Optimisation of laser welding of deep drawing steel for automotive applications by Machine Learning: A comparison of different techniques

Received: 24 October 2022 | Revised: 3 March 2023 | Accepted: 8 May 2023

DOI: 10.1002/qre.3377

WILEY

SPECIAL ISSUE ARTICLE

Optimisation of laser welding of deep drawing steel for automotive applications by Machine Learning: A comparison of different techniques

Giacomo Maculotti | **Gianfranco Genta** | **Maurizio Galetto**

Department of Management and Production Engineering, Politecnico di Torino, Torino, Italy

Correspondence
Giacomo Maculotti, Department of Management and Production Engineering, Politecnico di Torino, C.so Duca degli Abruzzi 24, 10129 Torino, Italy.
Email: giacomo.maculotti@polito.it

Abstract

Laser welding is particularly relevant in the industry thanks to its simplicity, flexibility and final quality. The industry 4.0 and sustainable manufacturing framework gives massive attention to in situ and non-destructive inspection methods to predict laser weld final quality. Literature often resorts to supervised Machine Learning approaches. However, selecting the ApTest method is non-trivial and often decision making relies on diverse and unclearly defined criteria. This work addresses this task by proposing a statistical comparison method based on nonparametric tests. The method is applied to the most relevant supervised Machine Learning approaches exploited in literature to predict laser weld quality, specifically, considering the optimisation of a new production line, hence focussing on supervised Machine Learning methods that do not require massive data set, that is, Generalized Linear Model (GLM), Gaussian Process Regression, Support Vector Machine, Classification and Regression Tree, and Genetic Algorithms. The statistical comparison is carried out to select the best-performing model, which is then exploited to optimise the production process. Additionally, an automatic process to optimise Machine Learning models and process parameters is resorted to, basing on Bayesian approaches, to reduce operator effect. This work provides quality and process engineers with a simple framework to compare Machine Learning approaches performances and select the most suitable process modelling technique.

KEYWORDS
deep drawing steel, Gaussian process regression, laser welding, process optimisation, supervised machine learning

1 | INTRODUCTION

Manufacturing in Industry 4.0 makes extended use of artificial intelligence (AI) to analyse the big data collected during manufacturing process to qualify, model, optimise and control the quality of the products.[1] The big data analytics supported by machine learning (ML) techniques are a pillar of the digital transformation of the manufacturing sector, which is finding realisation at different levels of manufacturing, spanning from pre-production and research and development to actual production lines.[2] In turn, this is essential for the creation of digital twins of manufacturing processes that are a key enabling technology to deploy Zero-Defect Manufacturing for sustainable production successfully.[3,4]

Welding is one of the most widely exploited joining technologies, finding applications in several sectors.[5] Amongst the several available welding techniques, laser welding (LW) is particularly interesting. In fact, with respect to other solutions, it features a simpler setup, which does not require mechanical contact with the components to be welded, and when correctly performed, it achieves better final quality in terms of penetration depth, mechanical properties and stability, at a greater welding speed.[5–7] Therefore, it allows higher flexibility, effectiveness and productivity, making it highly attractive for industrial applications, for example, in aerospace, automotive, military, shipbuilding and electronics.[8] However, the inherently chaotic nature of the laser system is liable to introduce several defects, for example, sputter and weld break-ins, and ultimately requires tight, adaptive and responsive quality controls. Currently, in situ controls allow achieving high informativity in real-time by means of non-destructive procedures and typically exploit high-speed optical and thermal cameras, X-ray computer tomography (X-CT), and acoustic and optical sensors.[8] The massive amount of data calls for AI processing by machine vision and ML techniques to extract relevant features for real-time quality control[9–11] and establish analytical and empirical models to predict the final quality of the component.[8,9,12,13] Indeed, post-process inspections are also essential. These aim at inspecting the weld geometry and detecting visible and internal defects. They are based on eddy current, X-CT and ultrasonic techniques and sometimes on destructive inspections that require cross-section and inspection by optical microscopes. Although highly expensive, the latter is mandatory when innovative materials are being developed or introduced in the manufacturing line. In particular, the penetration depth of the weld bead is one of the most critical parameters to ensure the final quality of the process and ensure adequate mechanical properties.[8]

Thus, establishing empirical models to predict and control it from process parameters is essential in pre-production and in R&D. Literature applies several ML techniques to model these kinds of relationships, ranging from supervised to unsupervised techniques.[7,8,14]

Design of Experiments (DOE) and relevant analysis methodology, for example, response surface methodology (RSM) and Generalized Linear Model (GLM), have been extensively adopted. Abioye et al.[15] applied DOE and GLM to maximise the penetration depth of disk laser welding of aluminium alloys. Sathish et al.[16] exploited Taguchi design to optimise laser welding of butt joints of aluminium alloy in terms of mechanical strength. Similarly, Torabi and Kolahan[17] optimised via RSM the weld bead to maximise the ultimate tensile strength for thin stainless steel. Ozkat et al.[18] deployed RSM to achieve a physics-driven model of the weld bead geometry and coupled it with FEM simulative models. Indeed, similar approaches have been deployed to other welding processes, for example, spot welding by Satpathy et al.[19] and friction stir welding by Gagliardi et al.[20]

Gaussian Process Regression (GPR) has also been adopted in the literature to model the effect of process parameters on weld bead geometry of stainless steel[21] and of aluminium alloys.[22]

Kernel-based regression models have also been adopted to study the effect of process parameters on penetration depth and mechanical properties.[7,8] For example, Petković[23] exploited support vector machine regression (SVM or SVR) to model the geometry and the resistance of the weld based on laser welding process parameters, including clamping conditions. Later on, Zhang and Zhou[24] exploited SVR to optimise the weld bead geometric and mechanical properties of stainless steel.

Classification and Regression Trees (CART) have also been adopted to model the relationship between process parameters and weld quality,[25] for example, by XGBoost algorithm for Al-Li alloys by Zhang et al.[26]

Moreover, in addition to explainable artificial intelligence,[27,28] genetic programming (GP) and neural networks (NN) have been adopted for several welding applications.[29,30] GP has been exploited by Wilson et al.[31] to model the LW of deep drawing coated materials, and by Nikolić et al.[32] for low carbon and stainless steel. As far as NNs usage is concerned, for example, Nikolić et al.[32] trained an artificial-NN to predict the geometry of the LW bead for low carbon and stainless steels, Schmoeller et al.[33] trained a variable autoencoder to predict the penetration depth of the LW of aluminium alloys.

Indeed, within this quite complex framework offering several alternatives, authors often investigate multiple ML approaches to model and optimise LW.[7,8,14] However, in several cases, the comparison relies only on root mean square error (RMSE) of predictions, sometimes neglecting Bayesian approaches upon which ML methods rely and often disregard-

ing hyperparameters optimisation.[23,24,29,32] Literature presents method to compare performances of different supervised ML modelling based on holistic and structured framework. However, they tackle problem specific modelling related to data-rich and knowledge poor scenarios.[34] Conversely, in the case of resource intensive manufacturing process setup, for example, LW, data tends to be scarce, due to the high monetary and environmental costs, but knowledge tends to be high, that is, main process parameters and their effect is already known, so that variable and dimension reduction is not necessary. This work proposes a simple methodology to cater for inherent stochastic nature of ML modelling approaches to compare performances of ML when deployed in technological application. Accordingly, this work compares performances of some of the most applied in the literature ML methods within a statistical framework, to provide practitioners guidelines in adopting ML techniques in modelling LW quality. The comparison will be limited to the models most largely applied in laser welding process optimisation, according to the literature as detailed above. The comparison will be performed on an industrially relevant case study, that is, the process setup of deep-drawing steel for automotive application. In particular, the paper considers as input parameters raw process parameters that can be directly set on the machine. The rest of the paper is structured as follows: Section 2 presents the considered case study, and the applied ML techniques and the comparison methodology, while Section 3 outlines and discusses the results, and Section 4 finally draws the conclusions.

2 | METHODOLOGY

2.1 | Materials and experimental setup

This work focuses on modelling the setup of a LW process for deep drawing steel for automotive applications to achieve an understanding of the process for optimisation. The case study is offered by AGLA Power Transmission, an industrial company operating in the automotive sector. The LW process has been carried out by state-of-the-art mass production equipment featuring an Ytterbium fibre laser source, with an adjustable power source up to 10 kW, single-mode beam. The LW targets the manufacturing of a support for the clutch discs of a CVT gearbox. The part consists of two components: a hub and a tonewheel, both out of standard deep drawing steel.[35] This work considers the penetration depth of the weld bead, S_n as the quality control variable. Such choice is not the uniquely possible, and is motivated by the customer requirement for the considered industrial case study, and supported, as far as its practical relevance, by the literature review, briefly outlined in Section 1, which relates it to the mechanical strength of the joining.[8]

The investigation of the process parameters is performed according to the literature and considers the effect of the welding speed v, the laser power P, the focal position (also referred to focus offset) F_O. Conversely to other studies,[36] such approach provides operator a direct indication about how to act on the machine, dispensing with the requirement of complex information on material properties, for example, absorptivity, and process, for example, the relationship between the focus and the beam area. For confidentiality, this works reports analysis based on normalised power with respect to the average laser spot area, that is, the power density P_d. The parameters are reported to have a well-defined effect on the weld geometry. In fact, increasing the power or decreasing the speed a deeper penetration can be obtained. Similar results can be obtained by decreasing the focal spot area (related to the focal position), for a given pair of power and speed.[15,37]

The effect of the process parameters is investigated by realising 88 specimens according to an unbalanced design resulting from the implementation of four investigative DOEs and the addition of some further sparse conditions to enrich the investigated space. Indeed, the implemented experimental design is inevitably linked to the involved companies' available resources. The choice of the parameters, shown in Table 1, and the resulting investigated conditions, reported in Table A1, are according to the industrial company's former experience in processing the materials with a different solid state laser source.

Once the LW has been performed, the component has been cross-sectioned in lubricated condition. The cross-section is then polished with grit paper (240, 320, 800 and 1200) and then with a diamond solution with decreasing grain size (6 μm, 3 and 1 μm). Optical inspection of the weld bead cross-section is performed after Nital etching by means of a metallographic optical microscope Laborlux 12 ME Leitz with 50× magnification to allow the measurement of the weld depth S_n.

2.2 | Machine Learning modelling approaches

According to literature, briefly reviewed in Section 1, the most commonly adopted ML methods are considered in this work. Because this work tackles the optimisation of the process setup, which is typically associated with the availability

TABLE 1 Considered parameters values in the implemented experimental design. Values of welding speed are in angular units as the welded component is axial symmetric. Power values are normalised to the laser spot area to avoid the disclosure of sensitive information and the power density is thus reported.

v / rad/s	P_d / W/mm^2	F_O / mm
1.5	12,025	−20
1.8	13,086	−16
2.1	14,147	−12
2.4	15,031	−8
2.7	15,915	−4
	16,623	0
	17,684	

of few data, neural networks are not considered. Supervised ML techniques are considered, both belonging to explainable AI, that is, GLM, GPR, SVR, CART and not, that is, GP. In the following, a brief overview of the considered approaches is provided, along with the optimisation approach of the relevant parameters. In general, any of the supervised machine learning method that will be discussed can be represented as a function that achieves an estimation of the output y, $\hat{y} = f(x, \vartheta)$, based on a set of predictors, $\boldsymbol{x}$, and a set of hyperparameters, ϑ. In the following, the main features of the methods are discussed functionally to the introduction of the hyperparameters that will be optimised. More in-depth discussion on the methods can be found in reference literature, for example[38,39] for GPR and[40] for GP.

When the models are validated, a method based on statistical inference is applied to compare performances. The best model will be then optimised to seek the process parameter set that maximises the weld bead length.

2.2.1 | Generalized Linear Model

Generalised Linear Model (GLM) is a ML technique that infers a statistical model between a set of predictors and one or more outputs, whose probability distribution is most typically assumed to be normally distributed. The model is a linear combination of the predictors, which may be passed through a nonlinear function,[38] and the coefficients of the linear combination are estimated by the GLM by the least square method, that is, by maximising the estimation of the log-likelihood.[38,41] Here, a third-order model is considered. ANOVA is usually exploited to identify statistically significant parameters while catering for the degrees of freedom of the estimation and of random errors. Therefore, including non-significant parameters in the model is liable to worsen prediction and increase RMSE. Consequently, variable reduction is essential and non-trivial. In this work, variable reduction is performed by the stepwise method, which obtains a model solely consisting of significant terms by adding and removing predictors in a sequence of steps according to selected *alpha-to-enter* and *alpha-to-remove* thresholds, here set at 15%.[42–44]

2.2.2 | Gaussian Process Regression

Gaussian Process Regression (GPR) is a stochastic regression method for interpolating and inferring models in sparse datasets, that is, investigating large portion of the domain in a non-necessarily structured nor densely filled way.[39,45,46] It relies on the assumption that the model response, that is, the prediction, depends on the correlation between the model response at two different evaluation points, and that the correlation is a function of the distance h between these evaluation points. In particular, under this assumption, the GP prediction can be written as:

$$\hat{y} = \boldsymbol{f}^T(\boldsymbol{x}_0)\boldsymbol{\beta} + \boldsymbol{r}_0^T\boldsymbol{R}^{-1}(\boldsymbol{Y}_n - \boldsymbol{F}\boldsymbol{\beta}) \tag{1}$$

where $\boldsymbol{f}^T(\boldsymbol{x}_0)\boldsymbol{\beta}$ is a regressive term and $\boldsymbol{r}_0^T\boldsymbol{R}^{-1}(\boldsymbol{Y}_n - \boldsymbol{F}\boldsymbol{\beta})$ a correction term. The solution, which minimises the mean squared prediction error, assumes the variable $Y(\boldsymbol{x}) = \boldsymbol{\beta}^T\boldsymbol{f}(\boldsymbol{x}) + \Psi(\boldsymbol{x})$ consists of a regression, in particular a linear combination of m function $f(\boldsymbol{x})$ with linear combination parameters $\boldsymbol{\beta}$, and the spatially correlated regression error $\Psi(\boldsymbol{x}) \sim N(0, \sigma_Y^2\boldsymbol{R}(\boldsymbol{h}; \boldsymbol{\theta}))$, having $\boldsymbol{R}(\boldsymbol{h}; \boldsymbol{\theta})$ the correlation matrix dependent on the distance h and a set of parameters $\boldsymbol{\theta}$. The

correction term depends on the residuals, with $\boldsymbol{F}$ is the matrix with entries $F_{ik} = f_k(x_i),\ i \in \{1, 2, \dots, n\},\ k \in \{1, 2, \dots, m\}$ and $\boldsymbol{Y}_n$ the training set with n data, weighted by the correlation, let $\boldsymbol{r}_0 = (\boldsymbol{R}(\boldsymbol{x}_0 - \boldsymbol{x}_1), \cdots, \boldsymbol{R}(\boldsymbol{x}_0 - \boldsymbol{x}_n))^T$.[39,45,46] When training a GP model, it is crucial to determine the regression model $\boldsymbol{F}$ and the correlation parameters. Thanks to the presence of the corrective term in the prediction, a possible solution is the ordinary kriging, which includes only a constant term rather than a linear combination of trend functions. The estimation of the spatial correlation can be performed by means of variogram $\gamma(h)$ and its empirical estimate $\hat{\gamma}_M(h)$, for example according to Matheron, that is,

$$\gamma\ (h) = \sigma_Y^2\ (1 - R\,(h;\theta)) \tag{2.1}$$

$$\widehat{\gamma}_M\,(\boldsymbol{h}) = \frac{1}{\#Q\,(\boldsymbol{h})} \sum_{Q(\boldsymbol{h})} \left(Y\,(\boldsymbol{x}_i) - Y\,\left(\boldsymbol{x}_j\right)\right)^2 \tag{2.2}$$

with $Q(\boldsymbol{h}) = \{(\boldsymbol{x}_i, \boldsymbol{x}_j) : \boldsymbol{x}_i - \boldsymbol{x}_j = h; i, j \in \{1, 2, \cdots, n\}\}$and the operator # is the cardinality.[47] The prediction of spatial correlation at different points and distances requires fitting the empirical variogram $\hat{\gamma}_M$, according to some kernel function, and several alternatives are available in the literature, for example, Matèrn, squared exponential, Gaussian.[48,49]

2.2.3 | Support Vector Machine Regression

Support Vector Machine is a machine learning classification algorithm that charts the input data in hyperspace and defines a hyperplane that can achieve binary classification.[50] The solution of a nonlinear classifier is enabled by space transformation through a kernel function, such that hypersurfaces nonlinear in the original hyperspace are hyperplanes in the transformed space.[38] The main parameters of an SVM are the parameters describing the hyperplanes and the tolerance ε. This defines a tolerance around the hyperplane for the classification. When deployed for regression, SVM aims at minimising the distance between the data point and the hyperplane, according to the tolerance. In particular, if the training data point y falls within the tolerance ε of the related estimate $\hat{y}$, the error $e(y_i, \hat{y}_i)$ is considered null,[38,51] that is:

$$e\ (y_i, \hat{y}_i) = \begin{cases} 0\ if\ \ |y_i - \hat{y}_i| < \varepsilon \\ |y_i - \hat{y}_i| - \varepsilon\ otherwise \end{cases} \tag{3}$$

2.2.4 | Regression Trees

Classification and Regression Tree (CART) graphically describe a regression model by a tree, where terminal leaves are regressors or constants and branch nodes indicate mathematical operations to be performed to the branches merging in that node. Trees' main parameters are the depth, that is, the number of nodes on the same branch, and the width, that is the number of branches. CART can be constructed by boosting. Boosting applies a weak learner that is a certain CART, to a weighted dataset iteratively. At each iteration, weights are updated so that data points with greater prediction error are associated with larger weights. The procedure aims to maximise the accuracy.[52] The accuracy can be expressed in several ways, depending on the specific boosting algorithm that is applied. Amongst the others, L2Boosting minimises squared error, Gradient boosting the absolute error, AdaBoost the exponential loss, LogitBoost the logloss. Further criticality in identifying the adequate CART is the excessive growth of trees, for it worsens readability and makes them liable to overfitting the data.[38] The creation of an ensemble of weak learners is particularly effective in relieving this issue,[53,54] for the ensemble results in greater simplicity, robustness and accuracy by aggregating several simpler weak learners. The decision rule across the weak learners is typically by a simple majority. Consequently, in addition to the parameters typical of the weak learner, and the boosting method, the base learner numerosity and the method to create their several instances define the ensemble. Bootstrap aggregating, that is, bagging, allows creating an ensemble of trees by taking bootstrap samples $\mathcal{L}_\mathrm{B}$ of m data from the learning set $\mathcal{L} = \{\boldsymbol{x}, y\}$, drawn randomly with replacement, and exploiting each of them to create a weak learner f_B, such that $\widehat{y} = f_B(\boldsymbol{x})$.[53] If the input data and the data dimension are subsampled with the same methodology, a random forest can be obtained. This solution has the advantage of constructing uncorrelated predictors for the different samples $\mathcal{L}_\mathrm{B}$.[38,54]

2.2.5 | Genetic Programming

Genetic Programming (GP) is an alternate route for constructing CART. This methodology identifies the most suitable realisation of a CART by means of stochastic investigation of several alternatives, modelling the population. The alternatives are stochastically generated, relying on genetic principles of crossover and mutation and survival.[55,56]

GP requires the creation of a first initial population of CART, randomly generated, of a certain size. Each CART fitness is evaluated according to a criterion. In this work, the RMSE was selected. Then, the population is updated iteratively; at each iteration, a new generation of CART is available for evaluation so that the fittest CART can be selected. Typically, the algorithm is stopped after a certain number of generations. In this work, a population of 500 individuals and 50 generations are considered. Each generation consists of the same number of individuals as the initial population. Genetic programming intervenes in how the new individuals are generated. These exploits either the *crossover*, that is, a new individual results from the random combination of branches of CART in the most recent population, and *mutation*, that is the new individual is generated by randomly modifying a branch of an existing CART. The mix of these two genetic operators is relevant, as well as the possibility that a new individual is *reproduced* from the population. Last, survival can limit the portion of newly generated individuals that will actually go through in the next generation. In particular, elitism principles can be applied. This can keep the fittest individual (*keep-the-best*) between both parents and children, while others are *replaced*, that is, selected by fitness by giving priority to children. Alternatively, *total-elitism* and *half-elitism* can be considered. The former takes in the new generation the absolute fittest individuals between parents and children, with no prioritisation. The latter selects half of the next generation's population as the fittest individuals between parents and children, while the other half is *replaced*. In this work, *total-elitism* is not considered, for it reduces the investigation capabilities of GP. In addition to GP specific hyperparameters, the usual dimension of CART, that is, width and depth, play a major role in complexity, readability, and computational effort.[40] Last, a further source of variability lies in the initial population random generation. Therefore, the literature suggests testing the same hyperparameters set on multiple initial population independent random generations. In this work, 60 runs are performed, and the model providing the minimum RMSE is considered the best GP model.[40,55,56]

Typically, the selection of hyperparameters is performed by trials and errors and largely relies upon operator expertise.[40,55,56] In this work, an automated optimisation based on a Bayesian algorithm is applied; the algorithm is described in Section 2.3. To the authors' best knowledge, the application of Bayesian optimisation to optimise GP hyperparameters is unreported.[40]

2.3 | Model optimisation and validation

The presented and considered supervised Machine Learning approaches, that is, GPR, SVR, CART and GP, are highly interesting tools to draw models describing relationship between output variables and input influence factors. In general, they can be represented as a function that achieves an estimation of the output y, $\hat{y} = f(\boldsymbol{x}, \boldsymbol{\vartheta})$, based on a set of predictors, $\boldsymbol{x}$, and a set of hyperparameters, $\boldsymbol{\vartheta}$. The hyperparameters are highly specific for the considered ML method, similar to the algorithm to evaluate such function f. However, the selection of the best $\boldsymbol{\vartheta}$ is a non-trivial task.[38] In fact, due to the numerous parameters and their nonlinear effect on the goodness of fit, optimisation in closed-form solutions is often computationally expensive. A heuristic alternative methodology exploits black-box models between the hyperparameters ϑ and a cost function modelling the prediction accuracy, $Acc = Acc(\boldsymbol{\vartheta})$. In the case of regression, RMSE can be chosen as (lack of) accuracy estimate. Indeed, this is not the unique alternative, for other metric to describe goodness of fit of regression could have been selected, for example, MSE, MLE.[57,58] However, RMSE is a conventional choice and is suitable to have a straightforward estimation of the effect on the model's prediction uncertainty. Under the considered assumption, the methodology exploits Bayesian optimisation algorithm[58] to maximise the accuracy, that is, minimise the RMSE, by finding $\boldsymbol{\vartheta}_{best} = \underset{\boldsymbol{\vartheta}}{\text{argmin}}(RMSE(\boldsymbol{\vartheta}))$. Table 2 summarises the $\boldsymbol{\vartheta}$ for the considered supervised ML methods, in accordance with the description in Section 2.2. Parameters' ranges were selected out of literature and best practices tackling similar problems, as summarised in the introduction. Some parameters are held fixed, that is in the case of GP, in accordance with the best practices present in literature and sensitivity analysis. Specifically, the selected values are minimum from empirical practices to provide robust results independent from their values.

The Bayesian optimisation is computationally less demanding and achieves a global optimisation.[59] The Bayesian optimisation assumes that the black-box accuracy cost function,$Acc(\boldsymbol{\vartheta})$, is a real Gaussian Process realisation, for it provides

TABLE 2 Considered hyperparameters ϑ to be optimised by the Bayesian algorithm.

ML method	ϑ
GPR	Trend function: constant (ordinary kriging), linear, quadratic, cubic
	Kernel function: Matèrn, exponential, squared exponential, Gaussian, rational quadratic
SVR	Kernel function: Linear, quadratic, cubic, Gaussian, ϵ (unbound)
CART	Weak learner: Depth, width (unbound)
	Boosting algorithm: L2Boosting, Gradient boosting, AdaBoost, LogitBoost
	Ensamble: number of weak learners (unbound)
GP	Population size: 500 (fixed)
	Independent initial population random generation: 60 (fixed)
	Number of generation: 50 (fixed)
	Genetic operators: crossover, mutation, reproduction
	Genetic operators mix (crossover to mutation ratio): $0.1 < p < 0.9$
	Elitism survival: keep-the-best, half-elitism
	Width: unbound
	Depth <6

suitable flexibility and regularity to the function. A Gaussian Process prior is hypothesised and maintained through updated posterior distribution that is new observation of the function. The particular choice of the prior allows tractable posterior and introduce a covariance term, dependent on the distance of probed points that allows improved exploitation and exploitation of the domain.[58,60,61] Thus, choosing the new evaluation point, $\boldsymbol{\vartheta}_{next}$, of the cost function is essential in the algorithm because it determines the posterior. Studying a certain acquisition function, $a = a(\boldsymbol{\vartheta}^*; \boldsymbol{\vartheta}, \{(\boldsymbol{x}, y)\})$, such that $\boldsymbol{\vartheta}_{next} = \underset{\boldsymbol{\vartheta}^*}{\text{argmax}}(a)$, allows identifying $\boldsymbol{\vartheta}_{next}$.[58,60] One of the criticalities is defining a suitable trade-off between the exploitation and the exploration of the hyperparameter space. In particular, the acquisition function has to guarantee that regions that provide minimisation of the cost function are thoroughly investigated, that is exploited, and that those with higher uncertainty, that is little explored, are appropriately investigated. Amongst the others, the *constrained overexploitation expected improvement per second* is a suitable choice for achieving a global optimisation considering the computational effort.[58,60] In particular, the method assumes the acquisition function as:

$$a_{EIpS}\left(\boldsymbol{\vartheta}^*, \boldsymbol{\vartheta}, \{(\boldsymbol{x}, y)\}\right) = \frac{\mathrm{E}\left[\max\left(0, \mu_Q\left(\boldsymbol{\vartheta}^*\right) - Acc\left(\boldsymbol{\vartheta}\right)\right)\right]}{\mu_s\left(\boldsymbol{\vartheta}^*\right)} \tag{4}$$

where $\mu_Q(\boldsymbol{\vartheta}^*)$ is the minimum posterior mean and $\mu_S(\boldsymbol{\vartheta}^*)$ the posterior mean of the Gaussian Process model describing the evaluation time. Under the assumption that the accuracy distributes as a Gaussian Process model with a predictive mean $\mu(\boldsymbol{\vartheta}^*; \boldsymbol{\vartheta}, \{(\boldsymbol{x}, y)\})$ and predictive standard deviation $\sigma(\boldsymbol{\vartheta}^*; \boldsymbol{\vartheta}, \{(\boldsymbol{x}, y)\})$, Equation (4) becomes:

$$a_{EIpS}\left(\vartheta^*, \vartheta, \{(x, y)\}\right) = \frac{\sigma\left(\vartheta^*, \vartheta, \{(x, y)\}\right)\left(\gamma\left(\vartheta^*\right)\Phi\left(\gamma\left(\vartheta^*\right)\right) + \mathcal{N}\left(\gamma\left(\vartheta^*\right); 0, 1\right)\right)}{\mu_s} \tag{5.1}$$

$$\gamma\left(\boldsymbol{\vartheta}^*\right) = \frac{Acc\left(\boldsymbol{\vartheta}_{best}\right) - \mu\left(\boldsymbol{\vartheta}^*; \boldsymbol{\vartheta}, \{(\boldsymbol{x}, y)\}\right)}{\sigma\left(\boldsymbol{\vartheta}^*, \boldsymbol{\vartheta}, \{(\boldsymbol{x}, y)\}\right)} \tag{5.2}$$

where $\Phi(\gamma(\boldsymbol{\vartheta}^*))$ is the standard normal density function.[58,59] Additionally, the posterior standard deviation must not be smaller than a certain fraction of the prior standard deviation. This constraint avoids overexploitation, that is, finding local minima. In fact, if it is not satisfied, the new hyperparameters set $\boldsymbol{\vartheta}_{next}$ belongs to a region with a small uncertainty, that is, it is between already tested points. If that is the case, a multiplication factor proportional with a factor multiple of 10 to the number of performed iterations of the Bayesian algorithm is applied to the $\boldsymbol{\vartheta}_{next}$ to correct the next evaluation point.[59] According to best practices, 30 iterations are performed.[58]

The Bayesian optimisation algorithm is applied to each ML modelling approach independently. The Bayesian optimisation algorithm is not applied to GLM because it would result in unnecessary complexity.

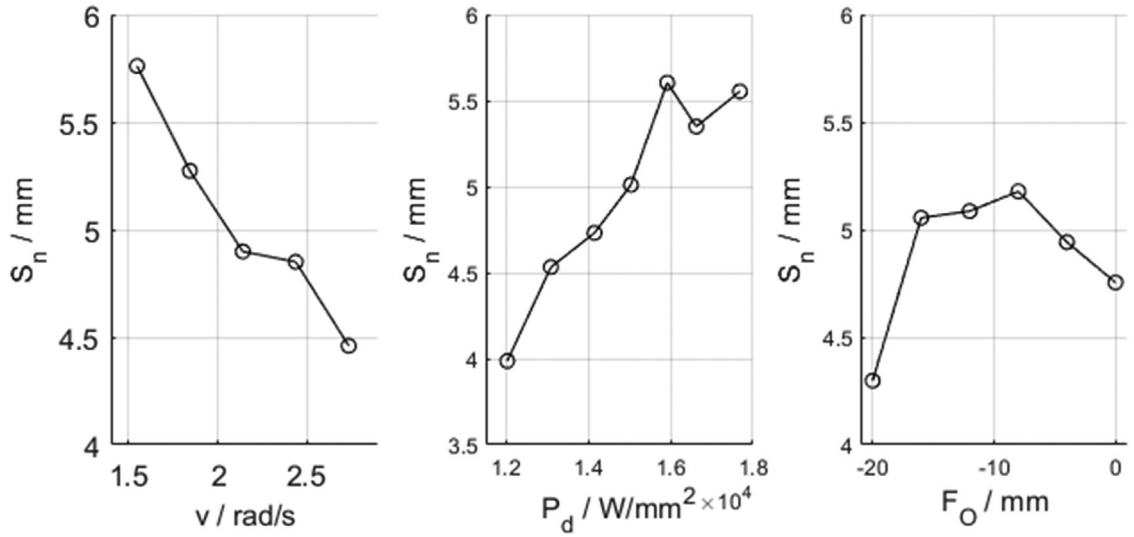

FIGURE 1 Main effect plot of penetration depth with respect to speed, power density and focus offset. Each data point represents the average of the 88 collected data grouped per each factor level.

Model validation to test for generalisation and robustness is performed, and accuracy is evaluated in terms of RMSE on a validation set obtained by a constrained bootstrap sampling of the training set. In particular, *k*-fold cross-validation is adopted in this work. *K*-fold cross-validation splits in *k* folds the data set, each of which in turn is used as a test set. Accordingly, each point is predicted once and used to build the classifier *k*-1 times. In this work a conventional 5-fold cross-validation is considered.[38] The accuracy is computed as the average accuracy of all the folds.[38] Per each k-fold, the model is trained, the parameters optimised by the Bayesian optimisation algorithm, and then validated.

Both supervised ML methods and related optimisation and validation are implemented in MATLAB 2019b. GP base algorithm is deployed by relying on the GPLAB Toolbox v3.0.[62]

2.4 | Performances comparison

The comparison between the sets of RMSE related to the different considered models is performed by means of the Wilcoxon rank-sum test.[63] This is a nonparametric hypothesis test to compare the median of two samples, which can have different sizes, under the null hypothesis that the medians are equal. Adopting a nonparametric test is useful in the case at hand for twofold reasons. It does not require performing a hypothesis on the distribution of the RMSE, and it enables the comparison of samples of different sizes.

3 | RESULTS AND DISCUSSION

The optical metallographic inspection allowed to identify the weld depth S_n. Results are reported for the sake of readability in the annex in Table A1. These are exploited to draw prediction models according to the methodology discussed in Section 2. While formal RMSE comparison will be discussed as per Section 2.4 exploiting k-fold validation, when model parameters optimisation results are discussed, synthetic indication of average RMSE and R^2 evaluated from the *k*-fold cross-validation will be reported.

3.1 | Generalized linear model

Figure 1 shows the main effect plot of the output variable S_n to the considered factors. Nonlinear effects and qualitative significance can be appreciated. Accordingly, the tentative choice of selecting a third-order polynomial model with complete interaction seems reasonable for the GLM.

The GLM is applied with stepwise variable selection, resulting in the model of Equation (6).

$$S_n = 4.77 \times 10^{-4} P_d + 6.67 \times 10^{-4} {F_O}^3 + 1 \times 10^{-6} {F_O}^2 P_d - 1 \times 10^{-7} v {P_d}^2 \quad (6)$$

The goodness of fit was evaluated by performing the Anderson–Darling normality test on the residuals, which with a risk of error of 5% could not reject the null hypothesis of normality. Cross-validation yielded a R^2 of 87% with a RMSE of 0.252 mm.

3.2 | Gaussian Process Regression

Bayesian optimisation selected a universal kriging model with Matèrn 5/2 kernel. This kernel expresses an exponential-like covariance function as:

$$\sigma_Y^2 \boldsymbol{R}(\boldsymbol{h}; \sigma_l) = \sigma_Y^2 \left(1 + \frac{\sqrt{5}\boldsymbol{h}}{\sigma_l} + \frac{5\boldsymbol{h}^2}{3\sigma_l^2}\right) e^{-\frac{\sqrt{5}\boldsymbol{h}}{\sigma_l}} \quad (7)$$

where the parameter σ_l is the correlation length. The Bayesian optimisation estimates the σ_l to 7.6534 mm, while the $\sigma_Y = 4.0297$ mm. The average RMSE of 0.226 with a R^2 of 82% resulted from cross-validation. As could have been expected, the Bayesian optimisation selected a universal kriging model for it has greater flexibility.

3.3 | Support vector machine regression

Bayesian optimisation selected a Gaussian kernel and a tolerance ϵ of 0.028. The Gaussian kernel achieves a space transformation of two regressors according to:

$$G\ (x_1, x_2) = e^{\gamma \| x_1 - x_2 \text{AptCommand2016;}^2} \quad (8)$$

where γ is related to scaling, and selected by Bayesian optimisation as 0.003. This optimisation, when cross-validated, results in an RMSE of 0.327 mm and an R^2 of 73%.

3.4 | Regression Trees

The Bayesian optimisation was performed to select between the parameters reported in Table 2. The CART was selected as an ensemble of 210 weak learners, built by L2Boosting. The Bayesian optimisation constrained the maximum depth to be smaller than 10, leaving the width free. The CART resulted in an average RMSE of 0.269 mm with R^2 of 93% from cross-validation.

3.5 | Genetic Programming

The Bayesian optimisation was performed to select the best hyperparameters. Because, to the best knowledge of the authors, this was not applied before to achieve an automatic selection of GP model parameters, more insights are offered in this case. Figure 2 shows the box plots of the RMSE resulting from the 60 random independent generations of the initial population of the 30 iteration steps of the Bayesian optimisation procedure.

According to Section 2.2.5 and 2.4, the model is selected as the one associated with minimum RMSE, that is the best set is the one identified in the 10th iteration of the Bayesian optimisation. This choice is validated by testing if there is a significant difference between the minimum RMSE and the median of the sample having the minimum median RMSE that is those got in the 23rd iteration of Bayesian optimisation. Systematic differences could be highlighted with a risk of error (p-value) largely smaller than 0.1%. Thus, although the set of hyperparameters associated with the 10th iteration is more sensitive to the initial random generation of CARTs, its performances yield the actual minimum RMSE. The selected

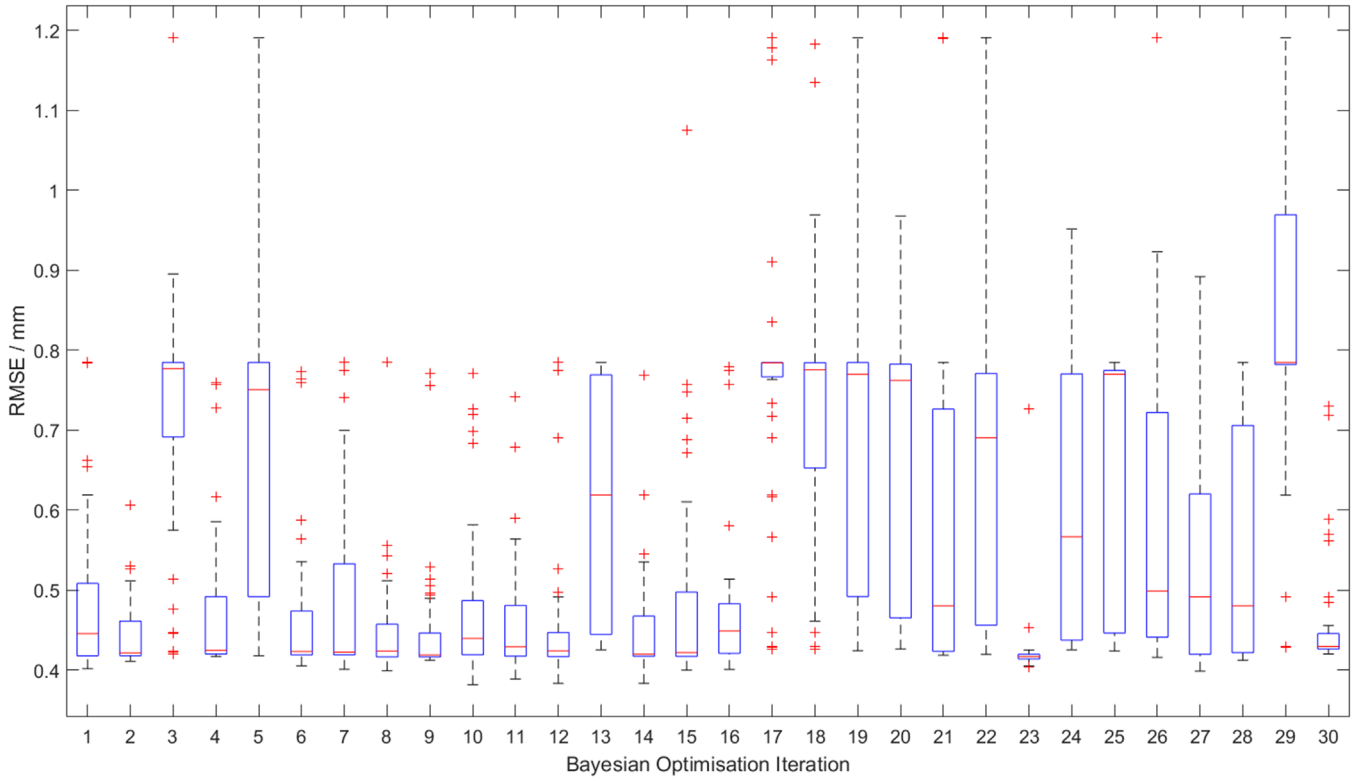

FIGURE 2 Genetic Programming Bayesian Optimisation result.

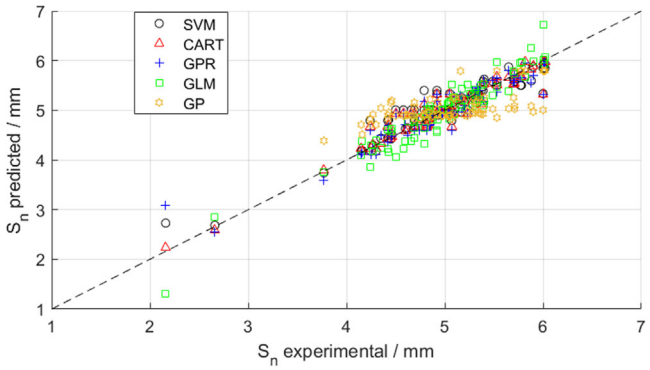

FIGURE 3 Predicted versus experimental value of weld depth S_n.

hyperparameters to generate the optimal GP model exploits a *keep-the-best* elitism operator that acts on a new generation created according to a mix of genetic operators of *crossover*, *mutation* and *replication* of 27%, 3% and 70%, which explains the greater sensitivity to the initial condition. The model resulted in a CART with a width of 16 leaves and a depth of 6 nodes, achieving a RMSE of 0.4 mm and a R^2 of 63%, after cross-validation.

3.6 | Performance comparison

Figures 3 and 4A show, for the considered machine learning approaches, the predicted value and the residuals as a function of the experimental values, respectively. Only in the case of GP, a significant trend can be appreciated in the residuals, suggesting the poor performance of the model, also already indicated by the RMSE and the R^2. This lack of fit of GP makes testing the normality of related residuals little meaningful. Poor GP performances can be explained considering that GP provides significative advantages when a large sample space has to be investigated, considering a high-dimensionality

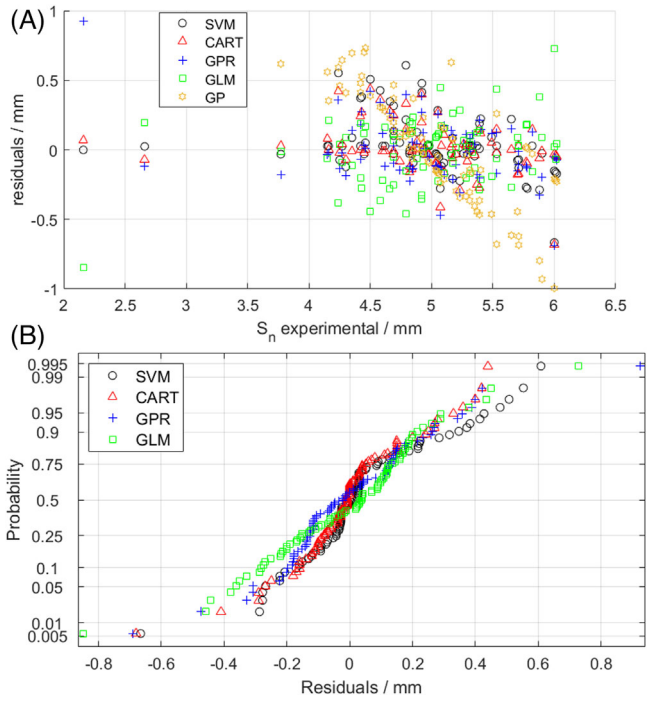

FIGURE 4 (A) Residuals versus experimental value of weld depth *Sn*; notice the trend in GP prediction. (B) Normal probability plot of the residuals; GP is excluded from the analysis due to the systematic trend in the residuals. No deviation from normality can be appreciated for the GLM residuals. Slight hyponormality is suggested for GPR residuals, but not identified by quantitative test.

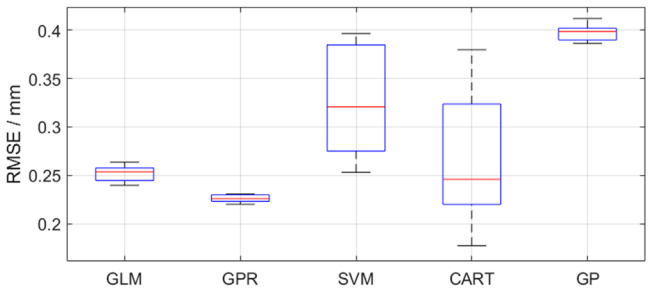

FIGURE 5 Box-plot of the RMSE of the different trained models of the results of *k*-fold cross-validation.

problem domain.[55,56,64,65] As far as other models are concerned, no systematic deviation from normality can be appreciated neither graphically by means of normal probability plot of the residuals in Figure 4B, nor identified by the Anderson–Darling test, with a risk of error of 5%, for the GLM and GPR residuals. In particular, GPR residual presents a slightly hyponormal NPP, which is not significant when tested by the quantitative normality test. Conversely, SVM and CART residuals distributions are significantly different from a normal distribution, with a relevant skewness and high kurtosis.

Bayesian optimisation's computational load engages for from 2′ to 5′ a high-end performance laptop (16 BG RAM, CPU Intel Core i7-8750H @2.2 GHz, GPU NVIDIA GeForce GTX 1060), which increases to 15′ in the sole case of the GP, due to its inherent training structure which increases the complexity of the operations.

Figure 5 shows the box-plot of the RMSE from the cross-validation of the considered models. These data are exploited to perform the Wilcoxon nonparametric test, as presented in Section 2.4. Pairwise comparison between the cross-validation RMSE performed at a risk of error of 5% by the Wilcoxon rank-sum test shows that the GPR performs better than the other trained models. In particular, Table 3 summarises the alternative hypothesis that cannot be rejected when the null hypothesis that the two sample medians are equal ($H_0 : \tilde{x}_1 = \tilde{x}_2$) is rejected with a risk of error of 5%.

TABLE 3 Results of the Wilcoxon rank sum test. Inequality indicates the unrejected alternative hypothesis with a *p*-value <0.05, empty cells mean H_0 could not be rejected.

RMSE1 \ RMSE2	GLM	GPR	SVM	CART	GP
GLM		>	<		<
GPR	<		<	<	<
SVM	>	>			<
CART		>			<
GP	>	>	>	>	

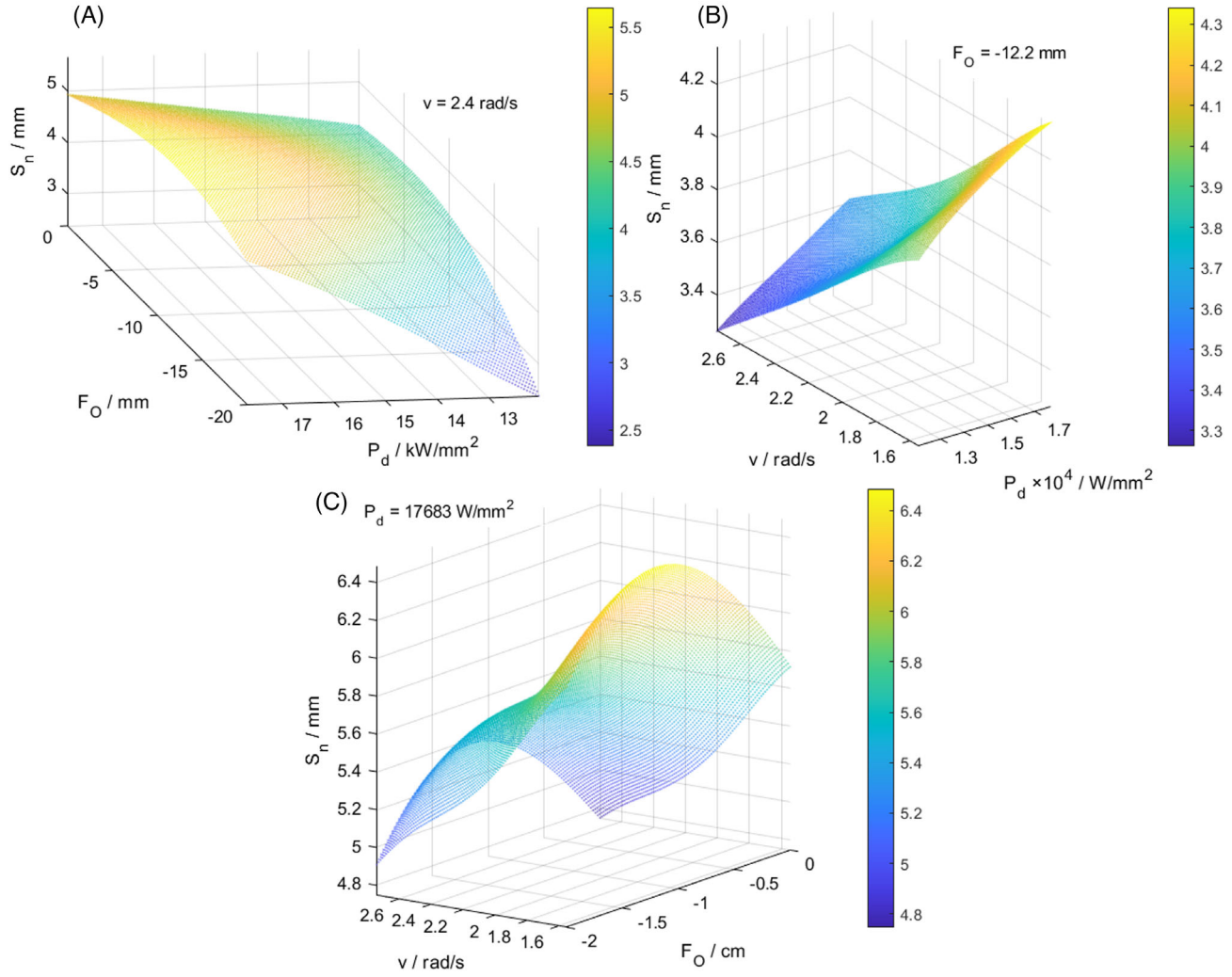

FIGURE 6 Surface plots of GPR model with respect to (A) focus offset-power density plane, (B) speed-power density plane, (C) speed-focus offset plane. Surfaces are drawn holding the third variable constant to the process optimum.

3.7 | Process optimisation

According to the former section, the best trained and cross-validated model is the Gaussian Process Regression. Figure 6 shows three representative surface plots. The GPR model is exploited to achieve process optimisation to maximise the weld depth. When addressing process optimisation, productivity and sustainability are essential. Amongst the considered process parameters, speed is associated with productivity; thus, it is better to have it constrained at a high value. In particular, considering the main effect plot in Figure 1, it was held at 2.4 rad/s. Another critical aspect is process sustainability,

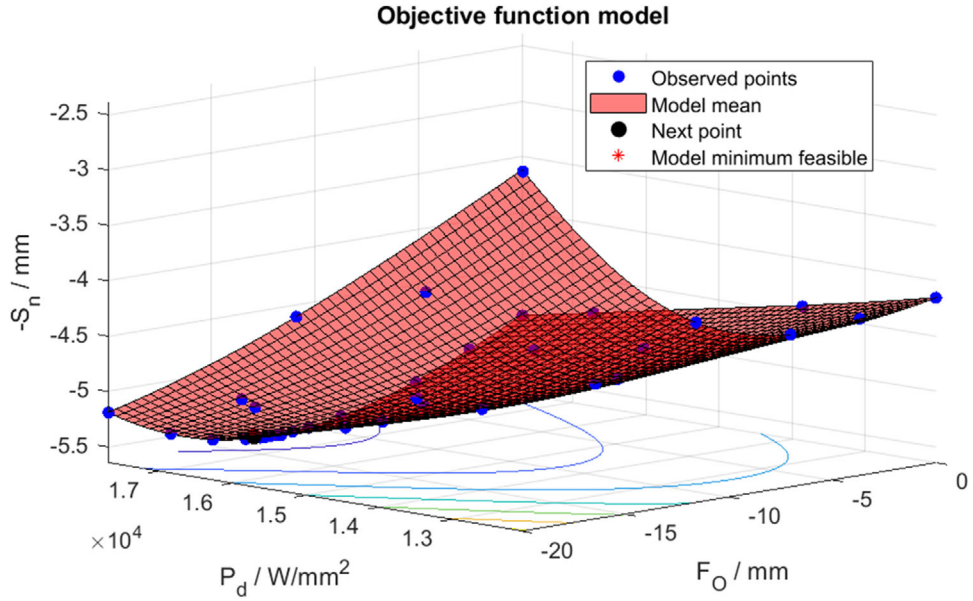

FIGURE 7 Cost function surface for weld depth S_n maximisation exploiting the Gaussian Process Regression model.

TABLE 4 Results of process optimisation for the two best prediction models.

Trained Model	P_d / W/mm^2	F_O / mm	v / rad/s	S_n / mm
GPR	17,683	−12.2	2.4	5.645
GLM	17,684	−12.9	2.4	5.537

which can also be associated with energy consumption, with which laser power density is associated. However, within the considered process window, no dramatic changes can be induced[66]; therefore, it is left unbound during the optimisation. The optimisation is performed again through Bayesian optimisation to avoid other computationally heavy methods. In this case, the cost function is $-S_n = f(F_O, P_d, v = 2.4\,rad/s)$, where the negative sign is required for the optimisation algorithm seeks to minimise the cost function, whilst the process optimisation targets S_n maximisation. Figure 7 shows the response surface of the cost function which allowed the identification of optimal process conditions, which are reported in Table 4. Consistently with physics-based models,[36,37] a deeper penetration is obtained by higher power, lower speed and best focus; however, productivity bounded optimisation derogates from absolute conditions to increase as much as possible the speed, while respecting the tolerance specifications. It is worth noting that such considerations are made possible thanks to the selection of explainable ML, as in the case of GPR and GLM. Those are in general much more difficult to be performed when other ML modelling approaches are considered, for example, CART, GP, SVM. This current limitation of some black-box ML modelling is currently tackled by Physics-based AI.[67,68]

For the sake of comparability, also the second-best trained model, that is the GLM, is exploited for process optimisation. This comparison is also considered because the GLM is a conventional regression method.[41] Results are compared by exploiting the prediction intervals in Figure 8.[42,45] The comparison shows that the average predictions are compatible. Moreover, it shows that despite the prediction of GPR model is more uncertain locally, for it includes a covariance estimation, the overall prediction interval is still less uncertain than the GLM approach. The covariance-estimation weight in the GPR prediction results in a better accuracy and overall precision, which is consistent with the GPR properties as introduced in Section 2.2.2. Thus, GPR is more robust and general, see Figures 3 and 4, where prediction and residuals of GPR show only one outlier related to a poor weld. Conversely, GLM shows two outliers: one at very large penetration depth, that is, in correspondence of a correct weld, and the one related to poor weld. Poor welds are those that resulted in a defective joining, mostly for $S_n < 4.5$ mm. GPR show smaller error (i.e., better accuracy) for these points, which is consistent with the predictive behaviour of GPR, at the cost of higher local greater prediction intervals,[45] which do not impact the overall behaviour, as shown in Figure 8, resulting in a relative prediction uncertainty of 8% for GPR (against a 10% of GLM). Consequently, the covariance-weighted error in the prediction allows a reduction of the expanded uncertainty of 20%, which can be essential in defect prediction and quality planning.[69]

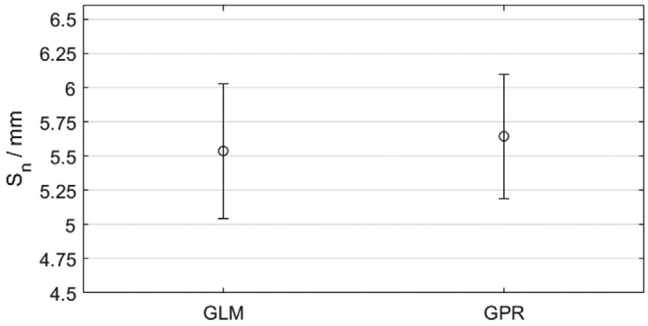

FIGURE 8 Prediction Intervals at 95% confidence level of the optimised weld depth for the best (GPR) and second-best (GLM) model.

4 | CONCLUSIONS

Industry 4.0 and sustainable manufacturing requires adopting non-destructive predictive models for process quality. Laser welding main quality indicator is the weld depth, which may be correlated to process parameters by means of supervised Machine Learning algorithms. This work has proposed a statistical comparison method to compare the performances of different modelling approaches. The method provides a straightforward tool to process and quality designers to select the most suitable model for the scenario at hand. The approach first optimises model hyperparameters via Bayesian optimisation and then compares performances by means of nonparametric median-based hypothesis tests. The necessity for a Bayesian approach to optimise Genetic Programming while catering for its generation method is demonstrated. Greater robustness of Gaussian Process Regression is shown with respect to other models, for it can predicts defects even when in the training dataset very few defects-related data are presented. Additionally, physics informed explanation on the obtained results is performed on the basis of some of the selected ML models, for example, Gaussian Process Regression and Generalized Linear Model. Other modelling approaches, for example, SVM, GP, that does not fall in the category of explainable-AI, are currently hindering such explanation. The selected model is then exploited to optimise the quality of the laser welding of a deep drawing steel for automotive application. Process optimisation is achieved by a bounded Bayesian optimisation of the selected best model, which in the specific case study is Gaussian Process Regression, to achieve tolerance specification, while catering for productivity through a computationally effective approach. The drawn model allows quality prediction, thus enabling a significant reduction, possibly to zero, of destructive quality inspections, for the considered variable. In particular, estimation of the probability of generated defects can be performed, both in optimised and variable process state, through the verification of compliance with the tolerance specification of the prediction interval.

ACKNOWLEDGEMENTS

The authors would like to thank S. Bonù from AGLA Power Transmission, L. Bonù from FORZA SMART INDUSTRY and Dr. R. Cagliero from LBN Ricerca for having provided the case study and the laboratory facilities, and Miss G. Di Paola and Mr. L. Bonamassa for the support in the laboratory activity and data preparation. No funding was received to support this research.

DATA AVAILABILITY STATEMENT

None.

ORCID

Giacomo Maculotti https://orcid.org/0000-0002-2467-0426

REFERENCES

1. Fahle S, Prinz C, Kuhlenkötter B. Systematic review on machine learning (ML) methods for manufacturing processes – Identifying artificial intelligence (AI) methods for field application. *Procedia CIRP*. 2020;93:413-418.
2. Kang Z, Catal C, Tekinerdogan B. Machine learning applications in production lines: a systematic literature review. *Comput Ind Eng*. 2020;149:106773.
3. Kim DH, Kim TJY, Wang X, Kim M, Quan YJ, Oh JW, Min SO, Kim H, Bhandari B, Yang I, Ahn SH. Smart machining process using machine learning: a review and perspective on machining industry. *Int J Precis Eng Manuf - Green Technol*. 2018;5:555-568.

4. Franciosa P, Sokolov M, Sinha S, Sun T, Ceglarek D. Deep learning enhanced digital twin for closed-loop in-process quality improvement. *CIRP Ann*. 2020;69:369-372.
5. Jäger M, Humbert S, Hamprecht F. A Sputter tracking for the automatic monitoring of industrial laser-welding processes. *IEEE Trans Ind Electron*. 2008;55:2177-2184.
6. Alippi C, Braione P, Piuri V, Scotti F. A methodological approach to multisensor classification for innovative laser material processing units. *Conf Rec - IEEE Instrum Meas Technol Conf*. 2001;3:1762-1767.
7. Cai W, Wang J, Zhou Q, Yang Y, Jiang P. Equipment and machine learning in welding monitoring: a short review. ICMRE 19 Proc. 5th Int. Conf. Mechatronics Robot. Eng., 2019:9-15.
8. Stavridis J, Papacharalampopoulos A, Stavropoulos P. Quality assessment in laser welding: a critical review. *Int J Adv Manuf Technol*. 2018;94:1825-1847.
9. Cai W, Jiang P, Shu LS, Geng SN, Zhou Q. Real-time monitoring of laser keyhole welding penetration state based on deep belief network. *J Manuf Process*. 2021;72:203-214.
10. Kershaw J, Yu R, Zhang YM, Wang P. Hybrid machine learning-enabled adaptive welding speed control. *J Manuf Process*. 2021;71:374-383.
11. Günther J, Pilarski PM, Helfrich G, Shen H, Diepold K. Intelligent laser welding through representation, prediction, and control learning: an architecture with deep neural networks and reinforcement learning. *Mechatronics*. 2016;34:1-11.
12. Stavridis J, Papacharalampopoulos A, Stavropoulos P. A cognitive approach for quality assessment in laser welding. *Procedia CIRP*. 2018;72:1542-1547.
13. Knaak C, Kolter G, Schulze F, Kröger M, Abels P. Deep learning-based semantic segmentation for in-process monitoring in laser welding applications. *Appl Mach Learn*. 2019;1113905:2.
14. Rishikesh Mahadevan R, Jagan A, Pavithran L, Shrivastava A, Selvaraj SK. Intelligent welding by using machine learning techniques. *Mater Today Proc*. 2021;46:7402-7410.
15. Abioye TE, Mustar N, Zuhailawati H, Suhaina I. Parametric analysis of high power disk laser welding of 5052-H32 aluminium alloy. *Mater Today Proc*. 2019;17:599-608.
16. Sathish T, Sevvel P, Sudharsan P, Vijayan V. Investigation and optimization of laser welding process parameters for AA7068 aluminium alloy butt joint. *Mater Today Proc*. 2020;37:1672-1677.
17. Torabi A, Kolahan F. Optimizing pulsed ND:YAG laser beam welding process parameters to attain maximum ultimate tensile strength for thin AISI316L sheet using response surface methodology and simulated annealing algorithm. *Opt Laser Technol*. 2018;103:300-310.
18. Ozkat EC, Franciosa P, Ceglarek DA. Framework for physics-driven in-process monitoring of penetration and interface width in laser overlap welding. *Procedia CIRP*. 2017;60:44-49.
19. Satpathy MP, Mishra SB, Sahoo SK. Ultrasonic spot welding of aluminum-copper dissimilar metals: a study on joint strength by experimentation and machine learning techniques. *J Manuf Process*. 2018;33:96-110.
20. Gagliardi F, Navidirad M, Ambrogio G, Misiolek WZ. Effect of material properties and process parameters on quality of friction stir forming. *J Manuf Process*. 2021;70:553-559.
21. Jiang P, Wang C, Zhou Q, Shao X, Shu L, Li X. Optimization of laser welding process parameters of stainless steel 316L using FEM, Kriging and NSGA-II. *Adv Eng Softw*. 2016;99:147-160.
22. Wu J, Zhang S, Sun J, Zhang C. Data-driven multi-objective optimization of laser welding parameters of 6061-T6 aluminum alloy. *J Phys Conf Ser*. 2021;1885:042007.
23. Petković D. Prediction of laser welding quality by computational intelligence approaches. *Optik (Stuttg)*. 2017;140:597-600.
24. Zhang F, Zhou T. Process parameter optimization for laser-magnetic welding based on a sample-sorted support vector regression. *J Intell Manuf*. 2019;30:2217-2230.
25. Köksal G, Batmaz I, Testik MC. A review of data mining applications for quality improvement in manufacturing industry. *Expert Syst Appl*. 2011;38:13448-13467.
26. Zhang Z, Huang Y, Qin R, Ren W, Wen G. XGBoost-based on-line prediction of seam tensile strength for Al-Li alloy in laser welding: experiment study and modelling. *J Manuf Process*. 2021;64:30-44.
27. Rudin C. Stop explaining black box machine learning models for high stakes decisions and use interpretable models instead. *Nat Mach Intell*. 2019;1:206-215.
28. Adadi A, Berrada M. Peeking inside the black-box: a survey on explainable artificial intelligence (XAI). *IEEE Access*. 2018;6:52138-52160.
29. Elangovan S, Anand K, Prakasan K. Parametric optimization of ultrasonic metal welding using response surface methodology and genetic algorithm. *Int J Adv Manuf Technol*. 2012;63:561-572.
30. Kim KY, Ahmed F. Semantic weldability prediction with RSW quality dataset and knowledge construction. *Adv Eng Informatics*. 2018;38:41-53.
31. Wilson D, Kaur D, Forrest M, Lu F. A grammatical evolution approach to system identification of laser lap welding. *SAE Tech Pap*. 2006;2006. https://doi.org/10.4271/2006-01-1614
32. Nikolić V, Petković D, Jocić D, Savić M. Parameters forecasting of laser welding. *FU Mech Eng*. 2018;16:193-201.
33. Schmoeller M, Stadter C, Wagner M, Zaeh MF. Investigation of the influences of the process parameters on the weld depth in laser beam welding of AA6082 using machine learning methods. *Procedia CIRP*. 2020;94:702-707.
34. Rendall R, Reis MS. Which regression method to use? Making informed decisions in "data-rich/knowledge poor" scenarios – The Predictive Analytics Comparison framework (PAC). *Chemom Intell Lab Syst*. 2018;181:52-63.
35. ISO 10111:2008 Continuously hot rolled low carbon steel sheet and strip for cold forming — Technical delivery conditions. ISO, Genève.

36. King WE, Barth HD, Castillo VM, Gallegos GF, Gibbs JW, Hahn DE, Kamath C, Rubenchik AM. Observation of keyhole-mode laser melting in laser powder-bed fusion additive manufacturing. *J Mater Process Technol*. 2014;214:2915-2925.
37. Abioye TE, Farayibi PK, Clare AT. A comparative study of Inconel 625 laser cladding by wire and powder feedstock. *Mater Manuf Process*. 2017;32:1653-1659.
38. Murphy KP. *Machine Learning: A Probabilistic Perspective*. The MIT Press; 2012.
39. Cressie NAC. *Statistics for Spatial Data*. John Wiley & Sons Inc.; 1993.
40. Poli R, Langdon WB, McPhee NF. A Field Guide to Genetic Programing. Published via http://lulu.com and freely available at http://www.gp-field-guide.org.uk, 2008. (With contributions by Koza JR); 2008.
41. Montgomery DC. *Design and Analysis of Experiments*. John Wiley & Sons; 1991.
42. Galetto M, Genta G, Maculotti G, Verna E. Defect probability estimation for hardness-optimised parts by selective laser melting. *Int J Precis Eng Manuf*. 2020;21:1739-1753.
43. Sharma MJ, Yu SJ. Stepwise regression data envelopment analysis for variable reduction. *Appl Math Comput*. 2015;253:126-134.
44. Wagner JM, Shimshak DG. Stepwise selection of variables in data envelopment analysis: procedures and managerial perspectives. *Eur J Oper Res*. 2007;180:57-67.
45. Maculotti G, Pistone G, Vicario G. Inference on errors in industrial parts: kriging and variogram versus geometrical product specifications standard. *Appl Stoch Model Bus Ind*. 2021;37:839-858.
46. Maculotti G, Genta G, Quagliotti D, Galetto M, Hansen HN. Gaussian process regression-based detection and correction of disturbances in surface topography measurements. *Qual Reliab Eng Int*. 2021; 38:1501-1518.
47. Cressie NAC. Spatial prediction and ordinary kriging. *Math Geol*. 1997;20:407-421.
48. Genton MG. Classes of kernels for machine learning: a statistics perspective. *J Mach Learn Res*. 2002;2:299-312.
49. Ginsbourger D, Helbert C, Carraro L. Discrete mixtures of kernels for Kriging-based optimization. *Qual Reliab Eng Int*. 2008;24:681-691.
50. Cortes C, Vapnik V. Support-vector networks. *Mach Learn*. 1995;20:273-297.
51. Coleman TF, Li Y. A reflective newton method for minimizing a quadratic function subject to bounds on some of the variables. *SIAM J Optim*. 1996;6:1040-1058.
52. Schapire RE. *Boosting: Foundations and Algorithms*. The MIT Press; 2012.
53. Breiman L. Bagging predictors. *Mach Learn*. 1996;24:123-140.
54. Breiman L. Random forests. *Mach Learn*. 2001;45:5-32.
55. Gervasi R, Azzali I, Bisanzio D, Mosca A, Bertolotti L, Giacobini MA. Genetic programming approach to predict mosquitoes abundance. *Eur Conf Genet Program*;19AD:35-48.
56. Azzali I, Vanneschi L, Mosca A, Bertolotti L, Giacobini M. Towards the use of genetic programming in the ecological modelling of mosquito population dynamics. *Genet Program Evolvable Mach*. 2020;21:629-642.
57. Murphy KP. *Machine Learning: A Probabilistic Perspective*. MIT Press; 2012.
58. Snoek J, Larochelle H, Adams RP. Practical Bayesian optimization of machine learning algorithms. *Adv Neural Inf Process Syst*. 2012;4:2951-2959.
59. Bull AD. Convergence rates of efficient global optimization algorithms. *J Mach Learn Res*. 2011;12:2879-2904.
60. Gelbart MA, Snoek J, Adams RP. Bayesian optimization with unknown constraints. *Uncertain Artif Intell - Proc 30th Conf UAI*. 2014;2014:250-259.
61. Brochu E, Cora VM, de Freitas N. A tutorial on Bayesian Optimization of Expensive Cost Functions, with Application to Active User Modeling and Hierarchical Reinforcement Learning. *arXiv*. 2010; 1012.2599.
62. Silva S. *A Genetic Programming Toolbox for MATLAB*. Recherche Google; 2007.
63. Gibbons JD, Chakraborti S. *Nonparametric Statistical Inference*. 5th ed. CRC press; 2011.
64. Luciani S, Feraco S, Bonfitto A, Tonoli A, Amati N, Quaggiotto M. A machine learning method for state of charge estimation in lead-acid batteries for heavy-duty vehicles. *Proc ASME Des Eng Tech Conf*. 2021;1:1-8.
65. Gu P, Zhu C, Mura A, Maculotti G, Goti E. Grinding performance and theoretical analysis for a high volume fraction SiCp/Al composite. *J Manuf Process*. 2022;76:796-811.
66. Lunetto V, Galati M, Settineri L, Iuliano L. Unit process energy consumption analysis and models for Electron Beam Melting (EBM): effects of process and part designs. *Addit Manuf*. 2020;33:101115.
67. Karniadakis GE, Kevrekidis IG, Lu L, Perdikaris P, Wang S, Yang L. Physics-informed machine learning. *Nat Rev Phys*. 2021;3:422-440.
68. Gunasegaram DR, Murphy AB, Barnard A, DebRoy T, Matthews MJ, Ladani L, Gu D. Towards developing multiscale-multiphysics models and their surrogates for digital twins of metal additive manufacturing. *Addit Manuf*. 2021;46:102089.
69. Maculotti G, Genta G, Verna E, et al. Minimization of defects generation in laser welding process of steel alloy for automotive application. *Procedia CIRP*. 2022:115:48–53.

How to cite this article: Maculotti G, Genta G, Galetto M. Optimisation of laser welding of deep drawing steel for automotive applications by Machine Learning: A comparison of different techniques. *Qual Reliab Eng Int*. 2023;1-18. https://doi.org/10.1002/qre.3377

ANNEX

TABLE A1 Details of investigated conditions (in terms of welding speed v, power density P_d, and focal position F_O) and related weld penetration depth S_n results.

#	1	2	3	4	5	6	7	8	9	10	11
v / rad/s	1.5	1.5	1.8	1.8	2.1	2.1	2.1	2.1	1.5	1.5	1.5
P_d / W/mm^2	15,915	15,915	15,915	15,915	15,915	15,915	14,147	14,147	14,147	15,915	15,915
F_O / mm	−16	−16	−16	−16	−16	−16	−16	−16	−16	−16	−16
S_n / mm	6.01	5.82	5.7	5.4	5.39	5.37	4.42	4.24	5.16	6	6.02
#	12	13	14	15	16	17	18	19	20	21	22
v / rad/s	2.1	2.1	1.8	1.8	1.5	1.5	2.1	2.1	2.1	2.1	1.8
P_d / W/mm^2	15,031	15,031	14,147	14,147	15,031	15,031	15,031	15,031	14,147	14,147	15,031
F_O / mm	−16	−16	−16	−16	−16	−16	−16	−16	−16	−16	−16
S_n / mm	5.23	4.97	5.17	5.08	5.53	5.78	4.79	4.69	4.69	4.74	5.29
#	23	24	25	26	27	28	29	30	31	32	33
v / rad/s	1.8	2.1	2.1	2.1	2.1	2.1	2.1	2.1	2.1	2.1	2.1
P_d / W/mm^2	15,031	15,031	15,031	14,147	14,147	15,031	15,031	15,031	15,031	15,915	15,915
F_O / mm	−16	−16	−16	−16	−16	−16	−16	−16	−16	−16	−16
S_n / mm	5.17	4.58	4.5	4.82	5.07	4.88	5.05	5.1	5.23	4.79	4.92
#	34	35	36	37	38	39	40	41	42	43	44
v / rad/s	2.1	2.1	2.1	2.1	2.1	2.1	2.1	2.1	2.1	2.1	2.1
P_d / W/mm^2	15,031	16,623	13,086	13,086	15,031	16,623	16,623	13,086	15,031	13,086	16,623
F_O / mm	−12	−12	−12	−16	−16	−16	−16	−12	−12	−16	−12
S_n / mm	6	5.39	4.43	4.3	5.04	5.28	5.35	4.84	4.92	4.25	5.71
#	45	46	47	48	49	50	51	52	53	54	55
v / rad/s	2.1	2.1	2.1	2.1	2.1	2.1	2.1	2.1	2.1	1.5	1.8
P_d / W/mm^2	15,031	13,086	15,031	15,031	13,086	16,623	16,623	15,031	15,031	15,031	15,031
F_O / mm	−16	−12	−12	−16	−16	−12	−16	−20	−8	−16	−16
S_n / mm	5.06	4.82	5.31	4.79	4.15	5.488	5.31	4.16	5.26	5.77	5.1
#	56	57	58	59	60	61	62	63	64	65	66
v / rad/s	2.4	2.7	2.1	2.1	2.1	2.1	2.1	2.1	2.1	2.1	2.1
P_d / W/mm^2	15,031	15,031	12,025	17,684	15,031	15,031	15,031	13,086	13,086	13,086	13,086
F_O / mm	−16	−16	−16	−16	−4	0	−16	−20	−8	−4	0
S_n / mm	4.85	4.46	3.77	5.9	4.93	4.6	5.05	2.16	4.69	4.69	4.62
#	67	68	69	70	71	72	73	74	75	76	77
v / rad/s	2.1	2.1	2.1	2.1	2.1	2.1	2.1	2.1	2.1	2.1	2.1
P_d / W/mm^2	16,623	16,623	16,623	16,623	12,025	12,025	12,025	12,025	12,025	17,684	17,684
F_O / mm	−20	−8	−4	0	−20	−12	−8	−4	0	−20	−12
S_n / mm	4.84	5.88	5.31	4.93	2.66	4.27	4.35	4.45	4.42	5.53	5.65
#	78	79	80	81	82	83	84	85	86	87	88
v / rad/s	2.1	2.1	2.1	2.1	2.1	2.1	2.1	2.1	2.1	2.1	2.1
P_d / W/mm^2	17,684	17,684	17,684	15,031	15,031	15,031	15,031	15,031	15,031	15,031	15,031
F_O / mm	−8	−4	0	−16	−16	−16	−12	−16	−16	−16	−16
S_n / mm	5.71	5.33	5.2	4.98	5.1	5.033333	5.05	4.66	5.04	4.98	5.05

AUTHOR BIOGRAPHIES

Giacomo Maculotti received the Master of Science Degree in Automotive Engineering from Politecnico di Torino, Italy, in 2017 and the PhD Degree in "Management, Production and Design" from Politecnico di Torino in 2021. He is currently Assistant Professor at Politecnico di Torino—Dept. of Management and Production Engineering (DIGEP). He is Fellow of A.I.Te.M. (Associazione Italiana di Tecnologia Meccanica) and E.N.B.I.S. (European Network for Business and Industrial Statistics) and EUSPEN (European Society for Precision Engineering and Nanotechnology). His current research interests are Industrial Metrology, Technological Surfaces Characterisation, and Quality Engineering.

Gianfranco Genta received the Master of Science Degree in Mathematical Engineering from Politecnico di Torino, Italy, in 2005 and the PhD Degree in "Metrology: Measuring Science and Technique" from Politecnico di Torino in 2010. He is currently associate professor at the Department of Management and Production Engineering (DIGEP) of the Politecnico di Torino, where he has been teaching "Experimental Statistics and Mechanical Measurement" since 2012. He is Research Affiliate of CIRP (The International Academy for Production Engineering) and Fellow of A.I.Te.M. (Associazione Italiana di Tecnologia Meccanica) and E.N.B.I.S. (European Network for Business and Industrial Statistics). He is author and coauthor of 3 books and more than 80 publications on national/international journals and conference proceedings. His current research focuses on Industrial Metrology, Quality Engineering and Experimental Data Analysis.

Maurizio Galetto received the Master of Science Degree in Physics from University of Turin, Italy, in 1995 and the PhD Degree in "Metrology: Measuring Science and Technique" from Politecnico di Torino, Italy, in 2000. He is currently Full Professor and Head of the Department of Management and Production Engineering (DIGEP) of the Politecnico di Torino, where he teaches "Quality Engineering" and "Experimental Statistics and Mechanical Measurement". He is Associate Member of CIRP (The International Academy for Production Engineering) and Fellow of A.I.Te.M. (Associazione Italiana di Tecnologia Meccanica) and E.N.B.I.S. (European Network for Business and Industrial Statistics). He is Member of the Editorial Board of the scientific international journal Nanomanufacturing and Metrology and collaborates as referee for many international journals in the field of Industrial Engineering. He is author and coauthor of 4 books and more than 100 published papers in scientific journals, and international conference proceedings. His current research interests are in the areas of Quality Engineering, Statistical Process Control, Industrial Metrology and Production Systems. At present, he collaborates in some important research projects for public and private organisations.