IMEKO 24th TC3, 14th TC5, 6th TC16 and 5th TC22 International Conference
11 – 13 October 2022, Cavtat-Dubrovnik, Croatia

AN INTERACTIVE APPLICATION FOR COMPARISON OF FIT APPROACHES IN BALANCES PRESSURE CALIBRATION

M. Morales[1], S. Higuera[2]

Instituto Nacional de Metrología, Colombia, [1] mmorales@inm.gov.co, [2] sjhiguera@inm.gov.co

Abstract:
The calibration of a pressure balance implies to determine the values of the zero-pressure effective area A_0 and the pressure distortion coefficient λ that defines the piston-cylinder assembly effective area A_{pi}. In order to analyze this behavior, this work compares the method of ordinary least squares with the generalized least squares and weighted least squares to calculate the contributions to type A uncertainty, in particular the residual deviation. For this purpose, an interactive application has been built in R- Shiny, which allows for getting the results mentioned.

Keywords: Ordinary least squares (OLS); weighted least squares (WLS); generalized least squares (GLS); pressure balances; interactive application

1. INTRODUCTION

Pressure balances are widely used to calibrate pressure sensing devices and to realise the SI unit of pressure, the Pascal. The calibration procedure of a piston gauge is known as "**crossfloat**" and means the determination of its effective area as a function of pressure. The piston gauge effective area is obtained from the fit of the calibration results to a function of the pressure. For example, when we have a linear function of pressure $y = a \cdot p + b$ the effective area at zero pressure will be $A_0 = b$ and the distortion coefficient $\lambda = a/b$, then the effective area in function of the pressure will be $A_{\text{pi}} = A_0(1 + \lambda \cdot p_i)$. Now, there are several ways to make the fit of the data, the simplest and easy to use is the Ordinary Least Squares (OLS), and others are the Weighted Least Squares (WLS), and Generalized Least Squares (GLS) the differences between the regression schemes have been analysed in several works [1].
One of the key assumptions of the Ordinary Least Squares method is that the residuals are distributed with equal variance at each level of the predictor variable; this assumption is known as homoscedasticity. In the case of effective area estimation A_{pi} and the pressure distortion coefficient λ, these fit assumptions could affect the contribution to statistical uncertainty and to increase the residual standard error, especially in the lowest pressure points, because these points deteriorate the determination of the fitting due to their higher uncertainties [2]. For this, a comparison of the regression methods mentioned is approached, focused on the comparison of the standard deviations of the estimators and the residual error of the regression, it is highlighted WLS and GLS methods, and we evaluate their validity and applicability.

Due to the inclusion of residual analysis and other regression techniques to ensure that the right model is established, a specific issue with data processing in Excel has been identified. In order to make many of the calculations and definitions of interest easier to develop, we have created a web application-type tool. This application incorporates various features of the very powerful R programming language designed for statistical calculation and high-quality graphical representations.

2. DESCRIPTION OF THE WORK

Calculations for effective area A_{pi} and pressure distortion coefficient λ were reproduced for an oil operated 98 mm^2 piston-cylinder assembly Desgranges et Huot, Range: from 0.1 MPa to 10 MPa [3].

We used R-Shiny, which is an open source R package that provides an elegant and powerful web framework for building web applications using R for Windows R version 4.2.0 (2022-04-22 ucrt); we chose R because it is a highly flexible and powerful tool for analysing and visualizing data, then, Shiny is the perfect companion to R, making it quick and simple to share analysis and graphics from R that users can interact, [4].

In this case, our goal is to obtain the study results of the pressure balance calibration findings. The application has been designed for linear regression analysis for any pair of numerical variables.

3. APPLICATION OVERVIEW

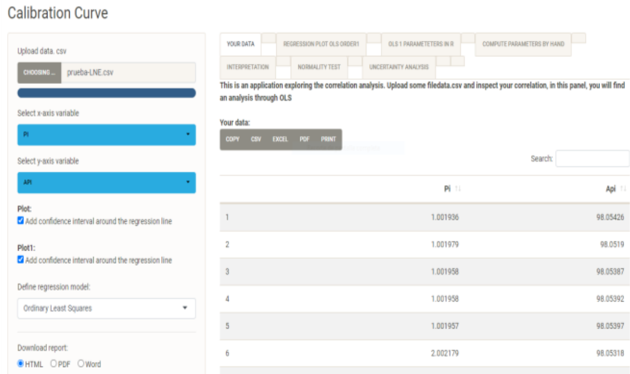

Figure 1: Data loading and data preview.

The home page of the developed app is displayed in Figure 1. Data of pressure and the area are uploaded in CSV format; the user must choose the independent and dependent variable.

To analyse the results, three distinct modules were developed. The main objective of this article is to show how the application works in the calibration of pressure balances.

3.1. Ordinary Least Square Module (OLS)

Once the data is loaded on the application home page, the user can find in the next tab a graphical result with the linear adjustment using the ordinary least squares method, the function used in R is lm (), it is used to fit linear models to data frames in the R Language. It can be used to carry out regression, single stratum analysis of variance, and analysis of covariance to predict the value corresponding to data that is not in the data frame. Figure 2 presents the linear fit $A_{\mathrm{pi}}(p_i)$ with confidence interval around the regression line; the R plotting package ggplot2 has an awesome function called "stat_smooth" for plotting a regression line (or curve) with the associated confidence band.

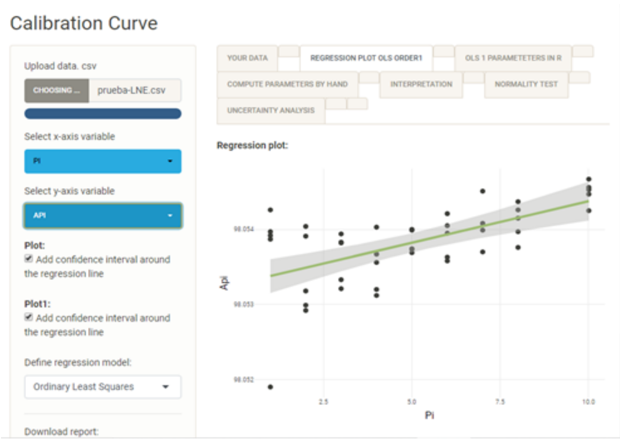

Figure 2: Regression line display with confidence interval around the regression line.

The percentile for the confidence interval (should fall between 0 and 1). The default is 0.95, which corresponds to a 95 percent confidence interval.

This OLS method is the most commonly used by the metrology laboratories because it is simple to implement. It assumes that the uncertainty of the effective areas is unknown and constant (homoscedasticity) [2]. However, this method is not adapted because the uncertainties $u(Ap)$ are known and vary with the pressure.

Calibration of pressure balances at the highest metrological level consists in determining the effective areas A_p of the piston-cylinder assembly as a function of the pressure. The values of these effective areas are then analysed to determine the effective area at null pressure A_0 and the distortion coefficient λ if it is significant, [5].

The coefficients are calculated taking into account a linear model:

$$A_{pi} = (A + B\, p_i)\,. \tag{1}$$

The equation can also be written:

$$A_{pi} = A_0(1 + \lambda\, p_i)\,. \tag{2}$$

Comparing equations (1) and (2):

$$A = A_0 \tag{3}$$

$$\lambda = \frac{B}{A_0}\,. \tag{4}$$

$A_{\mathrm{pi}}, A_0, \lambda$ type A uncertainties are taken as data deviations from A_{pi}, with respect to curve fitting are calculated from variances $V(A_0), V(B)$ and covariance $\mathrm{cov}(A_0, B)$. The problem with the OLS method focused on the fact that it is assumed that there is no correlation between the analysed variables, results, can be found in Table 1.

Theoretically, it is then assumed that the errors are uncorrelated with each other and have equal variance [2]. The application has a residual analysis with graphs and an assumption test on residual normality. Figure 3 allows for analysis if the model built $A_{\mathrm{pi}}(p_i)$ is the indicated one, also allows for visualizing abnormal or extreme observations, which can greatly affect the final result; the residuals behaviour for the lower pressure points shows that the variance is not the same throughout the pressure interval, as well as the detection of outliers that can affect the regression results.

The OLS algorithm has assumptions that are regularly ignored or taken for granted without prior verification [6]. The application performs a verification of most of these assumptions.

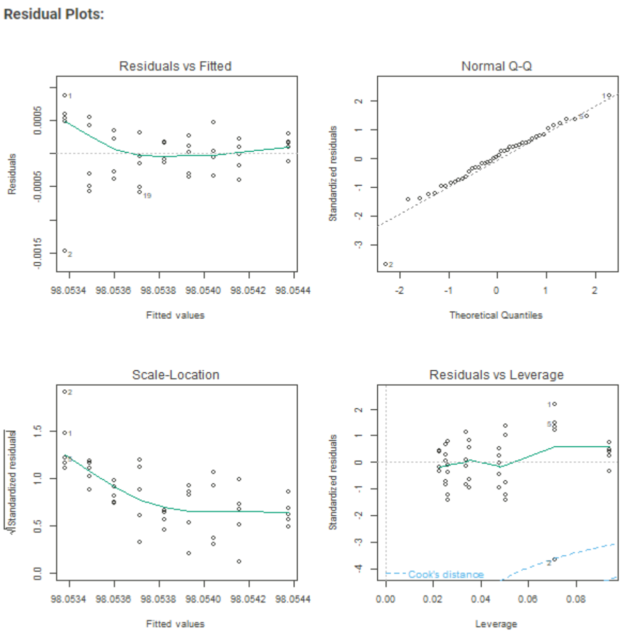

Figure 3: Residual analysis from module 1

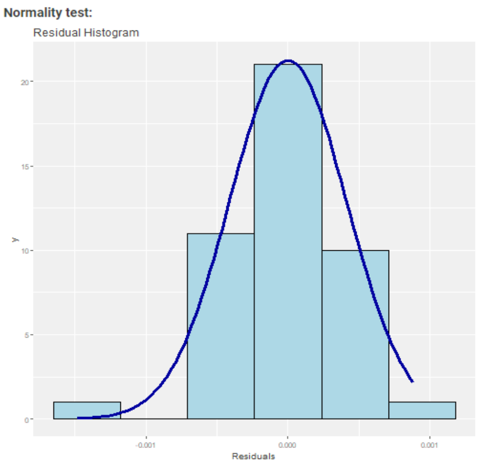

Figure 4: Residual Histogram to verify residuals normality hypothesis

Figure 4 verifies whether the residuals could be fitted to a normal distribution whose mean should be around zero, the fit shows a typical Gaussian curve of the normal distribution. On the other hand, the user will find the Shapiro Wilk test for normality.

Finally, in this module the linear fit of Figure 5 was constructed, the distance between the observed points and the estimated points on the line can be estimated.

For this work we want to highlight the comparison that has been developed regarding the results of other regression methods, therefore, we present the results directly related to the formalisms implemented in the R-Shiny web application. The remaining modules implemented in the application contain the results of the WLS and GLS regression methods investigated.

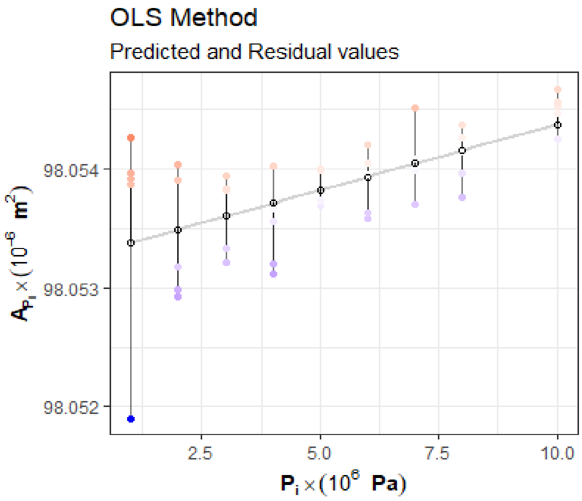

Figure 5: Predicted and Residual values from OLS method

3.2. Weighted Least Square Module (WLS)

Weighted least squares regression is useful for estimating the values of model parameters when the response values have differing degrees of variability over the combinations of the predictor values.

It is adopted to take into account the fact that the effective areas have different uncertainties [2].

The weight is related to minimize a weighted sum of the squared residuals, in which each squared residual is weighted by the reciprocal of its variance [7]. The weighted least squares method, derived from OLS, is adopted by some laboratories, mainly NMIs, to account for the change in uncertainty $u(Ap_i)$ as a function of pressure [6].

We use the term w_i for the weighting of each pair of measured values (A_{pi}, p_i)

$$w_i = \frac{1}{(\sigma_{A_{Pi}})^2}. \tag{5}$$

To find the $\sigma_{A_{Pi}}$ values:

Fit an OLS model, calculate fitted values from a regression of absolute residuals vs fitted values, fit a WLS model using weights:

$$w_i = \frac{1}{(fitted\ values)^2}.$$

Since each weight is inversely proportional to the error variance, it reflects the information in that observation so, an observation with small error variance has a large weight, since it contains relatively more information that an observation with large error variance.

With respect to the calculation of the regression coefficients, the weighting factors are introduced. Residual variance is calculated and standard deviations from the intercept A_0 and slope B according to the following equations:

$$\sigma^2 = \frac{\sum w_i (A_0 + BP_i - A_{\mathrm{pi}})^2}{n-2}. \tag{6}$$

$$\sigma_B = \sqrt{\frac{\sigma^2}{\sum w_i (p_i - \tilde{p})^2}} \quad (7)$$

$$\sigma_{A_0} = \sqrt{\frac{\sigma^2 \sum w_i p_i^2}{n \sum w_i (p_i - \tilde{p})^2}} \quad (8)$$

with:

$$\tilde{p} = \frac{\sum w_i p_i}{n}. \quad (9)$$

The linear fit with weighted least squares is shown in Figure 6.

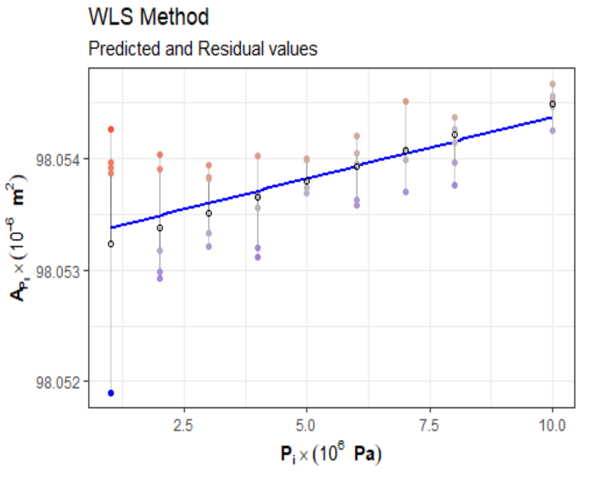

Figure 6: Predicted and Residual values from WLS method

3.3. Generalized Least Square Module (GLS)

It is a method which takes the uncertainties and the correlation of the effective areas into account. The function weight to be minimized is related to a sum of squared residuals, from the OLS method [7].

The generalized least squares method, adopted by few laboratories because of its complexity, considers the set of variances and covariances associated with areas A_{pi}. The weight attributed to each point (A_{pi}, p_i) comes from a decomposition of the variance-covariance matrix of the areas A_{pi} *[6]*.

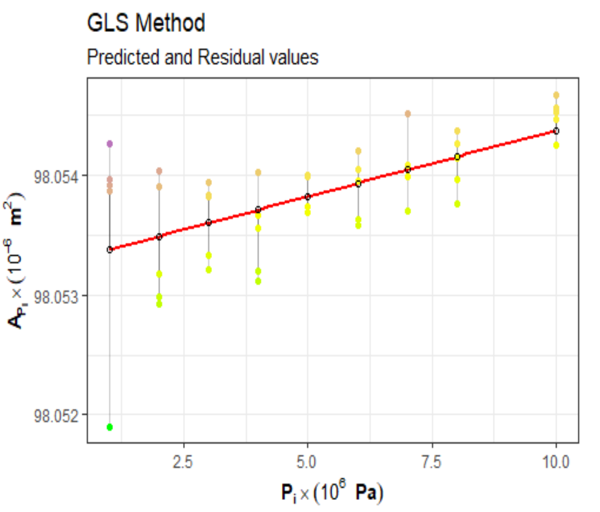

Figure 7: Predicted and Residual values from GLS method

In the application build in R, we used the fgls() function: Jointly estimates the fixed-effects coefficients and residual variance-covariance matrix in a generalized least square model by maximizing the (multivariate-normal) likelihood function, via "optim" in the Rbase distribution [8].

The residual variance-covariance matrix is block diagonal sparse, constructed with "bdsmatrix" from the bds matrix package [8].

The linear fit with generalized least squares is shown in Figure 7.

4. ANALYSIS

We considered the residual deviation (RSE), which stands for the equation's measurement of uncertainty

$$RSE = \sqrt{\frac{1}{n-2} \sum (Ap_i - A\widehat{p_i})^2} \; . \quad (10)$$

Table 1: Results and comparison using the OLS, WLS and GLS methods.

	OLS	WLS	GLS
A_0 (m^2)	9.805327E-6	9.805310E-6	9.805327E-6
$u(A_0)$ (m^2)	1.309E-10	1.641E-10	1.309E-10
λ (MPa^{-1})	1.130E-6	1.418E-6	1.130E-6
$u(\lambda)$ (MPa^{-1})	2.307E-7	1.759E-7	2.307E-7
RSE (m^2)	4.177E-10	1.767E-10	1.704E-10

Calculations for effective area A_{Pi} and pressure distortion coefficient λ were reproduced from the calibration certificate for an oil operated 98 mm^2 piston-cylinder assembly Desgranges et Huot, Range: From 0.1 MPa to 10 MPa. The results presented in Table 1 show that the estimated effective areas at zero pressure A_0 are consistent with the associated uncertainties $u(A_0)$ for the three methods used. However, the OLS and WLS method present the same deviations and uncertainties, being for this case the WLS method that presents greater uncertainty $u(A_0)$. The GLS method shows satisfactory results, the three methods show similar relative uncertainties.

Concerning the distortion coefficient λ, the results presented in Table 1 show a good result for the OLS method: the deviation from the true value is the same for GLS method, the other method WLS exhibits similar results for the determination of this parameter and the estimation of its uncertainty, but it results a better uncertainty $u(\lambda)$.

For the uncertainty due to the fitting, the GLS adjustment has been better, however, all three methods present uncertainties of the same order of magnitude.

5. SUMMARY

To estimate the calibration function with three different regression methods, we had developed a software with R-Shiny. It was used for processing the calibration data presented in this document to determine the calibration function parameters A_0 and λ. The results obtained are shown. Table 1 shows the values of effective area at null pressure and the distortion coefficient estimated by each method. It is important to denote, that using the generalized least square method we obtained the best value for Residual deviation (RS).

The results indicate that some contrast statistics can be improved with the generalized squares method, the residual standard error is particularly useful for comparing the fit of different regression models.

On the other hand, the application with R-Shiny allows an easier and more direct interaction with the fit methods for the analysis results of the pressure balances calibration.

4. REFERENCES

[1] V. Ramnath, Comparison of straight line curve fit approaches for determining parameter variances and covariances, Int. Journal of Metrology and Quality Engineering, 11, 14 (2020) DOI: 10.1051/ijmqe/2020011

[2] P. Otal, I. Morgado, D. Steindl, N. Medina Martín, A. Lefkopoulos, A. Altintas, L. Grgec Bermanec, S. Bursic, J. Setina, M. Bergoglio, I. Spohr, C. Wüthrich, M. Rantanen, C.Vámossy, W. Sabuga, I. Koças, Evaluation of cross-float measurements with pressure balances – Results of EURAMET project 1125, PTB-Mitteilungen 121, pp. 294-297 (2011), Heft 3.

[3] Calibration certificate N° P153180/1.

[4] J. Doi, G. Potter, J. Wong, I. Alcaraz, P. Chi, Web Application Teaching Tools for Statistics Using R and Shiny DOI: 10.5070/T591027492

[5] P. Otal, C. Yardin, Modelling methods for pressure balance calibration, Meas. Sci. Technol. 2020 31 034004. DOI: 10.1088/1361-6501/ab5910

[6] A. Possolo, National Research Council Canada, & Sistema Interamericano de Metrologia. (2020). Measurement uncertainty: A reintroduction. http://publications.gc.ca/collections/collection_2021/cnrc-nrc/NR16-330-2020-eng.pdf

[7] M. Taboga, Generalized least squares, Lectures on probability theory and mathematical statistics, Kindle Direct Publishing. Online [Accessed 20221209] https://www.statlect.com/fundamentals-of-statistics/generalized-least-squares

[8] X. Li, S. Basu, M. B. Miller, W. G. Iacono, M. McGue, A Rapid Generalized Least Squares Model for a Genome-Wide Quantitative Trait Association Analysis in Families, Hum Hered 2011;71:67-82. DOI: 10.1159/000324839