Key Engineering Materials
ISSN: 1662-9795, Vol. 926, pp 2201-2210
doi:10.4028/p-jpf626

Submitted: 2021-12-09
Revised: 2022-01-31
Accepted: 2022-01-31
Online: 2022-07-22

A Method for Benchmarking of FEM Packages for Multi-Stage Sheet Metal Forming Simulations

Matteo Strano[1,a*], Quirico Semeraro[1,b] and Matteo Panzeri[2,c]

[1]Dipartimento di Meccanica, Politecnico di Milano, via La Masa 1, 20159 Milan (Italy)

[2]P&C Automotive, Via 1° Maggio 16, Missaglia (Italy)

[a]matteo.strano@polimi.it, [b]quirico.semeraro@polimi.it, [c]matteo.panzeri@outlook.com

Keywords: sheet metal, simulation, calibration, benchmarking, variability.

Abstract. Computer simulation plays a crucial role in the designing of sheet metal stamping processes for the prediction of process output, before try-out die sets are manufactured. Different commercial software packages are available on the market for sheet forming simulation, but their accuracy can vary, depending on the selection of the pre-processing parameters and on their formulation. Software benchmarking can be used to select the most appropriate package for a given application. Calibration, i.e. the inverse determination of the correct set of pre-processing parameters, can be used for improving the prediction accuracy.

The scientific literature on numerical simulations of sheet metal forming processes presents some examples of software calibration and very few examples of benchmarking. The literature generally neglects a critical and important issue: the inherent variability of real forming processes.

In this work, the experimental results of two similar multi-stage deep drawing processes are presented and compared to the simulation output of two popular software packages used in the industry. Statistical methods for benchmarking and calibration are proposed. The paper demonstrates how benchmarking can be misleading if process variability is not considered.

Introduction

Computer simulation plays a crucial role in the designing of sheet metal stamping processes for the prediction of process output, before try-out die sets are manufactured, to increase productivity, compress time-to-markets and improve product quality [1]. Most companies in the automotive industry perform sheet stamping simulations on a regular basis [2], and the Finite Element Method (FEM) is the dominant technology in this field. Many software packages are commercially available for sheet metal forming simulation, both general purpose (e.g. Ls-Dyna and Abaqus/Explicit) and specially designed (e.g. Autoform, Pam-Stamp, Optris, Indeed, Stampack). The accuracy of FEM packages can vary, depending on the selection of the pre-processing parameters and on their mathematical formulation [3]. For this reason, benchmarking of alternative software solutions has been the focus of a few studies in the scientific literature. As an example, the springback predictions with Optris and Ls-Dyna were compared already in year 2000 in a SAE technical paper [4]. The Numisheet conference regularly organizes benchmark comparisons to verify the state of the art in the prediction of complex phenomena such as: blank draw-in and springback [5], forming limits [6], mechanics of incremental forming [7], etc. Other benchmark geometries have been proposed by Roberts at el. [8]. Most examples in the literature focus onto the geometrical definition of the benchmark and on the mechanical response variable of interest, but little attention is generally given to the mathematical or statistical treatment of the results.

Recently, Pimentel et al. [9] proposed a comprehensive study based on the Numisheet 2008 Benchmark #2, to compare three commercial packages. The authors conclude that the accuracy of the FEA tools is roughly the same. Amaral et al. [10] used the Numisheet 2016 springback benchmark to discuss some critical numerical issues in the prediction of springback in sheet metal forming.

In the Numisheet benchmarks and in the above cited papers, the results are not compared according to a quantitative method and, besides, they are not confronted to the variability of the process (process capability).

(a) Part A *(b) Part B*

Fig. 1. Final components manufactured by the die sets.

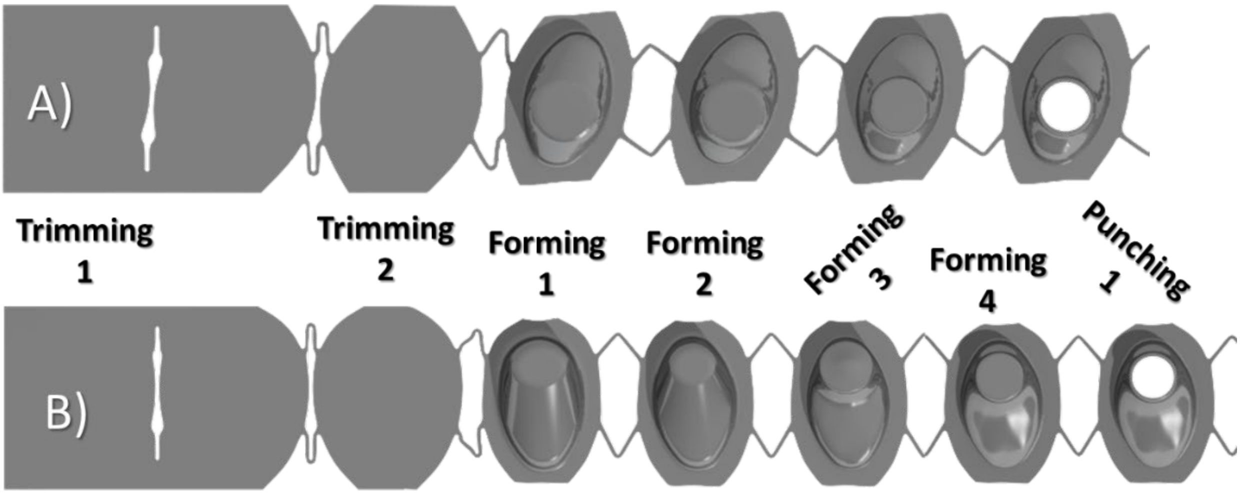


Fig. 2. Processes for the production of the two components. It can be seen that, although processes are very similar, Part B is produced through a higher number of forming stages.

In this paper, we propose a method for benchmarking of different codes, which uses a statistical test in order to take the process capability into account and we demonstrate how benchmarking can lead to misleading results if the real process capability is not considered. The paper will compare the results of the two codes (AutoForm and Pam-Stamp), which have a great industrial diffusion. However, the purpose of the paper is not to perform the benchmark, but to propose a methodology for benchmarking. The results cannot be taken as an indication to compare the two software packages in general terms for any applications, but they can only be limited to the present study.

In the following Section, the two multi-stage deep drawing experimental test cases will be described. Then, the numerical setups of two FEM models developed with two different FEM commercial codes are described. In the last Section, the benchmarking between the two codes is performed with statistical tests, both for the draw-in and for the thickness.

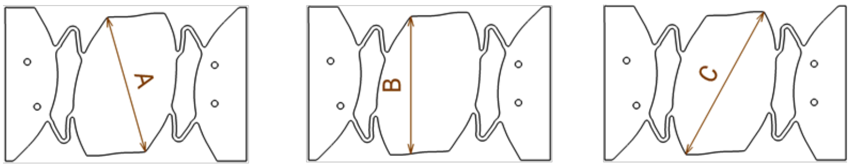
Experimental Test Cases

The test cases are two stamping processes of stainless steel AISI 304 exhaust components for automotive applications. The components, shown in Figure 1, are characterized by the same initial sheet thickness (t_0=1.75 mm), similar dimensions and shape, same level of requested tolerances. The main geometrical difference is a more severe asymmetry in the shape of the second component, called part B.

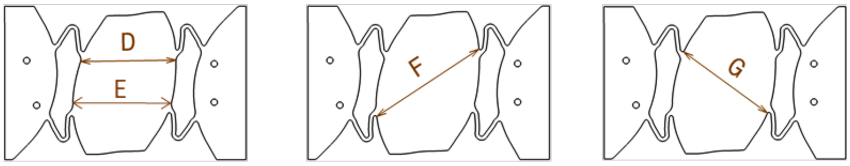
The stamping process for both parts is performed out of a steel strip (width w0= 240 mm) and using progressive dies. The whole cycle includes several stages for trimming, forming, punching, flanging and calibration, but only the forming operations will be simulated. The first part of the production cycles is shown in Fig. 2, until the first hole punching operation. The cycle for part B requires one additional forming stage. For this study, 4 replicates for each forming stage have been stamped and measured for comparison with simulations results. There are 3 forming stages for production of Part A, and 4 for Part B. As a consequence, 28 physical samples are obtained and measured in total, 12 samples for component A and 16 for component B.

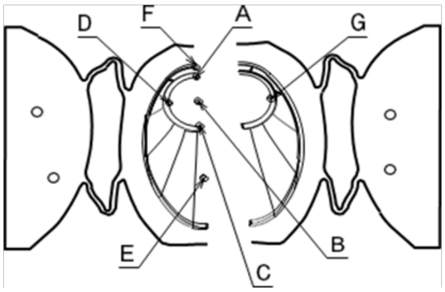
Response variables. The most relevant responses for these parts are the thinning of the part, especially in the collar and the draw-in. Both responses play an important role in the design stage of the die set. When thinning exceeds a limit prescribed by the customer, the part is defective. When excessive draw-in occurs during forming, the outer profile might cross the external trim line, and the part cannot match its designed shape after the flanging operation. Both risks must be reduced thanks to FEM simulations, which should be accurate with respect to the draw-in and thickness prediction.

On each of the 28 samples, 7 measurements of draw-in and 7 measurements of thickness have been taken, on the locations (or sites) shown in Fig. 3. The draw-in is not measured as conventionally, i.e. as a displacement of the flange contour, but the distances at corner locations on the parts (indicated in Fig. 3 too) have been taken. This unusual method of measuring the draw-in has been chosen because it significantly reduces the measurement error, i.e. it reduces the measured process variability.



Distribution of the 7 draw-in measurements





7 Thickness measurement sites

Fig. 3. Sites of experimental measurements of draw-in and thickness performed at each forming stage and for each part. Dimensions “A” and “C” measure about 100 mm.

FEM Simulation Setups

FEM simulations have been performed using AutoForm plus R6 by AutoForm Engineering GmbH and Pam-Stamp 2015.1 by ESI Group.

Positioning and constraint of the sheet were set using physical pilots and blankholder force was set as variable, considering the stiffness of the gas springs that apply their load onto the blankholder. The benchmarking simulations were run using the default input parameters suggested by the two codes. The material in both cases was modeled as elastic-plastic with no dependence on strain rate nor kinematic hardening. Since AISI 304 is a very common material, both software codes provide a default set of material parameters within their built-in data-bases. The default associated flow rules and hardening laws have not been changed. The only modification to the default values has been done to the anisotropy Lankford’s coefficients, because a preliminary sensitivity study has shown that the thickness and draw-in results are very significantly influenced by these parameters and the default values were not correct. Therefore, the sheet metal has been tested according to the ASTM E517 procedure and the correct experimental Lankford's coefficients have been used. Values of main parameters can be found in Table 1.

The simulations were performed on the same desktop computer, with significantly different CPU times. Each run required about 60-80 minutes with Pam-Stamp and about 5-8 minutes with AutoForm.

In post-processing, measurements of the thickness have been taken by creating an auxiliary geometry with the measurement points using CAD software and then importing it in the simulation packages, for having precise references of the measurement points. Draw-in results instead are measured exporting the boundaries of the deformed parts into a CAD software, and then measuring the distances into the CAD environment.

Table 1. Main parameters selected in the benchmarking simulation setups.

			Pam-Stamp (explicit solver)	**AutoForm** (implicit solver)
Parameter		**type**	Value	Value
Blank meshing parameters	Initial element size	default	23 mm	20 mm
	Maximum refinement level	default	5	6
	Maximum element angle	default	7.5°	22.5°
Tool meshing parameters	Meshing tolerance	default	0.1 mm	0.1 mm
	Maximum element size	default	10 mm	50 mm
Coulomb Friction coefficient		default	0.12	0.15
Plasticity parameters	Yield locus	default	Hill '48	Hill
	Hardening law	default	Krupkosky	Krupkosky
	r_0	experim.	0.92	0.92
	r_{45}	experim.	1.18	1.18
	r_{90}	experim.	1.02	1.02

Post processing of results. For each part A or B, the error between simulation and experiments can be calculated as:

$$\%\Delta t_{sjkm} = \frac{t_{sjkm}^{exp} - t_{sjkm}^{sim}}{t_{sjkm}^{exp}}$$
$$\%\Delta d_{sjkm} = \frac{d_{sjkm}^{exp} - d_{sjkm}^{sim}}{d_{sjkm}^{exp}} \quad (1)$$

where *t* is the thickness, *d* is the draw-in and the subscripts indicate:

s = AutoForm, Pam-Stamp software packages
j = 1, 2, 3, 4 forming stages
k = A, B, C, D, E, F, G measurement sites
m = 1, 2, 3, 4 experimental replicates

The errors can also be measured as absolute percentage values $|\%\Delta t_{sjkm}|$, $|\%\Delta d_{sjkm}|$ or as absolute differences:

$$|\Delta t_{sjkm}| = t_{sjkm}^{exp} - t_{sjkm}^{sim}$$
$$|\Delta d_{sjkm}| = d_{sjkm}^{exp} - d_{sjkm}^{sim} \quad (2)$$

A total of 168 values (2 sw packages x 3 stages x 7 sites x 4 experimental replicates) are therefore available for part A and 224 for part B, which has 1 more forming stage.

Benchmarking Methodology and Results

A benchmarking method is here proposed, based on the statistical analysis of the numerical-experimental errors, using the ANOVA (Analysis of Variance) technique, which performs multiple tests of hypothesis, with the Fischer's F statistics. The statistical software package Minitab has been used for computations and graphs. This method can be useful whenever, as in the present case, multiple measurements, multiple experimental replicates and multiple forming stages are available. The software package can be therefore statistically tested as one of the factors of the ANOVA.

Benchmark on the draw-in prediction. As a first, most general comparison, a general test has been conducted using all the 392 values of errors on the draw-in. The ANOVA table on the error with sign $\%\Delta d_{sjkm}$ is reported as Table 2. It clearly indicates that there is a difference between the two software packages, as testified by the p-value being equal to zero.

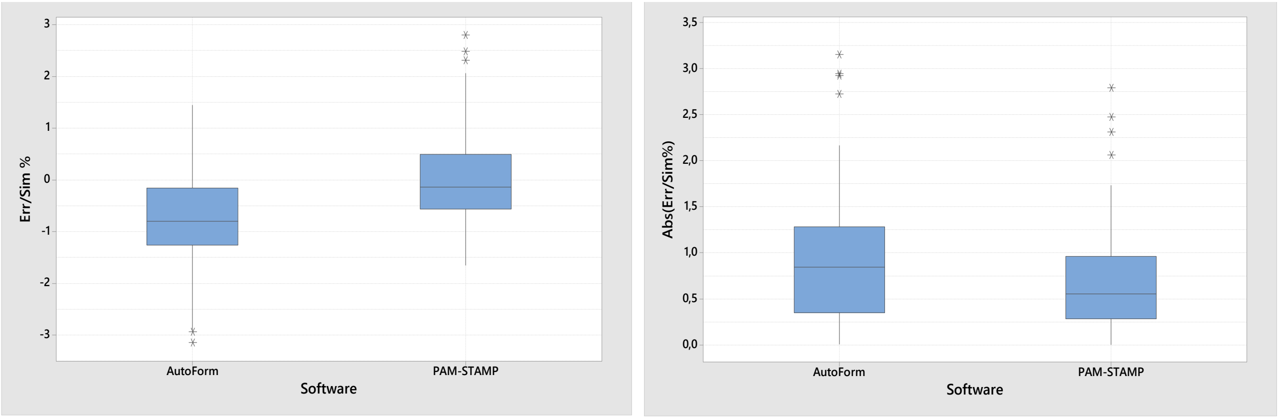
Table 2. ANOVA table for the error on the draw-in $\%\Delta \mathrm{d}_{\mathrm{sjkm}}$.

Factor	**Type**	**Levels**	**Values**
Software PACKAGE	Fixed	2	AutoForm; Pam-Stamp

Source	**DF**	**Adj SS**	**Adj MS**	**F-Value**	**P-Value**
Software PACKAGE	1	46,85	46,8468	72,95	**0,000**
Error	390	250,45	0,6422		
Total	391	297,30			

Model Summary

S	**R-sq**	**R-sq(adj)**	**R-sq(pred)**
0,8014	15,76%	15,54%	14,89%



(a) boxplot of $\%\Delta d_{sjkm}$ *(b) boxplot of* $|\%\Delta d_{sjkm}|$

Fig. 4. Boxplot of errors on the draw-in, calculated on the 392 available values.

The corresponding boxplot of data, grouped by software type, is given in Figure 4. The mean error is -0.71% for Autoform and -0.023% for Pam-Stamp. In conclusion, the percentage prediction error is extremely small for both software codes, but Autoform underestimates, on average, the material draw-in.

The same kind of analysis can be run again using the absolute errors $|\%\Delta d_{sjkm}|$ instead of the errors with sign. The ANOVA has been run by performing a so called "Box-Cox" transformation of the response variables. This transformation is required when the residuals on the analysis are not normally distributed, in order to improve their normality. The corresponding Table 3 presents the results. Here again the software package is statistically significant, with an average absolute error for Autoform equal to 0.74% and for Pam-stamp equal to 0.55%. From an engineering point of view, the predicting capability is not that different, if looking at the absolute percentage error.

Table 3. ANOVA table for the for Box-Cox transformed draw-in response $\left|\%\Delta d_{sjkm}\right|$.

Analysis of Variance for Transformed Response

Box-Cox transformation	
Rounded λ	**0,5**
Estimated λ	0,46695
95% CI for λ	(0,383450; 0,553450)

Factor	**Type**	**Levels**	**Values**
Software	Fixed	2	AutoForm; Pam-Stamp

Source	**DF**	**Adj SS**	**Adj MS**	**F-Value**	**P-Value**
Software PACKAGE	1	1,305	1,3054	11,31	0,001
Error	390	45,024	0,1154		
Total	391	46,330			

Model Summary for Transformed Response

S	**R-sq**	**R-sq(adj)**	**R-sq(pred)**
0,3398	2,82%	2,57%	1,82%

A deeper analysis can be done, adding additional factors to the ANOVA and making different benchmarks for the two parts A and B. The other factors are the forming stage and the measurement location. Two different ANOVA analyses have been performed using the absolute difference $\left|\Delta d_{sjkm}\right|$ (see equation 2) as the response variables, respectively for parts A and B. The ANOVAs for the draw-in are reported in Table 4. The advantage of the ANOVA is that it performs a simultaneous benchmark test over all sites and stages. The analysis shows that all factors and all first and second order interactions are statistically significant, since all the p-values are zeros. Therefore, one software package significantly predicts the draw-in better that the other. However, the presence of interactions means that the difference between the two software packages is not uniform over all forming stages and measurement sites.

In combination with an interaction plot (Figure 5), the ANOVA table allows to effectively perform a benchmark. For part A, Figure 5 shows that Pam-Stamp overperforms AutoForm in most measurement sites and stages, except sites E and F. For part B, Figure 5 shows that Pam-Stamp overperforms AutoForm in stages 1 and 4, and in sites A to D.

In this case, all factors and interactions are statistically significant. Therefore, a non-statistical comparison of the two packages based on the graphical snooping of the errors $\Delta \mathrm{d}_{\mathrm{sjkm}}$ or a comparison of average errors would have led to similar conclusions.

From a technological point of view, it must be noted that the errors in location A for part A and in locations B and D for part B have the largest values for both codes, but they are larger for AutoForm.

Table 4. ANOVA table for the absolute error on the draw-in $\left|\Delta \mathrm{d}_{\mathrm{sjkm}}\right|$, including the factors “forming stage” and “measurement site”.

	part A				**part B**			
Source	DF	Adj MS	F-Value	P-Value	DF	Adj MS	F-Value	P-Value
software PACKAGE	1	12.76	386.8	**0.00**	1	5.249	74.38	**0.00**
forming STAGE	2	17.76	538.5	**0.00**	3	1.358	19.25	**0.00**
measurement SITE	6	3.559	107.9	**0.00**	6	7.743	109.7	**0.00**
PACKAGE*STAGE	2	2.921	88.56	**0.00**	3	22.48	318.6	**0.00**
PACKAGE * SITE	6	1.729	52.42	**0.00**	6	3.645	51.65	**0.00**
STAGE * SITE	12	0.861	26.12	**0.00**	18	0.778	11.03	**0.00**
PACKAGE * STAGE * SITE	12	1.420	43.07	**0.00**	18	4.822	68.32	**0.00**
Error	126	0.033			168	0.071		
Total	167				223			

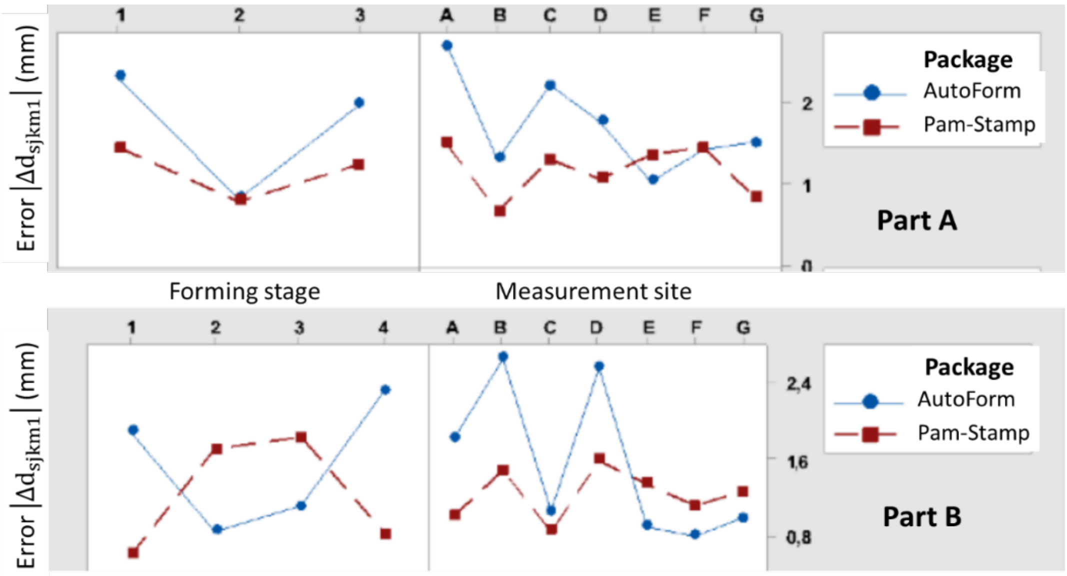


Fig. 5. Interaction plots for absolute differences on the draw-in $|\Delta d_{sjkm}|$, in mm.

Benchmark on the thickness prediction. A similar approach has been followed for the comparison on thickness. As a first general comparison, an overall test has been conducted using all the available data. The ANOVA table on all 392 values of $\%\Delta t_{sjkm}$ (Table 5) clearly indicates that there is a difference between the two software packages, as testified by the p-value being equal to 0.

Table 5. ANOVA table for the error on the thickness $\%\Delta t_{sjkm}$.

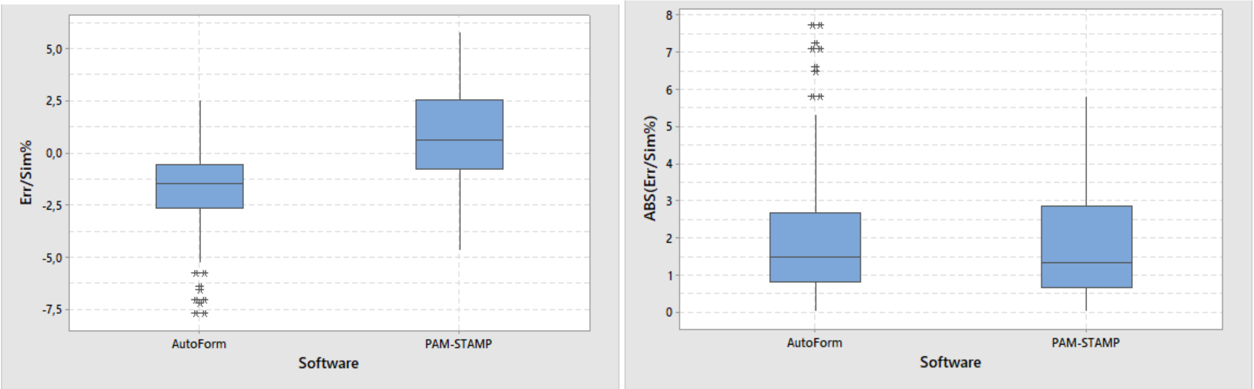
Factor	Type	Levels	Values
Software PACKAGE	Fixed	2	AutoForm; Pam-Stamp

Source	DF	Adj SS	Adj MS	F-Value	P-Value
Software PACKAGE	1	614,0	614,024	152,51	0,000
Error	390	1570,2	4,026		
Total	391	2184,2			

Model Summary

S	R-sq	R-sq(adj)	R-sq(pred)
2,00653	28,11%	27,93%	27,37%

The corresponding boxplot of data, grouped by software type, is given in Figure 6. The mean error is negative (-1.81%) for Autoform and +0.70% for Pam-Stamp. In conclusion, the percentage prediction error is very small for both software codes, but Autoform underestimates the thickness, on average, while Pam-stamp yields an overestimation, although closer to zero.



(a) boxplot of $\%\Delta t_{sjkm}$ *(b) boxplot of* $|\%\Delta t_{sjkm}|$

Fig. 6. Boxplot of errors on the thickness prediction, calculated on the 392 available values.

The same analysis can be run using the absolute errors $|\%\Delta t_{sjkm}|$, rather than the signed percentage error. The ANOVA has been run by performing a square root transformation of the response variables, to improve normality of the residuals. This transformation is required to improve the normality of regression residuals. The corresponding Table 6 presents the results. Here there is no statistical nor practical difference between the two software packages. While the average absolute error in thickness for pam-stamp is 1.79% and it is 1.95% for Autoform, this difference is not statistically significant.

Table 6. ANOVA table for the transformed thickness response $\sqrt{|\%\Delta t_{sjkm}|}$.

Factor	**Type**	**Levels**	**Values**
Software	Fixed	2	AutoForm; PAM-STAMP

Source	**DF**	**Adj SS**	**Adj MS**	**F-Value**	**P-Value**
Software PACKAGE	1	0,402	0,4018	1,25	0,264
Error	390	125,022	0,3206		
Total	391	125,424			

Model Summary for Transformed Response

S	**R-sq**	**R-sq(adj)**	**R-sq(pred)**
0,566189	0,32%	0,06%	0,00%

A deeper analysis can be done, adding additional factors to the ANOVA and making different benchmarks for the two parts A and B. The benchmark analyses in this case is more complicated than for the draw-in estimation.

For part A, the main factor “stage” is not statistically significant, i.e. no package has an overall better performance. However, there is a significant interaction with the measurement site. As the interaction plot in Figure 7 shows, Pam-Stamp has a significant error on thickness at site F, while AutoForm has a larger error at site A. The errors in other sites and across the three forming stages are different but these differences are not statistically significant, i.e. they are not larger than the natural scatter of the experimental data. In this case, a non-statistical comparison of the two packages based on the graphical snooping of the errors Δt_{sjkm} or on a comparison of average or maximum errors would have provided misleading conclusions.

For part B, the main factor “stage” is statistically significant with its first order interactions, and this is well explained by the bottom part of Figure 7. The figure shows that Pam-Stamp generally overperforms AutoForm on all locations, except for measuring site B.

Table 7. ANOVA table for the absolute error on the thickness $|\Delta t_{sjkm}|$, including the factors "forming stage" and "measurement site".

	part A				part B			
Source	DF	Adj MS	F-Value	P-Value	DF	Adj MS	F-Value	P-Value
software Package	1	0.000062	1.26	0.263	1	0.002405	10.28	**0.002**
forming Stage	2	0.000335	6.82	**0.001**	3	0.000682	2.91	0.036
measurement Site	6	0.002887	58.88	**0.000**	6	0.006144	26.26	**0.000**
Package*Stage	2	0.000233	4.75	0.010	3	0.013435	57.43	**0.000**
Package * Site	6	0.002676	54.58	**0.000**	6	0.004765	20.37	**0.000**
Stage * Site	12	0.000491	10.02	**0.000**	18	0.000468	2.00	0.012
Package * Stage * Site	138	0.000049			168	0.000234		
Error	167				223			
Total	1	0.000062	1.26	0.263	1	0.002405	10.28	**0.002**

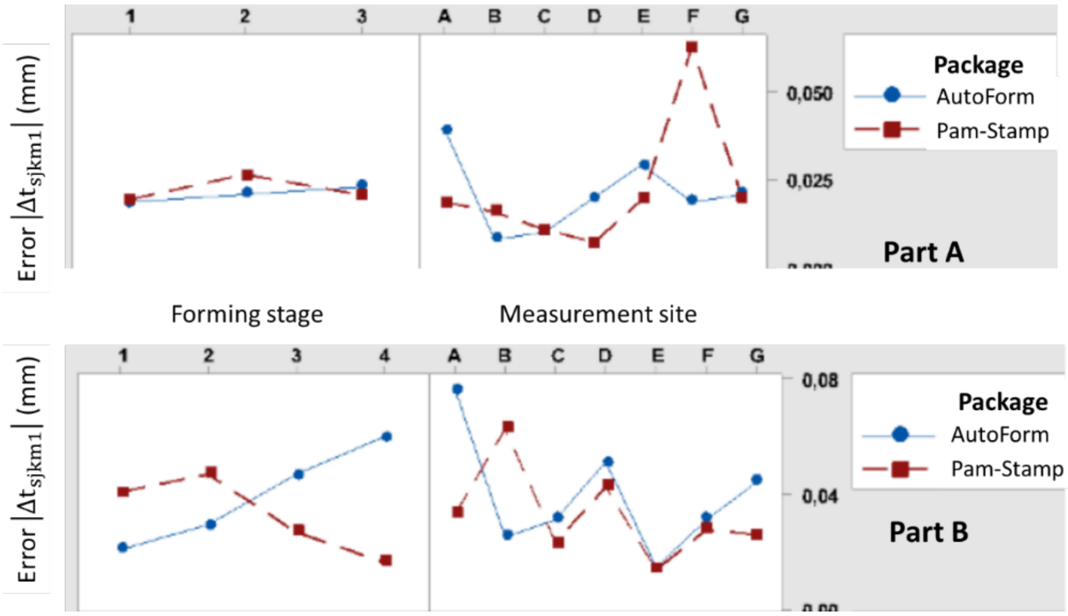


Fig. 7. Interaction plots for absolute differences on the thickness, in mm.

Conclusions

In this paper, a method has been proposed for statistical benchmarking of different software packages. The method is based on an ANOVA test which includes the "software" as one of multiple factors. Hence, the method is particularly suited when response variables are obtained at multiple forming stages in multiple locations. The method has been demonstrated with two similar multi-stage progressive die operations (called A and B). The software packages have been benchmarked with respect to the prediction of thickness and draw-in. The paper shows how a non-statistical method could have led to a misleading conclusion about the predicting ability of the part thickness.

While the proposed methodology holds a general validity, the specific benchmark results cannot be taken as general comparison of the two software packages for any applications. The purpose of the paper was not to compare or to select software products, but to describe a statistical benchmarking technique.

Acknowledgements

The authors are thankful to the company Nuova Stame S.p.A. of Sirtori (LC) in Italy, which allowed the use for experimental purposes of their presses and progressive die sets for parts A and B.

References

[1] Lee MG, Kim C, Pavlina EJ, Barlat F. Advances in Sheet Forming—Materials Modeling, Numerical Simulation, and Press Technologies. J Manuf Sci Eng 2011;133:061001. https://doi.org/10.1115/1.4005117.

[2] Banabic D. Sheet Metal Forming Processes. Berlin, Heidelberg: Springer Berlin Heidelberg; 2010. https://doi.org/10.1007/978-3-540-88113-1.

[3] Ablat MA, Qattawi A. Numerical simulation of sheet metal forming: a review. Int J Adv Manuf Technol 2017;89:1235–50. https://doi.org/10.1007/s00170-016-9103-5.

[4] Demeri MY, Lou M, Saran MJ. A Benchmark Test for Springback Simulation in Sheet Metal Forming. vol. 1. 2000. https://doi.org/10.4271/2000-01-2657.

[5] Zhang L. Background and Tryout Report for BM2: Underbody Cross Member. AIP Conf. Proc., vol. 778, AIP; 2005, p. 888–93. https://doi.org/10.1063/1.2011334.

[6] Volk W, Illig R, Kupfer H, Wahlen A, Hora P, Kessler L, et al. Virtual Forming Limit Curves. Part A: Physical Tryout report. Numisheet benchmark 1, Zurich: 2008.

[7] de Sena JI V., Guzman CF, Duchene L, Habraken AM, Valente RAF, Alves de Sousa RJ. Numerical simulation of a conical shape made by single point incremental, 2013, p. 852–5. https://doi.org/10.1063/1.4850104.

[8] Roberts SM, Hall FR, Van Bael A, et al (1992) Benchmark tests for 3-D, elasto-plastic, finite-element codes for the modelling of metal forming processes. J Mater Process Technol 34:61–68. https://doi.org/10.1016/0924-0136(92)90090-F

[9] Pimentel AMF, de Carvalho Martins Alves JL, de Seabra Merendeiro NM, Vieira DMF. Comprehensive benchmark study of commercial sheet metal forming simulation softwares used in the automotive industry. Int J Mater Form 2018:1–21. https://doi.org/10.1007/s12289-018-1397-4.

[10] Amaral RL, Neto DM, Wagre D, et al (2020) Issues on the Correlation between Experimental and Numerical Results in Sheet Metal Forming Benchmarks. Metals (Basel) 10:1595. https://doi.org/10.3390/met10121595