УДК 51-74
Асадуллин Э.З. – кандидат технических наук
E-mail: env60@yandex.ru
Казанский государственный архитектурно-строительный университет
Адрес организации: 420043, Россия, г. Казань, ул. Зеленая, д. 1

Повышение точности определения исчисленных данных с использованием комплексного учета условий измерений

Аннотация

На основе ряда фактов было сделано предположение о возможности повышения точности определения данных для расчетов на основе комплексного учета условий измерений, были намечены основные пути решения этой задачи. Гипотеза проводимого исследования, как и все гипотезы, в силу своего вероятностного характера требует проверки, доказательства. В настоящей статье рассматривается порядок проверки общей гипотезы исследования о повышении точности определения исчисленных данных для расчетов на основе комплексного учета условий измерений.

Ключевые слова: комплексный учет, дисперсия, математическое ожидание, математическая модель, методы регрессионного анализа.

1. Применение методов математического моделирования для воспроизведения условий определения исчисленных данных

Характеристики условий измерений воспроизводились на специально созданной для их изучения математической модели. Потребность в моделировании возникла в связи с тем, что исследование непосредственно самого процесса невозможно, так как дорого и требует слишком длительного времени. Подобие между моделью и объектом исследования заключается в тождестве математического описания «поведения» объекта и модели. Модель описывает некоторые, существенные в данном исследовании, свойства и функции объекта исследования. Исследуемые стороны модели описывались теми же математическими формулами, что и моделируемые свойства объекта. Таким образом, математическая модель объекта исследования подобна, иначе говоря, в достаточной степени соответствует самому объекту исследования.

Для исследования процесса определения исчисленных данных при различных условиях измерений и выработки рекомендаций применен опытно-теоретический метод. Расчеты, полученные при теоретическом методе оценки эффективности, проверялись в ходе реальных полевых измерений.

При теоретическом методе применялись точный способ статистических испытаний и приближенный – графический. Моделирование случайных величин осуществлялось путем преобразования независимых значений случайного числа, распределенного равномерно в интервале (0…1). Последовательность частных значений случайного числа получена на ЭВМ с помощью датчика-генератора случайных чисел.

Для нахождения закона распределения случайных величин (СВ) необходимо располагать достаточно обширным статистическим материалом, порядка нескольких сотен опытов. Однако на практике приходится иметь дело со статистическим материалом ограниченного объема – два-три десятка наблюдений, часто даже меньше. Это связано с дороговизной и сложностью каждого опыта. В исследуемой задаче вид закона распределения случайной величины – ошибки измерения, известен заранее – нормальный закон распределения [1, 2], (в нашем случае это даже несущественно), требуется найти ее числовые характеристики. Для этого нет необходимости проводить неограниченное количество опытов.

Если число опытов n невелико, то замена математического ожидания средним арифметическим приводит к какой-то ошибке. Значение искомого параметра, вычисленное на основе ограниченного числа опытов, всегда будет содержать элемент случайности. Такое приближенное значение называется оценкой параметра $\tilde{a}$. Любая оценка, вычисляемая на основе наблюдаемых СВ, представляет собой функцию величин $X_1, X_2, \ldots, X_n$ и сама является величиной случайной:

$$\tilde{a} = \tilde{a}(X_1, X_2, \ldots, X_n). \tag{1}$$

Нахождение статистических оценок $\tilde{a}$ параметров законов распределения в данном исследовании проводилось с помощью моментов эмпирического распределения, которые являются состоятельными оценками соответствующих моментов теоретического распределения.

Эмпирический начальный момент *k-го* порядка для дискретной СВ определяется равенством:

$$a_k\left[X\right] = \sum_{i=1}^{n} x_i^k \cdot p_i, \tag{2}$$

где x_i – значения СВ X;

p_i – соответствующие вероятности.

Начальный момент первого порядка есть ни что иное, как математическое ожидание СВ:

$$m_x = M\left[X\right] = a_1\left[X\right]. \tag{3}$$

Оценкой математического ожидания $\tilde{m}_x$ СВ X является среднее арифметическое ее наблюдаемых значений:

$$\tilde{m}_x = \frac{1}{n}\sum_{i=1}^{n} X_i. \tag{4}$$

Требование состоятельности оценки $\tilde{a}$ сводится к тому, чтобы при увеличении числа опытов *n,* она приближалась (сходилась по вероятности) к искомому параметру a ($\tilde{a} \xrightarrow[p]{n\to\infty} a$). Оценка математического ожидания m_x СВ X при увеличении числа опытов *n,* согласно закона больших чисел, сходится по вероятности к математическому ожиданию m СВ X ($\tilde{m}_x \xrightarrow[\partial]{n\to\infty} m_x$).

Оценка для дисперсии $\tilde{D}$ при небольшом числе опытов ($n \le 40 \ldots 50$, в проведенном исследовании $n=30$) не может быть приравнена к статистической дисперсии $D^* = \frac{1}{n}\sum_{i=1}^{n}(X_i - \tilde{m})^2$. В этом случае возникает систематическая ошибка, для того чтобы ее исправить, достаточно ввести коэффициент $\frac{n}{n-1}$ и тогда:

$$\tilde{D} = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \tilde{m})^2. \tag{5}$$

Выбранная несмещенная оценка $\tilde{a}$, должна обладать минимальной дисперсией по сравнению с другими оценками:

$$D\left[\tilde{a}\right] = \min. \tag{6}$$

Степень близости статистической оценки $\tilde{a}$ к ее характеристике, a теоретического распределения может быть определена с помощью равенства:

$$P\left\{\tilde{a} - a_2 < a < \tilde{a} - a_1\right\} = \gamma, \quad (a_2 > a_1), \tag{7}$$

которое означает: вероятность того, что случайный интервал ($\tilde{a} - a_2,\ \tilde{a} - a_1$) содержит в себе достоверную, но неизвестную характеристику a, равную γ. Вероятность γ называется доверительной вероятностью, а интервал ($\tilde{a} - a_2,\quad \tilde{a} - a_1$) – доверительным интервалом.

Вероятность γ характеризует надежность статистической оценки $\tilde{a}$. В том случае, если a_1 и a_2 равны по абсолютной величине и противоположны по знаку, то можно говорить о точности статистической оценки:

$$\alpha = \frac{1}{2}(a_2 - a_1) = a_2 = \left|a_1\right|. \tag{8}$$

В настоящем исследовании для оценки доверительного интервала задана доверительная вероятность или надежность 95 % (0,95) [1, 2].

Коэффициенты корреляции r_x, r_z между ошибками измерения дальности и направления по реперной точке и предмету близки к 1, т.е. рассеяние экспериментальных точек малое, а протяженность поля точек достаточно большое. Поэтому для усреднения несовместных решений системы уравнений применялся один из методов регрессионного анализа – метод наименьших квадратов (МНК). Решение задачи, получаемое МНК по экспериментальным точкам, содержащим случайные ошибки, само также случайно, но благодаря усреднению многократных расчетов оно становится более определенным, более устойчивым.

При анализе области разброса исходных экспериментальных данных во внимание принимались ошибки способов определения исчисленных данных, неадекватность принятых вариантов решения задачи, а также учитывалась невоспроизводимость от опыта к опыту, или диффузность исследуемого процесса.

Разброс исходных данных складывается из трех составляющих: $\sigma_{Диф.}$ – диффузности условий измерений, $\sigma_{Вар.}$ – ошибки адекватности принятых вариантов, $\sigma_{Сп.}$ – ошибки способов измерений. Эти составляющие можно считать некоррелированными, тогда:

$$\sigma_{ИД} = \sqrt{\sigma^2_{Диф} + \sigma^2_{Вар} + \sigma^2_{Сп}}\,. \tag{9}$$

Для упрощения принято условие, что варианты решения задачи адекватны ($\sigma_{Вар.} << \sigma_{Диф.}$ и $\sigma_{Вар.} << \sigma_{Сп}$), и тогда размером $\sigma_{Вар}$ можно пренебречь. В этом случае, при уменьшении ошибок способов измерений (величиной $\sigma_{Сп}$ можно пренебречь) результирующий разброс исходных данных будет определяться только диффузностью условий измерений. Для усреднения разброса необходимо провести большое количество опытов, что требует увеличения затрат времени. Если же в этих условиях принять менее точные способы измерений, до тех пор, пока $\sigma_{Сп} < \sigma_{Диф}/3$, погрешность обработки останется неизменной, а эффективность эксперимента повысится, за счет уменьшения затрат времени.

Таким образом, при $\sigma_{Сп.} << \sigma_{Диф.}$ точность определения данных не может быть значительно повышена, за счет более точных способов измерений. Единственный путь повышения точности – статистическая обработка многократных отсчетов.

В качестве условия принято следующее утверждение – для обеспечения наибольшей эффективности эксперимента нет смысла уменьшать ошибку способа определения установок больше чем до $\sigma_{Сп.} < \sigma_{Диф.}/3$, и увеличивать объем выборки (количество опытов) до тех пор, пока величина $\sqrt{(\sigma^2_{Сп} + \sigma^2_{Диф})/n}$ не будет сопоставима с погрешностью адекватности принятых вариантов или систематической составляющей ошибки способа определения установок.

При рассмотрении вопроса о повышении точности путем статистического усреднения полагают, что точность с увеличением числа n усредняемых значений возрастает, как $\sqrt{n}$, потому что $\sigma_{\bar{x}} = \sigma_i / \sqrt{n}$. Это справедливо лишь при полном отсутствии систематических погрешностей и абсолютной независимости значений между собой, т.е. при полном отсутствии их взаимной корреляционной связи. Очевидно, что у каждого из способов измерений есть свои систематические ошибки и многие способы зависят друг от друга. Следовательно, повышение точности, в соответствии с соотношением $E_{\bar{x}} = E_i / \sqrt{n}$, будет происходить лишь в ограниченном диапазоне значений числа n усредняемых значений [3].

Для того чтобы компенсировать влияние существующих корреляционных связей, при планировании и проведении эксперимента необходимо учесть возможно более полно различные условия измерений, т.е. варианты условий для проведения эксперимента должны быть адекватны (тождественны) реальным условиям полевых измерений.

При проведении математического моделирования условий измерений и определения установок для измерений использовался метод имитационного моделирования.

2. Планирование проведения эксперимента

Под экспериментом понимается совокупность операций, совершаемых над объектом исследования с целью получить информацию о его свойствах. В данном исследовании проводится активный имитационный эксперимент с использованием как эмпирических зависимостей, так и математических описаний. Цель планирования эксперимента – получение максимального объема информации об исследуемой системе в каждом эксперименте.

Спланирован полный факторный эксперимент вида ПФЭ2^k, где $k=2$.

Составлено уравнение вида:

$$y = b_0 + b_1x_1 + b_2x_2 + b_{12}x_1x_2 + b_{11}x_1^2 + b_{22}x_2^2 \tag{10}$$

Общее число различных комбинаций уровней в ПФЭ для k факторов можно вычислить как:

$$N = 2^k + 2k + 1 \tag{11}$$

План ПФЭ2^k для этого уравнения представлен в виде (табл. 1).

Таблица 1

План проведения эксперимента ПФЭ2^k

i	0	1	2	3	4	5	y
U	x_0	x_1	x_2	$x_3=x_1{\times}x_2$	$x_4 = x_1^2$	$x_5 = x_2^2$	
1	+1	-1	+1	+1	+1	+1	y_1
2	+1	+1	-1	-1	+1	+1	y_2
3	+1	-1	+1	-1	+1	+1	y_3
4	+1	+1	+1	+1	+1	+1	y_4
$\sum_{U=1}^{N} x_{iU}$	4	0	0	0	4	4	

Геометрическое отображение плана ПФЭ2^k с указанием номеров плана представлено на рис. 1. Точки плана располагаются в вершинах квадрата.

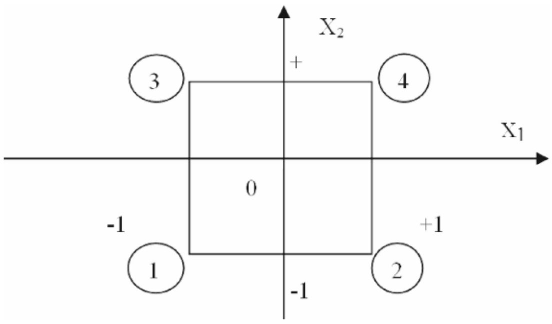

Рис. 1. Геометрическое отображение плана ПФЭ2^k в факторном пространстве

В ходе эксперимента по математическому моделированию условий измерений были получены следующие результаты:

- срединные ошибки предлагаемого способа по дальности и направлению становятся меньше, чем ошибки наиболее точных способов измерений в два-три раза;
- предлагаемый способ можно использовать в пределах всего диапазона измерений дальности.

Срединные ошибки способов равноточных и неравноточных измерений (РИ и НРИ) унифицированного способа (УС) составляют по дальности E_X=(0,35…0,63) %Д, по направлению E_Z=(1,61…2,47) ед. изм. углов.

Таблица 2

Срединные ошибки измерения расстояния установок по дальности (в %$Д_{изм}$)

Количество реперных точек	Унифицированный способ		Наиболее точный способ
	РИ	НРИ	
1	0,57	0,63	1,04
2	0,38	0,45	
3	0,35	0,41	

Таблица 3

Срединные ошибки измерения направления (в ед. изм. углов)

Количество реперных точек	Унифицированный способ		Наиболее точный способ
	РИ	НРИ	
1	2,26	2,47	5,01
2	1,91	1,87	
3	1,81	1,61	

При проведении эксперимента была исследована зависимость между срединными ошибками по дальности и по направлению для следующих вариантов: когда в расчет принимаются коррелированные поправки и когда эти поправки не учитываются.

Срединные ошибки способов РИ и НРИ (УС) при учете коррелированных поправок: по дальности составляют E_X=(0,32…0,74) %Д, по направлению Е=(1,60…2,97) ед. изм. углов; соответственно, без учета коррелированных поправок - E_X=(0,39…0,79) %Д, Е=(1,52…2,85) ед. изм. углов; (табл. 4).

Очевидно, что при практическом применении нет необходимости рассматривать и учитывать степень корреляционной связи вышеперечисленных способов.

Таблица 4

Зависимость срединных ошибок по дальности (в %$Д_{изм.}$) и по направлению (в ед. изм. углов) от учета коррелированности поправок при равноточных измерениях

$Д/Д_{max}$	По дальности		По направлению	
	С учетом коррелир. поправок	Без учета коррелир. поправок	С учетом коррелир. поправок	Без учета коррелир. поправок
0,15	0,42	0,39	1,71	1,61
0,30	0,38	0,40	1,60	1,58
0,45	0,36	0,42	1,58	1,52
0,60	0,32	0,34	1,52	1,62
0,75	0,36	0,39	1,66	1,64
0,9	0,40	0,41	1,86	1,69
1,0	0,42	0,39	2,31	1,98

3. Проверка адекватности регрессионной модели

Для обобщения полученных результатов проведен регрессионный анализ. В ходе регрессионного анализа получены регрессионная зависимость функции отклика (срединной ошибки) от факторов (времени и количества реперных точек). Рассчитаны доверительные интервалы прогноза, проведена проверка адекватности и точности модели.

Для экспериментального диапазона данных (0≤T≤6 и 1≤N≤4) составлена аппроксимирующая формула:

$$E = K_3T^3 + K_2T^2 + K_1T + K_0, \tag{12}$$

где $K_3 = A_3N^3 + B_3N^2 + C_3N + D_3$,

$K_2 = A_2N^3 + B_2N^2 + C_2N + D_2$,

$K_1 = A_1N^3 + B_1N^2 + C_1N + D_1$,

$K_0 = A_0N^3 + B_0N^2 + C_0N + D_0$.

Таблица 5

Значения коэффициентов в аппроксимирующей формуле

i	A_i	B_i	C_i	D_i
0	0,725	5,66	-12,60	7,82
1	0,428	-3,32	7,14	-4,01
2	0,08	0,67	-1,29	0,71
3	0,01	-0,038	0,08	0,04

N – количество реперных точек, принятых для обработки;
T – промежуток времени, ч.

Рассчитаны значения срединных ошибок определения данных по дальности и направлению для интервала времени Δt до 11 часов и составлен интервальный временной ряд (табл. 2, 3).

Исходя из опыта полевых измерений и практической целесообразности, увеличивать промежуток времени свыше 11 часов не имеет смысла. Кроме того, нет необходимости увеличивать количество реперных точек, принимаемых для обработки, свыше 4-х, так как в этом случае точность будет повышаться очень незначительно. Для выявления аномальных уровней временных рядов использован метод Ирвина [4]. При этом применены следующие формулы (13, 14):

$$\lambda_t = \frac{|y_t - y_{t-1}|}{\sigma_y}, \quad t = \overline{1,n}, \qquad (13)$$

σ_y – среднеквадратическое отклонение временного ряда:

$$\sigma_y = \sqrt{\frac{\sum (y_t - \bar{y})}{n-1}}, \; \bar{y} = \frac{\sum y_t}{n} \qquad (14)$$

Таблица 6

Срединные ошибки определения данных по дальности (в %Д)

Δt, ч	Количество реперных точек								
	1		2		3		4		СП
	РИ	НРИ	РИ	НРИ	РИ	НРИ	РИ	НРИ	
0-3	0,57	0,63	0,38	0,45	0,35	0,41	0,26	0,32	1,39
4	0,62	0,68	0,52	0,57	0,38	0,45	0,31	0,37	1,45
5	0,66	0,7	0,6	0,64	0,47	0,51	0,39	0,42	1,48
6	0,71	0,78	0,63	0,68	0,53	0,57	0,46	0,52	1,52
7	0,78	0,82	0,74	0,78	0,59	0,62	0,63	0,67	1,57
8	0,90	0,98	0,82	0,85	0,75	0,77	0,69	0,74	1,61
9	1,09	1,11	0,91	0,93	0,82	0,85	0,76	0,80	1,65
10	1,27	1,29	1,05	1,09	0,88	0,91	0,83	0,87	1,69
11	1,68	1,75	1,12	1,15	0,95	0,94	0,90	0,93	1,74

Таблица 7

Срединные ошибки определения установок по направлению (в единицах измерения углов)

Δt, ч	Количество реперных точек								СП
	1		2		3		4		
	РИ	НРИ	РИ	НРИ	РИ	НРИ	РИ	НРИ	
0-3	2,26	2,47	1,91	1,70	1,81	1,61	1,38	1,32	6,68
4	3,15	2,92	2,53	2,20	2,25	1,95	1,72	1,75	7,06
5	4,37	3,82	2,85	2,25	2,52	2,05	2,10	2,25	7,40
6	4,70	3,95	3,20	2,30	2,90	2,12	2,52	2,35	7,74
7	5,76	3,96	3,67	2,58	3,26	2,34	2,88	2,82	8,10
8	6,61	4,42	4,09	2,76	3,61	2,50	3,26	3,12	8,45
9	7,46	4,87	4,51	2,95	3,96	2,67	3,64	3,40	8,81
10	8,3,2	5,33	4,93	3,13	4,32	2,83	4,02	3,78	9,16
11	8,97	5,79	5,14	3,98	4,60	3,80	4,45	4,57	9,51

Рассчитанные значения критерия Ирвина λ_2, λ_3… сравнивались с табличным критерием Ирвина λ_a (табл. 8), все они оказались меньше табличных и, следовательно, заданные значения временного ряда являются нормальными.

Таблица 8

Значение критерия Ирвина для уровня значимости α=0,05

n	2	3	10	20
λ	2,8	2,3	1,5	1,3

Установлено, что значения остаточной компоненты, выделенное из исследуемого ряда тренда, удовлетворяют свойствам случайности, независимости и она подчиняется нормальному закону, следовательно, исследуемая модель адекватна. Средняя относительная ошибка прогноза составляет $|\delta\Delta t| = 1,8\ \%$, оба исследуемых фактора: время Δt и количество реперных точек N являются значимыми.

Заслуживает интереса тот факт, что при определении ошибки рекомендуемого способа установлено: способ равноточных измерений дает меньшую ошибку по дальности и несколько меньшую ошибку по направлению, чем способ неравноточных измерений. Это связано с тем, что при неравноточной обработке результатов измерений были приняты одинаковые срединные ошибки как для коррелированных способов, так и для некоррелированных. Исследования показали, что расчет уточненных коэффициентов корреляции, при своей сложности, приводит к увеличению точности на (0,01…0,04) %Д и на (0,1…0,3) единиц измерения углов, по сравнению со способом равноточных измерений. С практической точки зрения, такое несущественное увеличение точности не может компенсировать значительное усложнение расчетов, особенно если нет возможности использовать специализированные ЭВМ. Кроме того, как утверждалось ранее, «нет смысла уменьшать ошибку способа определения установок больше чем до $\Delta_{Сп} < \Delta_{Диф}/3...$» [4], в противном случае – систематическая ошибка способа будет нивелировать повышение точности. С практической точки зрения это означает, что использование двух-трех менее точных результатов дает меньшую ошибку, чем учет одного-двух более точных.

Таким образом, при использовании способа комплексного учета условий измерений срединные ошибки определения данных в 1,5…3 раза меньше, чем ошибки рекомендованных способов определения данных. Использование этого способа не требует дополнительных материальных затрат и новых приборов для проведения измерений.

Список литературы

1. Коваленко И.Н., Филиппова А.А. Теория вероятностей и математическая статистика. – М.: Высшая школа, 1982. – 256 с.
2. Пугачев В.С. Теория вероятностей и математическая статистика. – М.: Наука, 1979. – 495 с.
3. Вентцель Е.С., Овчаров Л.А. Теория вероятностей и ее инженерные приложения. – М.: Академия, 2003. – 464 с.
4. Новицкий П.В., Зограф И.А. Оценка погрешностей результатов измерений. – Л.: Энергоатомиздат, 1985. – 248 с.

Asadullin E.Z. – candidate of technical sciences
E-mail: env60@yandex.ru
Kazan State University of Architecture and Engineering
The organization address: 420043, Russia, Kazan, Zelenaya str., 1

Increase the accuracy of the calculated data with using an integrated accounting measurement conditions

Resume

In this article examines the procedure of verification of the General hypothesis of the study on improving the accuracy of calculated data for calculations on the basis of a comprehensive accounting of the measuring conditions. The study was conducted on the basis of the analysis of the scientific literature and guidance documents in the field of processing the measurements, theory of probability and mathematical statistics. The study used a combination of theoretical and experimental methods: of mathematical data processing, analysis and generalization of practice, comparison, statistical modeling, observation and experiment, graphical display of the results of the study. Reliability of the results is confirmed by convergence of theoretical and practical research results. The results of the study do not contradict the theoretical work and recommendations of the guidelines. Officials receive a large amount of information about the conditions of measurements. The guidance document on the definition of calculated data suggests that the accuracy of implementation of the measures of preparation of the data must constantly grow, to be more precise measurement of the various factors and etc. But in this case, the results are less accurate measurements are dropped, that is, deliberately not used all the available information. Calculations show that the complex accounting and processing of information obtained in the performance of particular activities preparation of measurements, allows increasing the accuracy of measurements in average by 0,5-4 %. Put forward a proposal for the identification data of measurements with the use of all available information on the conditions of measurements with the use of equal to the accuracy of measurements (EM), and not equal to the accuracy of measurements (NEM).

Keywords: complex accounting, dispersion, the variance of mathematical expectation, mathematical model, methods of regression analysis.

References

1. Kovalenko I.N., Filippova A.A. Probability theory and mathematical statistics. – M.: Higher school, 1982. – 256 p.
2. Pugachev V.S. Probability theory and mathematical statistics. – M.: Science, 1979. – 495 p.
3. Wentzel E.S., Ovcharov L.A. Theory of probability and its engineering applications. – M.: Academy, 2003. – 464 p.
4. Novitsky P.V., Zograf I.A. Evaluation of errors of measurements. – L.: Energoatomizdat, 1985. – 248 p.