GENERAL PROBLEMS OF METROLOGY AND MEASUREMENT TECHNIQUE

DEFINITIONAL UNCERTAINTY AND ERROR OF INADEQUACY

S. F. Levin

An interpretation of definitional uncertainty as a scattering parameter (standard deviation) of the contour estimate of the error of inadequacy is given for the equations of the method of indirect measurement. The problem of calibration of a thermometer from the Guide to the Expression of Uncertainty in Measurement is considered. In solving the calibration problem, an estimate of the definitional uncertainty obtained this way amounts to 74.1% of the standard uncertainty of the corrections.

Keywords: *error of inadequacy, definitional uncertainty, reliability, confidence probability.*

On April 19, 2017, the Federal Service of Accreditation of the Russian Federation approved the *Guide for Applicants and Accredited Individuals for the Creation of Domains of Accreditation of Calibration Laboratories with Allowance for Uncertainty*. There is no mention in the *Guide* of All-Russia State Standard GOST R 54500 3–2011/Guide ISO/MEK 98-3–2008, *Measurement Uncertainty. Part 3. Guide to the Expression of Uncertainty in Measurement* (GUM), or of the similarly titled GOST 34100.3–2017. However, in the course of accreditation calibration laboratories are assigned the task of expressing measurement capabilities "in accordance with GUM," i.e., whether in accordance with [1] or with the English-language original and, moreover, with allowance for the precision of the standards employed on the basis of the recommendations of the measurement chains. It is also necessary to specify the minimal values of the expanded uncertainty of measurements in calibration obtained by multiplication of the standard uncertainty by the conformance factor $k = 2$ corresponding to a confidence level of approximately 95% under the assumption of a normal distribution.

In fact, the measurement equation

$$Y = F(X_1, ..., X_N) \tag{1}$$

is the subject of the study [1], where Y is the desired quantity, or output variable and $F(\cdot)$ a given functional dependence or mathematical model of the measurement object; the values of the input variables $X_1, ..., X_N$ are obtained in measurements or from outside sources.

As a result of a calibration of a measurement instrument an equation of form (1) must represent, by RMG 28–2013, *State System for Ensuring the Uniformity of Measurements (GSI). Metrology. Basic Terms and Definitions*, a calibration curve, where Y and X_1 are the corrected value and the measured value, respectively, of a measurable quantity, while the other input variables are auxiliary quantities needed for construction of the calibration curve $F(\cdot)$. The requirements associated with the minimal value of the expanded uncertainty of measurements in calibration must be referred to the calibration diagram.

By [1] (sec. 3.3.2), "in actual practice, there exist many possible sources of uncertainty in measurement, including: (a) incomplete determination of a measurable quantity; (b) imperfect realization of a measurable quantity; (c) nonrepresentative sample – the measured model may not represent a definite measurable quantity; (d) *inadequate* knowledge of the effects of environmental conditions that influence the measurement or imperfect measurement of the environmental conditions; (e) a subjective systematic error of the operator or in taking the readings of similar devices; (f) finite resolution of a device or

finite sensitivity threshold; (g) inaccurate values assigned by the standards used in the measurements and by standard samples of substances and materials; (h) inaccurate values of constants and other parameters obtained from outside sources and used in the data processing algorithm; (i) approximations and assumptions used in the measurement method and in the measurement procedure; (j) variations in repeated observations of a measurable quantity under clearly identical conditions. These sources are optionally independent, and some of the sources in the list from (a) to (i) may introduce a contribution to source (j). Of course, an *unknown systematic effect cannot be taken into account in an estimate of the uncertainty of the result of measurements, though it may introduce a contribution to its error.*"

The main point in [1] (sec. D.1.1) concerning the quantitative measure of "inadequacy" is that "The first step in performing a measurement is to decide on the measurable quantity, i.e., the quantity which is to be measured. A measurable quantity thus cannot be determined in terms of some value but only in terms of its *description*. In principle, however, a *complete description* of a measurable quantity requires an unlimited amount of information. Incompleteness of the description of a measurable quantity leaves room for different interpretations and, thus, introduces into the uncertainty of the result of a measurement a component that may be, and, in fact, may not be *significant* by comparison with the precision required of the measurement. ... At any level of specification of a determination of a measurable quantity, the latter will possess an *inherent uncertainty* that, *in principle*, may be estimated in one way or another. This uncertainty characterizes the extreme accuracy with which a measurable quantity may be known and every measurement in which such uncertainty is attained may be considered the best possible measurement of the given quantity. In order to obtain a result of a measurement with lesser uncertainty, it will be necessary to determine the measurable quantity with greater completeness."

In a comment on this point it is stated that "although a measurable quantity must be determined in sufficient detail in order that any uncertainty caused by incompleteness of its determination be negligibly small by comparison with the required measurement precision, it must be acknowledged that this is not always *practical* (i.e., realized in practice – revision to GOST R 54500.3–2011)."

Moreover, in GOST 34100.1–2017/ISO/IEC Guide 98-1:2009, *Uncertainty in Measurement. Part 1. Introduction to Guide to the Expression of Uncertainty in Measurement*, it is stated:

"7.2.3. For many measurement situations reliable results are obtained by the method used to calculate uncertainty based on GUM [cf. JCGM 100 (sec. 5)]. If the measurement function is linear relative to the input quantities and if these quantities are distributed by a normal law, the method used to estimate uncertainty on the basis of GUM yields exact results [cf. JCGM 101 (5.7)]. But even if these conditions are not observed, the method may function well enough in practice [cf. JCGM 101(5.8)].

7.2.4. However, there exist measurement situations for which the method of estimation of uncertainty based on GUM cannot be considered satisfactory. This is particularly the case if (a) the measurement function is nonlinear; (b) the probability distributions for the input quantities are asymmetric; (c) $|c_1|u(x_1)$, ..., $|c_N|u(x_N)$ yield a contribution to uncertainty (cf. 4.14) that are not quantities of roughly the same order of magnitude [cf. JCGM 100 (G.2.2)]; (d) the probability distribution for the output quantity is either asymmetric or differs substantially from a normal distribution or *t*-distribution. Sometimes it is difficult to decide in advance whether a given measurement problem permits the use of a GUM-based method of estimation of uncertainty."

But at least one of the conditions (a)–(d) will *always* be satisfied in practice [2]. This special case at once acquired the name , "drama of inadequacy" [3], since, relative to reliability, it is known that the conformance probability, confidence level, and confidence probability are not all the same [4]. The problem is that no individual specific method (let alone a principle) of estimation of *"inherent" uncertainty* was specified. It was supposed that uncertainty due to incompleteness of determination of a quantity, i.e., an equation of form (1), is negligibly small by comparison with the required measurement precision, a risky preposition which a calibration laboratory cannot dispense with (in view of the above *Guide for Applicants and Accredited Individuals* of April 19, 2017).

Thus, there arose the problem of definitional uncertainty under extremely strange formulations that lack any real solution, a problem that also became central for the *Guide* [1]. At the same time, the compositional approach of R 50.2.004–2000, *GSI. Determination of the Characteristics of Mathematical Models of Relationships between Physical Quantities in the Solution of Measurement Problems. Basic Assumptions*, represents an alternative feature-based approach to estimation of the

precision of solutions of measurement problems based on the error of inadequacy for models of type (1). In fact, definitional uncertainty and error of inadequacy relate to the same property of the measurement equation (1) in the basic measurement problem of the *Guide* [1] which, from the qualitative point of view, is directly related to the analytic expression of this equation and is referred to as "inadequacy."

Definitional uncertainty and inadequacy. The term "definitional uncertainty" does not occur in [1] or in its English-language original, but there exists the word "definition" [1] (sec. 3.1.3), the Russian-language term "neadekvatnost" [1] (sec. 4.1.2), and the English-language term "inadequacy" in the original. Nevertheless, detailed explanations are presented in [1].

In the first place, "since a mathematical model may be incomplete, ranges of variability of the influencing quantities corresponding to what occurs under practical measurement conditions must be available in order to estimate uncertainty on the basis of observation data. In order to obtain *reliable* estimates of uncertainty, it is recommended that, if at all possible, empirical mathematical models based on long-term measurements of quantitative quantities should be used. Comparison standards and control charts in order to decide whether a measurement is under *statistical control* should also be used. If the data of observations, including results of statistically independent measurements of one and the same measurable quantity, attest to incompleteness of the model, the model must be reviewed. The *reliability* of estimates of uncertainty may be substantially improved with the use of well-designed experiments, hence experimental design must be considered an important component of the technique of performing measurements" [1] (sec. 3.4.2). But in the theory of experimental design adequacy or inadequacy is considered as the conformance of nonconformance to some statistical criterion at a given significance level.

Secondly, a number of revisions are given in [1], Appendix G:

"G.1.4. If the probability distributions of the input quantities X_1, X_2, ..., X_N (their mathematical expectations and variance and, if these quantities are not normal, moments of higher orders (cf. C.2.13 and C.2.22)) on which the measurable quantity Y depends are known and if Y is linear function of the input quantities, $Y = c_1 X_1 + c_2 X_2 + ... + c_N X_N$, the probability distribution of Y may be obtained by a convolution of the probability distributions of the input quantities (cf. [5][1]). Thus, the values k_p forming intervals with specified confidence level p_1 may be calculated from this convolution.

G.1.5. If the functional relationship between Y and the input quantities is nonlinear and if the limitation to terms of first order of an expansion in a Taylor series of this relationship cannot be considered an acceptable approximation (cf. 5.1.2 and 5.1.5), the probability distribution of Y is not a convolution of distributions of the input quantities. In such cases other types of analytic or numerical methods of calculation must be used."

Thirdly, however, "in actual practice a convolution procedure in the calculation of intervals with given confidence levels is not used or used extremely rarely because of the following reasons: the parameters of the distribution of the input quantity are usually not known precisely and are only estimates; it is not easy to expect that the confidence level for a given interval may be known with a high degree of precision; and implementation of this procedure is complicated from the mathematical point of view. In addition, approximations based on the Central Limit Theorem are used" [1] (sec. G.1.6).

Obviously, we are considering only the methods of [5] and "other analytic or numerical methods of calculation" known to developers [1], moreover, verification of the conditions of applicability of the Central Limit Theorem is hardly open to question.

The essence of the special case of the "drama of uncertainty" is set forth in the dictionary [6] (sec. 2.27) with the use of new terms as follows: "the uncertainty of the very definition of a measurable quantity (*definitional uncertainty*) establishes a minimal limit of uncertainty of measurements." But it is precisely the *absence* of any indication of a *particular method of* estimation of the minimal limit of "inherent" uncertainty that turns "measurement uncertainty" into an *unknowable and unknown quantity*.

The above minimal level of uncertainty of a measurement belongs exclusively to the description of the quantity $Y = F(X_1, ..., X_N)$ and is not related to the uncertainty of measurements of the arguments of Eq. (1) that may be performed in the course of solving a measurement problem in [1]. And since by [1] (sec. D.1.1) "in principle, a *measurable* quantity may be *completely* described only where there is an unlimited amount of information," this means that definitional uncertainty of any mathematical model *always* exists and that the *significance* of definitional uncertainty by comparison with the required

---

[1] Here and below, the enumeration of the sources in the citations were changed in accordance with the list of references to the present article.

precision of the result of the solution of a measurement problem must *always* be verified. But it is precisely this which is missing from [1] and from the scheme of the revision [7]. Moreover, in [1], despite its claim in sec. 3.4.8 regarding "critical thinking and intellectual integrity and competence," it is not only "inherent" uncertainty which is considered to be negligibly low by comparison with the other components of measurement uncertainty, but also the desired quantity in the basic measurement problem, which is assumed to be uniquely defined.

However, the principal problem [1] is stated thus: "true value" and "error" are declared to be *highly likely* unknowable and unknown. And in the international dictionary VIM-3 [6] certain terminological improvements appear *deus ex machine*. Whereas in [1] the method of substitution and the differential and null methods are classified as measurement methods, in [6] this list is complemented with the method of direct measurement and method of indirect measurement, with the latter coinciding with the method of solving the basic measurement problem [1] by means of Eq. (1).

A number of circumstances spoil the pattern of "coincidence." By [6] (sec 2.1) a "measurement is a process of experimental derivation of one or more values of a quantity, all of which may be justifiably assigned to a quantity," while by RMG 29–2013, *GSI. Metrology. Basic Terms and Definitions*, in Comment 1 to the term "4.19 direct measurement" it is noted that "strictly speaking, a measurement is always a *direct* measurement." And it is not clear what is more important in RMG 29–2013 – a remark or a definition.

But neither the "method of indirect measurements" nor, even more so, the "indirect method of measurement" (cf. [6], sec. 2.1, Remark 1) are measurement methods, since they contain a *nonexperimental* procedure for deriving a *computed value* of the output variable of a mathematical model (1) from the data of measurements of the input variables. A computational "experiment" by the Monte-Carlo method is a *nonexperimental* procedure when the lack of data is compensated for by a conjecture concerning these data. Nevertheless, "Use of a method of estimation based on GUM is complicated in the search for partial derivatives (or their numerical approximations) for a complex model of measurements, which is necessary for application of the *law of transformation of uncertainties, particularly if it is necessary to calculate derivatives of higher orders* (cf. JCGM 100 (sec. 5). In such cases the Monte-Carlo method is more appropriate and easier to use (cf. sec. 7.4)" [1] (sec. 7.2.5).

Without mentioning problems of sensors of pseudorandom numbers, let us again recall that the Monte-Carlo method is not a true source of measurement data.

"Analytic methods by means of which an algebraic formula for the probability distribution of an output quantity may be obtained do not contain any approximations but may be used only in comparatively simple cases. The potential to be gained from the use of such methods is demonstrated in [8. 9]. Among the measurement problems for which an analytic derivation is possible, there are problems where the output quantity is a linear function N of the input quantities ... which are all distributed by a normal law or by a rectangular law with the same set of boundaries. An example for two quantities ... with rectangular probability distribution which yield a trapezoidal distribution of the output quantity (cf. [5]) ...) [1] (sec. 7.3.1). But *analytic formulas for probability distributions* have their own uncertainty and the physical sense and method of identification of this uncertainty are not disclosed in [1] or even mentioned. This component of the error of inadequacy may also be identified by the method of contour estimation by MI 2916–2005, *GSI. Identification of Probability Distributions in the Solution of Measurement Problems*. And the claims [1] of several "generally recognized interpretations of probability" (more than four are known today) do not suddenly explain anything.

Though, *conceptually speaking*, definitional uncertainty is a standard deviation, it should be recalled that uncertainty, *in the broad sense*, is a probability distribution and, in the narrow sense, a scattering parameter of the same distribution [10]. Moreover, it is not the "variance of an observable quantity which is an appropriate measure of the uncertainty of the result of a measurement, but rather the variance of the arithmetic mean of a sample of observations. It is necessary to clearly distinguish the variance of the random variable z and the variance of its arithmetic mean $\bar{z}$" [1] (sec. C.3.2). And the question arises, can a known scheme for determining the error of inadequacy be used to estimate an uncertain definitional uncertainty?

Error of inadequacy of a mathematical model. Practically simultaneously with the publication of the translation [1], recommendations related to metrology R 50.2.004–2000 were introduced that presented a definition of the terms "measurement problem" and "error of inadequacy," a classification of methods of solution of measurement problems was specified, and methods and criteria for statistical identification of errors of inadequacy for mathematical models of functional-type measurement objects were specified with examples.

Traditional criteria for testing hypotheses of sufficiency of completeness or specification of mathematical models were previously considered within the framework of dispersion analysis [11] on the basis of the noncentral distribution χ^2 and Fisher distribution F [12].

"Suppose, for example, it is necessary to approximate the *experimental* points of some curve and we are using for the approximation the following form:

$$y(x) = \alpha_0 + \alpha_1 x + \alpha_2 x^2 + ... + \alpha_r x^r + \varepsilon.$$

Here ε is the residual error (error of approximation [S. F. L.]) with variance σ^2 while we do not know the value of r. The criterion F by means of which the *significance* of the coefficient α_r is verified is a powerful method for the solution of the problem" [12].

This example illustrates an approach of the theory of experimental design to the concept of the adequacy of mathematical models. In GOST 24026–80, *Experimental Design. Terms and Definitions*, the "adequacy of a mathematical model" is defined as the "degree of conformance of a mathematical model to the experimental data according to a selected criterion" with the remark that "Fisher's F-criterion is often used to verify the adequacy of a model." But it is precisely the arbitrary choice of level of *significance* in statistical hypothesis testing of the structure of a model even in the range 0.05–0.1 which is a fundamental drawback not only of the Fisher criterion. In this sense, adequacy may exist and, in fact, may not exist. On the other hand, the inadequacy of the description of a quantity *always* exists. And this is closer to the essential nature of definitional uncertainty from the point of view of [1] (sec. D.1.1).

In regression analysis the cross validation scheme is a method of estimating the inadequacy of mathematical models that has been used since 1949. It has been considered by M. Kenny, J. Tukey, A. G. Ivakhnenko, H. Akaike, H. Waba, M. Stone, B. Efron, V. Ya. Katkovnik, and others. A more effective variant of the scheme based on the relation of the Kolmogorov distance between probability distribution functions for a trial $F_t(x)$ sample and a control $F_c(x)$ sample of the data of joint measurements with the probability of their conformance relative to the fraction of the intersection of the probability distributions $f_t(x)$ and $f_c(x)$ is used in R 50.2.004–2000 as a measure of reproducibility:

$$\int_{-\infty}^{\infty} \inf_f \{f_t(x), f_c(x)\} dx \equiv 1 - 0.5 \int_{-\infty}^{\infty} |f_t(x) - f_c(x)| dx\big|_{x=x_0} \equiv 1 - \sup_F |F_t(x) - F_c(x)| = 1 - D(x_0).$$

Here x_0 is the unique point where $f_t(x_0) = f_c(x_0)$. Identity is generalized to the case of a finite number of points of intersection of the densities $f_t(x)$ and $f_c(x)$, while the control window "1/2" [13] is generalized to "1/(M + 1)" relative to the number M of parameters of the model. Thus, models of different degrees of complexity, including models of probability distribution, may be compared.

With this variant of the cross observation scheme

– the remaining portion of a sample may be used as a trial sample for the control windows and for this sample the parameters of models of a given structure estimated by given traditional methods, such as the method of least squares (MLS), method of least moduli (MLM), method of median interpretation (MMI), and others;

– the resulting models may be extrapolated to the data of control windows, thus forming an extrapolation functional of a model of given structure with given method of estimation of its parameters;

– from the distribution of the deviations of the data of measurements of an output variable from an extrapolation functional, its probabilistic estimate, what is known as the compactness function, may be obtained and the mean modulus of the random component of the error of inadequacy (MMEI) and the mean modulus of the error of closure (MMEC) calculated as estimates of the nonexcluded systematic component. The dimension of the control window is then related to the prediction interval and to the binary code θ of the structure of the model while the number of digits in the code, to a model of maximum complexity, while the sum of its units is equal to the number M of parameters of the model.

Through use of this variant of the cross observation scheme, the question of distance in the Feller variation [14] becomes a closed question and the conformance probability is introduced as a "*measure (criterion)*" of the *reproducibility* of the results of a solution of measurement problems.

The error of inadequacy of the mathematical model of a measurement object is defined in R 50.2.004–2000 as "a computed quantity, or difference of the computed value of the output variable of a model of the object relative to data of measurements of the input variables and the results of its measurement under conditions that correspond to the calculation." Based on the original sources, the components of the error of inadequacy of mathematical models of measurement objects are divided into three groups – dimensional components, parametric components, and structural components.

Dimensional components: errors of given measurements of the input variables relative to the output variable; rounding errors; errors due to interruptions in the calculations with representation of functions by means of series; errors of transformation of the data of measurements and calculations.

Parametric (observable) components: errors of approximation by a model of the data of joint measurements due to errors of estimation of the parameters of a model based on sample data, parameterization of the variables, and implementation of a computational scheme.

Structural components: errors of prediction; errors of selection of the structure of a model of a functional dependence or model of probability distribution; error of selection of method of estimation of the parameters of a model; errors associated with discontinuity of functions; errors due to statistical nonuniformity of measurement data.

These components of the error of inadequacy correspond to the "inherent uncertainty of measurement equations" of form (1) in the course of calibration.

The error of inadequacy of a model of a measurement object is estimated in the course of metrological validation, which is the first item in the list of basic problems of metrological appraisal in accordance with the recommendations of RMG 63–2004, *GSI. Assurance of the Effectiveness of Measurements in Control of Production Processes. Metrological Appraisal of Technical Documentation*.

By type of mathematical description (constant quantities, known functions, statistical and probabilistic distributions, intervals, and discrete sets) and the form of the domain of definition of the error of measurements, in accordance with GOST 8.401–80, *GSI. Classes of Precision of Measurement Instruments. General Requirements* (sec. 2.3.1), errors of adequacy of described by a unified mathematical apparatus.

For given values of the input variables of a mathematical model of a measurement object, the error of inadequacy may be represented as a component of a convolution in accordance with MI 1317–2004, *GSI. Results and Characteristics of a Measurement Error. Forms of Representation. Methods of Use in Tests of Samples of Products and Control of their Parameters* (Appendix D) and MI 2916–2005. In the general case, a multidimensional probability distribution of deviations of the data of measurements of an output variable Y from the position characteristic of a model constitutes a tolerance zone [15], the boundaries of which correspond to a determination of a statistical tolerance (P, γ)-interval found from a general population relative to a random sample in such a way that with confidence probability P not less than a fraction γ of this population is identified [4]. The errors of inadequacy of probability distribution functions for these deviations are determined by the method of contour estimation in accordance with MI 2916–2005 and in measurement problems $P = \gamma$ is adopted in accordance with the norms of confidence probability.

There is a special feature of structuro-parametric identification of models in the function of the number of parameters, i.e., complexity in the sense of Kolmogorov [16]. To a first approximation, the errors of measurements which are only supposed to be used for the construction of a functional dependence naturally do not depend on the number of parameters of the future model. On the other hand, however, the errors of approximation by a model for a sample of measurement data of fixed size exhibit a tendency to decrease with increasing number of parameters of the model, and when this number is equal to the size of the sample, these errors may become equal to zero, whereas under these conditions the errors of prediction exhibit a tendency to increase. Thus, there also arises an "awkward situation" referred to in [1] (sec. 3.4.2) consisting in a test to determine "whether a measurement is under statistical control." As the size of a sample of measurement data increases the parameters of the model must be recalculated or the structure of the model supplemented, otherwise new data will be interpreted in most cases as outliers. And in this sense the cross observation scheme serves as a natural source of "new" data which had not been used for the construction of the model. In fact, in order to obtain reliable estimates of uncertainty it is recommended in [1] (sec. 3.4.2) that "if possible, empirical mathematical models based on long-term measurements of quantitative values as well as comparison standards and control charts should be used," i.e., in accordance with RMG 63–2004 and

R 50.2.004–2000, a model of the "measurement equation" (1) must first be validated. This must also be related to the "first step in performing measurements" (cf. [1], sec. D.1.1).

Traditional methods of estimation of parameters (method of least squares, method of least moduli, method of median interpretation, etc.) of regression models *that disregard inadequacy* prove to be methods of "smoothing" of data by means of models of order no higher than third order. This is due to the increase in the sensitivity of the model to random deviations, the manifestation of multicollinearity, and the appearance of "stray" correlation in the course of data rounding [17]. Therefore, a combination of the above traditional methods in R 50.2.004–2000 with the cross observation scheme and modular criteria such as the method of maximum compactness (MMC) is denoted by the algorithms MMCMLS, MMCMLS, MCMLM, MMCMMI, etc. [18].

The existence of a model of optimal complexity with structure corresponding to the balance of the parametric and structural components of the error of inadequacy is a special feature of the above cross observation scheme. Moreover, the MMEI minimum depends on the errors of the data of joint measurements; that is, with increasing measurement error it shifts towards models with fewer parameters that are also more primitive, while in the opposite case, to more complex models. Meanwhile, complementing a model of optimal complexity may, from the point of view of [1] (sec. 3.4.2), lead only to an increase in its error of inadequacy.

A definition of the error of inadequacy as an analog of the " minimal limit of uncertainty of the description of a quantity" desired in a measurement problem, i.e., with definitional uncertainty [1], is related to the analytic description of mathematical models of measurement objects by functional relationships (by R 50.2.004–2000) and by probability distributions (by MI 2916–2005) [1].

Specification. By [6] and RMG 29–2013, "2.27. Definitional uncertainty is a component of measurement uncertainty which is the result of limited *specification* in the determination of a measurable quantity.

Remark 1. Definitional uncertainty is the practical minimum of the uncertainty of measurements in *any* measurement of a given quantity.

Remark 2. *Any* change in specification in the determination of a quantity leads to a different definitional uncertainty."

The definition contains undetermined parts that require some explanation from the point of view of the theory of measurement problems in accordance with R 50.2.004–2000 and MI 2916–2005.

1. The binary code of the structure $\vartheta = \vartheta_0 \vartheta_1 ... \vartheta_r ... \vartheta_{R-1} \vartheta_R$, for example,

$$y(x) = \vartheta_0 \alpha_0 + \vartheta_1 \alpha_1 x + ... + \vartheta_r \alpha_r x^r + ... + \vartheta_{R-1} \alpha_{R-1} x^{R-1} + \vartheta_R \alpha_R x^R + \varepsilon,$$

determines a *specification* of a model $y(x)$ of complexity R.

2. An exhaustive search of codes yields variants of the structure of the model in the cross observation scheme and for a given set of measurement data is directed towards a search for a model with minimal error of inadequacy and model of optimal complexity. In the case of several arguments, the situation of the error of inadequacy is complicated exclusively as regards the cross observation scheme, in that now the space of input variables must be divided into hypercubes.

3. In both cases, the logic of statistical inference is preserved: in the method of maximum compactness this presupposes successive validation of a system of null hypotheses of degeneracy $\mathbf{H_0}$, continuity $\mathbf{H_{00}}$, and compositional uniformity $\mathbf{H_{000}}$. Hypothesis $\mathbf{H_0}$ corresponds to the absence of any dependence. Hypotheses of nonzero values of the parameters with integral nonzero degrees of the input variables are alternative hypotheses. Hypotheses of piecewise-continuous characteristic of a position separated by "change points," where the position characteristic experiences structuro-parametric variation, are alternatives to hypothesis $\mathbf{H_{00}}$. A unified model of the position characteristic of an ensemble of the data of joint measurements corresponds to hypothesis $\mathbf{H_{000}}$. Hypotheses of a division of a statistical ensemble of the data of joint measurements into subsets to each of which a different model of the position characteristic corresponds are alternatives.

4. A selection of a method of estimation of the parameters of a model represents a *specification* of a model.

Let us now turn to the "awkward situation" [1] (sec. 3.4.2), *testing* whether "measurements are under statistical control," a step that requires preliminary metrological validation of "measurement equation" (1) and to the following definition from the international dictionary [6]: "2.9 (3.1) result of a measurement – a set of values of a quantity assigned by a measurable quantity in addition to any other accessible and essential information.

Remark 1. ... This may be expressed by the probability density function ...”

With this hint in mind we are easily led to suspect that the "set of values of a quantity assigned to a measurable quantity" is nothing other than the probability distribution.

In 1925, P. Levy proved that the central difference of the probability distribution function $F(x)$ for the characteristic function of the probability density yields, together with the uniform distribution on the interval $\pm h$, the density [19, 20]

$$f_{*R}(x) = \left[F_*(x+h) - F_*(x-h)\right]/(2h).$$

This result is easily generalized for the sum $X = \Xi + \Psi$ of an observable component Ξ with probability distribution function $F_*(\xi)$ and an unobservable component Ψ on the closed interval $[a, b]$ with probability density function

$$f_\Psi(\psi) = \left[\mathbf{1}(\psi - a) - \mathbf{1}(\psi - b)\right]/(b-a) = f_R(\psi),$$

where $\mathbf{1}(\psi)$ is the Heaviside function.

And the following equality is valid at the boundaries of the interval of uncertainty:

$$\int\limits_{-\infty}^{+\infty} f_*(x-\psi)f_\Psi(\psi)d\psi = \left.\int\limits_{-\infty}^{+\infty} f_*(x-\psi)\left[1(\psi-a) - 1(\psi-b)\right]d\psi \middle/ (b-a)\right. = \left.\int\limits_{a}^{b} f_*(x-\psi)d\psi \middle/ (b-a)\right..$$

We now perform the substitution of variables $x - \psi = z$, $d\psi = -dz$, obtaining

$$\left.\int\limits_{a}^{b} f_*(x-\psi)d\psi \middle/ (b-a)\right. = \left.\int\limits_{x-b}^{x-a} f_*(z)dz \middle/ (b-a)\right.;$$

$$f_{*R}(x) = \left[F_*(x-a) - F_*(x-b)\right]/(b-a). \tag{2}$$

A *contour estimate* of the statistical distribution function is a corollary of formula (2), while the closed interval $[a, b]$ in the case where, by MI 2916–2005, the Smirnov statistics are used to determine its boundaries, has the sense of an interval of uncertainty for the error of inadequacy of the adopted function $F_*(\xi)$.

Recall that Smirnov statistics (the greatest statistic is the Kolmogorov distance) constitute the extreme terms of the variational series of the greatest deviations of the characteristic points (accumulated relative frequencies) of the statistical distribution function from the hypothetical probability distribution function [5]. The contour estimates are discussed by A. N. Kolmogorov, H. Cramer, F. P. Tarasenko, and others. These types of estimates of the error of inadequacy of probability distributions derived from a worst-case calculation, therefore, will not allow us to consider alternative hypotheses for the most plausible distributions, which corresponds to the principles of a confidence, and not a "realistic" estimation. But with the use of an extrapolation functional for a model of the probability distribution function, it becomes possible to also perform statistical testing of nonparametric hypotheses relative to its error of inadequacy as well [21].

Thus, the *standard deviation of the error of inadequacy* defined by R 50.2.004–2000 corresponds, within the framework of a compositional approach, to estimation of the precision of the results of a solution of measurement problems [22] in the form of convolutions or compositions of probability distributions of components proper of the *definitional uncertainty* by RMG 29–2013. Meanwhile the recommendations of MI 2916–2005 may be used to calculate the definitional uncertainty of the probability distribution of possible values of the "corrected result of a measurement" in a calibration performed "in accordance with GUM" [1].

Problem of calibration of a thermometer from *Guide to the Expression of Uncertainty in Measurement*. By [1] (sec. H.3), a "thermometer is calibrated through a comparison of readings t_k, $k = 1, ..., 11$, of a thermometer possessing negligibly low uncertainty, with corresponding reference values of the temperature $t_{R,k}$ in the range from 21 to 27°C in order to obtain values of the corrections $b_k = t_{R,k} - t_k$ to the readings. *Measured* corrections and *measured* temperatures t_k are the input quantities for the estimation. The linear calibration characteristic

$$b(t) = y_1 + y_2(t - t_0) \tag{H.12}$$

TABLE 1. Verification of Solution of Problem of Calibration of a Thermometer by the GUM Technique Using the MMCMLS Algorithm

k	t_k, °C	$b_k = t_{R,k} - t_k$, °C	$b(t_k)$, °C	$b_k - b(t_k)$, °C	$b_{11}(t_k)$, °C	$b_k - b_{11}(t_k)$, °C
1	2	3	4	5	6	7
1	21.521	–0.171	–0.1679	–0.0031	–0.1684725801	–0.00252741993
2	22.012	–0.169	–0.1668	–0.0022	–0.1668117508	–0.00218824916
3	22.512	–0.166	–0.1657	–0.0003	–0.1657205358	–0.00027946416
4	23.003	–0.159	–0.1646	+0.0056	–0.1646489627	**0.00564896271**
5	23.507	–0.164	–0.1635	–0.0005	–0.1635490180	–0.00045098201
6	23.999	–0.165	–0.1625	–0.0025	–0.1624752624	–0.00252473757
7	24.513	–0.156	–0.1614	+0.0054	–0.1613534934	0.00535349341
8	25.002	–0.157	–0.1603	+0.0033	–0.1602862851	0.00328628514
9	25.503	–0.159	–0.1592	+0.0002	–0.1591928877	0.00019288771
10	26.010	–0.161	–0.1581	–0.0029	–0.1580863957	–0.00291360430
11	26.511	–0.160	–0.1570	–0.0030	–0.1569929983	**–0.00300700173**

adjusts the corrections and temperatures (relative to the measurement data) by the method of least squares. The parameters y_1 and y_2, correspondingly, the free term and the angular coefficient of the calibration characteristic, are two measurable (output) quantities. The temperature t_0 is selected by stipulation as some fixed point, hence it is not included among the independent parameters that are subject to a determination by the method of least squares. After the estimates y_1 and y_2 and their variance and covariance have been determined, formula (H.12) may be used to calculate the correction, which must be introduced into the readings of the temperature t of the thermometer and its standard uncertainty."

With $t_0 = -20$°C, $y_1 = -0.1712$°C, $s(y_1) = 0.0029$°C, $y_2 = 0.00218$, and $s(y_2) = 0.00067$°C, the sample standard deviation of the corrections $s = 0.0035$°C [1] (sec. H.3).

This problem was previously analyzed in [3], and the initial data and results of a solution of the problem obtained in [1] (sec. H.3) are presented in columns 1–5 of Table 1 of the present study, while in columns 6 and 7 may be found the results of a verification of the calculation [1] (sec. H.3) by the MMCMLS algorithm, i.e., the computed values of the MMCMLS-correction $b_{11}(t_k)$ with code $\vartheta = 11$ and deviations from the correction $b_k - b_{11}(t_k)$.

Using these data, a model of the correction function with structure code $\vartheta = 11$ for MMEI = 0.007066667°C was identified for purposes of control in accordance with R 50.2.004–2000 using the MMC-stat 2.0 program:

$$b_{\text{MMCMLS}}(t) = -\mathbf{0.2148}514 + \mathbf{0.002182}436t. \tag{3}$$

Digits that coincided with the digits of the estimates in [1] (sec. H.3) are identified in (3). Statistical testing of hypotheses was not adopted in [1]. Therefore, an identification in accordance with MI 2916–2005 of the probability distribution of possible deviations of the corrections b_k from model (3) based on uniform distributions of the components of the convolution in accordance with MI 1317–2004 in light of statistical control [23] of the parameter $\Omega_r \le$ MMEC = MMEI – MAD = = $7.066667 \cdot 10^{-3} - 2.579371621 \cdot 10^{-3} = 4.487295379 \cdot 10^{-3}$ [°C], where MAD denotes the mean absolute deviation, $|b_k - b_{\text{MMCMLS}}(t_k)|$, relative to the definition given in MI 187–86, *Procedural Specifications. GSI. Measurement Instruments. Reliability Criteria and Parameters of Verification Techniques*, yields the trapezoidal distribution (Fig. 1*a*) cited in [1] though not used there. This corresponds to the reliability boundary of the error of first-order working standards (GOST 8.558–2009, *GSI. State Measurement Chain for Instruments for Temperature Measurement*) with 0.95 confidence probability.

The procedure implemented by the two programs MMK-stat and MMI-verification and described in R 50.2.004–2000 and MI 2916–2005 is a key element in metrological certification of functional models. Complementing this procedure with a stability study of a calibrated measurement instrument solves the problem of "statistical control" [1] (sec. 3.4.2).
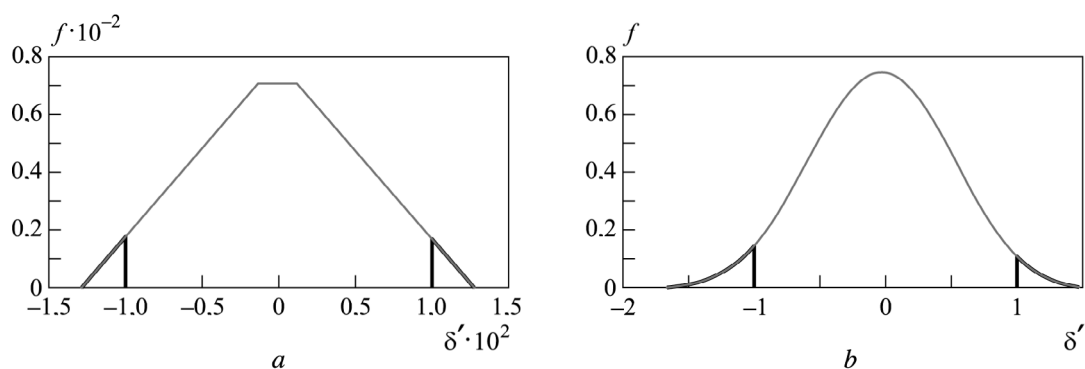
Fig. 1. Identification of distribution of deviations from models of the correction function in the MMI-verification program [15]: *a*) model (3), convolution interval [–0.011853; +0.011867]°C; *b*) model (4), normalizing value $\Delta_{0.95} = 0.00507$°C, convolution interval [–0.00846736644; +0.00753892269]°C.

The analysis would have been incomplete had it not specified that a procedure by means of an MMCMLS model of optimal complexity relative to the minimal criterion MMEI with MMEI = **2.7**12152·10^{-3}; MAD = 1.843555225·10^{-3}, and MMEC = 8.685967747·10^{-4} leads to the expression

$$b_{11011}(t) = -0.4409122 + 0.01379538t + 5.303979 \cdot 10^{-6} t^3 - 3.728615 \cdot 10^{-6} t^4. \tag{4}$$

For model (4) with structure code equal to 11111, the convolution of the truncated Gaussian probability with uniform distribution (Fig. 1*b*) proves to the most likely convolution within the boundaries [–0.00846736644; +0.00753892269]°C, which is **48.1**917502% narrower, while the "equivalent standard relative to Ω_r" is **5.166**1391318 times more precise than in the example [1] (sec. H.3).

Distributions of the observable component of the deviations of the corrections b_k from model (3) with scattering parameter 0.005194°C, nonparametric component of the inadequacy of a uniform distribution with scattering parameter 0.0021795°C, and parametric component of the inadequacy of a model with scattering parameter 0.004487295379°C form a convolution on the interval [–0.011853; +0.011867]°C. Moreover, the definitional uncertainty of the result of a GUM-based calibration amounts to (0.0021795°C + 0.004487295379°C)/√3 = **0.003849**076107°C) or 74.1% of the total standard uncertainty specified in [1] (sec. H3). The obtained result for a "model example of a GUM-based calibration" represents a response to the essential question, namely whether definitional uncertainty may be ignored in practical applications and whether statistical control of definitional uncertainty is needed for equations of form (1).

However, it should be noted that an "appropriate definition of measurement uncertainty" is not mentioned in this example and is indicated only in the Appendix [1] (sec. C.3.2) and differs from the basic definition [1] (sec. 2.2.3) as "simply" the standard deviation. And though from the point of view of an "estimation of Type A measurement uncertainty" the difference of the "standard deviation of the *arithmetic mean over a sample of observations*" desired in the basic problem [1] of a quantity from "simply" its standard deviation is essential, there is something else more important. Despite the constant reminder [1] that "error" and "uncertainty" are essentially different concepts, in order to arrive at the best estimate of the measurement equation (1) proper and of its mathematical expectation as a systematic component, the fact that the convolutions of the distributions of the components of the budget of uncertainty estimated relative to Type B and the uniformly distributed components of the budget of errors obtained from "outside sources" coincide remains indisputable [24]. And in that case, as noted in [1] (sec. E.5.2–E.5.4), the estimate of error and estimate of uncertainty will, in fact, coincide.

For this reason, Appendix B together with a sample calculation of measurement uncertainty the result of which coincided with the result of a calculation of error in Appendix A was eliminated from the scheme R 50.2.038–2004, *GSI. Direct Repeated Measurements. Estimation of Errors and the Uncertainty of the Result of a Measurements*. In other words, the statement of the objective of a measurement problem expressed in terms of the characteristics of a mathematical model of the measurement object, in accordance with R 50.2.004–2000, from the point of view of the principal objective of the Federal

Law, *On Ensuring the Uniformity of Measurements*, in the case where the concern is with a function of corrections, while in accordance with MI 1317–2004, the "probability distribution law of this random variable constitutes a mirror image of the probability distribution law of the measurement error," already presupposes the possibility of using the concept of the "error of inadequacy" to interpret the concept of "definitional uncertainty."

Calibration concludes once a relationship for the corrections to the readings of a measurement instrument has been obtained that indicate the corresponding confidence boundaries. For this purpose the boundaries of the shortest tolerance (P, γ)-interval in accordance with GOST R ISO 16269-6–2005, *Statistical Methods. Statistical Representation of Data. Determination of Statistical Tolerance Intervals*, are found on the basis of a convolution of distributions of corrections and their errors of inadequacy and errors of the standard. The problem involving a search of a model of optimal complexity based on the minimum criterion MMEI is solved within the framework of the MMI-verification program [15] in accordance with MI 2916–2005.

And the international standard ISO/IEC DIS 17025:2017, *General Requirements for the Competence of Testing and Calibration Laboratories*[2] supporting the "introduction of thinking based on estimation of the level of risk relative to nonconforming studies and claims of conformance, such as false adoption and false deviation as well as statistical presuppositions" then appeared at an inopportune time. Laboratories must constantly and objectively determine these risks, moreover, the "principal risk is that an unreliable result may be obtained in the course of a calibration or in the course of tests."

In an example of calibration of a thermometer, hypotheses of the structure and parameters of functional-type metrological characteristics, such as transformation functions, calibration characteristic, functions of corrections, and functions of the errors of a standard, along with probability density and probability density functions that are subject to statistical identification relative to sample data are all classified as "statistical presuppositions." Models of the drift of metrological characteristics as functions of time are also classified as functional-type characteristics. Moreover, article 7.2 of ISO/IEC DIS 17025:2017 also introduced the concept of verification of a technique implemented by competent personnel for the purpose of "obtaining confirmation that a laboratory possesses required characteristics." In addition, the standard also specified the following:

"7.6.1. A laboratory that performs calibration, including calibration of its own equipment, must estimate the error of measurements for all calibrations.

7.6.2. A laboratory that performs sampling or tests must estimate the uncertainty of measurements."

With that in mind, how should we understand the requirements of the *Guide for Applicants and Accredited Individuals with Regard to Construction of Areas of Accreditation of Calibration Laboratories with Respect to Uncertainty* relative to all calibrations of measurement instruments where it is necessary to estimate the reliability of results of calibrations and tests?

Expressed in terms of VIM-3 [6]: "2.5 measurement method" and " 2.6 measurement technique" as well as in the duplicate terms "4.1 measurement" and "4.23 measurement problem" of RMG 29–2013, techniques of "indirect measurements" are classified as "measurements." For these terms, the "definitional uncertainty of measurement equations" also determines the measurement capabilities of calibration laboratories. Of course, not to mention the method of joint measurements, the method of indirect measurements is an essential element of techniques used in the calibration of measurement instruments, and in order to determine the "measurement capabilities" of calibration laboratories these techniques must, in fact, undergo validation.

Now it only remains to cite VIM-3 [6]:

"2.45 **validation**, or certification – verification in which already established requirements are related to a proposed use.

2.44 **verification** (of a measurement instrument) – provide objective evidence that a given object fully satisfies established requirements.

Remark 2. A process, measurement technique, material, substance or measurement system, for example, may be an object. *Remark 5*. Verification must not be confused with calibration. Not every verification is a validation."

Conclusion. Metrology is the fundamental science of methods and means of describing physical reality by means of mathematical models. For practical applications, however, only those methods and means that assure a required precision of coincidence of the results of calculations with the results of measurements are appropriate. This property of mathematical

---

[2] GOST ISO/MEK 17025–2009, *General Requirements Imposed on the Competence of Testing and Calibration Laboratories*, was in force prior to the introduction of this standard.

models of measurement objects is related directly to the concept of inadequacy and the need for preliminary testing of the conditions of applicability of mathematical models in concrete measurement problems. The latter circumstance turns the concept of "definitional uncertainty" into a base concept of a feature-based approach to the estimation of precision. The question of the observability of mathematical models corresponding to this concept is central here. And only its solution will "in one way or another" give us a basis for asserting the comprehensibility or incomprehensibility of the phenomena of physical reality.

REFERENCES

1. *Guide to the Expression of Uncertainty in Measurement* [Russian translation], Mendeleev VNIIM, Moscow (1999).
2. S. F. Levin, "Is there a need for the Explanations of Measurements Uncertainty, appendix to Guide to the Expression of Uncertainty in Measurement," *Sovet. Metrologa*, No. 2, 49–56 (2011).
3. S. F. Levin, "Metrology. Mathematical statistics. Legends and myths of the twentieth century," *Partn. Konkurenty*, No. 1, 13–25 (2001).
4. Yu. V. Prokhorov (ed.), *Probability and Mathematical Statistics: Encyclopedia*, Bol'shaya Ross. Entsiklop., Moscow (1999).
5. C. F. Dietrich, *Uncertainty, Calibration, and Probability*, Adam-Hilger, Bristol (1991).
6. JCGM 200:2008, *International Vocabulary of Metrology. Basic and General Concepts and Associated Terms (VIM-3)* [Russian translation], Mendeleev VNIIM, BelGIM, St. Petersburg (2010), 2nd ed.
7. S. F. Levin, "Guide to the Expression of Uncertainty in Measurement: revision – a change of paradigm or a new sanction?" *Zakonodat. Prikl. Metrol.*, No. 5, 31–44 (2016).
8. G. C. Casella and R. L. Berger, *Statistical Inference*, Duxbury Press, Pacific Grove (2001).
9. M. G. Cox and P. M. Harris, *SSfM Best Practice Guide No. 6. Uncertainty Evaluation*, Tech. Rep. DEM-ES-11, National Physics Laboratory, Teddington, UK (2006).
10. S. F. Levin, "Uncertainty of the results of a solution of measurement problems in the broad and in the narrow sense," *Metrologiya*, No. 9, 3–24 (2006).
11. H. Scheffer, *Dispersion Analysis* [Russian translation], Mir, Moscow (1963).
12. D. Hudson, *Statistics for Physicists: Lectures on Probability Theory and Elementary Statistics* [Russian translation], Mir, Moscow (1967).
13. A. G. Ivakhnenko, "Method of grouped calculation of arguments – a competitor to the method of stochastic approximation," *Avtomatika*, No. 3, 58–72 (1968).
14. W. Feller, *Introduction to Probability Theory and Its Applications* [Russian translation], Mir, Moscow (1984), Vol. 2.
15. S. F. Levin, "The measurement problem of identification of the error function," *Zakonodat. Prikl. Metrol.*, No. 4, 27–33 (2016).
16. A. N. Kolmogorov, "On tables of random numbers," *Semiot. Informat.*, Iss. 18, 3–13 (1982), deposited in VINITI.
17. I. Vuchkov, L. Boyadzieva, and E. Solakov, *Applied Linear Regression Analysis* [Russian translation], Finansy i Statistika, Moscow (1987).
18. S. F. Levin, "Method of maximum compactness and complex measurement problems," *Izmer. Tekhn.*, No. 7, 15–21 (1995).
19. P. Levy, *Calcul des Probabilités*, Gauther-Villars, Paris (1925).
20. H. Cramer, *Mathematical Methods in Statistics* [Russian translation], Mir, Moscow (1975), 2nd ed.
21. S. F. Levin, "Identification of probability distributions," *Izmer. Tekhn.*, No. 2, 3–9 (2005).
22. S. F. Levin, "Guide to the Expression of Uncertainty in Measurement: problems, unrealized capabilities, and revisions. Part 1. Terminological problems," *Izmer. Tekhn.*, No. 2, 3–8 (2018), DOI: 10.32446/0368-1025it.2018-2-3-8.
23 S. F. Levin, "Mathematical theory of measurement problems: Applications. Statistical control procedures in precision measurements," *Kontr.-Izmer. Prib. Sistemy*, No. 3, 8–11 (2018).
24. S. F. Levin, "Guide to the Expression of Uncertainty in Measurement: problems, unrealized capabilities, and revisions. Part 3. Reduction to a common terminological denominator," *Izmer. Tekhn.*, No. 7, 14–22 (2019), DOI: 10.32446/0368-1025it.2019-7-14-22.