехсреднего значения по сравнению со средним арифметическим. Для хлоридов в водах Кузнечихи стандартное отклонение для средних значений на порядок превышает величины этого параметра для робастных статистик. В итоге за счет влияния наблюдений, составляющих не более 6 % от общего объема, можно сделать неправильный вывод о сильном фоновом осолонении вод на рассматриваемом участке р. Кузнечиха, что может, в свою очередь, привести к ошибке, например, в случае выбора технологической системы для обессоливания дельтовых вод при необходимости их забора. Так, для кутовой части Унской губы при семиразовом отборе проб воды в течение суток можно получить ошибочный вывод о превышении среднесуточной величины уровня ПДК.

Таблица 2

Статистические характеристики солености и насыщенности вод кислородом в устье р. Кемь в зависимости от дискретности отбора проб воды в течение суток (октябрь 1988 г.)

Число наблюдений, n	Соленость, ‰			Насыщенность кислородом, %		
	C_c	C_m	C_t	C_c	C_m	C_t
3	9,24	10,63	10,08	77	85	82
4	15,38	16,18	15,78	87	86	87
5	14,97	10,78	13,83	82	87	87
7	15,95	15,37	15,91	84	87	87
9	14,14	12,09	12,58	84	87	87
13	15,58	13,42	15,08	85	87	87
25	15,43	13,28	15,12	85	87	87
	Стандартное отклонение σ					
	2,34	2,13	2,11	3,2	0,8	1,9
	Коэффициент вариации K_v					
	16,3	16,3	15,0	3,8	0,9	3,7
	Размах значений P					
	6,71	5,55	5,83	10	2	5

Таблица 3

Статистические характеристики солености и БПК₅ в водах Унской губы у мыса Красный в зависимости от дискретности отбора проб воды в течение суток (июнь 1996 г.)

Число наблюдений, n	Соленость, ‰			БПК ₅ , мг О ₂ /л		
	C_c	C_m	C_t	C_c	C_m	C_t
3	6,87	6,98	6,94	1,81	1,82	1,81
4	6,40	6,50	6,45	1,96	1,81	1,88
5	6,60	6,62	6,62	1,81	1,82	1,82
7	6,26	6,62	6,53	2,13	1,82	1,92
13	6,42	6,39	6,50	1,98	1,82	1,84
	Стандартное отклонение σ					
	0,24	0,22	0,20	0,13	0,04	0,05
	Коэффициент вариации K_v					
	3,7	0,3	0,3	6,9	0,2	2,4
	Размах значений P					
	0,61	0,59	0,49	0,32	0,01	0,11

Сравнение медианы и трехсреднего значения не выявило явных преимуществ ни одной из статистик. С другой стороны, при прикладных исследованиях качества устьевых вод целесообразнее ориентироваться на медиану, как на стандартный показатель, легко определяемый по программе Excel на компьютере. Для оценки масштаба изменчивости исследуемых данных рекомендуется находить квартили (25 %-е точки) и интерквартильный размах.

Полученные результаты позволяют говорить о необходимости применения медианы, как обязательного компонента информационной системы, связанной с обработкой и использованием результатов мониторинга качества вод устьев рек приливного моря. Помимо более адекватного отражения состояния водной среды медиана и другие робастные статистики дают возможность проводить корректные сравнения экологических ситуаций, обеспеченных различным числом наблюдений. Реализация подобного подхода в приложении к экологическим проблемам многих предприятий лесной и целлюлозно-бумажной промышленности несомненно поможет их эффективному решению.

СПИСОК ЛИТЕРАТУРЫ

1. *Благовещенский Ю.Н., Самсонова В.П., Дмитриев Е.А.* Непараметрические методы в почвенных исследованиях. – М.: Наука, 1987. – 95 с.
2. *Боголицын К.Г.* Разработка научных основ экологически безопасных технологий комплексной химической переработки древесного сырья // Лесн. журн. – 1998. – № 2-3. – С. 40–52. – (Изв. высш. учеб. заведений).
3. *Компьютерная биометрика* // Учеб. пособие. – М.: Изд-во МГУ, 1990. – 232 с.
4. Метод анализа сезонной изменчивости неэквидистантных временных рядов гидролого-гидрохимических данных / С.М. Микулинская, Н.Н. Михайлов, В.А. Рожков, И.С. Шпаер // Режимобразующие факторы, информационная база и методы ее анализа. – Л.: Гидрометеиздат, 1989. – С. 109–132.
5. *Микулинская С.М., Рожков В.А.* Обработка малых выборок // Режимобразующие факторы, информационная база и методы ее анализа. – Л.: Гидрометеиздат, 1989. – С. 167–176.
6. *Павелко В.Л.* Пути совершенствования методов расчета оценок качества вод // Гидрохимические материалы. – 1981. – Т. LXXVIII. – С. 32–41.
7. Робастность в статистике. Подход на основе функций влияния / Ф. Хампель, Э. Ронгетти, П. Рауссеу, В. Штаэль. – М.: Мир, 1989. – 512 с.
8. *Справочник по гидрохимии.* – Л.: Гидрометеиздат, 1989. – 392 с.
9. *Тьюки Д.* Анализ результатов наблюдений. Разведочный анализ / Пер. с англ. – М.: Мир, 1981. – 688 с.
10. *Харлеман Д.Р., Инпен А.Д., Кейлеган Д.Т.* Гидродинамика береговой зоны и эстуариев. – Л.: Гидрометеиздат, 1970. – 394 с.
11. *Хьюбер П.* Робастность в статистике. – М.: Мир, 1984. – 303 с.

Институт экологических проблем Севера УрОРАН
Архангельский государственный
технический университет

Поступила 20.10.2000 г.

I.V. Miskevich, K.G. Bogolitsyn

Some Special Features of Statistical Analysis for Water Quality Parameters of Tidal River Mouth

The use of robust parameters is shown to be expedient under the statistical analysis of water quality characteristics of tidal river mouth, median being recommended to be used instead of arithmetical mean value.
