PROFICIENCY TESTING FOR INTERLABORATORY COMPARISONS ON INDUSTRIAL COMPUTED TOMOGRAPHY

Herminso Villarraga\textsuperscript{1,3}, Robert Hocken\textsuperscript{1,2,3}, Edward Morse\textsuperscript{1,2}, and Stuart Smith\textsuperscript{1,2,3}

\textsuperscript{1}Center for Precision Metrology
\textsuperscript{2}Department of Mechanical Engineering and Engineering Science,
\textsuperscript{3}Department of Physics and Optical Science,
University of North Carolina at Charlotte
Charlotte, NC, USA

This paper discusses proficiency testing (PT) for inter-laboratory comparisons on industrial computed tomography (CT) measurements using the normalized error ($E_n$ number) as the statistical parameter to assess agreement between measured data. When using these metrics the results obtained in the analysis of such round robin tests should be interpreted with caution. We take two different sets of data coming from two recent CT comparisons [1, 2]. By scrutinizing some of the reported data, a laboratory can measure one parameter with large deviations from an assigned, or reference, value, but nevertheless be positively assessed ($|E_n| \leq 1$). Oppositely, some laboratories can receive an unsatisfactory performance ($|E_n| > 1$) even if their reported measurements are close to the assigned value. This happens when participants’ reported estimates of measurement uncertainties are not determined in a consistent way. $E_n$-scores not only provide a measure of how closely a reported laboratory result agrees with the assigned value (a reference), but also require uncertainty claims in both the laboratory result and the reference value. In that way, the normalized error value really gives a representative impression about the consistency of the results in regards to the claimed laboratory uncertainties. It is tempting then to review the uncertainty budgets from the industrial CT users, but there is not yet a reliable way of uncertainty determination for dimensional measurements with CT machines due to the complexity of the physical phenomena involved in the CT principle. A complete quantitative estimation of CT influencing factors is not available to guide development of a task-specific measurement uncertainty complying to the GUM approach [3]. Additionally, there is currently a lack of standard procedures and guidelines for users of CT coordinate metrology.

ANALYSIS

To aid interpretation, the PT results usually need to be transformed into statistics that allow a performance judgment [4]. The usual statistics for PT performance assessment when participants are required to report the uncertainty estimates is the normalized error,

$$E_n = \frac{x - X}{\sqrt{U_{lab}^2 + U_{ref}^2}}$$

(1)

where $U_{lab}$ and $U_{ref}$ are the expanded uncertainties associated with the laboratory (or with its result $x$) and with the measurand (or the assigned value $X$), respectively. For a result to be acceptable, the $E_n$ number must satisfy $|E_n| \leq 1$; unsatisfactory performance is obtained if $|E_n| > 1$. Unsatisfactory results can be split into those caused by considerable errors in the measurement and those caused by an overly optimistic uncertainty estimate. An overly optimistic uncertainty estimate either means that some uncertainty contributions are underestimated or missed altogether or that there are mathematical errors in the uncertainty estimate. For the proficiency test to be valid, $U_{lab}$ must contain all uncertainty components of importance for a given
measurement. If $U_{\text{lab}}$ is underestimated, then it may lower the value of the denominator of equation (1), which increases the value for $|E_n|$ turning the performance criterion to register unsatisfactory of the measurement precision and process calibration. On the other hand, an overestimated uncertainty would conduct a satisfactory result even if the deviations between the measurement and reference values are relatively large. Thus since overestimated $U_{\text{lab}}$ values may reduce the $|E_n|$ and underestimated values may increase it, the use of such metric to assess the performance would eventually lead to distorted results causing a false failure for the $E_n$-criterion based PT.

CONCLUSIONS

The use of the normalized error ($E_n$) for performance assessment of PT does not reflect the precision and calibration of a process. Specifically, an overly large uncertainty estimate with a measurement process of relatively poor precision will be more highly rated than a very precise process carefully calibrated for which a relatively small error in uncertainty places the error bar beyond the reference value. This is observed in measurements representing the current state of the art of CT metrology. Participant 9 in Figure 1 overestimated $U_{\text{lab}}$ and gets $|E_n|<1$ (acceptable). The same happens with participants 13 and 20 on Figure 2. By other hand participants 11 and 12 with precise measurements on Figure 2, underestimated their uncertainties, getting $|E_n|>1$ (unsatisfactory). For the PT performance assessment to be valid, $U_{\text{lab}}$ must contain all uncertainty contributors of importance in the given situation. However, even though uncertainty budgets are difficult estimate, the inter-laboratory comparisons still have value for industrial CT because they offer an opportunity to increase understanding of quality issues in dimensional metrology and to determine possible reasons for variations. One of our goals will be to assemble a database of sources of uncertainty to allow better estimation of $U_{\text{lab}}$ in CT metrological measurements.

REFERENCES


