Measurement Science
and Technology

PAPER • **OPEN ACCESS**

Characterising and propagating the sources of error and uncertainty in the optical measurement of gears using designed experiments

To cite this article: Denis Sexton *et al* 2025 *Meas. Sci. Technol.* **36** 065014

OPEN ACCESS
IOP Publishing
Measurement Science and Technology
Meas. Sci. Technol. **36** (2025) 065014 (14pp)
https://doi.org/10.1088/1361-6501/addd40

Characterising and propagating the sources of error and uncertainty in the optical measurement of gears using designed experiments

Denis Sexton[1,*], **Sofia Catalucci**[1,2], **Andy Sharpe**[3], **Robert Frazer**[4] and **Samanta Piano**[1]

[1] Manufacturing Metrology Team, Faculty of Engineering, University of Nottingham, United Kingdom
[2] Department of Industrial Engineering (DII), University of Padua, Italy
[3] Department of Metrology and NDT, Manufacturing Technology Centre (MTC), United Kingdom
[4] National Gear Metrology Laboratory (NGML) Newcastle University, United Kingdom

E-mail: denis.sexton@nottingham.ac.uk

Received 22 October 2024, revised 19 May 2025
Accepted for publication 27 May 2025
Published 6 June 2025

Abstract

In recent years, the growth of optical techniques has introduced the possibility of allowing several alternative methods for non-contact gear measurement to be utilised. Optical methods can offer many advantages over tactile, such as the potential to evaluate delicate surfaces quickly and measure the whole area of the gear tooth flank at the sub-micron level. However, to maximise their potential, the magnitude of error and characterisation of the sources of error and uncertainty need to be understood. By utilising a series of designed experiments with known size gear artefacts, the effects caused by the change of specific key instrument parameters can be evaluated. These measurement trials demonstrate how the results from experimental methodologies can be used to determine the statistical significance of any predetermined instrument variables under study. When correctly planned, designed experiments allow the identification of the sources of error. By applying statistical methods, we can determine if these sources are significant or not. This will allow determination of which parameters need to be defined when optimising the conditions for measurement by comparing the results with those from the UK National Gear Metrology Laboratory. This comparison can also provide guidance on developing measurement uncertainty.

Keywords: investigation, errors, uncertainties, optical, measurements, gears, designed experiments

* Author to whom any correspondence should be addressed.

1. Introduction

Used to transmit motion and power, and to transfer torque and speed in mechanical and hydraulic systems, gears have a wide and growing use in many industrial and commercial applications as an essential part of our modern world. For equipment to work efficiently, gears must be well designed and used with suitable bearings and lubrication in a robust gearbox. From a metrology standpoint, the main requirement is that gears are measured as accurately as possible, so a good understanding of the most suitable measurement methods is required including a thorough study of all the consequent error and uncertainty sources. To comprehensively investigate dimensional errors and sources of uncertainty, knowledge of the manufacturing process is very advantageous. Traditional gear machining methods such as hobbing have been joined by modern computer numerical controlled (CNC) technologies, with software that can generate standard tooth profiles and modify geometries using only standard tools [1]. By reducing the need for specialised cutting tools, and taking advantage of CNC cycle times, costs can be considerably reduced. In order to analyse how gear deviations are affected by the geometric errors of a measuring machine, physical artefacts have been developed by, and shared between, the UK National Gear Metrology Laboratory (NGML) and the German Physikalisch-Technische Bundesanstalt (PTB). Artefacts for the evaluation of contact instruments in the measurement of gears are defined in the BS ISO 18653:2003 [2]. The software required to evaluate simulated geometric errors is available, and it is worth noting that by applying a reversal technique, it was found that more than 80% of the gear deviations stemming from geometric errors can be compensated for [3]. This could considerably reduce the errors in the measurement, and the need to develop and ship large physical artefacts.

In addition to well-established tactile solutions, other methods for measurement include the polar coordinate method [4], structured light systems [5], and focus variation by means of a confocal microscope [6]. Some recently deployed multi-sensor instruments offer the ability to measure interchangeably with hybrid (combined tactile and optical) methods. Other optical measurement methods include line structured light sensors generating three-dimensional (3D) point clouds in conjunction with a floating rotary table [7]. The captured data are then analysed utilising statistical methods specifically for the evaluation of measurement error [8]. While traditional methods for the evaluation of measurement uncertainty in gears are defined in ISO 18653, the general reference document for uncertainty is JCGM Guide 100:2008 [9] supported by the UKAS guide to uncertainty [10]. It is possible to assign task-specific measurement uncertainty through simulation software as defined in ISO/TS 15530-4:2008 [11] and ISO 14253-1:2017 [12]. Gear measurement uncertainty simulations can make use of a virtual coordinate measuring machine and employ Monte Carlo methods for uncertainty propagation [13]. The latter are considered as the most common solutions in CMM measurements, as they can handle very complex measurement sequences. While the ISO gear inspection standard [14] does not currently define any optical techniques, these methods can offer several advantages over tactile ones, such as potentially higher throughput allowing for more of the surface (flank) of the tooth to be measured. If the whole flank is measured, a predictor model can be generated for wear on gear teeth and could be utilised as a digital twin to aid gear design [15].

There are several types of optical systems for gear measurement, and some can create areal deviation 3D maps of the whole tooth surface and use a meshed point cloud [16]. One commercial example of hybrid technology is offered by the Hexagon® HP-O optical sensor. This system combines a frequency-modulated interferometric optical distance measurement sensor within a Leitz coordinate measuring machine (CMM). With this system, no point cloud is created, but rather the tactile probe is replaced with a sensor which gathers points from a single beam of laser light. For the work in this paper, a cylindrical spur gear was provided by the NGML, and a series of measurement trials were undertaken with the HP-O sensor to explore the relationship between some key optical instrument parameters. The parameters considered were point density (pd) and scan speed (ss). Designed experiments (DOE/DOX) [17] were chosen as the central methodology to explore error and potential sources of uncertainty by studying the effects caused by the deliberate changes to the aforementioned parameters. The statistical methodologies presented here can determine not only the error in the measurement of gears but can be utilised for further research beyond the scope of this paper. Results gathered from these studies have been verified by correlation back to the results obtained from the gear artefacts provided and measured by the UKAS approved tactile measurement method at the NGML (UKAS registered laboratory 2363).

2. Applying DOE methodologies to gear measurement

The use of designed experiments to explore specific uncertainty components has previously been investigated [18], but much less so for the optical measurement of larger module (larger than 3 mm) gears, and not with the system chosen. The novelty of this work is the addition of experimental control charts 'analysis of means' (ANOM) & 'analysis of range' (ANOR) [19] to explore the nature of induced changes in the instrument settings and to determine the nature of the variation on the measurement results. A key to developing uncertainty budgets is not only to understand the particular sources (or input variables) within the measuring system, but also to explore the nature of any applicable sensitivity coefficients. While it is common to make use of partial differentiation for this purpose [10], the assumptions of changing one input while keeping all others constant is often not practical because it is not possible to guarantee that all inputs are kept constant. Designed experiments may be more suitable to determine the

effects of the sources of variation as DOE can investigate multiple input (and output) variables, and these methods do not require the assumptions of partial differentiation to be met. Any assumptions that are made can be statistically tested *post-hoc*, as shown in section 3.2.1.

By utilising designed experimental theory with various instruments across known artefacts, the equations which define the relationship between the instrument settings or the key process input variables (KPIV) can be determined, and the measurement results or key process output variables (KPOV) [19] will allow the measurement process optimisation of any instrument under study through traceability of the results back to the primary source (NGML). By removing any sources of instrument variation which are not statistically significant, a reduced or simpler uncertainty budget can be developed containing only those key instrument variables (also called sources or factors) which were identified as potentially significant. Once determined, the key sources defined as the major sources of error can be populated in an initial design matrix. This procedure and the definitions of statistical significance are described in section 3.

The HP-O sensor was chosen in this experimental setup. In the frequency modulated interferometric optical distance measurement method [20] on which this sensor is based, the measured distance L can be estimated by:

$$L = c \cdot \Delta\Phi \,/\, (2 \cdot B \cdot n_g) \tag{1}$$

where B is the optical frequency scanning range of the laser, $\Delta\Phi$ is the phase change of the interference signal, c is the speed of light in vacuum, and n_g is the refractive index of air. The 3 mm diameter of the HP-O (HPOAL) sensor and the measurement range of up to 20 mm can achieve repeatability of under 0.3 µm, when used with the Leitz PMM-C CMM. The sensor can measure difficult-to-access features at a scanning speed of 1000 points per second and a resolution of 0.9 nm [21]. Optical measurements can be captured in single points or in scanning mode for rapid throughput. The sensor is easy to mount on an indexable head, as the weight is less than 190 g. One disadvantage is given by the relatively small acceptance angle. For a mid-range sensor with a working distance of 10.5 mm, the acceptance angle is approximately $\pm 30°$ and for a reflecting surface it is $\pm 1°$. For this reason, this system might not be suitable for some types of surfaces (i.e. sharp threads). For these gear studies, a fairly large module (3.9 mm), non-polished gear was chosen, while the CMM provided by the Manufacturing Technology Centre (MTC) incorporated an integrated rotary table.

2.1. Gear setup on the CMM

A 29-tooth spur gear with a module of 3.9 mm (supplied by the NGML) was mounted on a Leitz PMM-C CMM at the UK MTC as shown in figure 1.

The first series of trials were designed to look at the repeatability of the optical sensor. To set a baseline, the gear was first aligned and measured twice with a conventional tactile ruby sphere of 3 mm diameter and best practice was followed where

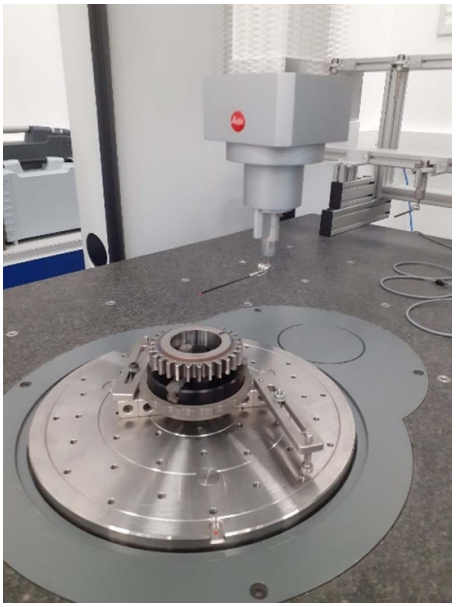

Figure 1. NGML spur gear measured with a Hexagon HP-O sensor on a Leitz PMM-C CMM at the MTC.

applicable [22]. The first measurement set was made with the tactile probe but without the use of the rotary axis, while the second tactile run was completed with the rotary axis. The profile scan speed was initially set at 2 mm s^{-1} with 0.005 mm distance between points, while the lead (or helix) was scanned at 1 mm s^{-1} with a maximum distance of 0.0005 mm between points. The order of tooth scan was as follows; left flank profile, left flank lead, then right flank profile followed by right flank lead. The line of scan for profile, lead and pitch was the same for both tactile and optical sensors and is shown in figure 2(a), while the line of contact for meshing gears (both helical and spur contact) is shown in figure 2(b).

Each sensor was utilised in a single orientation (tactile sensor A axis $= 0°$, B axis $= 0°$ and optical sensor A axis $= 90°$, B axis $= 0°$). The rotary axis (R) was required during all optical trials, due to indexing head constraints with the optical sensor. Initially, a tactile/optical comparison was not possible across the full depth of the tooth, but only from the reference diameter $d = 113.10$ mm to the tip diameter $d_2 = 120.90$ mm. The width of the teeth was 20 mm, and a lead evaluation range of 14 mm was chosen to allow 3 mm of clearance at each end of the gear tooth. Following successful repeatability trials for both tactile and optical methods, experimental work involving the optical scans was undertaken as presented in this paper. The gear software utilised was Quindos® (available from Hexagon Metrology). It is noteworthy that a paradigm shift from line orientated measurements and evaluations to a holistic area orientated gear inspection method has been proposed [23], however this work is ongoing and outside the scope and inspection methodology defined in this paper.

2.2. Gear characteristics

The mandatory gear characteristics for reporting include profile (FHα, ffα and Fα), lead (FHβ, ffβ and Fβ), individual and

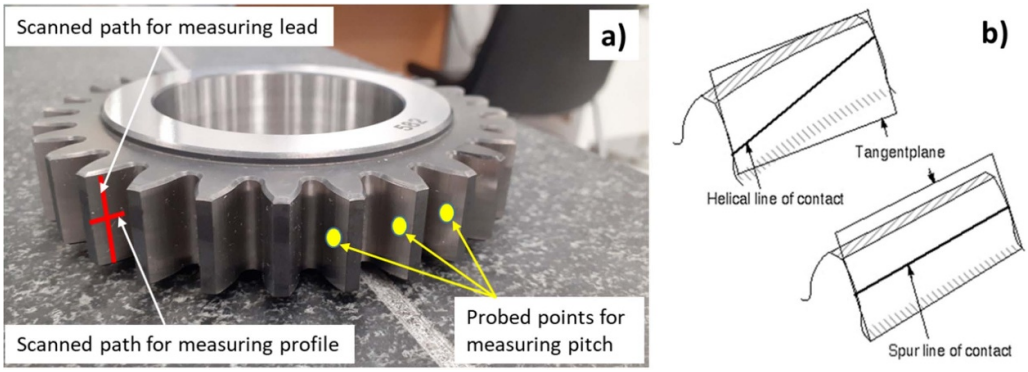

Figure 2. (a) Gear profile (involute), lead (helix) and pitch, (b) line of contact for helical and spur gear.

cumulative pitch (Fp,fp), and radial runout (Fr) as defined in ISO 1328-1 [24]. Profile and lead are measured on both flanks of four teeth approximately 90 degrees apart (on teeth identified as 1, 8, 16 and 23). Individual and cumulative pitch are measured across both flanks of all teeth. Runout is measured across 360 degrees of the gear diameters back to the datum axis. When measuring profile and helix, a scanned line is utilised with a minimum number of 300 points as recommended by the NGML. This requirement does not cause any issues for most modern CMM's. In addition, a guidance on data filtering is defined [14] and was followed (as indicated in section 4.5). However, when considering pitch or runout this is not a necessary consideration, since gear pitch is not normally measured as a scanned line but as a series of points (refer to figure 2(a)). For this study, the pitch was measured in an unconventional manor as a series of multiple scans across all teeth, allowing point density and scan speed to be changed and evaluated during the experiments. It seemed sensible to consider pitch and concentricity/runout characteristics as a starting point since some of the initial results for profile and lead were inconsistent and because the preferred HP-O (HPOAM) sensor was damaged, and no compatible replacement was available.

2.3. ANOM and ANOR charts

When evaluating the repeatability of measured results (i.e. data recorded in time series under the same input settings), statistical process control (SPC) charts are most commonly used [25]. However, when analysing experimental data, a particular pair of charts called ANOM and ANOR are utilised [19]. The detection limits applied on the ANOM & ANOR charts look like the control limits on a SPC chart, but they differ slightly. SPC charts are used for data characterised by routine while experimental data are characterised by uniqueness. This is why SPC charts have a potential shortcoming as tools for analysing experimental data. Industrial experiments will generally involve the exploratory analysis of a limited amount of data that is, *a priori,* thought to contain real differences. SPC is set up for the analysis of ongoing streams of data that, hopefully, contain no real differences. So, if a SPC chart is used to analyse experimental data, those differences identified as potential signals are likely to represent real effects (though some real differences may be overlooked). In conclusion, the ANOM/ANOR charts differ from traditional SPC charts in two physical aspects: (a) They are limited to a finite number of subgroups, and (b) they require the specification of an overall alpha level for the procedure (in this case 0.1 or 10% for ANOM and 0.05 or 5% for the ANOR). The first of these differences prevents the user from using these techniques with production data. The second of these differences lets the user adjust the sensitivity of the procedure. ANOM tests whether the ten (in this case) individual treatment means differ from the overall mean (also called the grand mean). As with most other SPC charts, the first chart ANOM plots central tendency or location, while the lower ANOR chart plots dispersion or spread.

3. Results

3.1. ANOM and ANOR example

The maximum individual pitch error (fpi) from the left flanks of a 29-tooth spur gear with a normal module of 3.9 mm was measured three times under ten different experimental conditions on the CMM with the HPOAL sensor. Two different point densities and five different scan speeds were chosen to evaluate the effects of change observed upon the results. The three measurements were recorded at each of the ten coded combinations of point density (pd) and scan speed (ss). Point density was coded with A and B levels (where A is 100 points mm^{-1} and B is 20 points mm^{-1}), while scan speed was identified with five levels coded 1, 2, 3, 4, and 5 (where level 5 is 5 mm s^{-1}, down to level 1 which is 1 mm s^{-1}). The recorded results are shown in table 1.

From the obtained data, three questions can be asked:

1. How does the point density affect the values of maximum individual pitch error (fpi)?
2. How does the scan speed affect the values of maximum individual pitch error (fpi)?
3. Does any identified interaction between point density and scan speed affect the results?

Table 1. Maximum individual pitch error recorded (μm).

Point density (pd)	A	A	A	A	A	B	B	B	B	B
Scan speed (ss)	5	4	3	2	1	5	4	3	2	1
Result fpi 1 (μm)	3.50	3.10	3.00	2.80	1.90	3.80	3.60	3.90	3.50	3.00
Result fpi 2 (μm)	3.90	3.70	3.10	2.00	1.80	4.60	4.40	3.20	4.70	3.80
Result fpi 3 (μm)	3.60	3.60	3.30	2.30	2.20	4.10	3.90	3.80	4.00	3.10
Mean fpi (μm)	3.67	3.47	3.13	2.37	1.97	4.17	3.97	3.63	4.07	3.30
Range (μm)	0.40	0.60	0.30	0.80	0.40	0.80	0.80	0.70	1.20	0.80

The data in table 1 was populated into the ANOM/ANOR charts as shown in figures 3 and 4 respectively. The calculation of the detection limits is similar to the computation of the control limits for a traditional SPC chart. The grand average obtained from the data reported in table 1 is equal to 3.373 μm (as shown on the ANOM chart), and the average range is 0.68 μm (as shown on the ANOR chart). With an overall alpha level of 10%, the ANOM scaling factor for $k = 10$ subgroups of size $n = 3$ (taken from ANOM statistical tables) is 0.893. The ANOM detection limits are calculates as follows:

$$\begin{aligned}&\text{ANOM Detection Limits (LDL \& UDL)}\\&\quad = \text{Grand Average} \pm \text{ANOM}_{.10}\,(\text{Average Range})\\&\quad = 3.373\,\mu\text{m} \pm 0.893\ (0.68)\\&\quad = 2.766\ \mu\text{m}\ (\text{LDL})\ \text{and}\ 3.981\,\mu\text{m}\ (\text{UDL})\end{aligned} \tag{2}$$

The ANOR proceeds in a similar manner. With an alpha level of 5%, and with $k = 10$ and $n = 3$, the ANOR scaling factor for $k = 10$ subgroups of size $n = 3$ (taken from ANOR statistical tables) is 0 2.519. The ANOM detection limit is as follows:

$$\begin{aligned}&\text{ANOR Upper Detection Limit (UDL)}\\&\quad = \text{ANOR}_{.05}\ (\text{Average Range})\,\text{UDL}\\&\quad = 2.519\ (0.68) = 1.713\ \mu\text{m}\,.\end{aligned} \tag{3}$$

There is no lower detection limit for this ANOR chart.

It can be observed from the range ANOR chart that each group has similar 'within' subgroup, variation, meaning only common or random cause variation is present [19]. We observe the ten group subgroup ranges, while the mean of all ranges is 0.68 μm with the upper detection limit of 1.713 μm. There is no lower detection limit for this ANOR chart. On the ANOM chart, each data point reported is the mean of the three readings for each of the ten subgroups as indicated in table 1. It can be further observed that the overall mean is 3.373 μm, and while subgroups 2B and 5B are above the upper detection limit (3.981 μm), subgroups 1A and 2A are below the lower detection limit (2.766 μm). This indicates that the two groups identified as A and B (point density) are different. Group A has significantly lower values than group B meaning higher point density results in lower values, while the lower scan speed (the numbered subgroups) generally results in lower values. Therefore, both point density and scan speed are significantly different between levels. Significance is calculated and defined numerically in section 3.2.1. Since only two predictor variables are being considered, only one interaction is present. This interaction effect can be exploited or avoided as necessary only after it has been visualised. In the ANOM chart, the interaction can be visualised by looking at the line between 2A and 3A and comparing it with the line between 2B and 3B. These lines are almost perpendicular showing that interaction is present at this point. If numerical values (p values) are required to define the significance of the factors and the interaction, this can be conducted via an analysis with ANOVA (refer section 3.2 onwards).

So, to answer the questions:

- How does the point density affect the results? *Moving from higher (A) to lower (B) point density levels leads to significantly higher values for maximum individual pitch error.*
- How does the scan speed affect the results? *Moving from higher (5) to lower (1) scan speeds levels leads to significantly lower values for maximum individual pitch error.*
- Does the interaction between point density and scan speed affect the results? *There is one point of interaction between the two-point densities, and it occurs between scan speeds of levels 2 A and 3 A when compared to 2B and 3B (see* figure 3 *and bottom left of* figure 5).

To determine which reported values are more representative of the true size under study, correlation was obtained from the NGML measurements via their Klingelnberg P65 instrument. This instrument is UKAS approved with known error and uncertainties. The 'true' maximum pitch error was obtained and is discussed in 3.2.3. The initial purpose of this study was simply to determine if point density and scan speed (and their interaction) were statistically significant on the results obtained from one specific gear artefact, measured with one specific sensor HPOAL, on one instrument (a Leitz CMM), at one location (the MTC), and at one point in time.

3.2. Initial orthogonal screening matrix

The ANOM/ANOR charts test whether the treatment means differ from the grand average (or overall mean). When utilising an orthogonal matrix in conjunction with ANOVA techniques [17], it is possible to test whether multiple treatment means differ significantly from each other. To explore the effects of changing the parameters on measurement results, an orthogonal matrix was developed, and is illustrated in table 2. Since the output results from gear trials include many gear characteristics, the maximum individual pitch error data (fpi) from the previous study shall again be considered.

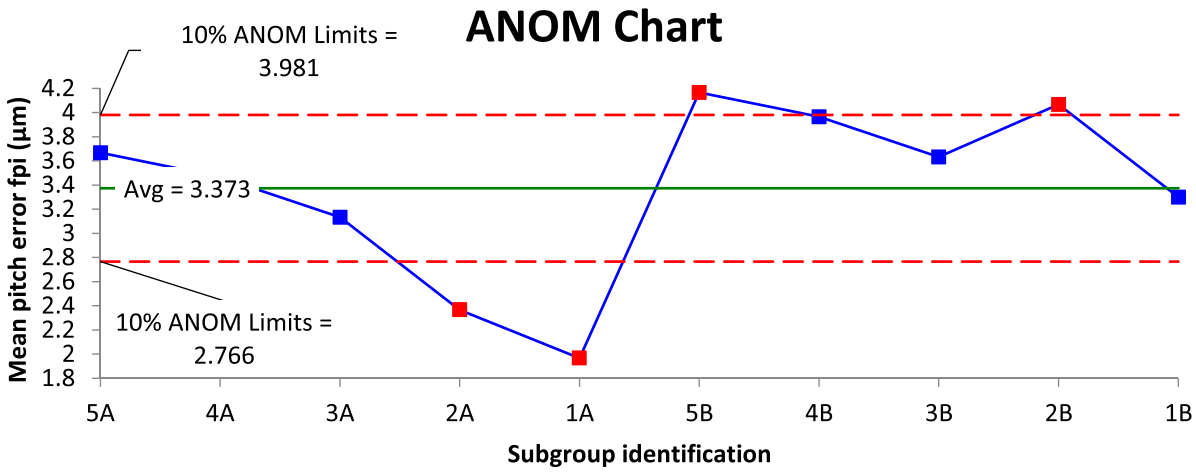

Figure 3. ANOM Chart for subgroup mean values (groups 2A, 1A, 5B, and 2B are outside the detection limits).

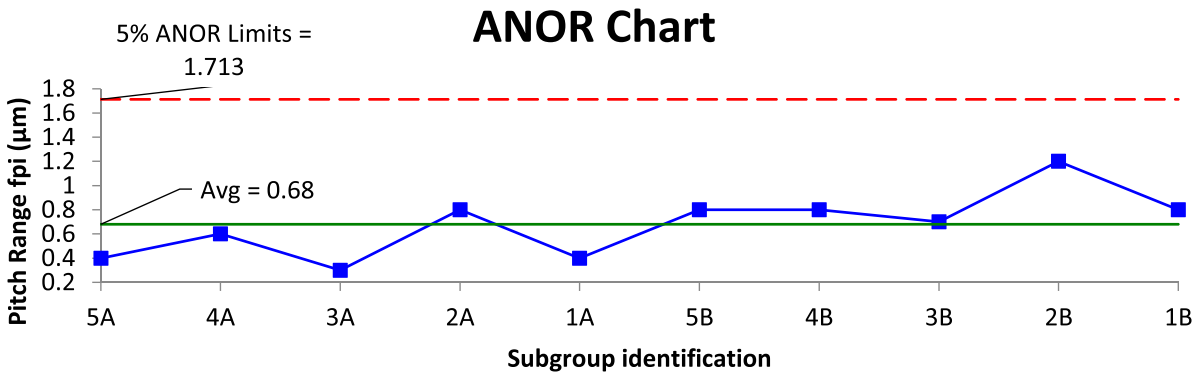

Figure 4. ANOR dispersion Chart for individual pitch subgroup range values.

3.2.1. Analysis via factorial ANOVA. Table 2 has the same output values as in table 1 but it was generated in Minitab® statistical software. It is only necessary to specify the number of factors, the levels of factors, and the number of replications for the matrix to be created. The blocks represent a complete set of readings for each trial. Since there were three readings per trial, we have three blocks of data. Normally when changing the experimental conditions, it is recommended to do this in a random order, and this order can be generated by the software. In this study, the existing data from table 1 was copied into the appropriate cell in the final column of table 2. Table 3 shows the numerical analysis from Minitab statistical software utilising ANOVA [17].

The analysis model in table 3 shows significant difference (defined as P values < 0.05) for point density, scan speed, and their interaction confirming the findings of the ANOM and ANOR charts. P values are used extensively with ANOVA (and sometimes to determine the normality of data) so it may be helpful to briefly explain how they are calculated. In table 3, the sum of squares for all sources and interaction (Adj SS) are added in quadrature to give a total of 16.8387 at 29 $(n-1)$ degrees of freedom. Next, consider the two-way interaction source between point density and scan speed. The F-Value is calculated by first dividing the adjusted mean square (Adj MS) for this source (0.4903) by the adjusted mean square assigned to error (0.1286). This provides a value of 3.81 as indicated. Statistical significance at the defined confidence interval can be determined either by comparing this value to the F ratio (F#) indicated in a set of F statistical tables, or by calculating the value directly with the F distribution. Since our source has 4 $(n-1)$ degrees of freedom, and our error term has 18, we can input these with our calculated value into a cell in Microsoft Excel® with command ' = FDIST (3.81,4,18)'. This function calculates the P value. The calculated value is 0.02 (rounded). All P values are statistically significant when they are smaller than 0.05 (the 95% confidence interval). Since both point density and scan speed have values lower than 0.01 (the 99% confidence interval) they are both highly significant in terms of their influence on the measurement results. The data gathered clearly demonstrates how results from various measurement sources and methods could be compared. To see if the results of the optical and tactile results differ significantly, the graphical and numerical techniques outlined here would be utilised to make this determination. Designed experiments could be applied to investigate various sources of instrument variation to see if such sources influence results and explore new sensitivity coefficients. DOE/DOX has an advantage over the partial

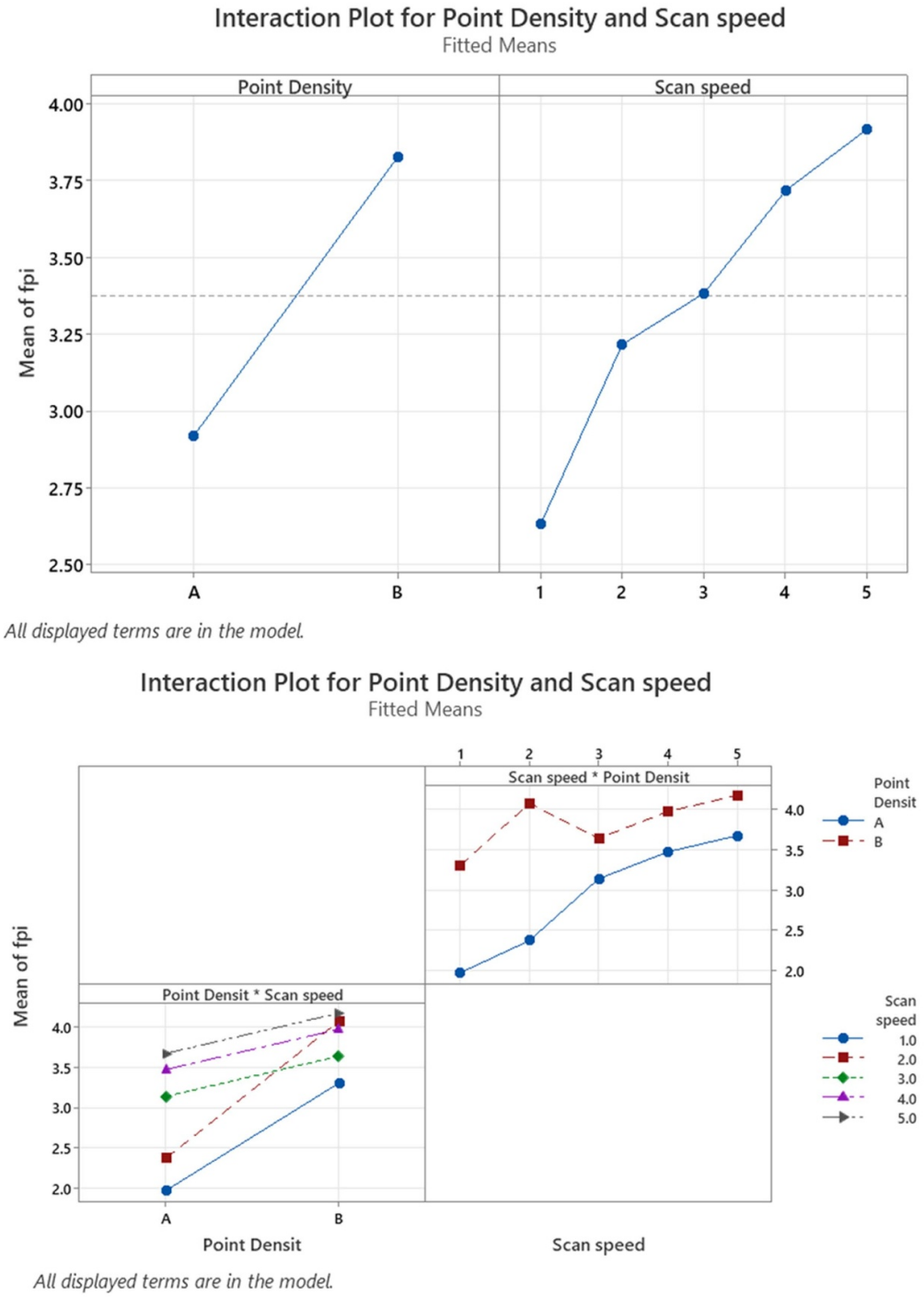

Figure 5. Graph of the mean main effects and interaction analysis in Minitab® software.

derivative method [10], since the various assumptions associated with ANOVA can (and always should) be tested using various *post hoc* analysis options [17] as shown in the analysis in figure 6. These assumptions include that the variation within each subgroup is homogeneous, no serial correlation is observed between means and standard deviations, and that the residuals are normally distributed within (but not necessarily between) groups. These conditions were all met during early trials.

3.2.2. Regression analysis. Regression analysis is applied as a part of the output statistics from DOE [17]. The key output from this is the *R*-Square value (or R^2) and is calculated from the sum of squares for the model. Our model in table 3 has 11 degrees of freedom consisting of the sum of squares for point density (6.1653) scan speed (5.9120) and the interaction between them (1.9613). We combine these terms (14.0386) and divide by the total sum of squares (16.8387). This ratio is then multiplied by 100 to obtain a percentage

Table 2. Screening matrix for CMM measurement trials on spur gear.

Trial	Blocks	Point density	Scan speed	fpi (μm)
1	1	B	1	3.00
2	1	B	2	3.50
3	1	A	1	1.90
4	1	A	5	3.50
5	1	B	5	3.80
6	1	A	3	3.00
7	1	B	3	3.90
8	1	A	4	3.10
9	1	B	4	3.60
10	1	A	2	2.80
11	2	B	3	3.20
12	2	B	2	4.70
13	2	A	1	1.80
14	2	A	2	2.00
15	2	B	5	4.60
16	2	A	5	3.90
17	2	A	3	3.10
18	2	B	4	4.40
19	2	B	1	3.80
20	2	A	4	3.70
21	3	B	4	3.90
22	3	A	4	3.60
23	3	A	3	3.30
24	3	A	5	3.60
25	3	B	3	3.80
26	3	B	1	3.10
27	3	B	2	4.00
28	3	A	2	2.30
29	3	B	5	4.10
30	3	A	1	2.20

Table 3. ANOVA analysis in Minitab statistical software.

Source	DF	Adj SS	Adj MS	*F*-Value	*P*-Value
Model	11	14.5233	1.3203	10.26	0.000
Blocks	2	0.4847	0.2423	1.88	0.181
Linear	5	12.0773	2.4155	18.78	0.000
Point density	1	6.1653	6.1653	47.93	0.000
Scan speed	4	5.9120	1.4780	11.49	0.000
2-Way interactions	4	1.9613	0.4903	3.81	0.020
Point density*scan speed	4	1.9613	0.4903	3.81	0.020
Error	18	2.3153	0.1286		
Total	29	16.8387			

value. In this case the value is 83.37%. R^2 is the coefficient of determination. This measures that proportion of total variation about the output mean (y) explained by the regression (or 83.37% of the total variation about the average or mean). Since each comparison consumes 1 degree of freedom ($n-1$), we have 29 degrees of freedom in total. It is worth noting, that had we defined the model with all the available degrees of freedom and all the sources, we would have a 'perfect' theoretical model with a R^2 value of 100%, but then we would not be able to distinguish between the source and any error. Those sources and interactions with the greatest influence (or values) are normally assigned to the model, while the rest are assigned as error. Where our R^2 value remains reasonably high (>60%) this decision may be justified. If the R^2 value is lower than expected, unknown or unassigned sources may be present inside the error term. Sometimes lower R^2 values are observed when trying to fit a linear regression to a data set which non-linear. This can sometimes be the case where only two levels for each source (x) are applied (as is common in DOE). Refer to section 3.2.4 on testing for linearity.

3.2.3. Regression for process optimisation. One of the most useful purposes of designed experiments is the ability

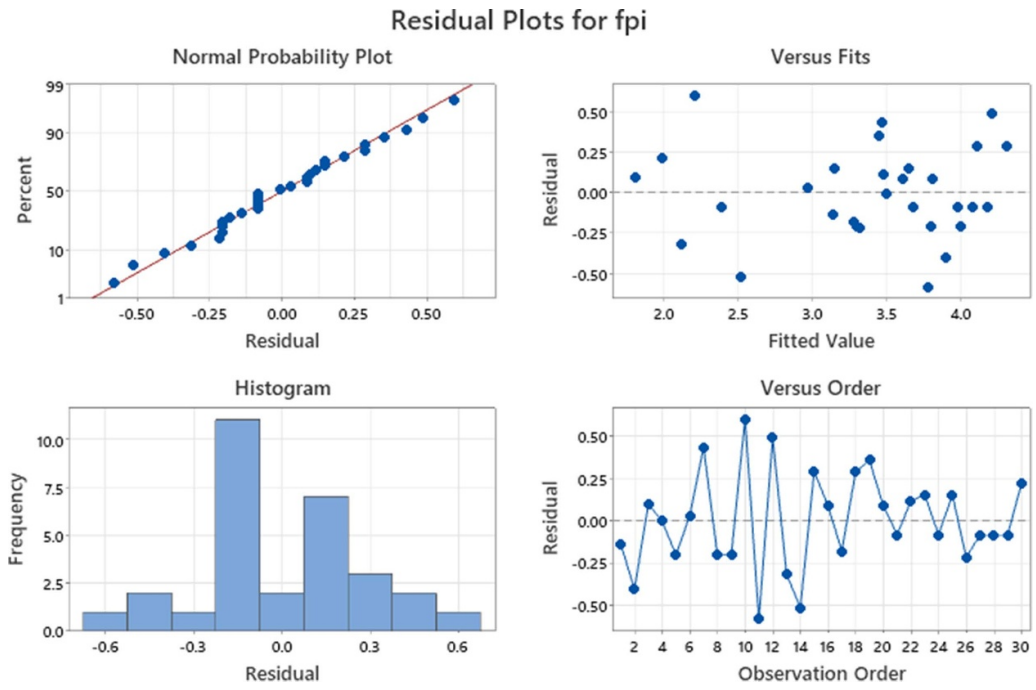

Figure 6. Post hoc residual analysis.

to fully understand the cause-and-effect relationship between the inputs and output(s). Having determined that scan speed and point density are statistically significant (defined as P values < 0.05) we can explore this relationship. One output equation from a DOE is a multiple regression which aims to find a relationship between variables in situations where there are several independent variables. The independent (or input) variables (x) can be either continuous or qualitative (or both) however, the dependent (or response) variable (y) must be measured on a continuous scale. As explained in [26], a multiple regression model with k independent variables will fit a regression surface in a $k+1$ dimensional space. The least squares regression line for multiple regression of n independent variables. Since y is defined as a function of x in this relationship, we can write the equation (4):

$$y = a + b_1x_1 + b_2x_2 + \ldots b_nx_n \tag{4}$$

where:

- a is a constant,
- x_n is the nth independent variable,
- b_n is the coefficient of the nth independent variable.

As the name implies, a bivariate model is a model with two independent variables, as expressed in equation (5). This also describes a multiple regression model. As explained in [26], the intercept a predicts where the plane will cross the y-axis. The value b_1 is gradient of the variable x_1, this predicts y with every change in unit of x_2 whilst x_1 is constant. The gradient of the variable x_2, b_2, predicts y with every change in unit of x_2 whilst x_1 is constant. This applies to our model in exactly the same way,

$$y = a + b_1x_1 + b_2x_2. \tag{5}$$

In our example, the pitch (fpi) of a master gear was measured on a CMM with an optical sensor. If pd represents the first independent variable (point-density) and ss represents the second independent variable (scan speed), then:

- 3.3733 µm is the grand average or overall mean (refer to figure 3)
- pda is the first independent variable at level A
- pdb is the first independent variable at level B
- ss_1 is the second independent variable at level 1
- ss_2 is the second independent variable at level 2
- ss_3 is the second independent variable at level 3
- ss_4 is the second independent variable at level 4
- ss_5 is the second independent variable at level 5

$$\begin{aligned} y = {} & 3.3733 - 0.4533\text{pda} + 0.4533\text{pdb} - 0.740\text{ss}_1 \\ & - 0.157\text{ss}_2 + 0.01\text{ss}_3 + 0.343\text{ss}_4 + 0.543\text{ss}_5 \\ & - 0.213\text{pdssA}_1 - 0.397\text{pdssA}_2 + 0.203\text{pdssA}_3 \\ & + 0.203\text{pdssA}_4 + 0.203\text{pdssA}_5 + 0.213\text{pdssB}_1 \\ & + 0.397\text{pdssB}_2 - 0.203\text{pdssB}_3 - 0.203\text{pdssB}_4 \\ & - 0.203\text{pdssB}_5. \end{aligned} \tag{6}$$

This knowledge can allow optimisation of the inputs for any desired output. The regression equation could be used when adjusting the predictor settings of point density and scan speed to instrument parameters that would most closely align with NGML results. The value of the maximum pitch error (fpi)

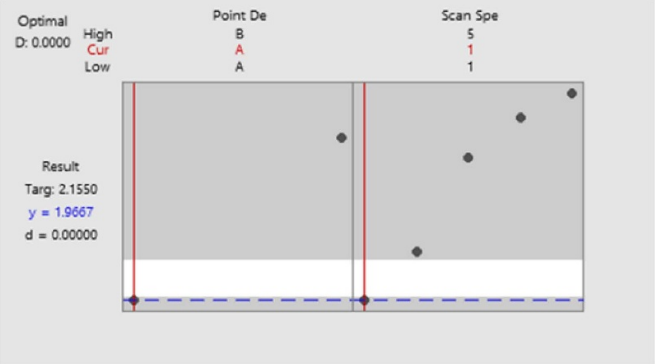

Figure 7. Process optimisation utilising regression model in Minitab®.

for the gear when measured on the Klingelnberg P65 tactile instrument at the NGML provided a result of 2.155 µm. We can observe that the closest value to the NGML result was 1.97 µm. This occurs at point density A and at scan speed 1 (table 1 and figure 3). Closer adjustment could not be achieved because the predictor variables (scan speed and point density) were coded as discrete (or categorical) variables. If the inputs were recorded as variables on a continuous scale (as was the output variable), more exact values for point density and scan speed which would give the desired output of 2.155 µm could have been determined. Experiments such as response surface methodologies [17] are useful for this purpose, and they could be executed in any future phases of experimental analysis if required. Had the ideal settings not been as obvious as in this case (for example, if the input parameters were multiple levels on a continuous scale), process optimisation analysis can be determined by the regression model as shown in figure 7. Here Minitab statistical software confirms the result we determined in our earlier example.

3.2.4. Testing for linearity. When only two levels of a variable (i.e. point density) are utilised, it may not always be obvious to see if any relationships are non-linear. For example, consider the following relationship:

$$y_o = 0,\ x_o \approx 0, \text{ and:}$$
$$y_4 = 4,\ x \approx 8.$$

Then if the relationship is linear, we would expect to see:

$$\text{When } y_2 = 2,\ x \approx 4$$

Most statistical software can check this mathematically quite easily. If the relationship is suspected to be non-linear, we would require more than two levels per source variable to determine if non-linearity was present (and what type) when developing a non-linear regression model to maximise our R^2 value.

3.2.5. DOE for the determination of sensitivity coefficients. UKAS M3003 [10] states that the sensitivity coefficient associated with each input estimate *xi* is represented by *ci*. It is the partial derivative of the model function (X) with respect to Xi, evaluated at the input estimates *xi*. It is given by,

$$c_i = \frac{\partial f}{\partial x_i}\,|_{X_i = x_i} \approx \frac{\partial y}{\partial x_i}. \tag{7}$$

In other words, it describes how the output estimate y varies with a corresponding small change in an input estimate *xi*. If the functional relationship is not well known for a particular measurement system, or it cannot easily be differentiated, the sensitivity coefficients can usually be obtained by the practical approach of changing one of the input variables by a known amount, whilst keeping all other inputs constant, and noting the change in the output estimate. As described earlier, the problem with this approach is that in practice it may not be possible to keep all other inputs constant, and it is also not possible to know the extent of the relationship (or interactions) between multiple input variables. This is not a problem if we utilise DOE theory, since we can explore both the effects of each input both independently and in any interactions with other inputs. This would allow a determination as to if the effects are significant at a known confidence interval, as well as develop a mathematical relationship between inputs and outputs.

3.2.6. Development of sensitivity coefficients for point density and scan speed on displacement. The most common sensitivity coefficient is the coefficient of thermal expansion (CTE). Here, the relationship of temperature of the measurand (commonly in units of °C) are converted into dimensional deviation or displacement (often in µm). The linear CTE for steel typically falls within the range of 10.8–12.5 $\times$ 10^{-6}/°C. CTE is just one of a number of sensitivity coefficients to consider for metrology applications. Using DOE methodologies in these studies, it was possible to develop the relationship between point density (pd) and scan speed (ss) on the measurement variation in gear geometry via traceability to the NGML UKAS registered laboratory. It may be desirable to develop the significant terms for other parameters generated from the regression equation individually as sensitivity coefficients. While different measurement systems will have potentially differing sources of variation and differing parameters which could affect the input to output relationship, the methodologies demonstrated here could assist in determining such relationships. Future studies to investigate such relationships in this and other optical systems are ongoing.

4. Conclusions and future work

4.1. Exploring the uncertainty of a sensor within the measurement system

Having determined the optimal conditions for point density and scan speed (and assuming we do not identify any other significant variables within the system), it would be desirable to consider how an uncertainty budget could be developed. This would need sensor specific results to be added

in quadrature with the other established uncertainty sources from the measurement system. In one sense this may be easier for the HP-O system than other optical systems for two reasons. Firstly, methods to determine the sources of volumetric error and uncertainty within the *X, Y* and *Z* axes and rotary axis of (CMM's have already been established [13, 27–31], and secondly, as this sensor does not create a point cloud, any such uncertainty considerations would not be applicable. It would still be necessary to explore the variation in measurement by measuring the same output characteristic multiple times (perhaps thirty times), to establish the repeatability of results in time series. Plotting these results onto a standard SPC chart [25] would establish if random or special cause variation was present, and this would determine the scale of variation as well as which type of uncertainty to assign. The methods already developed for calculating maximum permissible errors in CMM probing systems [32] could be helpful for this purpose.

4.2. Advantages of using DOE, ANOM, ANOR and ANOVA

The results discussed in section 3 report optical measurements with the HP-0 (HPAOL) sensor. Regardless of the measurement system, the statistical methodology presented demonstrates how the results from any measurements could be compared. To see if the results of the optical and tactile sensors differ significantly, the graphical and numerical techniques outlined here can be utilised to make this decision. To see which methodology gave results closer to the true value, assistance from the NGML was required. The proposed methodology for comparisons between point density and scan speed within one sensor shows the potential to be applied in any tactile/optical measurement comparisons, as well as a method to determine and compensate for any errors by verification back to master gear artefacts. In addition, those sources (or independent instrument input variables) which have a statistically significant effect on the displacement response(s) can be observed, and thus allow determination of which inputs need to be controlled to ensure the results which would most closely align with the results from the primary source. Finally, when the mathematical relationship between scan speed and point density with the measurements taken with the optical probe are understood, it should be possible to develop sensitivity coefficients for the dimensional displacement observed on the measured gear (or any other) artefact.

4.3. Probing limitations for the study

Due to physical constraints of the setup, it was not initially possible to scan the full length of each gear tooth as required by the standard [24], hence these results are a comparison between tactile NGML measurement and optical (HP-O) measurement over the same tooth area (as much as was achievable within the instrument constraints). For this trial, the HPOAL sensor was used, and some results were inconsistent for some output parameters. It was felt that the HPOAM sensor would give more consistent results, and this will be the focus of future trials. It should be noted that to keep the paper within constraints, limited data was included to demonstrate how DOE/DOX investigate the effects of how changing the instrument parameters can be utilised to correlate results with those obtained from the NGML. When using DOE and ANOVA to explore sensitivity coefficients, there are advantages, but the assumptions of ANOVA should always be tested using the various *post hoc* analysis. These conditions were tested for and met during early trials and demonstrated graphically in figure 6.

4.4. Larger module gear trials

At the time of writing, further trials are ongoing with a 9-tooth spur gear with a module of 5.744 mm. These trials are utilising the HP-OS-9010T optical sensor mounted on a Leitz PMM-C (HP-S-X1H) CMM as shown in figure 8. Like the tactile sensor, the optical sensor chosen will be utilised in a single orientation (A axis $= 0°$, B axis $= 0°$) and as with the MTC machine, this CMM has an integrated rotary axis. As per the previous trials, the gear is first aligned and measured twice with a conventional tactile ruby sphere (5 mm diameter), and again best practice followed. In addition to the X1H sensor, the larger module of this gear will allow greater access for light and thus can be better utilised for tactile/optical comparisons. And, as with the previous gear, the NGML will provide the measurement traceability. Initially studies will explore radial characteristics at various speeds and point densities. Following this, future work is planned around lead/helix.

4.5. Data filtering

Measurement data resides in the spatial domain, but measurement consists of instrument signals of various frequencies. The exclusion of certain portions of the measurement data frequency spectrum is called filtering. Gear profile and helix measurement data are usually conditioned by low pass (high frequency) filtering prior to analysis procedures. Filtering is normally instrument dependent and can be either mechanical, electrical or mathematical. Most modern filtering is applied via the software. The filtering applied to gear profile (involute) and lead (helix) requires a high 50% gauss filter [14]. This can be applied through the Quindos® CMM software. Since only pitch was considered in this paper, filtering was not relevant.

4.6. Observations

As the measurement results recorded in this work represent variable data on a continuous scale, analytical and enumerative statistical techniques were applied. The results shown in this paper (and in other experiments) proved that the measurement process was statistically stable. When looking at utilising ANOM and ANOR charts, it was possible to observe if the scanning speed and point density of the CMM probing had an influence on the measurement results. It was clear graphically that both variables had an influence, and that interaction was present. This means that both factors need to be considered as significant sources when developing sensitivity coefficients

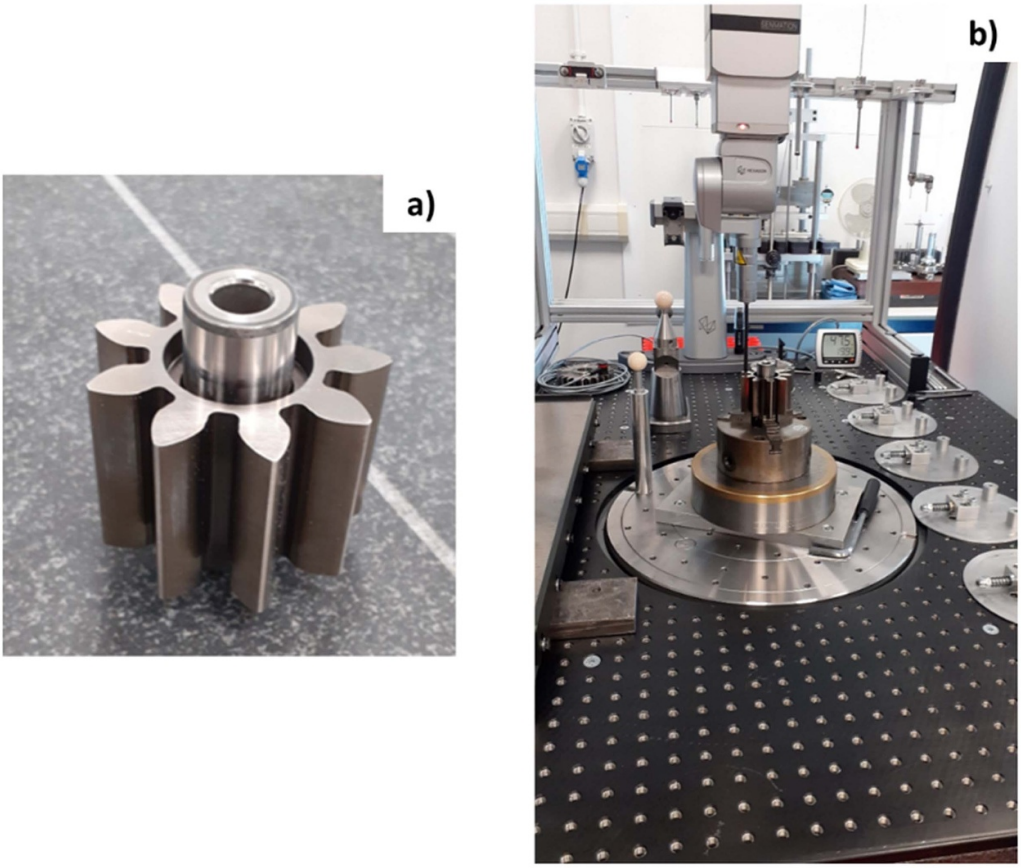

Figure 8. (a) The 9-tooth module 5.744 mm spur gear and (b), mounted in chuck of the rotary table of a Leitz, PMM-C (HP-S-X1H) CMM with HP-OS-9010T optical sensor.

(*ci*) for any uncertainty budget. Again, this can be determined in conjunction with known size artefacts. When looking at the orthogonal matrix and with the use of modern statistical software, various graphical and numerical outputs can be created and studied. Multiple inputs and outputs can be evaluated within one data set. The influence of multiple sources of variation within the instrument could be tested to determine if they were significant (or not) using p values in ANOVA. The DOE/DOX process allows for process optimisation which could be utilised to correlate the results obtained back to any primary sources. Also, by removing non-significant error sources (and interactions) from any experimental results, the uncertainty budget can be simplified because if the effects of any input source (or interactions) are not significant, then the uncertainty of that source cannot be significant either.

4.7. Digitalisation of metrology

Digitisation is the process of converting information from a physical format into a digital one, ultimately improving business competitiveness. Like many other sectors in business, industry and society, digitisation applies to metrology. In this context, there are many advantages to the way that measurement data is exploited and disseminated, extracted and utilised in the same way as in any design and manufacturing process. Traditional processes are still widely employed and most have already been digitised, including SPC, quality planning, FMEA, DOE and many others. Thanks to the drive of industry 4.0, measurement procedures and quality control processes have been optimised by enabling real-time execution of metrological analyses focused on critical parameters [33]. Digital manufacturing and smart measuring technologies support the zero-defect manufacturing concept, by moving from off-line metrology to in-line measurements and automated inspection systems [34, 35]. Compared to conventional off-line approaches, in-line measurement provides significant benefits, including decreased inspection durations through the elimination of laborious steps, real-time process quality monitoring, improved production speed, and the integration of fully automated manufacturing systems [36]. The use of robotic systems has added to the capabilities of measurement devices which can address a range of tasks autonomously. Automated inspection has also allowed for multiple features and characteristics to be measured by a single instrument thus removing the need for many different offline technologies. The shape and surface texture information of a measured part are representative of the process characteristics and the actual performances of the machine tool employed. As such, as reported by Gao *et al* [34], the integration of manufacturing and measuring operations is beneficial to the production process.

Data availability statement

All data that support the findings of this study are included within the article (and any supplementary files).

Acknowledgments

We would like to thank the Manufacturing Technology Centre (MTC) and the UKRI Research England Development (RED) Fund for funding this work via the Midlands Centre for Data-Driven Metrology. We would also like to acknowledge the NGML for their ongoing support for this project.

Conflict of interest

The authors declare that they have no known competing financial interests or personal relationships that may have influenced the work reported in this study.

CRediT authorship contribution statement

Denis Sexton: Writing—original draft, conceptualization, Investigation. **Sofia Catalucci**: Writing, review and editing. **Andy Sharpe**: Supervision and review. **Robert Frazer**: Supervision and review. **Samanta Piano:** Supervision, review, editing, funding.

ORCID iDs

Denis Sexton https://orcid.org/0000-0002-8187-9058
Sofia Catalucci https://orcid.org/0000-0003-3558-5902
Robert Frazer https://orcid.org/0009-0002-8103-9734
Samanta Piano https://orcid.org/0000-0003-4862-9652

References

[1] Gołębski R and Boral P 2021 Study of machining of gears with regular and modified outline using cnc machine tools *Materials* **14** 2913
[2] ISO 18653 2003 *Evaluation of Instruments for the Measurement of Individual Gears* (International Organisation for Standardisation)
[3] Lin H, Keller F and Stein M 2020 Influence and compensation of CMM geometric errors on 3D gear measurements *Measurement* **151** 107110
[4] Yin P, Han F, Wang J, Lu C and Du H 2022 Influence of coordinate system establishment error on tooth profile deviation measured by standard polar coordinate method on gear measuring center *Measurement* **187** 110344
[5] Shang Z, Wang J, Zhao L, Du H, Yin P and Zhang Y 2022 Measurement of gear tooth profiles using incoherent line structured light *Measurement* **189** 110450
[6] Danzl R, Helmli F and Scherer S 2011 Focus variation—a robust technology for high resolution optical 3D surface metrology *Strojniški vestnik—J. Mech. Eng.* **2011** 245–56
[7] Guo X, Shi Z, Yu B, Zhao B, Li K and Sun Y 2020 3D measurement of gears based on a line structured light sensor *Precis. Eng.* **61** 160–9
[8] Wang Y, Catalucci S, Senin N and Piano S 2024 Assessing the spatial distribution of positional error associated to dense point cloud measurements using regional Gaussian random fields *Measurement* **227** 114194
[9] JCGM 100 2008 GUM 1995 with minor corrections *Evaluation of Measurement data-Guide to the Expression of Uncertainty in Measurement* (BIPM)
[10] UKAS 2024 M3003 the expression of uncertainty and confidence in measurement
[11] ISO/TS 15530 part 4 2008 *Geometric Product Specifications (GPS)—Coordinate Measuring Machines (CMM): Technique for Determining the Uncertainty of Measurement—Part 4: Evaluating Task-Specific Measurement Uncertainty Using Simulation* (International Organization of Standardization)
[12] ISO 14253 part 1 2017 *Geometric Product Specifications (GPS) Inspection by Measurement of Workpieces and Measuring equipment—Part 1: Decision Rules for Verifying Conformity or Nonconformity with Specifications* (International Organization of Standardization)
[13] Rost K, Wendt K and Härtig F 2016 Evaluating a task-specific measurement uncertainty for gear measuring instruments via Monte Carlo simulation *Precis. Eng.* **44** 220–30
[14] ISO 10064 part 1 2019 *Code to Inspection Practice—Part 1: Measurement of Cylindrical Gear Tooth Flanks* (International Organization of Standardization)
[15] Lu C, Wang J, Yin P and Wang L 2021 Error identification of measurement software based on digital twin of gear measuring center *Measurement* **173** 108666
[16] Peng Y, Ni K and Goch D G 2018 Areal evaluation of involute gear flanks with 3D surface data (Gear solutions) (available at: https://gearsolutions.com/features/areal-evaluation-of-involute-gear-flanks-with-3d-surface-data/)
[17] Del Vecchio R J 1977 *Understanding Design of Experiments* (Carl Hanser Verlag GmbH Co KG)
[18] Gauder D, Gölz J, Biehler M, Diener M and Lanza G 2021 Balancing the trade-off between measurement uncertainty and measurement time in optical metrology using design of experiments, meta-modelling and convex programming *CIRP J. Manuf. Sci. Technol.* **35** 209–16
[19] Wheeler D J 1990 *Understanding Industrial Experimentation* (SPC Press)
[20] Shi G, Hei K, Wang W and Bhattacharya N 2019 Frequency-modulated continuous-wave laser distance measurement system using Fabry-Perot cavity as measuring reference (arXiv:1901.01131v1)
[21] Lavrinov D S, Khorkin A I and Privalova E A 2020 Comparative analysis of automatic methods for measuring surface of threads of oil and gas pipes *Advances in Automation, RusAutoCon 2019, (Lecture Notes in Electrical Engineering*) ed A Radionov and A Karandaev (Cham)
[22] Flack D 2011 *Good Practice Guide No. 42 CMM Verification* (National Physical Laboratory)
[23] Goch G, Ni K, Peng Y and Guenther A 2017 Paradigm change in gear inspection based on a holistic description, measurement and evaluation of gear flanks *ASPE Conf. (Charlotte, NC USA*)
[24] ISO 1328 part 1 2013 *ISO System of Flank Tolerance Classification—Part 1: Definitions and Allowable Values of Deviations Relevant to Flanks of Gear Teeth* (International Organization of Standardization)
[25] BS 5702 part 2 2008 *Guide to Statistical Process Control (SPC) Charts for Variables-Part 2: Charts for Individual Values* Permission Except As Permitted By Copyright Law British Standard
[26] Online Resources Numeracy, maths and statistics—academic skills kit (University of Newcastle) (available at: www.ncl.ac.uk/webtemplate/ask-assets/external/maths-resources/statistics/regression-and-correlation/multiple-regression.html)

[27] Wang Q and Goch G 2017 Geometric error measurement and compensation of a rotary table using a circular ball plate artefact *ASPE Conf.* (*Charlotte, NC USA*)
[28] Wang Q, Miller J, Von Freyberg A, Steffens N, Fischer A and Goch G 2018 Error mapping of rotary tables in 4-axis measuring devices using a ball plate artefact *CIRP Ann.* **67** 559–62
[29] Linares J M, Goch G, Forbes A, Sprauel J M, Clément A, Haertig F and Gao W 2018 Modelling and traceability for computationally-intensive precision engineering and metrology *CIRP Ann.* **67** 815–38
[30] Wang Q *et al* 2019 A performance investigation of a large-scale rotary table under loads *Doctoral Thesis* University of North Carolina at Charlotte (available at: http://hdl.handle.net/20.500.13093/etd:2190)
[31] Wang Q, Peng Y, Wiemann A-K, Balzer F, Stein M, Steffens N and Goch G 2019 Improved gear metrology based on the calibration and compensation of rotary table error motions *CIRP Ann.* **68** 511–4
[32] ISO 10360 part 5 2020 Geometrical product specifications (GPS) *Acceptance and Reverification Tests for Coordinate Measuring Systems (CMS)—Part 5: Coordinate Measuring Machines (CMMs) Using Single and Multiple Stylus Contacting Probing Systems Using Discrete Point and/or Scanning Measuring Mode* (International Organization of Standardization)
[33] Catalucci S, Thompson A, Piano S, Branson D T and Leach R 2022 Optical metrology for digital manufacturing: a review *Int. J. Adv. Manuf. Technol.* **120** 4271–90
[34] Gao W, Haitjema H, Fang F Z, Leach R K, Cheung C F, Savio E and Linares J M 2019 On-machine and in-process surface metrology for precision manufacturing *CIRP Ann.* **68** 843–66
[35] Leach R K 2020 Integrated metrology 10-year roadmap for advanced manufacturing (HVM Catapult)
[36] Takaya Y 2013 In-process and on-machine measurement of machining accuracy for process and product quality management: a review *Int. J. Autom. Technol.* **8** 4–19