
Article:
Obajemu, O., Mahfouf, M. orcid.org/0000-0002-7349-5396, Papananias, M. et al. (2 more authors) (2021) An interpretable machine learning based approach for process to areal surface metrology informatics. Surface Topography: Metrology and Properties, 9 (4). 044001. ISSN 2051-672X

https://doi.org/10.1088/2051-672x/ac28a7

PAPER • **OPEN ACCESS**

An interpretable machine learning based approach for process to areal surface metrology informatics

To cite this article: Olusayo Obajemu *et al* 2021 *Surf. Topogr.: Metrol. Prop.* **9** 044001

Surface Topography: Metrology and Properties

PAPER

An interpretable machine learning based approach for process to areal surface metrology informatics

Olusayo Obajemu[1], **Mahdi Mahfouf**[1], **Moschos Papananias**[1], **Thomas E McLeay**[2] and **Visakan Kadirkamanathan**[1]

[1] Department of Automatic Control and Systems Engineering, The University of Sheffield, Mappin Street, Sheffield S1 3JD, United Kingdom

[2] Sandvik Coromant, Mossvägen 10, Sandviken 811 34, Sweden

E-mail: m.papananias@cranfield.ac.uk

Keywords: fuzzy logic, manufacturing systems, industry 4.0, surface metrology

OPEN ACCESS

RECEIVED
7 June 2021

REVISED
13 August 2021

ACCEPTED FOR PUBLICATION
20 September 2021

PUBLISHED
4 October 2021

Abstract

Surface metrology parameters represent an important class of design variables, which can be controlled because they represent the DNA or fingerprint of the whole manufacturing chain as well as form important predictors of the manufactured component's function(s). Existing approaches of analysing these parameters are applicable to only a small subset of the parameters and, as such, tend to provide a narrow characterisation of the manufacturing environment.This paper presents a new machine learning approach for modelling the surface metrology parameters of the manufactured components. Such a modelling approach can allow one to understand better and, as a result, control the manufacturing process so that the desired surface property can be achieved whilst manipulating the process conditions. The newly proposed approach utilises a fuzzy logic based-learning algorithm to map the extracted process features to the areal surface metrology parameters. It is fully transparent since it employs IF...THEN statements to describe the relationships between the input space (in-process monitoring variables) and the output space (areal surface metrology parameters). Furthermore, the algorithm includes a ridge penalty based mechanism that allows the learning to be accurate while avoiding over-fitting. This new machine-learning framework was tested on a real-life industrial case-study where it is required to predict the areal parameters of a manufacturing (machining) process from in-process data. Specifically, the case study involves a full factorial experimental design to manufacture seventeen (17) steel bearing housing parts which are fabricated from heat-treated EN24 steel bars. Validation results showed the ability of this new framework not only to predict accurately but also to generalise across different types of areal surface metrology parameters.

1. Introduction

Surface metrology, defined as the science of measurement of small-scale characteristics (such as amplitude, spacing and shape of features) in manufactured parts [1], forms an important part of the manufacturing processes for two main reasons. The first relates to the fact that surface metrology can be thought of as the fingerprint of the whole manufacturing chain. This fact can be used for control of the manufacturing process [2, 3]. The second reason is that surface metrology can directly correlate with the manufactured components function. Such information is useful for quality assessment and function prediction. Predicting the quality or how a manufactured component will function is particularly valuable in helping to meet today's ever tighter budgetary and time constraints as well as the drive for right-first-time production of materials [4]. Indeed, a mechanism for controlling the surface metrology parameters can represent a valuable asset as evidenced by the plethora of research studies which have sought to design algorithms for this purpose [1, 5, 6]. However, before such a control can take place, a mapping from the process conditions to the surface metrology variables must be found. Such a mapping has formed the topic

of many research studies for several decades as will be discussed in the next section. The majority of these research studies focus on very simple mappings typically involving the creation of a limited list of input features from the process data. A data model is then found to map these features to selected surface metrology parameters (usually profile parameters). One notable example is the prediction of the surface roughness heights (Ra) from process conditions [5–7]. It should be noted, however, that these existing studies have mainly focussed on predicting the profile parameters and the application of modelling algorithms for predicting areal parameters which are arguably more important is limited [8]. The areal parameters provide a characterisation for the full 3D surface of the manufactured part and have been shown to be more descriptive of the surface as well as being better related to its function [8]. Therefore, mappings from process conditions to areal parameters can provide better value for the manufacturing process. This research study will therefore mainly focus on the modelling of the areal surface metrology parameters. Existing research studies also typically focus on very small subsets of areal parameters whilst neglecting the others. They also tend to derive coarse scale features extracted from the process data [9, 10]. However, as discussed in [5], many areal surface metrology variables can correspond to a particular function and as such it is often imperative that these areal parameters be combined in a systematic way for function prediction. The surface metrology variables can vary in a very different and sometimes unpredictable manner; an approach formulated for predicting one areal parameter might not be applicable for predicting another areal parameter. As the algorithms hitherto developed have only been validated on one or two areal parameters, it is difficult to make a concrete statement on how such modelling approaches perform across the many areal parameters. Consequently, validating the published algorithms on the other areal parameters (which may perhaps be of equal or more importance depending on the use of the variable) may prove to be problematic. The study in this paper proposes a new framework to predict areal surface metrology parameters based on features extracted from process conditions. The proposed approach is shown not only to generalise across unseen data, but is also robust enough to be utilised for all the twenty four (24) areal surface metrology parameters on which the proposed approach is tested upon. To validate the developed algorithms, a full factorial experimental design was carried out to manufacture seventeen (17) steel bearing housing parts as a case study. The sparse and highly uncertain multidimensional data obtained during this case study represent real manufacturing processes where components are manufactured in low volume. Therefore, the main contribution of this paper is the development of a modelling methodology which can generalise to a large number of manufacturing variables using a limited quantity of data. The details of the experimental design as well as process conditions are discussed in section 3. The proposed framework presents methodology that can aid the drive towards manufacturing automation and data exchange [11]. The review paper by [12] describes state-of-the-art in terms of algorithms, industry uptake and investments across a wide-range of manufacturing industries. For different materials and manufacturing processes, machine learning approaches, such as artificial neural networks, have also been developed with limited experimental data for predictive modelling of properties of manufactured components [13]. The properties of the components can be dictated by the properties of the material, mechanical or microstructural, but also via surface metrology parameters within a synergetic framework. There is a plethora of applied research works relating to the causality between process and material data and mechanical and microstructural properties, but there is little work on such causality with respect to surface metrology parameters. This holistic approach should improve our understanding of how the final properties of manufactured components may be optimised for right-first-time production. The remainder of the paper is organised as follows: section 2 presents a detailed literature review of existing techniques which have been used for mapping process conditions to surface metrology variables. The section details the strengths and weaknesses of these approaches to the overall manufacturing informatics system. As already mentioned, section 3 provides a detailed description of the experimental procedure for which the data has been derived. Section 4 discusses the proposed interpretable fuzzy-based machine learning approach for the surface metrology informatics system. Section 5 presents and discusses the results while section 6 provides the conclusion which can be drawn from the studies conducted from the paper as well as providing suggestions for future research.

2. Existing literature

The book by Whitehouse [1] may perhaps be described as the most important piece of literature where the use of surface metrology in manufacturing for function prediction and quality control is perfectly detailed. The book forms the foundation of many research studies which have investigated the use of surface metrology components to predict manufactured components function and consequently to control the manufacturing process. Controlling the manufacturing process is typically achieved by the manipulation of the process parameters. To achieve such a control framework, it is apparent that a model indicative of how the process parameters affect the surface metrology parameters must be identified [14]. Such a mapping framework has been the subject of many research studies as already discussed in [5, 6]. Surface

Table 1. Selected areal parameters as defined in the ISO documents. The derivations of some of these parameters are shown in figure 3. It should be noted that the data is sampled uniformly along the x and y axes. Z(x, y) represents the measured height at location (x, y).

Symbol	Name	Formula	Notes
Sa	Arithmetic Mean Height	$\frac{1}{A}\int\int_A \lvert z(x,\ y)\rvert dxdy$	This is defined as the arithmetic mean of the absolute of the ordinate values within a definition area (A). This parameter can correlate with friction of manufactured components.
Sq	Root Mean Square Height	$\sqrt{\frac{1}{A}\int\int_A z^2(x,\ y)dxdy}$	This is the root mean square value of the ordinate values within a definition area (A). Sq can relate the way light scattering effects from a surface.
Ssk	Skewness	$\frac{1}{Sq^3}\frac{1}{A}\int\int_A z^3(x,\ y)dxdy$	This is useful for the measurement of surface symmetry about the mean line.
Sku	Kurtosis	$\frac{1}{Sq^4}\frac{1}{A}\int\int_A z^4(x,\ y)dxdy$	This is the quotient of the mean cube value of the ordinate values and the cube of Sq within a definition area (A). It measures the profile symmetry about the mean line.
Sdq	Root Mean Square Gradient	$\sqrt{\frac{1}{A}\int\int_A \left(\frac{\partial z^2}{\partial x}+\frac{\partial z^2}{\partial y}\right)dxdy}$	This parameter is particularly useful in sealing applications.
Sdr	Developed Interfacial Area Ratio	$\frac{\sum\sum A_{ij}-A}{A}$	The Sdr parameter has a direct correlation with surface adhesion. ISO 25 178 part 2 defines the Sdr with integrals instead of summations.

profile parameters account for the majority of surface metrology variables utilised for understanding the manufacturing chain. Of the profile parameters defined in the ISO standards [15], the surface heights (Ra) is the most widely used because its derivation is simple, fast and its meaning is widely understood by manufacturing technologists. For example, a high value of Ra indicates a visually rougher surface. Predicting the Ra accounts for the majority of the surface profile predicted variable studies. Some of these studies include the prediction of surface roughness parameter (Ra) for a computer numerical controlled (CNC) milled surface using linear regression [16] and the assessment of surface roughness using time and frequency domain features for a polished surface [17]. In particular, the studies conducted in [18] have shown that the Ra strongly correlates with the mean and root-mean-square (RMS) of the vibration signals for the polishing process. However, one of the main limitations of the approach is that predicting the Ra may not be sufficient to fully characterise the manufacturing informatics system. This is because the Ra value is very simplistic and may not account for the variation across the surfaces [17]. One solution to this, which has been proposed in the literature, involves creating a distribution of Ra values but this has not been widely adopted by both academia and industry perhaps due to the complexity involved [19]. A better and recent approach relates to characterising the full surface as opposed to using profile parameters. This recent approach is known as the areal surface and it is the main subject of this paper. One of the most prominent studies in attempting to predict the areal surface parameters relates to the prediction of the Sa parameter for a rotating machined process from process variables as included in [19]. The areal parameters characterise the full 3D surface and have been standardised in the ISO25 178 documents [20]. These documents contain a comprehensive industry standard areal parameters. The parameters as well as their use are shown in table 1. Many of the algorithms which have been formulated for the prediction of areal surface parameters have only been applied to one or two of the areal parameters [8]. Validation of such approaches on the parameters on which they have not been tested upon may not be feasible. This paper presents a fuzzy modelling approach for the prediction of surface area metrology parameters. The proposed approach is tested on 24 areal parameters in order to show that the proposed approach can be generalised across the various surface metrology parameters. The paper in [21] provides an excellent overview of the use of fuzzy models in areal surface metrology predictions. Fuzzy logic systems provide a unique modelling approach of leading to interpretable but non-linear input/output mapping when predicting the surface metrology parameters. Manufacturing systems are in the middle of a revolution where different components and stages of the manufacturing process are increasingly becoming 'intelligent'. This intelligence stems from the fact the many components involved in this process are increasingly able to inter-communicate from upstream to downstream. This special ability is embedded in the concept of industry 4.0 which references the fourth industrial revolution in which machine components and processes are equipped with cyberphysical capabilities and are thus capable of tuning their process conditions in response to feedback from the environment and other manufacturing conditions. The promise of industry 4.0 is well discussed in [22]. Surface metrology represents a key enabling component of this revolution as surface metrology parameters play a key part in the inspection of manufactured components. The surface metrology parameters can provide insights for online decision making in a cyber-physically connected system. The Ra, for example, is a design variable and it is typically

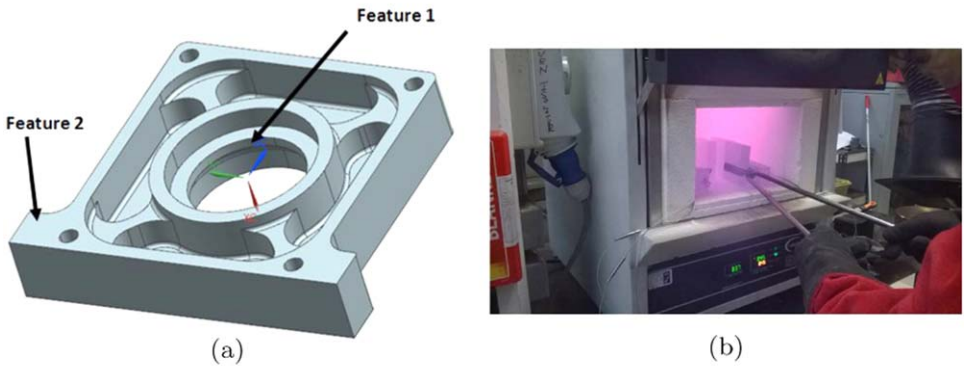

Figure 1. (a) CAD model of the manufactured part. Two features were measured for the purpose of surface metrology analysis. Each feature labelled is associated with one or two operations which correspond to the machining process component which produced the feature. (b) Heat treatment of the steel blocks.

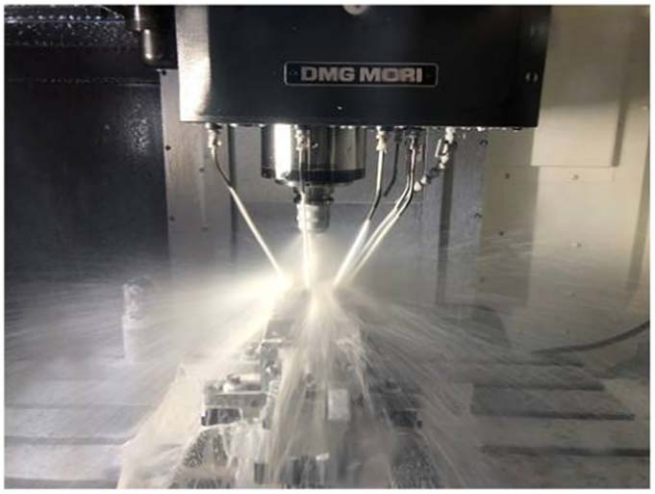

Figure 2. Machining process.

Table 2. Full factorial experimental design variables for 5 of the seventeen manufactured parts. Note that 'Rec' stands for recommended setting.

Run Order	Parts	Material Hardness	Feed	Spindle Speed	Datum Error
1	13	Hard	Rec	+ 20%	0 mm
10	24	Soft	Rec	Rec	0 mm
11	21	Soft	+ 20%	Rec	0.02 mm
15	6	Soft	+ 20%	Rec	0 mm
17	23	Hard	+ 20%	Rec	0 mm

required to not exceed a particular limit for the manufactured component to function as expected.

3. Experimental design

A full-factorial experimental design (see table 2) was performed on a steel bearing house [22]. The CAD model of the product to be manufactured is shown in figure 1a. Using a Vecstar furnace, the material blocks (steel EN24) were heat treated to approximately 845 °C (figure 1b) and then quenched in oil so that they can be hardened. The next step involved tempering at the selected design temperatures. Temperature gradients and variations during both heating and tempering were also measured using high temperature thermocouples. The surface hardness measurements of the blocks were obtained using a Rockwell device. The treated product was then machined (figure 2) using a DMG MORI NVX 5080 3-axis machine with variable controlling factors to arrive at the final manufactured component. During the machining

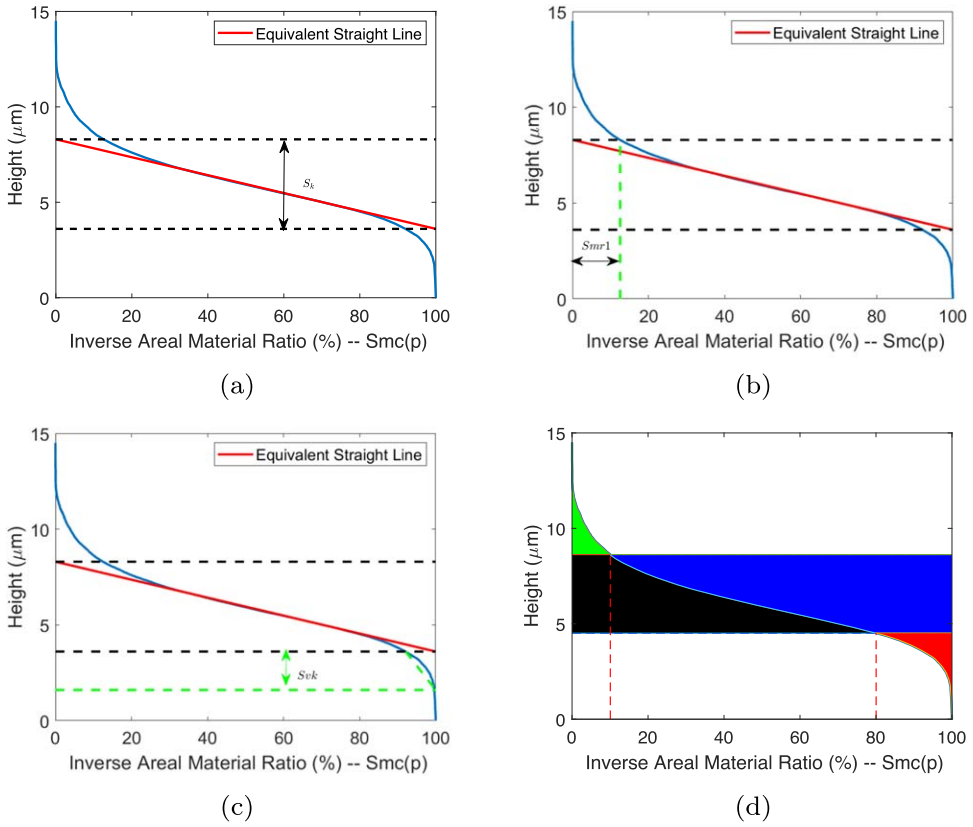

Figure 3. The process of calculating selected areal parameters. (a) Illustration of the core height (*Sk*). (b) Illustration of the material ratio at the first default point. (c) Calculation of the reduced valley height (*Svk*). (d) Important areas for calculating areal parameters: green for *Vmp*, black for *Vmc*, blue for *Vvc* and red for *Vvv*.

process, process data, such as vibration data, were measured along the three main axes of the work-piece. In particular, vibration data were obtained using an accelerometer sensor placed on the spindle which were then logged using LabView SIGNAL Express Software.

The areal surface measurements were obtained using an ALICONA interferometric instrument. Two surface measurements were obtained per part resulting in 34 measurements in total. The features measured per part are shown in figure 1a. This instrument records the height (z) at sampled locations (x, y) with uniform sampling and a sampling interval of 10 μm. The instrument measures the raw surface metrology data and preprocessing is needed to obtain the standardised surface metrology data. The procedure for obtaining the standardised surface metrology data is shown as follows.

1. Obtain the primary surface by the application of the S-Filter on the real Surface. The S-Filter utilised is the Gaussian filter and the standards recommended in the ISO 16 610-21 document [23] have been followed. For example, the wavelength of the S-filter is taken to be 15 times the sampling interval (150 μm).
2. If necessary (depending on the result obtained above), perform further surface filtering to obtain the scaled limited surface. It should be noted that this stage is entirely determined by expert knowledge.
3. Specify the evaluation area which is taken as 5 times the selected wavelength (750 μm).
4. Obtain the reference surface and calculate the parameters as described in figure 3.

A sample of the areal surface metrology measurements obtained following the procedure above is shown in figure 4.

4. Proposed fuzzy modelling approach

Fuzzy logic represents an extension of bivariate logic and was introduced in 1965 in Zadeh's seminal paper [24]. Since then fuzzy logic systems have found applications in a variety of domains including biomedicine [25], process control, manufacturing [26] and aerospace systems. The use of fuzzy systems in these applications offers a unique advantage of being able to model non-linear systems in an interpretable manner. The interpretability comes from the fact that a fuzzy logic system is a rule-based system and the rules are similar to the natural language of humans. These rules also allow for the incorporation of expert knowledge which can be valuable for the analysis of complex systems. Central to fuzzy logic systems are the fuzzy

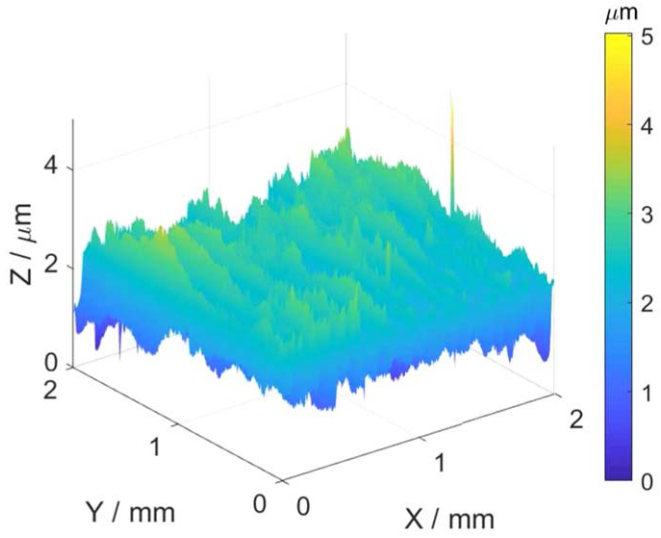

Figure 4. Surface metrology measurement of part 1 feature B. The figure includes a 3 mm × 2.5 mm surface patch which a sampling density along the two axis equals to 100 samples per mm. Hence sampling interval is 10 μm.

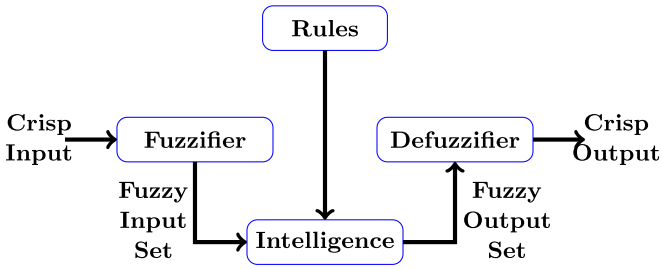

Figure 5. Fuzzy mapping block diagram.

sets. Fuzzy sets extend conventional sets in that they can provide to what extent an element belongs to a particular set. Mathemetically, a fuzzy set (type-1), **A**, may be expressed as follows:

$$\mathbf{A} = \{x,\ \mu_{\mathbf{A}}(x) | x \in X\} \tag{1}$$

where $\mu_{\mathbf{A}}(x)$ is the membership degree of the fuzzy set of an element x in the Universe of discourse X, $0 < \mu_{\mathbf{A}}(x) < 1$. The fuzzy logic system (FLS) can be considered to be a mapping from the input space (defined as X) to the output space (defined as Y) (figure 5). Such a mapping can be formulated by the following equation:

$$\hat{y} = \sum_{j}^{c} \phi_j(x)\,\lambda_j \tag{2}$$

where $\hat{y}$ is output of the fuzzy logic system, $\phi_j(x)$ represents the degree of validity for the jth rule (for a total number of c rules) for an input $x \in R^N$. $\phi_j(x)$ represents the normalised firing strength of a particular input in each input space. The nature of λ_j is what determines if the fuzzy system is of the Mamdani or of the Takagi Sugeno Kang (TSK) type. For the Mamdani type, λ_j represents the output/consequent fuzzy set of the jth rule while for the TSK type, λ_j represents a linear function ($\lambda_j = ax + b$).

4.1. Identifying Fuzzy Models

As discussed in the preceding section, the fuzzy model can be thought of as a nonlinear interpretable mapping from the input space to the output space. The fuzzy system is parameterised (the fuzzy sets can be represented by parameters) and such parameters can be learned from the data obtained from the system to be analysed via fuzzy logic. There exists a plethora of approaches for identifying the parameters of the fuzzy logic system such as optimisation of the cost function via gradient descent and iterated re-weighted least squares [27]. As the goal of this paper is to develop an approach that can generalise across the different areal parameters, it is imperative that a robust framework be found. Consequently, the proposed algorithm development follows a number of steps as discussed in the preceding sections.

4.2. Fuzzy Modelling Approach

The fuzzy model utilised here is of the Mamdani type because it can be shown to represent the most

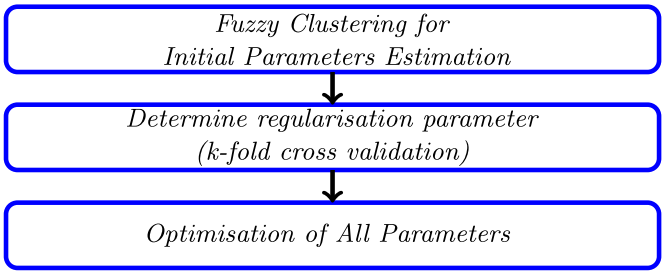

Figure 6. Block Diagram of the steps involved in obtaining the fuzzy model.

transparent of fuzzy models. The block diagram for the process of obtaining the fuzzy model from data is shown in figure 6.

The first step involves the use of fuzzy c-means data clustering of the product space which provides an initial good guess of the parameters of the fuzzy model and will later be optimised. As shown in [26], such an approach can help in preventing the optimisation algorithm from being stuck in a local optima. The number of clusters determines the number of fuzzy rules in the trained fuzzy models. To determine the optimal number of fuzzy rules (which is the same as the number of clusters), a crude search was carried-out to find out the region where the optimal number fuzzy rules is. The authors found that for very large number of fuzzy rules, the algorithm overfitted on the hold-out set and this gets progressively worse as the complexity of the model increases. The search for the optimal number of fuzzy rules was thus limited to between 2 and 12. The second step involves determining the regularisation parameter. This step involves defining a cost function—a penalised root mean square error (RMSE) defined by the following equation:

$$\hat{\boldsymbol{\beta}} = \arg\max_{\beta} = \underbrace{\|\mathbf{y} - \mathbf{f}(\mathbf{X}, \boldsymbol{\beta})\|_2^2}_{\text{Lossfunction}} + \underbrace{\lambda\|\boldsymbol{\beta}\|_2^2}_{\text{Penalty}} \quad (3)$$

where $\mathbf{f}(\mathbf{X}, \boldsymbol{\beta})$ represents the output of the fuzzy system, $\mathbf{y}$ is the vector representing the output data and λ is a penalty term that penalises for large values of the fuzzy model parameters. The value of λ is determined via a K-fold cross validation using the following steps:

Algorithm 1: *K*-fold cross validation algorithm for determining the regularisation term

1. Divide the training data set into K-folds. Note that there is a $70\% - 30\%$ split in training data to testing data. This resulted in a training data of 24 data points. The value of K was chosen to be 4 which means there were 6 data points per fold.
2. From 10^{-2} to 10^{6} (on the log scale), select a particular λ and train the fuzzy model on the three folds and test on the remaining one fold. The approach is repeated until when all the data folds have been tested. Record the λ value and corresponding RMSE.
3. Zoom in on the λ values and find the λ values with the lowest error (RMSE) and repeat procedure 1-2 if necessary.
4. Select the fuzzy model with the lowest RMSE (without the penalty term) and record the value of λ.

It should be noted that steps 2 and 4 above involve a training procedure which involves finding the parameters which minimise the error function as defined in equation (3). The procedure by which this has been done in algorithm 2 is based on the scaled conjugate gradient algorithm.

Algorithm 2: Scaled Conjugate Gradient algorithm for finding the optimal parameters

Given the objective function of equation (3), the parameters of the fuzzy models are obtained via the scaled conjugate gradient descent algorithm. The fuzzy sets for both the antecedent and the consequent variables are assumed to be defined by Gaussian membership functions with two parameters $\left(\mu(x) = \exp\left[-\frac{1}{2}\left(\frac{(x_j - v_{ij}}{\sigma_{ij}}\right)^2\right]\right)$. v and σ correspond to the centre and spread of the membership function. The output of a Mamdani fuzzy system is given by the following equation:

$$\hat{y} = \frac{\sum_{i=1}^{R}\lambda_i \prod_{j=1}^{n} \exp\left(-\frac{1}{2}\left(\frac{(x_j - v_{ij}}{\sigma_{ij}}\right)^2\right)}{\sum_{i=1}^{R}\prod_{j=1}^{n} \exp\left(-\frac{1}{2}\left(\frac{(x_j - v_{ij}}{\sigma_{ij}}\right)^2\right)} \quad (4)$$

where x represents the jth input for a total of n inputs and c fuzzy rules. The derivative of the antecedent and consequent parameters are given by the following equation:

$$\frac{\partial e}{\partial \theta_{ij}^{l}} = (\hat{y} - y)\left(\frac{\lambda_i - \hat{y}}{\mathbf{F}^{\mathrm{T}}\mathbf{1}}\right)\left(\left[\prod_{\substack{q=1 \\ q\neq j}}^{n} \mu_{iq}\right]\frac{\partial \mu_{ij}}{\theta_{ij}^{l}}\right) \quad (5)$$

where θ_{ij}^{l} is the lth parameter of the jth antecedent of the ith rule. for $j = 1, 2, \cdots, n, i = 1, 2, \cdots c$ and $l = v, \sigma$. For each parameter, it can be shown that, their

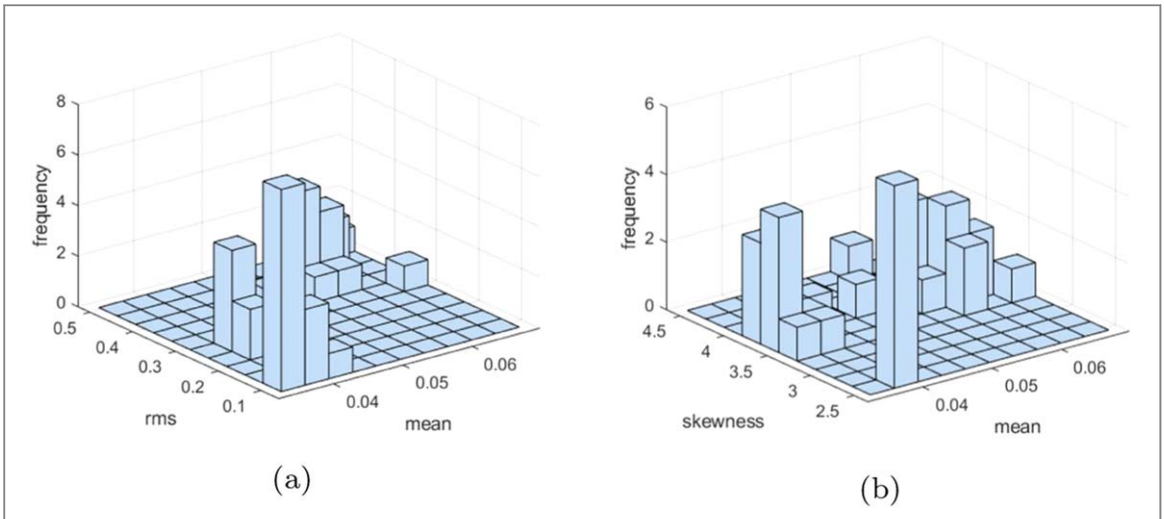

Figure 7. (a) Distribution of selected input variable's root mean square and mean (RMS) of the vibration data. (b) Distribution of selected input variable's skewness and mean of the vibration data.

derivatives with respect to the centre and spread of the membership functions can be given by the following equation:

$$\frac{\partial \mu_{ij}}{\partial v_{ij}} = \frac{(x_j - v_{ij})\mathbf{N}(v_{ij}, \sigma_{ij}; x_j)}{\sigma_{ij}^2} \tag{6}$$

$$\frac{\partial \mu_{ij}}{\partial \sigma_{ij}} = \frac{(x_j - v_{ij})^2\mathbf{N}(v_{ij}, \sigma_{ij}; x_j)}{\sigma_{ij}^3} \tag{7}$$

The derivative with respect to the consequent parameter is given by the following equation:

$$\frac{\partial e}{\partial \beta_i} = (\hat{y} - y)\left(\frac{f_i}{\mathbf{F}^{\mathrm{T}}\mathbf{1}}\right) \tag{8}$$

where β_i is the consequent parameter of the ith rule. It should be noted that **N** represents an un-normalised Gaussian function. **F** is a vector representing the firing strengths across all the rules and **1** is a vector of ones. It is worth emphasizing that the scaled gradient descent algorithm was utilised in this paper. At iteration k, the parameters are updated as follows:

$$\mathbf{P}_k + 1 = \mathbf{P}_k + \alpha_k \mathbf{d}_k \tag{9}$$

P is the vector of parameters, α is the step size and **d** is the search direction. $\psi_k = \alpha_k \mathbf{d}_k$ is given as follows:

$$\psi_k = \frac{g_k^{\top}\mathbf{d}_k}{\mathbf{d}_k^{\top}\mathbf{H} + \beta_k \|\mathbf{d}_k\|^2} \tag{10}$$

where **H** is the Hessian which can be approximated as discussed in [27]. It is worth emphasizing that equation (3) includes a loss function which can be utilised to control the interpretability of the elicited fuzzy model. The center of sets defuzzification method was employed in this research, but the proposed approach extends easily to other defuzzification methods.

5. Results

5.1. Data

The datasets utilised in this research study are the surface metrology data (an example is shown in figure 4) and the process vibration data. The vibration dataset is a time series data sampled at a frequency of 10KHz. Sets of vibration data in the x, y and z directions were obtained per feature in each of the parts. From the vibration data, feature extraction was performed. The features extracted included time and frequency domain features (for example mean [10], root mean square value [17] and the Fourier transform frequency components). A total of 206 features were obtained from the vibration data. A distribution of the vibration data as well as selected input features shown in figure 7 indicates that the data is sparse and multidimensional.

The 24 areal parameters from the surface metrology were also obtained using an in-house software developed by the authors. The procedure for deriving the parameters are as outlined in the ISO standard as well the studies performed in [20, 28].

It is worth emphasizing that the modelling problem is challenging because of the high dimensionality and sparseness of the data points. Specifically, there are 34 data points in all (25 training data points) which points to the fact that it is easy to overfit on the training data [26]. This phenomenon is representative of many manufacturing processes (such as in the manufacture of aerospace components) where parts are manufactured in low volume. It would be interesting to investigate how the proposed approach performs in this challenging modelling problem. It should be noted that a penalised error function coupled with K-fold cross validation is proposed for the modelling problem as discussed in section IV. There is a 70%–30% split between training and testing data sets. This split was performed after a random sampling of the

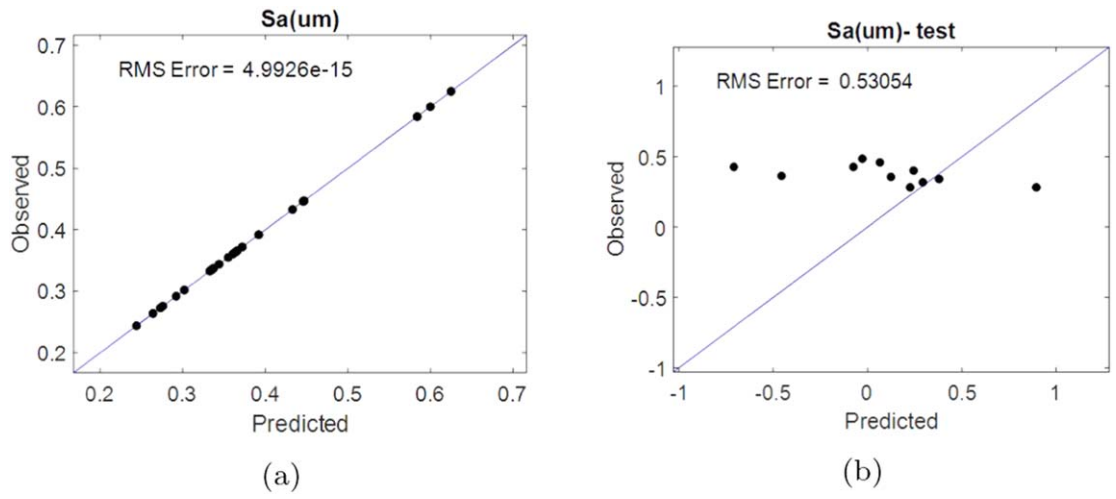

Figure 8. Linear regression performance on the training and testing data for a selected output variable (Sa). There is overfitting because the system is overdetermined.

full data set. The performance metric utilised for evaluating the developed models is the RMSE. The 206 features were extracted from the raw vibration data. New deep learning approaches make it possible to utilise raw time-series data in the modelling problem as shown in [29]. This line of thought was not pursued further because this may not be feasible for cases of low volume manufacture such as the one considered in this paper.

5.2. Linear Regression Modelling

Linear regression modelling is the work-horse of modelling in manufacturing. To test the proposed approach on other modelling problem, linear regression is chosen as a benchmark so that the results obtained from the proposed approach can be compared. The linear regression modelling is given by the following equation:

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \epsilon \tag{11}$$

where $\mathbf{X}$ represents the design matrix and $\boldsymbol{\beta}$ the corresponding parameters. ϵ represents a zero-mean Gaussian noise. For a sum of error square cost function, the solution to the optimisation problem is given by the following equation:

$$\hat{\boldsymbol{\beta}} = (\mathbf{X}^{\top}\mathbf{X})^{-1}\mathbf{X}^{\top}\mathbf{y} \tag{12}$$

It is worth noting that as there are significantly more features than data points, the linear regression modelling problem will be overdetermined and will result in overfitting on the modelling problem. This was indeed the case when a linear model was performed on the training data. These results are shown in figure 8.

As can be seen from the results of figure 8, the linear regression model fits the training data perfectly but does not generalise well to unseen data (as can be noted form the testing data set performance). To allow for better generalisation to unseen data, the linear regression cost function can be penalised as given by the following equation:

$$\hat{\boldsymbol{\beta}} = (\mathbf{X}^{\top}\mathbf{X} + \lambda\mathbf{I})^{-1}\mathbf{X}^{\top}\mathbf{y} \tag{13}$$

where λ is called the ridge parameter whose function is to penalise for large weights. As already mentioned, the penalty term (λ was determined by K-fold cross validation) as described in section 3. The penalised linear regression (ridge linear regression) results is as shown in figure 9.

As can be seen from figure 9, although the results of the testing datasets are more generalising when compared with ordinary linear regression results, the training data set is significantly much worse. This is as a result of the fact that the ridge parameter is able to find a compromise between the best training results (in the linear sense) and the best validation results (in the linear sense). The results suggest that a non-linear model is required to obtain a good mapping of the process parameters. It is for this reason that Mamdani fuzzy model is first considered as discussed in section 3. The first Mamdani model considered is not inclusive of any penalty term which has already explained can result in overfitting of the training model. Such a result is similar to the ordinary regression result (shown in figure 8). The fuzzy modelling result without any penalty term is shown in figure 10.

To allow for better generalisation, the same ridge linear regression training procedure (discussed in section 4) is also followed to train the Mamdani fuzzy model. The results of the ridge Mamdani fuzzy system is shown in figure 11. We have called this approach the ridge Mamdani fuzzy modelling approach to empasize its capability to penalise for large fuzzy weights in order to improve generalisation performance.

As can be seen in figure 11, the ridge fuzzy modelling framework provides a much improved performance and is able to map the process features to the surface metrology parameters. The result shown in

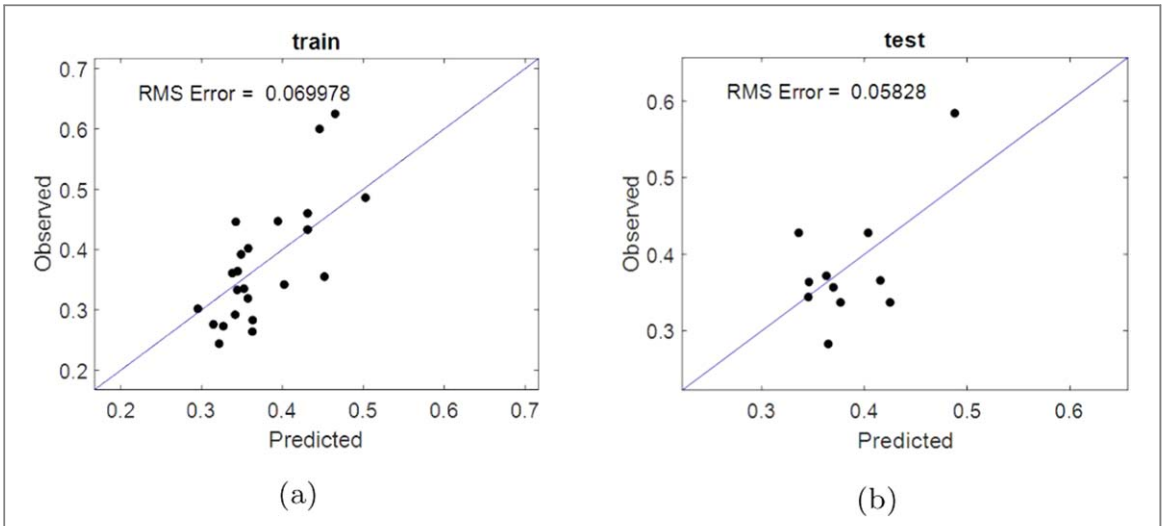

Figure 9. Penalised linear regression results.

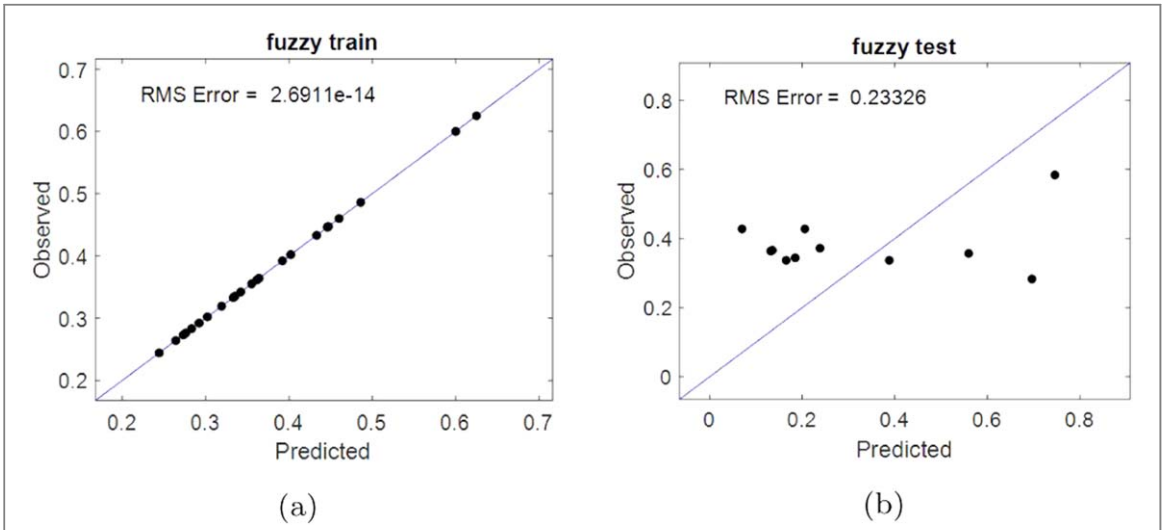

Figure 10. Mamdani-based fuzzy logic modelling results.

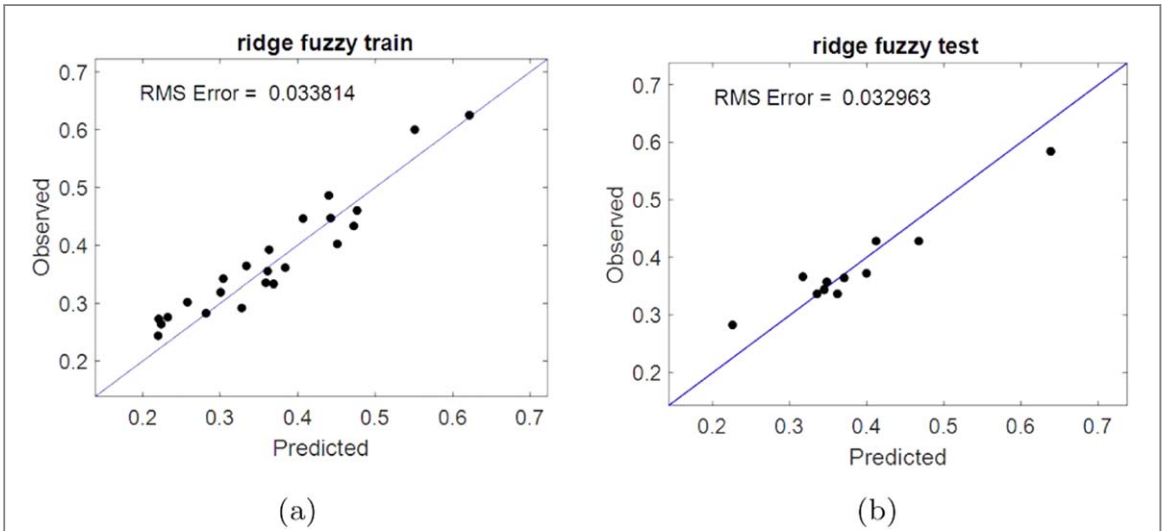

Figure 11. Ridge Fuzzy Modelling results.

Table 3. Linear model results in predicting 24 areal parameter values.

Output	Linear Model			
	Training (RMSE)		Testing (RMSE)	
	Ordinary	Ridged	Ordinary	Ridged
Sa (μm)	2e-15	0.070	0.531	0.058
S5z (μm)	9e-14	5.62	15.47	6.97
Std (deg)	1.7e-13	9.53	82.97	29.87
Smr2 (%)	3.5e-14	1.72	7.57	2.04
Smr1 (%)	1.6e-13	3.29	11.26	2.06
Svk (μm)	7.52e-14	0.236	1.013	0.124
Sk (μm)	9.38e-13	2.938	19.16	2.963
Spk (μm)	5.87e-14	0.357	0.907	0.426
Vvv ($\mu m^3/mm^2$)	3.99e-09	24148	183896	45 020
Vvc ($\mu m^3/mm^2$)	4.13e-08	200 705	803476	165 007
Vmc ($\mu m^3/mm^2$)	2.56e-08	108 141	342 019	83579
Vmp ($\mu m^3/mm^2$)	2.56e-09	20378	87845	26271
Sdr (%)	7.88e-15	0.0514	0.32129	0.033 43
Ssc ($1/\mu m$)	5.05e-16	0.00137	0.009 20	0.000 92
Sdq	3.27e-15	0.01	0.06993	0.00757
Sal (mm)	3.68e-15	0.0246	0.0712	0.0361
Str	1.07e-14	0.1842	0.30599	0.1697
Sds ($1/mm^2$)	5.42e-11	80.319	342	84.70
Sz (μm)	8.42e-13	7.1029	43.5086	8.7398
Sv (μm)	6.13-13	6.0068	58.2909	4.6297
Sp (μm)	2.93e-13	3.2261	20.5911	7.7376
Sku	4.20e-12	93.103	468.96	75.193
Ssk	1.34e-13	3.7123	33.302	4.8714
Sq (μm)	3.98e-14	0.1708	1.0248	0.1431

Table 4. Fuzzy model results in predicting 24 areal parameter values.

Output	Fuzzy Model			
	Training (RMSE)		Testing (RMSE)	
	Ordinary	Ridged	Ordinary	Ridged
Sa (μm)	3e-15	0.034	0.233	0.033
S5z (μm)	11e-14	2.98	17.21	3.24
Std (deg)	1.8e-15	12.13	93.07	13.80
Smr2 (%)	3.5e-14	1.34	6.54	1.86
Smr1 (%)	6.8e-14	2.12	12.08	2.86
Svk (μm)	2.88e-15	0.227314	0.72936	0.0409
Sk (μm)	13.59e-14	2.828	1.801	1.937
Spk (μm)	2.25e-15	0.343353	0.653594	0.1401
Vvv ($\mu m^3/mm^2$)	1.53e-10	23243.37	132405	14780.19
Vvc ($\mu m^3/mm^2$)	1.58e-09	193179.4	578503	54171.99
Vmc ($\mu m^3/mm^2$)	9.82e-10	104085.7	246254.1	27439.17
Vmp ($\mu m^3/mm^2$)	9.82e-11	19614.5	63249.06	25914.31
Sdr (%)	3.02e-16	0.0494	0.2313	0.0329
Ssc ($1/\mu m$)	1.94e-17	0.001323	0.00662	0.000908
Sdq	1.25e-16	0.009992	0.05035	0.007471
Sal (mm)	1.41e-16	0.023636	0.0044	0.0355
Str	4.11e-16	0.177273	0.0190	0.167451
Sds ($1/mm^2$)	2.08e-12	77.307	21.256	19.482
Sz (μm)	3.22e-14	6.8365	2.7018	2.010
Sv (μm)	2.35e-14	5.785	3.6198	1.0648
Sp (μm)	1.12e-14	3.101	1.2787	1.7796
Sku	1.61e-13	89.61	29.122	17.294
Ssk	5.13e-15	3.573	2.0680	1.1204
Sq (μm)	1.52e-15	0.1644	0.06364	0.0329

figure 11 can be replicated across all the other areal surface metrology parameter which indicates that the proposed modelling methodology predicts with accuracy regardless of the parameter of interest. Tables 3 and 4 respectively show the performances of linear/ridge regression method and the proposed fuzzy approach in predicting 24 areal parameters. The results from these tables indicate that the proposed approach is able to generalise across different areal parameters and provides consistent as well as robust modelling results.

As can be observed from tables 3 and 4, for the ordinary linear and fuzzy models (without penalising the weights), the models overfit significantly on the training data set and perform badly on the testing data set across all the 24 areal parameters. The training error is close to zero and this fact is - corroborated by figures 8 and 10. For ridge linear and fuzzy models, the results are better (improved modelling accuracy on the test data). For example, if one considers the Sa parameter in two tables mentioned, it can be seen that the training RMSE for both the ordinary linear and fuzzy models are negligible (2e-15 and 3e-15 respectively). The testing performance is respectively 0.531 and 0.233. Although the fuzzy model is better than the linear regression approach (for the ordinary model), there is overfitting on the training data set. The performance is much improved when utilising the proposed ridge approach. For example, the ridge ordinary fuzzy model has a training RMSE of 0.034 and a testing RMSE of 0.033 (shown in figure 11). The ridge approach is able to provide a balance in the accuracy of training and testing results. It should be noted that using the ridge approach on the testing data set, the fuzzy model is able to provide improvement on the modelling accuracy as compared to the linear modelling approach by approximately 75%.

6. Conclusion

This paper has presented a new framework which is based on ridge Mamdani fuzzy logic system for the mapping of process features to areal surface metrology parameters. The proposed approach represents a nonlinear but interpretable solution to the manufacturing informatics modelling problem. The main contribution of this paper is the development of a modelling solution which provides consistent accuracy across all the 24 areal parameters on which the results were tested. This is the first time such a framework has been validated across different areal parameters even in the face of a challenging, nonlinear, sparse, multidimensional modelling task. In particular, the validation results of the proposed strategy contrast existing areal parameters modelling methods where either results do not generalise across many areal parameters or validation results are difficult to obtain. The proposed approach may benefit from adding an extra layer of complexity (such as the use of type-2 fuzzy sets) so that uncertainties inherent in manufacturing systems can be adequately modelled as well as understood. This will be the main focus of future research studies.

Data availability statement

The data that support the findings of this study are available upon reasonable request from the authors.

Ethical Approval

Not Applicable.

Consent to Participate

Not Applicable.

Consent to Publish

Not Applicable.

Authors Contributions

Olusayo Obajemu analysed the datasets and carried-out the computer simulation experiments. **Moschos Papananias** and **Thomas E. McLeay** designed the laboratory experiments and architected the data acquisition/processing pipeline. **Mahdi Mahfouf** and **Visakan Kadirkamanathan** provided technical guidance and validation of the experiments/algorithms developed in the study.

Funding

The authors gratefully acknowledge funding for this research from the UK Engineering and Physical Sciences Research Council (EPSRC) under Grant Reference: EP/P006930/1.

Competing Interests

The authors have no conflicts of interest to declare that are relevant to the content of this article.

Availability of data and materials

Not Applicable.

ORCID iDs

Moschos Papananias https://orcid.org/0000-0001-7121-9681
Visakan Kadirkamanathan https://orcid.org/0000-0002-4243-2501

References

[1] Whitehouse D J 1994 *Handbook of Surface Metrology* (Boca Raton, FL: CRC Press)

[2] Jiang X J and Whitehouse D J 2012 Technological shifts in surface metrology *CIRP Ann.* **61** 815–36

[3] Mathia T G, Pawlus P and Wieczorowski M 2011 Recent trends in surface metrology *Wear* **271** 494–508

[4] Nicola S and Leach R 2018 Information-rich surface metrology *Procedia CIRP* **75** 19–26

[5] Grzesik W 2016 Prediction of the functional performance of machined components based on surface topography: State of the art *J. Mater. Eng. Perform.* **25** 4460–8

[6] Zhong Z W, Khoo L P and Han S T 2006 Prediction of surface roughness of turned surfaces using neural networks *The International Journal of Advanced Manufacturing Technology* **28** 688–93

[7] Grzesik W 2018 Prediction of surface topography in precision hard machining based on modelling of the generation mechanisms resulting from a variable feed rate *The International Journal of Advanced Manufacturing Technology* **94** 4115–23

[8] Lu W, Zhang G, Liu X, Zhou L, Chen L and Jiang X 2014 Prediction of surface topography at the end of sliding running-in wear based on areal surface parameters *Tribol. Trans.* **57** 553–60

[9] Hu Z-X, Wang Y, Ge M-F and Liu J 2019 Data-driven fault diagnosis method based on compressed sensing and improved multi-scale network *IEEE Trans. Ind. Electron.* **67** 3216–55

[10] Neuzil J, Kreibich O and Smid R 2013 A distributed fault detection system based on iwsn for machine condition monitoring *IEEE Trans. Ind. Inf.* **10** 1118–23

[11] Dartmann G, Song H and Schmeink A 2019 *Big Data Analytics for Cyber-Physical Systems: Machine Learning for The Internet of Things* (Amsterdam: Elsevier) (https://doi.org/10.1016/C2018-0-00208-X)

[12] Frank A G, Dalenogare L S and Ayala N F 2019 Industry 4.0 technologies: Implementation patterns in manufacturing companies *Int. J. Prod. Econ.* **210** 15–26

[13] Sharma A, Kumar S A and Kushvaha V 2020 Effect of aspect ratio on dynamic fracture toughness of particulate polymer composite using artificial neural network *Eng. Fract. Mech.* **228** 106907

[14] Worden K, Wieslaw S J and James H J 2011 Natural computing for mechanical systems research: A tutorial overview *Mech. Syst. Sig. Process.* **25** 4–111

[15] NFEN ISO. 4287. 1998. Geometrical Product Specifications (GPS). Surface texture: profile method. Terms, definitions and surface texture parameters, 2009.

[16] Lou M S, Chen J C and Li C M 1998 Surface roughness prediction technique for cnc end-milling *Journal of Industrial Technology* **15** 1–6

[17] Segreto T, Karam S and Teti R 2017 Signal processing and pattern recognition for surface roughness assessment in multiple sensor monitoring of robot-assisted polishing *The International Journal of Advanced Manufacturing Technology* **90** 1023–33

[18] Plaza E G, López P J N and González E M B 2019 Efficiency of vibration signal feature extraction for surface finish monitoring in cnc machining *J. Manuf. Processes* **44** 145–57

[19] Harris P M, Smith I M, Wang C, Giusca C and Leach R K 2012 Software measurement standards for areal surface texture parameters: part 2 comparison of software *Meas. Sci. Technol.* **23** 105009

[20] BSEN ISO. 25 178–2: Geometric product specification, surface texture (areal). part 2: Terms, definitions and surface texture parameters. British Standards Institute: London, UK, 2012

[21] Löberg J, Mattisson I, Hansson S and Ahlberg E 2010 Characterisation of titanium dental implants i: critical assessment of surface roughness parameters *The Open Biomaterials Journal* **2** 18–35

[22] Papananias M, McLeay T E, Mahfouf M and Kadirkamanathan V 2019 A bayesian framework to estimate part quality and associated uncertainties in multistage manufacturing *Comput. Ind.* **105** 35–47

[23] ISO 16 610-21. 16 610–21: Geometric product specification (gps)—filtration–part 21: linear profile filters: Gaussian filters. International Standards Organisation, 2011.

[24] Lotfi A 1965 Fuzzy sets *Inf. Control* **8** 338–53

[25] Obajemu O and Mahfouf M 2019 A dirichlet process based type-1 and type-2 fuzzy modelling for systematic confidence bands prediction *IEEE Trans. Fuzzy Syst.* **27** 1853–65

[26] Obajemu O, Mahfouf M and Catto J W F 2017 A new fuzzy modeling framework for integrated risk prognosis and therapy of bladder cancer patients *IEEE Trans. Fuzzy Syst.* **26** 1565–77

[27] Nabney I 2002 *NETLAB: Algorithms for Pattern Recognition* (Berlin: Springer)

[28] Richard L 2013 *Characterisation of Areal Surface Texture* (Berlin: Springer) (https://doi.org/10.1007/978-3-642-36458-7)

[29] Xueyi L I, Jialin L I, Yongzhi Q U and David H E 2019 Semi-supervised gear fault diagnosis using raw vibration signal based on deep learning *Chin. J. Aeronaut.* **33** 418–26