Automation of evaluation of the interlaboratory comparisons results by means of software

Velychko Oleh 1, Kursin Serhii 2, Haman Valentyn 3

1, 2, 3 SE «Ukrmetrteststandart», Ukraine
E-mail: kursin@gmail.com

Abstract

The number of interlaboratory comparisons of measurement results increase, and the number of their participants requires providers to automate data processing procedures. The article is devoted to developing software for the calculation automation and display of interlaboratory comparisons results. The article analyzes the existing software and identifies its main shortcomings. Peculiarities of the application of $E_n$, $z$, and $\xi$ criteria are considered. Features of using the Python programming language for calculations and data display are described. Using the example of conducted interlaboratory comparisons, the possibilities of the developed software for automating calculations and data analysis, and displaying results, are demonstrated.

Keywords: interlaboratory comparisons of measurement results, software, criteria of statistical functioning, calculations automation.

Abbreviations used: ILC – interlaboratory comparison of measurement results.

1. Introduction

Interlaboratory comparison of measurement results (ILC) is an important part of the metrological activity, which provides an opportunity to ensure the accuracy and reliability of measurements, to compare the results of different laboratories work, and to identify possible errors or deviations. The results of the laboratory’s participation in the MPR should be evaluated using the statistical performance criteria of $E_n$-criteria, $z$-criteria, $\xi$-criteria, which are given in DSTU EN ISO/IEC 17043 [1], and are considered in more detail in DSTU ISO 13528 [2], as well as recommendations on their application and visual display are given [3].

The condition for obtaining accreditation by a calibration or testing laboratory is participation in the ILC [4]. Also, the laboratories that have already received accreditation should take part in the ILC with other accredited laboratories during the accreditation cycle. Thus, there is a need for conducted MPRs for a large number of participants in various fields of calibration and testing.

Providers and the pilot laboratory are tasked with developing the methodology, organization, and implementation of the ILC. Since processing ILC results requires a lot of time, the implementation of calculations automation and analysis of ILC results, their visualization is an urgent task, the solution of which will significantly simplify and speed up the work of MPR providers.
2. Analysis of the latest research and publications in which the solution to the issue started

The use of computing capabilities of computers provides an opportunity to simplify data processing procedures, introduce automation of calculations and graphical display of results, make predictions, etc. [5].

General and specialized software that makes it possible to calculate the results of ILC, as well as to present them in tabular and graphical form exists. The most common general software is Microsoft Excel [6]. The advantages of Microsoft Excel are the flexibility of settings and the ability to process data that is usually already saved in table files with the format .xlsx and .xls. The main drawback can be defined as the lack of ready-made functions for evaluating the results of the ILC and constructing their diagrams — one needs to manually enter formulas for calculating indicators and adjust the diagrams taking into account the different number of participants of the ILC.

Software such as XLSTAT [7] provides wide opportunities for data analysis and display, including data processing in accordance with recommendations [2], namely the calculation of robust values. A special feature of XLSTAT is that it is an add-in to Microsoft Excel. The disadvantage of the software is the lack of statistical functioning calculations criteria and its high cost.

Among the specialized software, PROLab [8] should be noted as specially designed for planning, organizing, conducting, and analyzing ILC in accordance with requirements [1] and recommendations [2], qualification testing, and validation of methods. PRO-Lab provides an opportunity to calculate and visualize the evaluation of ILC results according to the $Z$, $Zu$, $Zeta$, $Z'$ or $E_n$ criteria. However, the disadvantages are the settings' complexity, the lack of an intuitive interface, and the price.

3. Selection of unresolved problem aspects and tasks formulation

The analysis of available solutions for automating the ILC results calculations, their evaluation, and graphical display allows us to conclude that there is an urgent need to create software that can provide an opportunity to automatically calculate the values of ILC results assessment criteria such as: $E_n$-criterion, $z$-criterion, $\xi$-criterion, and present them in tabular and graphical form.

4. Basic material

The Python language has powerful libraries for data processing, such as NumPy [9], Pandas [10], and SciPy [11]. These libraries can be used to analyze data obtained from measurement tools. Python also has libraries for data visualization, for example, Matplotlib and Seaborn [12], etc. These libraries can be used to display data in a convenient format that makes it easy to understand and analyze measurement results and evaluate results. Therefore, Python was chosen to develop the software's source code, using libraries for mathematical calculations, visualization of results, and providing a convenient user interface.

4.1. Data Input

Data of ILC results should contain information about the code of the laboratories that participated in the comparisons and the results of measurements with corresponding uncertainties. The data must be properly formatted and saved in tabular form in a *xls or *.xlsx file. xlrd, xlwt libraries are used to read and write data from a spreadsheet file. Using the xlrd.open_workbook function, a file is opened for reading, the name and path to which is contained in the variable filename. The measurement result with uncertainty for the reference laboratory is recorded in variables $X_{ref}$ and $U_{ref}$, respectively. In the future, they will be used in the calculation of criteria values. The measurement result with uncertainty for the reference laboratory is recorded in variables $X_{ref}$ and $U_{ref}$, respectively. In the future, they will be used in the calculation of criteria values. The measurement result with uncertainty for the reference laboratory is recorded in variables $X_{ref}$ and $U_{ref}$, respectively. In the future, they will be used in the calculation of criteria values. The measurement result with uncertainty for the reference laboratory is recorded in variables $X_{ref}$ and $U_{ref}$, respectively. In the future, they will be used in the calculation of criteria values.
The number of filled-in columns corresponding to the number of ILC participants is not limited and does not require special actions for its determination, which simplifies the input of result data with different numbers of participants for different ILCs.

4.2. Calculation of error values and indicators $E_n$, $z$ and $\xi$

Mathematical expressions for the criteria that must be calculated, as well as conditional expressions for evaluating the obtained results according to (1), according to which the algorithms were developed, were determined. The $E_n$ criterion evaluates the ratio of the differences between the values obtained by each participant and the assigned value provided by the reference laboratory to the differences in the corresponding expanded uncertainties of the measurement results and is determined by the expression:

$$E_n = \frac{x_{lab} - X_{ref}}{U_{lab}^2 + U_{ref}^2} = \frac{D_{lab}}{\sqrt{U_{lab}^2 + U_{ref}^2}}$$

where

$x_{lab}$ is the value obtained by the participant;
$X_{ref}$ is the value obtained by the reference laboratory;
$D_{lab}$ is the deviation of the participant’s readings;
$U_{lab}$ is the expanded uncertainty of the participant’s result;
$U_{ref}$ is the expanded uncertainty of the reference laboratory’s result.

The analysis of the results of each participant is evaluated according to the condition:

$|E_n| \leq 1.0$ – the result does not require correction or response measures;
$|E_n| > 1.0$ – the result requires corrective or response measures.

For the $E_n$ criterion, if the value is greater than 1, the value ‘r’ corresponding to the red color is written in the array of colors, and the code of the laboratory is written in the array of negative results, otherwise, the value ‘b’ corresponding to the color blue is written and the code of the laboratory is written in the array positive results. A fragment of the algorithm’s structure for evaluating the results according to the $E_n$ criterion is presented in Fig. 2, a.

It is obvious that the participating laboratory can compensate for the obtained large deviations by increasing the uncertainty value. To avoid this, it is advisable to use the $z$ criterion, which takes into account the mean square deviation of the results of all ILC participants:

$$z = \frac{x_{lab} - X_{ref}}{S}$$

where

$S$ is the mean squared deviation, or variance of the results of all ILC participants.

At the same time, if:

$|z| \leq 2.0$ – the result does not require correction or response measures;
2.0 < |z| ≤ 3.0 – the result requires certain correction or response measures;  
|z| > 3.0 – the result requires significant correction or response measures.

For the \( z \) criterion, if the value is greater than 3, the value ‘r’ corresponding to red is written in the array of colors, and the code of the laboratory is entered in the array of negative results, if the value is greater than 2 and less than or equal 3, the value ‘m’ corresponding to purple is written, and the code of the laboratory is entered into the array of doubtful results, otherwise, the value ‘b’ corresponding to the blue color is written, and the code of the laboratory is entered into the array of positive results.

A fragment of the algorithm’s structure for evaluating the results according to the \( z \) criterion is presented in Fig. 2, b.

The criterion \( \xi \) is used in the case of presenting standard uncertainties of measurement results and is determined by the expression:

\[
\xi = \frac{x_{\text{lab}} - X_{\text{ref}}}{\sqrt{u_{\text{lab}}^2 + u_{\text{ref}}^2}} = \frac{D_{\text{lab}}}{\sqrt{u_{\text{lab}}^2 + u_{\text{ref}}^2}}
\]

where

- \( X_{\text{lab}} \) is the value obtained by the participant;  
- \( X_{\text{ref}} \) is the value obtained by the reference laboratory;  
- \( D_{\text{lab}} \) is the deviation of the participant’s readings;
of 130 MHz are shown, as they demonstrate the peculiarities of the application of the $E_n$, $z$ criteria and their different values according to their possible variations.

$u_{lab}$ is the standard uncertainty of the participant’s result;

$u_{ref}$ is the uncertainty of the reference laboratory’s result.

The evaluation of the results according to the $\xi$ criterion is carried out under the same conditions as for the $z$ criterion. Negative results indicate that the laboratory still does not meet its own deviation requirements and/or the resulting uncertainty is underestimated and requires revision. Since the $\xi$ criterion is similar to the $E_n$ criterion, it is advisable to use $E_n$ for analysis, as required by [3].

The **math** library is used to calculate $E_n$, and $z$. The average value of the array of measurement results is calculated. The error values, the $E_n$ indicator value, and the squared deviation are calculated for all array elements, and the corresponding arrays are formed. The values dispersion of the values array and the $z$ indicator’s value is calculated.

The results are saved in a Microsoft Excel file and are displayed in tabular form on the tab of the working window of the software. An example of data display and calculation results in the program window is presented in Fig. 3.

![Fig. 3. An example of displaying calculation results in tabular form](image1)

### 4.3. Display of calculation and analysis results

For approbation of the program, the data of ILC conducted by SE «Ukrmetrteststandart» in 2018 on frequency measurement were used [13]. The calibration results of the OEM BPSG6 signal generator were compared for 130 MHz, 168 MHz, and 223 MHz. For example, the results for a frequency
With the help of the matplotlib.pyplot library, graphs are built that display the obtained results. So, the `errorbar` function is used for a graph showing deviations and uncertainties. An example of a window displaying graphs of the results of calculations of deviations and uncertainty in Hz is shown in Fig. 4, a.

The `bar` function is used to display graphs with $E_n$ and $z$ criteria. Variables -- arrays with corresponding criteria. The numerical values of the indicator are displayed above or below each bar, depending on whether the value is positive or negative. An example of the histogram display window with $E_n$ criteria in Fig. 4, b. An example of a histogram display window with $z$ criteria for a frequency of 130 MHz is shown in Fig. 4, c.

The same calculations and visualization of results can be performed with the help of the proposed software, in particular for the ILC given in [40].

5. Conclusions

The software has been developed, which provides an opportunity to automate calculations and analysis of criteria for evaluating the results of the ILC. The advantages of the software are the ability to read data and save the results in *.xls/*.xlsx format files, the ability to evaluate the results of the ILC according to the $E_n$-criterion, as required NAAU for providers and the $z$-criterion, the creation of a text file report based on the evaluation of the results of the ILC for all participants.

The obtained results of deviation and uncertainty estimates are displayed using graphs, and the calculated values of indicators are displayed using histograms. The results of the automatic analysis of the indicator values for each participant are presented with the corresponding color on the histogram. All this greatly simplifies the general analysis of the conducted ILC.

References


4. NAAU policy regarding the participation of conformity assessment bodies in professional level inspections and interlaboratory comparisons other than professional level inspections ZD–08.00.29 Edition 03 [Політика НААУ щодо участі органів з оцінки відповідності у перевірках професійного рівня та міжлабораторних порівняннях, відмінних від перевірок професійного рівня ZD–08.00.29 Редакція 03]. [In Ukrainian] https://naau.org.ua/userfiles/files/ZD-08_00_29_red_03_Politika%6060ka-NAAU-shhodo-uchasti-OOV-u-perevirkah-profesionogo-rivnya-ta-mizhlaboratoryr%60h-porivnyannya.pdf


