



On the Validation of Measurement Techniques in Solid Mechanics

M. Solaguren-Beascoa Fernández¹

Received: 6 October 2025 / Accepted: 31 May 2026
© The Author(s) 2026

Abstract

Currently, there is no established rigorous method for validating measurement techniques in solid mechanics. This paper presents the appropriate statistical approach for this purpose. Validation must compare measurements obtained using different techniques, considering their standard uncertainties. The use of the measurement uncertainties is essential for the metrological rigor of the validation, which is achieved through a statistical hypothesis test based on the chi-square distribution. A similar test is also possible to define for the identification of outliers. The method is simple and easy to apply, and is presented in the form of a procedural guide. Its applicability has been shown with an example of stress measurements by photoelasticity, which have been compared with the results of a finite element analysis.

Keywords Stress · Measurement · Metrology · Uncertainty · Photoelasticity

Introduction

In solid mechanics there are many measurement techniques, both to evaluate the different mechanical properties of materials and to know the stress/strain state in all types of mechanical or structural elements. Although many of these techniques were established many years ago, new methods, or improvements to existing ones (that promise greater simplicity, accuracy, range of application, etc.), continually appear.

When a new measurement technique is presented, it must be validated to ensure that it really works, that all possible sources of error have been considered and that, therefore, the technique measures without bias, or the latter is known and can be corrected. The validation process must consist of comparing the measurements obtained with their expected values, calculated theoretically or obtained by other measurement techniques. For this comparison to possess scientific rigor, it must be carried out using statistical tools. However, a review of published works on measurement techniques in solids mechanics [1–23] shows that the

validation methods employed are far from satisfactory¹. As a tradition, in all the works the comparison is made directly through a graph or Tables [1–13], by expressing the differences in terms of percentage [14, 15], or doing both [16–23]. After this basic comparison, the researchers conclude by saying that the measured values are consistent or in good agreement with theoretical ones [1–12, 15–17, 20–23], they highlight the good accuracy or reliability of the method [8, 13, 14] or they interpret the percentage of deviation as a measure of the error/accuracy of the technique [17–20].

While in other branches of science there are established comparison processes (such as analysis of variance with control groups in chemistry or medicine) and research papers in which a rigorous validation method is not carried out are rejected for publication, it is evident that in solid mechanics there is no an established comparison methodology that researchers know and use. In some of the papers discussed above researchers attempted to provide some rigor to the comparisons by calculating the mean and the standard deviation of several measurements; they interpret the standard deviation as a margin of error, but that is not strictly correct. The objective of this work is to present a method for comparing measurement results that, based on statistical concepts, allows rigorous validation of measurement techniques.

✉ M. Solaguren-Beascoa Fernández
msolaguren@ubu.es

¹ Structural Integrity Group, Escuela Politécnica Superior, University of Burgos, Avda. de Cantabria s/n, Burgos 09006, Spain

¹ The review conducted in this work included all papers in the solid mechanics literature from the past four decades in which a new measurement technique or procedure was presented or validated.

Mathematical Formulation

Without doubt, the best known and probably most used test to measure the discrepancy between an observed and a theoretical data distribution (goodness of fit) is the Pearson's chi-squared test [24, 25]. This test is applied to qualitative or categorical data through their observed and expected (theoretical) frequencies. Pearson proposed the following test-statistic to describe how different both groups of data are:

$$\chi^2 = \sum_{i=1}^n \frac{(O_i - E_i)^2}{E_i} \quad (1)$$

where data are classified into n mutually exclusive classes, O_i is the number of observations that falls into the i th class, and E_i is the expected count of each class. Under the assumption that both groups of data follow identical probability distributions and data are independent and normally distributed, the test-statistic follows a chi-squared distribution with $n - 1$ degrees of freedom. Due to the central limit theorem, the normality condition is not necessary if a sufficiently large number of data is taken.

Perhaps inspired by Pearson's chi-square test, in a paper in which six different models of measurement of nanoindentation fracture toughness are compared [26], the researchers took measurements of fracture toughness on different materials with the different measurement models and they evaluated the root-mean-square deviation between experimental and literature values for each model. Although this can serve as a comparison between techniques, since a smaller quadratic difference will express a better fit between theoretical and experimental values, it is not a rigorous hypothesis test in which the validity of the models is revealed.

A similar to Pearson's goodness of fit test can be established to quantitative data (such as measurements made in solid mechanics) through the definition of the chi-squared distribution. If we consider n independent normally distributed random variables $X_i \sim N(\mu_i, \sigma_i)$, then the sum of the squares of their standardized values is distributed according to a chi-square distribution with n degrees of freedom:

$$\sum_{i=1}^n \left(\frac{X_i - \mu_i}{\sigma_i} \right)^2 \sim \chi_n^2 \quad (2)$$

In a measurement process, due to measurement uncertainty, the measured values are only estimates of the exact values; and the standard uncertainties of these estimates are their standard deviations. This is established in the ISO 'Guide to the expression of uncertainty in measurement', commonly known as GUM [27], where the estimate of a

quantity Y_i is denoted by y_i and its standard uncertainty by $u(y_i)$. In the GUM the probability distribution assumed for $(y_i - Y_i)/u(y_i)$ is a t-distribution with ν degrees of freedom (obtained from the Welch-Satterthwaite formula). The t-distribution $t(\nu)$ resembles the standard normal distribution $N(0,1)$ and approaches to it as the number of degrees of freedom ν grows, so approximately it can be established that:

$$\frac{y_i - Y_i}{u(y_i)} \sim N(0,1) \Rightarrow y_i \sim N(Y_i, u(y_i)) \quad (3)$$

If, through another independent measurement method, another measurement of the same quantity is known (with estimate μ_i and standard uncertainty $u(\mu_i)$), in the same way it can be established approximately that:

$$\frac{\mu_i - Y_i}{u(\mu_i)} \sim N(0,1) \Rightarrow \mu_i \sim N(Y_i, u(\mu_i)) \quad (4)$$

Since the addition or subtraction of independent normal random variables also follows a normal distribution, it is true that:

$$y_i - \mu_i \sim N\left(0, \sqrt{u^2(y_i) + u^2(\mu_i)}\right) \Rightarrow \frac{y_i - \mu_i}{\sqrt{u^2(y_i) + u^2(\mu_i)}} \sim N(0,1) \quad (5)$$

Then, if we have n pairs of measurements, from (2) the following test-statistic can be formulated:

$$\chi^2(y, \mu) = \sum_{i=1}^n \frac{(y_i - \mu_i)^2}{u^2(y_i) + u^2(\mu_i)} \sim \chi_n^2 \quad (6)$$

This test-statistic follows a chi-squared distribution with n degrees of freedom and can be used to test the validity of the measurement method. Since a high $\chi^2(y, \mu)$ value indicates inconsistency between measurements obtained through the two independent measurement techniques, a hypothesis test based on the probability p of finding a χ_n^2 value greater than the observed $\chi^2(y, \mu)$ value can be established. This probability can be calculated from the chi-square distribution:

$$p = P[\chi_n^2 > \chi^2(y, \mu)] \quad (7)$$

If this probability is smaller than a certain value α , the hypothesis that the measured values are consistent with their expected values has to be rejected at a level of significance equal to α . Otherwise, if p is greater than α , it can be concluded that the measurement process obtains consistent values and, therefore, the measurement method is valid. Usually the hypothesis tests are run with a significance level $\alpha = 5\%$ because this value provides a good balance between type I and II errors [28].

One particularly useful case arises when the measurement technique used for comparison can be considered nearly exact (with standard uncertainty $u(\mu_i) \simeq 0$). In that case, from (6), the test-statistic to be used is:

$$\bar{\chi}^2(y, \mu) = \sum_{i=1}^n \frac{(y_i - \mu_i)^2}{u^2(y_i)} \sim \chi_n^2 \quad (8)$$

equation in which the chi-square value thus calculated has been denoted with an upper bar. Since this value, $\bar{\chi}^2(y, \mu)$, is greater than that obtained using (6), $\chi^2(y, \mu)$, the probability p for the hypothesis test using $\bar{\chi}^2(y, \mu)$ is always smaller than that obtained using $\chi^2(y, \mu)$:

$$\begin{aligned} \sum_{i=1}^n \frac{(y_i - \mu_i)^2}{u^2(y_i)} &> \sum_{i=1}^n \frac{(y_i - \mu_i)^2}{u^2(y_i) + u^2(\mu_i)} \\ \rightarrow P \left[\chi_n^2 > \bar{\chi}^2(y, \mu) \right] &< P \left[\chi_n^2 > \chi^2(y, \mu) \right] \end{aligned} \quad (9)$$

This means that the hypothesis test based on $\bar{\chi}^2(y, \mu)$ is more restrictive than the one based on $\chi^2(y, \mu)$; in other words, if the hypothesis test based on $\bar{\chi}^2(y, \mu)$, accepts the consistency of the data, the one based on $\chi^2(y, \mu)$ would also accept it. Therefore, this test can be used to verify the consistency of measurements obtained using a particular measurement method with other values for which the measurement uncertainties are unknown, such as those obtained through certain complex theoretical calculations.

On the other hand, from Eq. (5) it is also possible to define a statistic for the identification of outliers, which can be a very useful tool for the improvement of the measurement method if the validity test has led to its rejection. A measured value y_i is defined as an outlier if the deviation from its expected value $y_i - \mu_i$ is significantly different from zero taking into account the standard uncertainty of that difference, $u(y_i - \mu_i)$. By dividing, for each data, that deviation by its standard uncertainty, the corresponding normalized deviation is defined, which follows, from (5), a standard normal distribution:

$$d_i = \frac{y_i - \mu_i}{u(y_i - \mu_i)} = \frac{y_i - \mu_i}{\sqrt{u^2(y_i) + u^2(\mu_i)}} \sim N(0,1) \quad (10)$$

Then, a hypothesis test based on the probability p of finding a $N(0,1)$ value greater than the observed normalized deviation d_i value can be established. This probability can be calculated from the $N(0,1)$ distribution:

$$p = P[|N(0,1)| > d_i] \quad (11)$$

If this probability is smaller than a certain value α , the hypothesis that the i th data is consistent with its theoretical expected value has to be rejected at a level of significance equal to α ; therefore, that data would be an outlier. Since for a standard normal distribution $P(|N(0,1)| > 2) = 5\%$, a measurement with $|d_i| > 2$ is identified as an outlier at a level of significance $\alpha = 5\%$.

As with the consistency test, the particular case in which $u(\mu_i) \simeq 0$ is considered leads to d_i values greater than those obtained with $u(\mu_i) \neq 0$ and, therefore, to a more conservative outlier identification test: less refined but usable.

Methodology and Metrological Considerations

The mathematical formulation presented in the previous section can be implemented in a systematic manner in just a few steps, as shown in Fig. 1. Once the measurements have been taken, the process simply involves comparing them by evaluating the chi-square statistic (Eq. 6) and, based on that, calculating the probability given in Eq. 7; its value determines whether or not the methods are consistent. Furthermore, calculating the normalized deviations of the data (Eq. 10) makes it possible to determine whether there is consistency on a data-point-by-data-point basis.

The cornerstone of this process is the use of measurement uncertainties. The uncertainty of a measurement is the final piece in the chain of traceability, which connects the measurement to a primary standard through an unbroken chain of comparisons. For example, for optical techniques of strain measurement traceability is established to the unit of length [29]. This final link in the traceability chain is ensured through the calibration of measuring instruments. In the calibration process biases are estimated and corrected, the repeatability and reproducibility of measurements are ensured, and the calibration uncertainty is determined (which is necessary for evaluating the uncertainties of subsequent measurements). Calibration requires standardized procedures, and the Versailles Project on Advanced Materials and Standards (vamas.org) has been working for decades on standardization for the application of a wide variety of techniques in solid materials.

Therefore, although the validation process described in this paper is simple to implement, it requires the use of calibrated equipment and the evaluation of measurement uncertainties. However, this is a requirement that should always be met, as stated in the ISO/IEC 17025:2005 standard [30]: "Testing laboratories shall have and shall apply procedures for estimating uncertainty of measurement". It is precisely

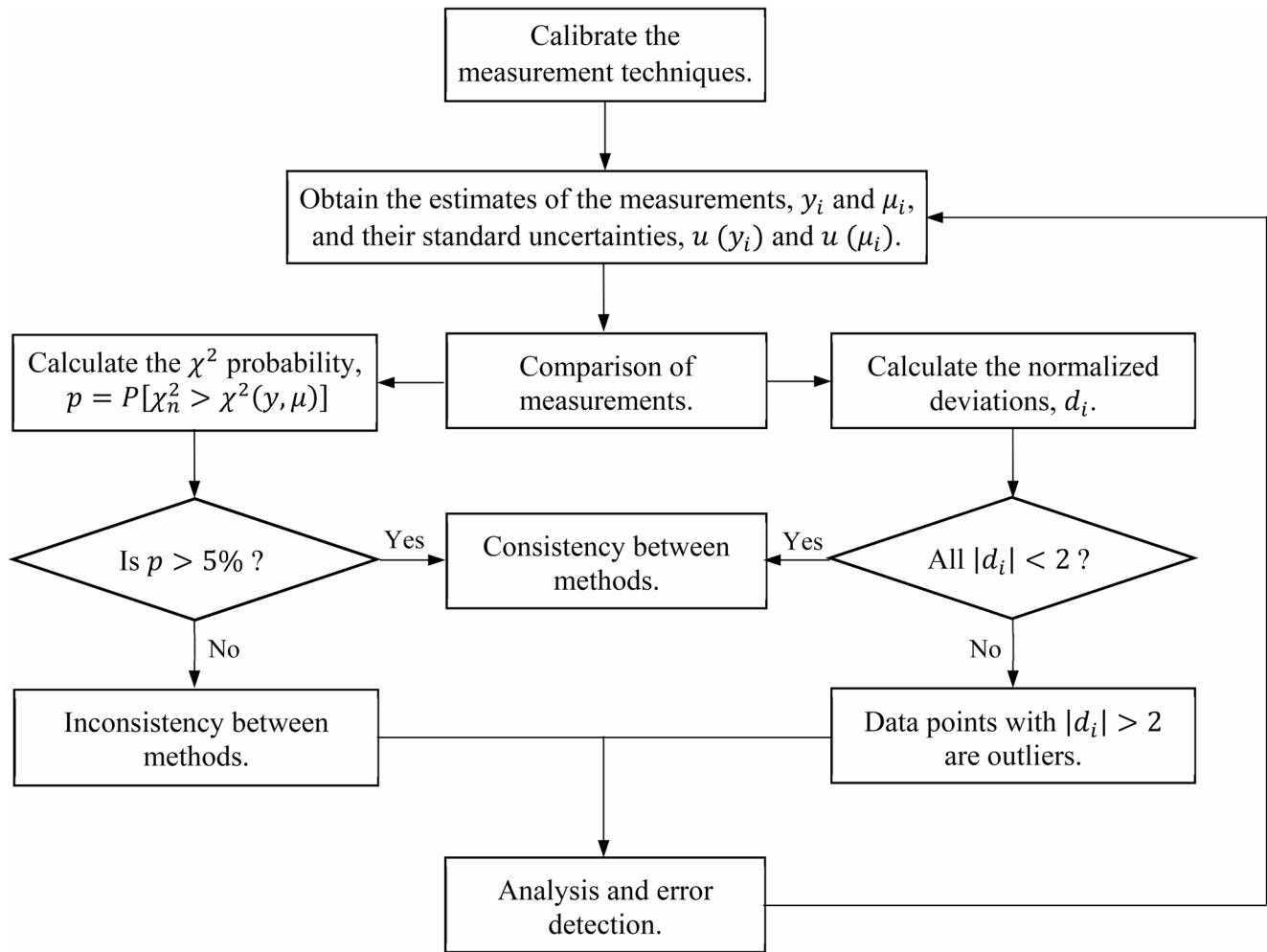


Fig. 1 Flowchart of the validation method

this connection between the measurements and the primary standards through the evaluation of measurement uncertainties that links the reliability of the proposed method with that of the rigorous and consolidated ISO GUM guide.

Results

As an example of the application of the validation method described below, it will be applied to a case of stress measurement by photoelasticity. The measuring equipment used is a calibrated [31–33] Vishay LF/Z-2 reflective polariscope, fitted with a Babinet compensator, which will be used to measure the isochromatic fringe order at points on the boundary of a piece. It is a simple and quite used measurement method, since it does not require a separation technique [34]. The results will be compared with those obtained theoretically by means of a Finite Element Method (FEM) calculation.

A 2.05 mm thick Vishay PS-1 photoelastic coating [35] was adhered to a 1.5 mm thick aluminum test piece, which can be easily loaded by means of a wing nut and a screw that pass through the 6 mm hole made in its back. Its shape and dimensions are indicated in Fig. 2. When the screw is tightened, the front face of the piece is subjected to a bending stress. In this face there are two holes, so that the stress distribution is not uniform, and gradients and stress concentration areas appear.

The piece was subjected to loading by tightening the loading screw and a displacement of 8.9 mm was measured between its ends. Then it was oriented perpendicular to the polariscope and illuminated with white light in order to measure the stress at a series of control points on the boundaries of the front face of the piece (Fig. 3). At a point on the boundary of a piece, the directions of the principal stresses are tangential and perpendicular to the boundary. The latter is null, whereas the tangential may be calculated from the value of the fringe order through the stress-optic law [36]:

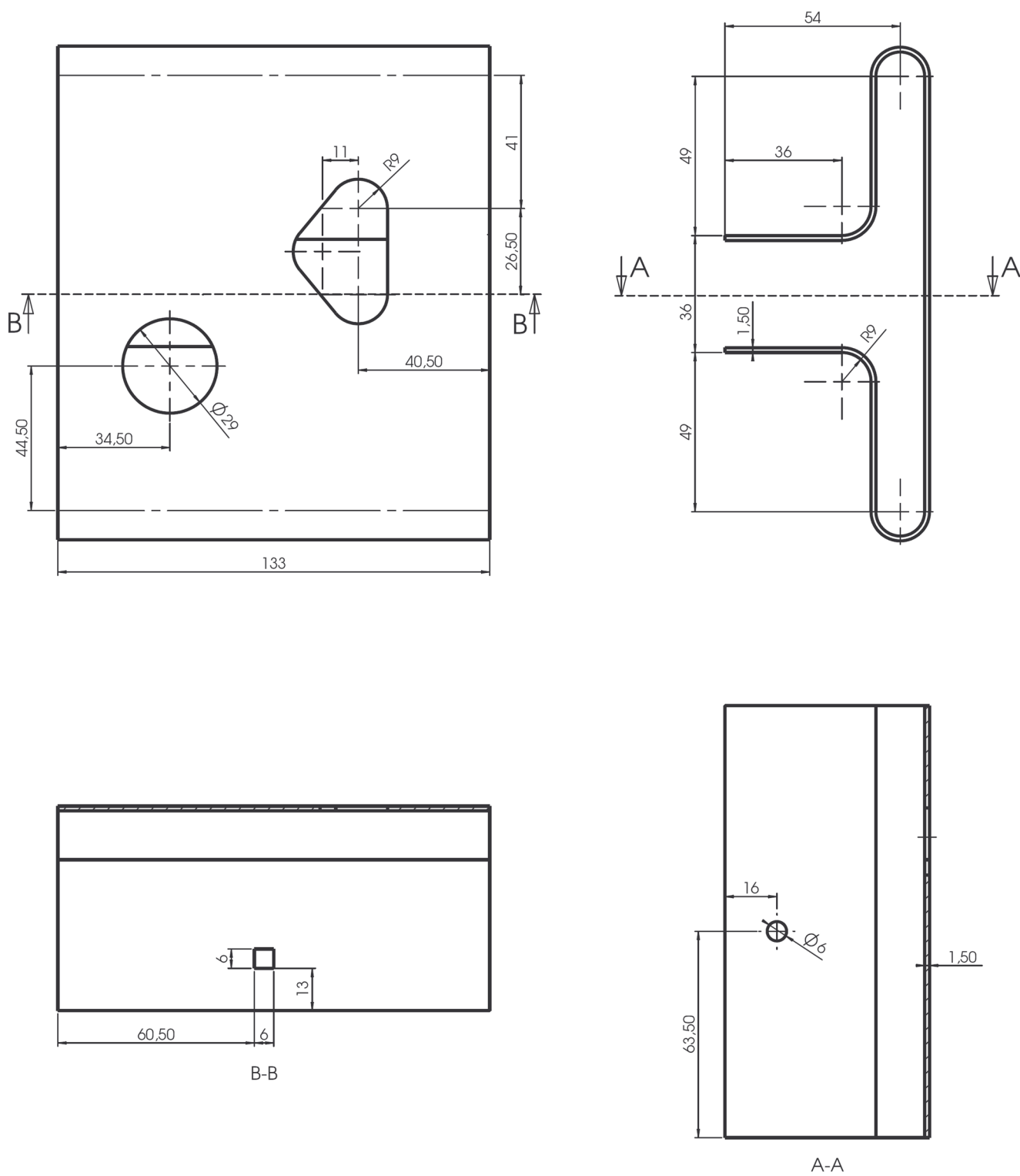


Fig. 2 Elevation and profile views of the piece to be analyzed, as well as two sections, A-A and B-B, to appreciate the rear holes used for the application of the load (dimensions in *mm*)



Fig. 3 Photograph of the experimental setup (the isochromatic fringe pattern observed through the polariscope is shown in the upper-left corner)

$$\sigma = \frac{E_s}{1 - \nu_s} \frac{N\lambda}{hK} \quad (12)$$

where E_s and ν_s are the Young's modulus and the Poisson's ratio, respectively, of the material of the specimen; N is the isochromatic fringe order measured with the compensator; λ is the average wavelength of the used light source; h is the width of photoelastic material crossed by the light; and K is the strain-optic coefficient of the photoelastic material.

Using the compensator, the fringe order N was measured with and without the load to eliminate the influence of parasitic or residual birefringence. After correcting the effects of the influence of the coating [37], the stresses

at the control points (Fig. 4), together with their respective measurement uncertainties, were calculated following the procedure set out in [38]. This calculation process requires, as established by the ISO GUM [27], preparing an uncertainty budget for each point, in which all quantities that contribute to the measurement uncertainty are considered. For each of them, the measurement estimate is calculated using Eq. (12) and its measurement uncertainty is evaluated using the law of propagation of uncertainty [27]. Data taken and calculations made to obtain the measurements at the control points are attached as supplementary material to this paper. Results are presented in Table 1.

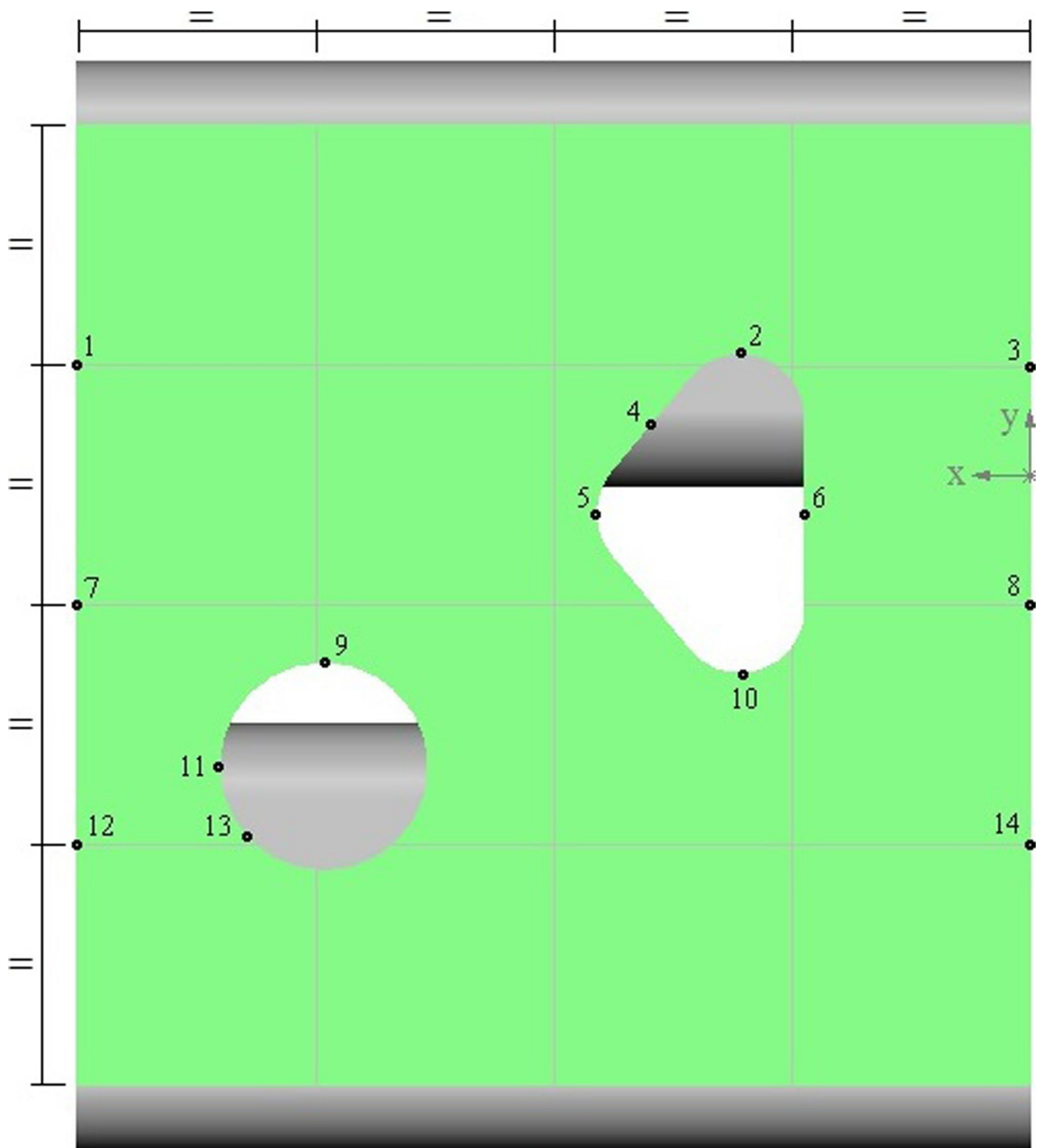


Fig. 4 Control points distributed on a grid on the flat surface of the test piece (the coordinate axes were taken at a rear corner of the piece)

Table 1 Coordinates of control points and stress measurements at them: stress estimation σ (tangential to the boundary) and its standard uncertainty $u(\sigma)$

Point	x (mm)	y (mm)	σ (MPa)	$u(\sigma)$ (MPa)
1	133	15.5	39.71	0.81
2	40.5	17	22.52	0.62
3	0	15.5	40.12	0.82
4	52.925	7.125	34.80	0.75
5	60.5	-5.25	70.83	1.24
6	31.5	-5.25	47.90	0.92
7	133	-18	40.53	0.82
8	0	-18	44.62	0.87
9	98.5	-26	15.56	0.56
10	40.5	-27.5	20.47	0.60
11	113	-40.5	72.05	1.26
12	133	-51.5	47.49	0.91
13	109.13	-50.36	42.17	0.84
14	0	-51.5	36.44	0.77

For the FEM analysis, the test piece was modelled using the 3D CAD software SolidWorks (Solidworks Corporation, solidworks.com), and its integrated finite element analysis module, Cosmosworks, was used for meshing and simulating its static behavior (Fig. 5). It is necessary to provide a series of data to the FEM program regarding the mechanical properties of the material, restrictions and loads. The material of the piece is aluminum, for which the Young's modulus and Poisson's ratio are taken as their estimated values, $E=73\text{GPa}$ and $\nu=0.33$ respectively. Regarding movement restrictions, it has taken that the cylindrical surface of the hole through which the load application screw is inserted is fixed (constraint on face: fixed surface). And as for loads, it has been taken that a constant downward pressure of 80 kPa acts on the surfaces of the square screw seat hole (pressure on face: pressure along the y-axis of -80000 Pa), which leads to a displacement of the ends of the piece of 8.9 mm , the same displacement as the one suffered by the real piece. To mesh the piece, the automatic mesh generator was configured to use high-quality meshing and parabolic tetrahedral elements. The resulting mesh consisted of 67,112 elements and 132,057 nodes, with a Jacobian check ratio of 4. The results obtained for each of the control points are summarized in Table 2.

Let's apply the validation process outlined in the previous section. In this case, the quantity Y_i measured is the component of the stress tangential to the boundary at each of the $n = 14$ control points, with estimate $y_i = \sigma_i$ and standard uncertainty $u(y_i) = u(\sigma_i)$, which has to be compared with its theoretically calculated values, $\mu_i = \sigma_{1i}$. As explained in Sect. 2 on mathematical formulation, given the complexity of calculating the uncertainty of values in a

FEM calculation², we will conservatively assume that the standard uncertainties $u(\mu_i) \simeq 0$. In Table 3, for each point, the first row contains the estimate of the photoelastic measurement, y_i ; the second row the corresponding standard uncertainty of that estimate, $u(y_i)$; the third row the estimate, μ_i , of the FEM value; the fourth row the corresponding standard uncertainty of that estimate, $u(\mu_i)$; the fifth row contains the normalized deviation, d_i ; and the sixth row the difference between both estimates expressed as a percentage, $\% \Delta = 100 \bullet (y_i - \mu_i) / \mu_i$. Note that this last row is not necessary for the proposed validation method, but it has been included in the table because this is the value that has traditionally been used to assess the consistency of the results. Although this percentage value provides a more intuitive way to quickly gauge the deviations between estimates obtained by one method or another, from a metrological standpoint it is the normalized deviations that allow us to determine, individually, whether the data are scientifically consistent, and the overall consistency of the data set is known using the chi-square statistic of Eq. (8).

In this case, the test-statistic (8) provides a value $\chi^2(y, \mu) = 12.8$. If both groups of data were consistent the test-statistic will follow a chi-squared distribution with $n = 14$ degrees of freedom. The probability that this value comes from a χ^2 distribution with 14 degrees of freedom is $p = P[\chi^2_{14} > 12.8] = 54.2\%$, greater than $\alpha = 5\%$, so the data are consistent with their theoretical values. This means that the measurement technique cannot be rejected with a 95% level of confidence: it is consistent with the theory on which the finite element analysis is based.

Discussion

The validation test has consisted of comparing experimental measurements with their expected values. If data are assumed to follow normal distributions and both groups of data were consistent, the random variable $\chi^2(y, \mu)$ given by (6) or (8) will follow a chi-square distribution with n degrees of freedom. Then it would be expected that the $\chi^2(y, \mu)$ value obtained was close to its expected value in case of consistency, which for a χ^2_n is n . The probability that, having consistency, a larger $\chi^2(y, \mu)$ value than

² Currently, MEF software do not include uncertainty evaluation, as this is a computationally infeasible process at present; it would require a Monte Carlo calculation involving thousands of simulations with random variations in all inputs (geometry, material behavior and properties, boundary conditions, etc.) to observe the variability of the results, while also accounting for discretization and truncation errors in the calculation.

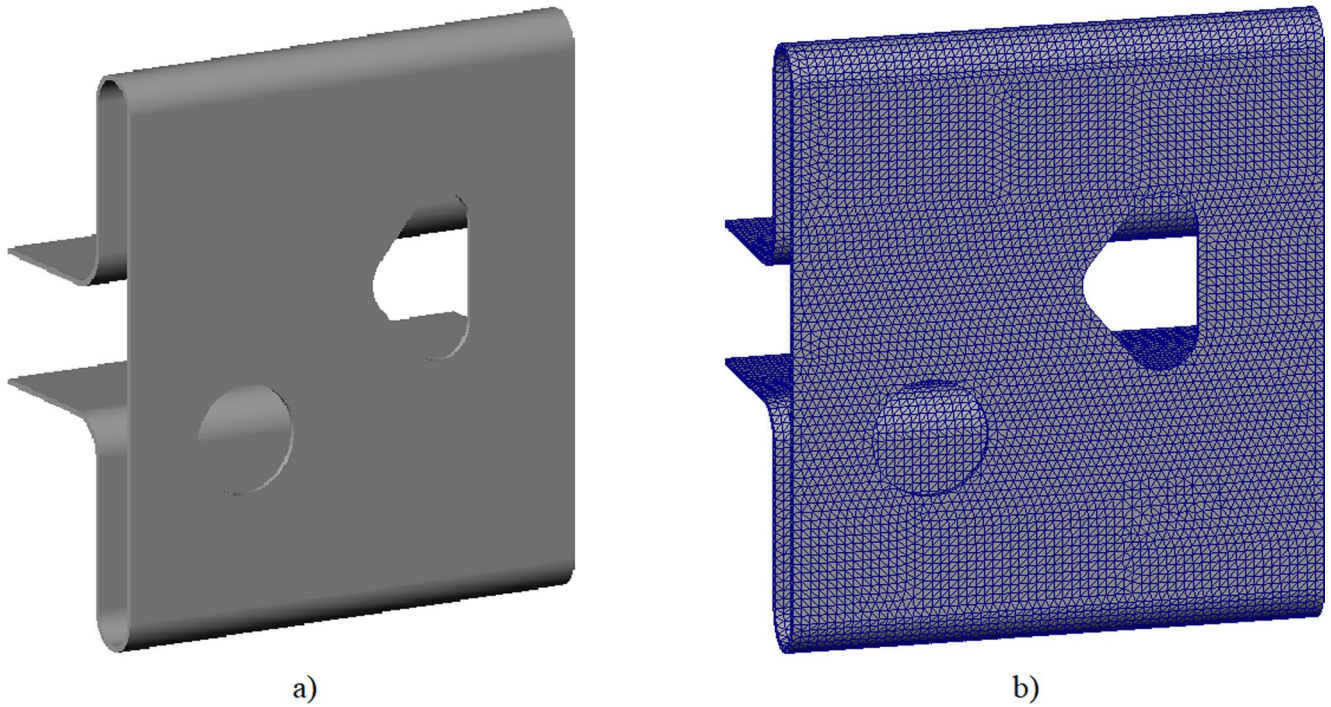


Fig. 5 (a) Model of the test piece created in Solidworks software and (b) mesh generated by the Cosmosworks module for its FEM analysis

Table 2 Results of the FEM analysis at control points: orientation θ of the major principal stress from the horizontal axis x ($^{\circ}$), normal and shear stress components σ_x , σ_y and τ_{xy} ; and principal stresses σ_1 and σ_2

Point	θ (degrees)	σ_x (MPa)	σ_y (MPa)	τ_{xy} (MPa)	σ_1 (MPa)	σ_2 (MPa)
1	90.00	0.00	38.99	0.00	38.99	0.00
2	0.00	22.46	0.00	0.00	22.46	0.00
3	90.00	0.00	41.09	0.00	41.09	0.00
4	60.00	8.92	26.00	15.44	35.66	0.00
5	90.00	0.00	72.24	0.00	72.24	0.00
6	90.00	0.00	46.38	0.00	46.38	0.00
7	90.00	0.00	41.14	0.00	41.14	0.00
8	90.00	0.00	43.94	0.00	43.94	0.00
9	0.00	15.98	0.00	0.00	15.98	0.00
10	0.00	20.00	0.00	0.00	20.00	0.00
11	90.00	0.00	72.45	0.00	72.45	0.00
12	90.00	0.00	46.63	0.00	46.63	0.00
13	-45.00	21.38	21.38	-21.38	42.75	0.00
14	90.00	0.00	37.37	0.00	37.37	0.00

expected is obtained is given by p . Since this probability has been greater than the significance level α , the measurement method has been accepted as valid because it cannot be rejected the hypothesis that both groups of data are consistent.

Regarding this test it is necessary to make some comments:

- First, as happens in all hypothesis tests, it emphasizes the rejection rather than the acceptance of the null

hypothesis. However, the structure of the test justifies acceptance of the null hypothesis (consistency of data) when the alternative hypothesis (inconsistency of data) is rejected, because specific alternative hypotheses with different power cannot be formulated. Consistency and inconsistency are complementary concepts: the non-existence of one implies, to a greater or lesser degree, the existence of the other.

- Secondly, the consistency is confirmed by the normalized deviations. These can be evaluated through (10)

Table 3 Comparison between measurements and theoretically calculated values for the stress at control points

	σ_1 (MPa)	σ_2 (MPa)	σ_3 (MPa)	σ_4 (MPa)	σ_5 (MPa)	σ_6 (MPa)	σ_7 (MPa)
y_i	39.71	22.52	40.12	34.80	70.83	47.90	40.53
$u(y_i)$	0.81	0.62	0.82	0.75	1.24	0.92	0.82
μ_i	38.99	22.46	41.09	35.66	72.24	46.38	41.14
$u(\mu_i)$	0.00	0.00	0.00	0.00	0.00	0.00	0.00
d_i	0.89	0.10	-1.18	-1.15	-1.14	1.65	-0.74
% Δ	1.85%	0.27%	-2.36%	-2.41%	-1.95%	3.28%	-1.48%
	σ_8 (MPa)	σ_9 (MPa)	σ_{10} (MPa)	σ_{11} (MPa)	σ_{12} (MPa)	σ_{13} (MPa)	σ_{14} (MPa)
y_i	44.62	15.56	20.47	72.05	47.49	42.17	36.44
$u(y_i)$	0.87	0.56	0.60	1.26	0.91	0.84	0.77
μ_i	43.94	15.98	20.00	72.45	46.63	42.75	37.37
$u(\mu_i)$	0.00	0.00	0.00	0.00	0.00	0.00	0.00
d_i	0.78	-0.75	0.78	-0.32	0.95	-0.69	-1.21
% Δ	1.55%	-2.63%	2.35%	-0.55%	1.84%	-1.36%	-2.49%

and they express the consistency of each measurement with its theoretical expected value. Normalized deviations are shown in Table 3 under the row heading d_i , and it can be seen that all of them are less than two, which is indicative that all data are consistent individually, that is, there are no outliers. The non-existence of outliers is one more evidence that confirms the validity of the measurement method.

- Thirdly, the assumption that data follow normal probability distributions is always fulfilled quite closely. As already mentioned before, in the GUM's context, the probability distribution for measurements is the t-distribution, which approaches the normal distribution as the number of degrees of freedom ν grows. In addition, the validation test is robust with respect to this condition of normality, because a high value of $\chi^2(y, \mu)$ indicates inconsistency regardless of the probability distribution of the data.
- Finally, it's important to emphasize that the purpose of a validation process is to verify whether there is agreement between the two methods used in the comparison. In the example from the previous section, the consistency between the data obtained using photoelasticity and the results of a FEM analysis has been confirmed. Thus, one could say that the experimental measurement method has been validated by the FEM calculation, or that the FEM calculation has been validated by the experimental measurements. In reality, the most accurate statement would be that the experimental and theoretical results are consistent.

The validation method described above, as well as the tool for the identification of outliers, demonstrate the importance of expressing measurement with their corresponding uncertainties. Although this is mandatory for the measurements to meet the validity, comparability and reproducibility

requirements that the scientific method requires [27, 31], the literature review carried out for this investigation reveals that it is rare to find research papers, especially in solid mechanics, in which measurement uncertainties are reported. It would be convenient that the scientific community, and in particular the editorial teams of the research reviews, to be aware of this situation and act accordingly.

Conclusions

Although it seems intuitive to validate the results of a measurement process by comparing measurements obtained with their expected values through a graph or table, this direct comparison lacks scientific rigor. For this, it is necessary that the comparison method be carried out by means of a statistical test that supports the assertion that both groups of data are consistent or not, and that takes measurement uncertainties into account. The combination of these two factors ensures the metrological rigor of the comparison. The method exposed simply consists in calculating a chi-square statistic from the measurements, their expected values and their corresponding uncertainties, and comparing it with its expected value in case of consistency. A similar test can be applied individually to each measurement, as a tool for identifying outliers, through the statistics known as normalized deviations.

In this paper, the method has been applied to a case comparing experimental measurements of linear static stress using photoelasticity with theoretical results obtained by finite element analysis. This is merely an example, as its scope of application is much broader. It could be applied to fatigue analysis, dynamic response analysis, contact interaction analysis, nonlinear behavior analysis, and so on; and not only in solid mechanics measurements, but in any other branch of science. In all cases, the method would yield a

comparison table similar to Table 3, a value for the chi-squared statistic, and values for the standardized deviations similar to those in the example used, all of which would allow for an assessment of the consistency of the results.

The validation method requires expressing the uncertainties of the measurements, which is mandatory in any measurement process. However, the literature review conducted in this study has shown that the validation methods used to date did not take measurement uncertainties into account. This indicates that there is still some permissiveness in scientific journals when publishing articles in which measurement uncertainties are not expressed or measurement techniques are not correctly validated. We hope that this paper serves to correct this lack of rigor when measurement techniques in solid mechanics are used or new ones are proposed.

Supplementary Information The online version contains supplementary material available at <https://doi.org/10.1007/s40799-026-00896-8>.

Funding Open access funding provided by FEDER European Funds and the Junta de Castilla y León under the Research and Innovation Strategy for Smart Specialization (RIS3) of Castilla y León 2021–2027.

Declarations

The author has no financial or proprietary interests in any material discussed in this article. Open access funding provided by FEDER European Funds and the Junta De Castilla y León under the Research and Innovation Strategy for Smart Specialization (RIS3) of Castilla y León 2021–2027.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

References

- Sanford RJ, Dally JW (1979) A general method for determining mixed-mode stress intensity factors from isochromatic fringe patterns. *Eng Fract Mech* 11(4):621–633. [https://doi.org/10.1016/0013-7944\(79\)90123-1](https://doi.org/10.1016/0013-7944(79)90123-1)
- Diaz FA, Patterson EA, Siegmund P (2010) A novel experimental approach for calculating stress intensity factors from isochromatic data. *Exp Mech* 50(2):273–281. <https://doi.org/10.1007/s11340-009-9282-1>
- Sanford RJ (1989) Determining fracture parameters with full-field optical methods. *Exp Mech* 29(3):241–247. <https://doi.org/10.1007/BF02321401>
- Yoneyama S, Morimoto Y, Takashi M (2006) Automatic Evaluation of Mixed-mode Stress Intensity Factors Utilizing Digital Image Correlation. *Strain* 42(1):21–29. <https://doi.org/10.1111/j.1475-1305.2006.00246.x>
- Yates JR, Zanganeh M, Tai YH (2010) Quantifying crack tip displacement fields with DIC. *Eng Fract Mech* 77(11):2063–2076. <https://doi.org/10.1016/j.engfracmech.2010.03.025>
- Ayatollahi MR, Nejati M (2011) An over-deterministic method for calculation of coefficients of crack tip asymptotic field from finite element analysis. *Fatigue Fract Eng Mater Struct* 34(3):159–176. <https://doi.org/10.1111/j.1460-2695.2010.01504.x>
- Zhang R, He L (2012) Measurement of mixed-mode stress intensity factors using digital image correlation method. *Opt Lasers Eng* 50(7):1001–1007. <https://doi.org/10.1016/j.optlaseng.2012.01.009>
- Shukla A, Nigam H (1985) A numerical-experimental analysis of the contact stress problem. *J Strain Anal* 20(4):241–245. <https://doi.org/10.1243/03093247V204241>
- Pankhawalla VA, Ramesh K, Dixit PM (1998) Evaluation of contact stress parameters by digital photoelasticity. In: Kalyanaraman, V., Sriram, P. (eds.) *Proceedings of 8th National Seminar on Aerospace Structures*, pp. 284–291. Allied Publishers Limited, New Delhi
- Krouskop TA, Dougherty DR, Vinson FS (1987) A pulsed Doppler ultrasonic system for making noninvasive measurements of the mechanical properties of soft tissue. *J Rehabil Res Dev* 24(2):1–8
- Sebastiani M, Bemporad E, Carassiti F, Schwarzer N (2011) Residual stress measurement at the micrometer scale: Focused ion beam (FIB) milling and nanoindentation testing. *Philos Mag* 91:1121–1136. <https://doi.org/10.1080/14786431003800883>
- Korsunsky AM, Sebastiani M, Bemporad E (2009) Focused ion beam ring drilling for residual stress evaluation. *Mater Lett* 63(22):1961–1963. <https://doi.org/10.1016/j.matlet.2009.06.020>
- Patel BD, Srinivas AR (2012) Validation of Experimental Strain Measurement Technique and Development of Force Transducer. *Int. J Eng Res* 3(10):1–4
- Di Maio D, Roberts S (2005) Measuring fracture toughness of coatings using focused-ion-beam-machined microbeams. *J Mater Res* 20(2):299–302. <https://doi.org/10.1557/JMR.2005.0048>
- Espinosa HD, Peng B (2005) A new methodology to investigate fracture toughness of freestanding MEMS and advanced materials in thin film form. *J Microelectromechanical Syst* 14:153–159. <https://doi.org/10.1109/jmems.2004.839013>
- McNeill SR, Peters WH, Sutton MA (1987) Estimation of stress intensity factor by digital image correlation. *Eng Fract Mech* 28(1):101–112. [https://doi.org/10.1016/0013-7944\(87\)90124-X](https://doi.org/10.1016/0013-7944(87)90124-X)
- Pastrama SD, Moreira PMGP, Castro PMST, Jiga G (2008) An overdeterministic algorithm for stress intensity factor calculation using finite element results. In: 11th Portuguese Conference on Fracture - PCF, Lisbon, Portugal
- Ramesh K, Gupta S, Kelkar AA (1997) Evaluation of stress field parameters in fracture mechanics – revisited. *Eng Fract Mech* 56(1):25–45. [https://doi.org/10.1016/S0013-7944\(96\)00098-7](https://doi.org/10.1016/S0013-7944(96)00098-7)
- Patil P, Vyasrayani CP, Ramji M (2017) Linear least squares approach for evaluating crack tip fracture parameters using isochromatic and isoclinic data from digital photoelasticity. *Opt Lasers Eng* 93:182–194. <https://doi.org/10.1016/j.optlaseng.2017.02.003>
- Tsakalakos T (1981) The bulge test: A comparison of the theory and experiment for isotropic and anisotropic films. *Thin Solid Films* 75(3):293–305. [https://doi.org/10.1016/0040-6090\(81\)90407-7](https://doi.org/10.1016/0040-6090(81)90407-7)
- Harilal R, Vyasrayani CP, Ramji M (2015) A linear least squares approach for evaluation of crack tip stress field parameters using DIC. *Opt Lasers Eng* 75:95–102. <https://doi.org/10.1016/j.optlaseng.2015.07.004>

22. Oliver W, Pharr G (1992) An improved technique for determining hardness and elastic modulus using load and displacement sensing indentation experiments. *J Mater Res* 7(6):1564–1583. <https://doi.org/10.1557/JMR.1992.1564>
23. Sebastiani M, Johanns KE, Herbert EG, Carassiti F, Pharr GM (2015) A novel pillar indentation splitting test for measuring fracture toughness of thin ceramic coatings. *Philos Mag* 95(16–18):1928–1944. <https://doi.org/10.1080/14786435.2014.913110>
24. Pearson K (1900) On the criterion that a given system of deviations from the probable in the case of a correlated system of variables is such that it can be reasonably supposed to have arisen from random sampling. *Lond Edinb Dublin Philosophical Magazine J Sci* 50(302):157–175. <https://doi.org/10.1080/14786440009463897>
25. Fisher RA (1922) On the Interpretation of χ^2 from Contingency Tables, and the Calculation of p. *J Roy Stat Soc* 85(1):87–94. <https://doi.org/10.2307/2340521>
26. Schiffrmann KI (2011) Determination of fracture toughness of bulk materials and thin films by nanoindentation: comparison of different models. *Philos Mag* 91(7–9):1163–1178. <https://doi.org/10.1080/14786435.2010.487984>
27. International Organization for Standardization (2008) ISO 98–3:2008 Evaluation of measurement data – Guide to the expression of Uncertainty in Measurement. ISO, Geneva
28. Manderscheid LV (1965) Significance Levels. 0.05, 0.01, or? *J Farm Econ* 47(5):1381–1385. <https://doi.org/10.2307/1236396>
29. Burguete R, Hack E, Kujawińska M, Patterson E, Saleem Q, Siebert T, Whelan M (2006) Standards project for optical techniques of strain measurement: Final report. *Proc 2006 SEM Annual Conf Exposition Experimental Appl Mech* 3:1486–1492
30. International Organization for Standardization (2005) ISO/IEC 17025:2005 General requirements for the competence of testing and calibration laboratories. ISO, Geneva
31. VAMAS TWA26, Draft (2007) VAMAS TWA26 pre-standard Calibration and assessment of optical strain measurements systems – Part I: Reference material for optical methods of strain measurement. University of Sheffield
32. VAMAS TWA26 (2007) Draft VAMAS TWA26 pre-standard Calibration and assessment of optical strain measurements systems – Part II: Standardised test materials for optical methods of strain measurement systems. University of Sheffield
33. VAMAS TWA26 (2007) Draft VAMAS TWA26 pre-standard Calibration and assessment of optical strain measurements systems – Part III: Good Practice Guide to (1) Geometric Moiré for in-plane displacement/strain analysis, (2) Grating (moiré) interferometry for in-plane displacement/strain analysis. Image correlation for in-plane displacement/strain analysis, (4) Thermoelastic stress analysis, (5) Reflection photoelasticity, (6) Transmission photoelasticity, (7) Electronic Speckle pattern interferometry for displacement/strain analysis. University of Sheffield
34. Solaguren-Beascoa Fernández M, Alegre Calderón JM, Bravo Díez PM, Cuesta Segura II (2010) Stress-separation techniques in photoelasticity: A review. *J Strain Anal Eng Des* 45(1):1–17. <https://doi.org/10.1243/03093247JSA583>
35. Vishay Measurements Group Inc. (2015) Bulletin S-116-H: PhotoStress Coating Materials and Adhesives
36. Vishay Measurements Group Inc. (2011) Technical Note 702-2: Introduction to stress analysis by the PhotoStress method
37. Vishay Measurements Group Inc. (2015) Technical Note 706-1: Corrections to photoelastic coating fringe-order measurements
38. Solaguren-Beascoa Fernández M (2010) Towards uncertainty evaluation in photoelastic measurements. *J Strain Anal Eng Des* 45(8):575–585. <https://doi.org/10.1177/030932471004500806>

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.