

Physics-guided machine learning for trochoidal milling: force and power prediction and limits of cross-diameter generalisation

Gorka Urbikain Pelayo

gorka.urbikain@ehu.es

Universidad del Pais Vasco <https://orcid.org/0000-0002-7159-8199>

Daniel Olvera-Trejo

Tecnológico de Monterrey: Tecnologico de Monterrey

Luis Norberto López de Lacalle

Alex Elías-Zuñiga

Tecnológico de Monterrey: Tecnologico de Monterrey

Research Article

Keywords: Trochoidal milling, Machine learning, Boosted regression trees, Physics-informed modelling, Cutting forces, Spindle power, Process monitoring.

Posted Date: October 16th, 2026

DOI: <https://doi.org/10.21203/rs.3.rs-11046489/v1>

License: © ⓘ This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

Abstract

The growing demand for lightweight aerospace structures has made trochoidal milling an increasingly important strategy for the high-efficiency machining of large pockets and thin-walled features in structural frames. Specifically mechanistic force prediction in trochoidal milling requires a continuously updated description of cutter–workpiece engagement and uncut chip thickness, which complicates its application across varying toolpath and tool geometries. This paper presents an experimental and data-driven analysis of trochoidal milling of AA7075-T651 aluminium alloy using three end-mill diameters (8, 10 and 12 mm). Cutting forces and spindle power were acquired synchronously with a piezoelectric dynamometer and the machine monitoring system across 72 tests, and post-processed in MATLAB to extract five robust indicators: mean in-plane force ($F_{xy,mean}$), maximum in-plane force ($F_{xy,max}$), robust peak-to-peak in-plane force ($F_{xy,ptp}$), maximum axial force ($F_{z,max}$) and mean spindle power (P_{mean}). Three physics-informed input structures – a minimal 5-variable model, a hybrid 5 + 2 model adding the maximum uncut chip thickness (h_{max}) and a simplified severity descriptor (MRR^*), and an enriched 5 + 4 model – were compared using boosted regression trees (LSBoost) with 5-fold cross-validation. The hybrid 5 + 2 model gave the best overall compromise, improving R^2 from 0.241 to 0.364 for $F_{xy,mean}$ and from 0.374 to 0.473 for P_{mean} relative to the minimal model, while the enriched model added redundancy (correlations up to $r = 0.93$) without consistent gains. Average, system-level responses were substantially more predictable than peak-based ones: all three structures returned negative R^2 for $F_{z,max}$, showing that test-level descriptors cannot capture the transient, local origin of axial force peaks. The results indicate that compact, physically motivated feature sets can support preliminary screening of trochoidal milling conditions without recourse to larger or less interpretable models.

Highlights

- Force–power monitoring characterises trochoidal milling of AA7075-T651.
- 72 tests across D8, D10 and D12 mm tools quantify diameter and engagement effects.
- Physics-informed descriptors h and MRR^* improve force and power prediction.
- A hybrid 5 + 2 boosted-tree model outperforms minimal and enriched input structures.
- Mean spindle power is best predicted ($R = 0.473$); axial force peaks resist modelling.

1. Introduction

Manufacturing companies face increasing pressure to raise productivity while satisfying tightening energy and resource-efficiency requirements, a tension central to the emerging paradigm of data-driven, sensor-instrumented "smart manufacturing" [1], and particularly visible in the roughing of high-strength aerospace alloys, where cycle time, tool loading and energy demand must be balanced simultaneously.

Comparing slot, plunge and trochoidal milling of representative aerospace roughing features, Urbikain Pelayo et al. showed that no single strategy simultaneously maximises material removal rate and minimises specific power consumption, and that trochoidal milling can yield a markedly lower specific-productivity-gain factor than serrated-tool slotting despite its favourable engagement control – underlining that the force and power behaviour of trochoidal milling cannot be assumed a priori and must be characterised experimentally [2].

Trochoidal milling was formalised analytically by Otkur and Lazoglu, who coupled an engagement model with a cutting-force model and validated predicted forces for both circular and more complex trochoidal geometries [3]. Because the cutter follows a looped path alternating cutting and non-cutting arcs, trochoidal strategies substantially reduce radial engagement relative to full-immersion slotting while sustaining high material removal rates, widely exploited for high-speed pocketing of difficult-to-cut materials [4]. Building on this foundation, Pleta and Mears characterised cutting forces in trochoidal milling of a nickel-based superalloy [5] and, with Akhavan Niaki, showed that undeformed chip thickness evolves continuously along the loop and requires a dedicated kinematic treatment rather than a circular approximation [6]; cutting-force coefficients identified from conventional slot milling were further shown not to transfer accurately to trochoidal conditions unless reference parameters are chosen to reproduce the trochoidal chip thickness [7], since coefficients calibrated under the constant engagement of conventional slotting do not represent the continuously evolving chip thickness intrinsic to the trochoidal loop. Ducroux et al. extended this analysis to Inconel 718 [8], while Rodriguez-Alabanda et al. showed that a stretched trochoidal toolpath with a variable-helix tool reduces cutting forces and specific energy in deep-slot milling of the aerospace alloy EN AW 2024-T3, extending these benefits to aluminium alloys [9]. A complementary research direction has addressed toolpath design and engagement control: Li et al. extended trochoidal-like motion to five-axis flank milling of deep cavities [10], Kozový et al. quantified the influence of trochoidal parameters on tool load [11], and Shah et al. linked in-process edge-force coefficients to tool-wear area estimation [12]. More recently, Li et al. proposed a parameter-variant trochoidal toolpath that adapts engagement angle to remain chatter-free while improving efficiency [13], and Nguyen et al. combined machine learning with multi-criteria optimisation to select trochoidal parameters for AISI 4140 steel, improving surface roughness, tool wear and cutting force simultaneously [14]. A comprehensive 2026 review by Guo et al. confirms that trochoidal milling of difficult-to-machine materials continues to be of scientific and industrial relevance, but notes that most studies remain confined to a single tool diameter and material, limiting the generality of reported trends and motivating campaigns that explicitly span several diameters [15].

The physical interpretation of trochoidal force and power signals rests on classical milling mechanics, where cutting speed, feed per tooth, axial depth of cut and radial engagement conventionally describe productivity, tool loading and machining outcomes [16–18], and cutting forces relate to instantaneous chip thickness through empirically identified coefficients. Engin and Altintas generalised this mechanistic approach to arbitrary helical end mills [19] and inserted cutters [20] by integrating local edge forces along the flute while accounting for true milling kinematics, the same framework underlying the stability-lobe methodology later introduced by Altintas and Budak for chatter-free milling [21]. For

toolpaths with continuously varying engagement, such as trochoidal loops, the cutter-workpiece engagement region must be recomputed along the path; Zhang et al. recently proposed a fast, local-Boolean-operation method for extracting this engagement in multi-axis milling, reducing the computational cost of force prediction for such trajectories [22]. These engagement- and chip-thickness-based quantities supply the mechanistic quantities – radial and axial engagement ratios, orbital-diameter ratio, maximum uncut chip thickness – used later in this paper to build physically interpretable descriptors of the data-driven model (Section 2.1).

In parallel with force modelling, synchronised sensor-based monitoring of cutting force, spindle power and related process signals has become a mature enabling technology for process characterisation, as documented in two CIRP Annals keynote reviews issued a decade apart: Teti et al. surveyed sensor technologies, signal-processing methods and decision-making strategies for machining process monitoring in 2010 [23], and a 2022 follow-up confirmed that increasingly intelligent, multi-sensor monitoring systems remain central to machining research, now closely integrated with data-driven analytics [24]. This is particularly relevant to trochoidal milling, where cutter-workpiece engagement and chip thickness vary continuously within each loop, so that synchronised, high-resolution monitoring is needed to resolve a process severity that a single steady-state measurement would average away. Duro et al. proposed a multi-sensor data-fusion framework that automatically identifies the most informative sensor locations and signal features for CNC monitoring [25]. Spindle power is a particularly attractive signal because it is system-level and low-instrumentation-cost: Denkena et al.'s CIRP Annals keynote on energy-efficient machine tools positions power monitoring as both an operational-efficiency indicator and an energy-accounting tool [26], and Pawar et al. proposed a physics-based model of spindle energy consumption in CNC milling linking spindle load directly to cutting severity [27]. These findings motivate the joint acquisition of force and power signals adopted here, in which power supports robust segmentation of consecutive tests while force provides the primary mechanical-load indicators.

Data-driven modelling increasingly complements these mechanistic and sensor-based approaches, though Wuest et al.'s review of machine learning in manufacturing identifies small, unbalanced experimental datasets as a persistent obstacle [28]. This limitation is most directly relevant to the present dataset: Wu et al. found random-forest and boosted-tree ensembles particularly robust on small datasets for tool-wear prediction [29], a robustness echoed by Ebrahimi and Niknam, who trained tree-based and deep-learning models to predict surface roughness and cutting force in aluminium milling under explicitly data-scarce conditions and found ensemble tree methods competitive with, or superior to, larger neural architectures when few tests are available [32], closely matching the scale, material and boosted-tree approach adopted here. Force predictability itself is not uniform across targets or conditions: Xie et al. showed that cutting-force predictability varies substantially with cutter-workpiece engagement stage and signal complexity across different modelling methods [40], a pattern directly echoed in Section 5.6 for the present dataset. Beyond force/power prediction specifically, machine learning has also been applied to related machining problems under similarly constrained data, including tool-wear classification in sustainable milling of additively manufactured steel [41], instantaneous cutting-force prediction by deep learning [30], extreme gradient boosting for cutting-force prediction [31],

remaining-useful-life estimation for milling cutters [42], and sensor-integrated tool-condition monitoring [33]; the boosted-tree ensembles used throughout these studies, and in the present work, build on the gradient-boosting framework formalised by Friedman [34]. What remains largely unaddressed by this body of work, however, is a systematic comparison of input representations, rather than of learning algorithms alone, across more than one tool geometry or size: even the aluminium-specific study closest to the present campaign reports results for a single tool diameter [32]. Physics-informed machine learning, which uses mechanistic knowledge to select, engineer or constrain a data-driven model's inputs rather than discarding it in favour of a purely black-box approach, is an increasingly effective way to narrow this gap under limited data, as identified by a 2026 review from Ra et al. [35]; directly relevant examples in milling include Pashmforoush et al., who integrated a thermo-mechanical force model with machine learning for tool-wear prediction, reporting a sizeable accuracy gain over a purely data-driven model [36], and Araghizad et al., who combined physics-based simulation with machine learning to predict milling forces accurately across previously unseen materials and tool types [37]. At the machine-tool level, Bergs et al.'s CIRP Annals keynote on digital twins for cutting processes [38] and Kim and Okwudire's physics-based, data-driven digital twin for feedrate optimisation [39] illustrate the same principle at a systems scale: a physical model combined with a learned correction tends to outperform either component alone while remaining more interpretable and data-efficient than a fully black-box model.

Taken together, this body of work indicates that (i) trochoidal milling is an established but still actively developed high-performance roughing strategy whose force and power response depends jointly on tool diameter, engagement and chip-load severity [3–15]; (ii) mechanistic and engagement-based descriptors provide a physically grounded vocabulary for interpreting this response, especially for toolpaths with continuously varying engagement [16–22]; yet mechanistic coefficients calibrated under one engagement regime do not automatically transfer to another [7], and recomputing engagement continuously along a trochoidal path is itself computationally demanding [22], motivating a complementary data-driven approach that learns the aggregate response directly from measurements; (iii) synchronised force and power monitoring is a mature but still evolving enabling technology [23–27]; and (iv) machine learning, and physics-informed machine learning in particular, is an increasingly effective way to model machining processes from limited experimental data, but has so far been applied mainly to tool wear, surface roughness or single-diameter force prediction rather than to a systematic, multi-diameter comparison of physically motivated input structures in trochoidal milling [28–42]. This gap is the specific target of the present study: to analyse and model high-performance trochoidal milling of AA7075-T651 aluminium alloy using synchronised cutting-force and spindle-power measurements across three tool diameters, with the specific objectives of acquiring and post-processing a consistent multi-diameter dataset, extracting robust and comparable force/power indicators, constructing and comparing minimal, hybrid and enriched physics-informed input structures for boosted-regression-tree models, and identifying which combination of descriptors offers the best compromise between predictive accuracy and interpretability for preliminary process screening. Section 2 describes the experimental methods; Section 3 examines the influence of tool diameter and physics-informed

descriptors on the measured responses; Section 4 details the machine learning framework; and Section 5 presents and discusses the results, before Section 6 concludes the paper and outlines future work.

2. Materials and experimental methods

2.1 Trochoidal milling parameterization

Trochoidal milling generates the tool-centre path as a cyclic combination of a circular motion and a forward translation, so that the cutter alternates between engaged and non-engaged arcs instead of remaining in full-immersion contact as in conventional slotting. The strategy is described through a set of four geometric parameters: the trochoidal (orbital) radius R , or equivalently the orbital diameter $D_{orb} = 2R$ (the width of the slot is $w = D + D_{orb}$), the feed per tooth f_z , the axial depth of cut a_p and the radial engagement a_e , which is the advanced length between two successive cycles (or the difference between the forward step s and backward step s'). Because engagement evolves continuously along each loop, the uncut chip thickness also varies continuously, reaching a maximum value h_{max} that is used throughout this work as a compact indicator of local chip-load severity.

Figure 1 illustrates four trochoidal cycles or repetitions, with the tool progressing along the slot direction, X . Each cycle consists of four basic movements: (1) a G2 circular interpolation, during which the tool periphery generates the envelope curve and performs the active cutting; (2) a backward linear movement (distance $s's'$) using G1 interpolation, with the tool rubbing against the right-hand wall of the slot; (3) a lateral repositioning movement in the Y direction (also G1); and (4) a forward linear movement (distance ss) using G1 interpolation, with the tool rubbing against the left-hand wall.

To compare tests performed with different tool diameters on a common basis, three dimensionless descriptors are defined: the normalised axial engagement λ (as commonly the total maximum cutting length is $3D$), the normalised radial engagement μ (as tool diameter D is the maximum ae), and the orbital-diameter ratio δ (as it should be selected according to D):

$\lambda = a_p / (3D)$	Eq. 1a
$\mu = a_e / D$	Eq. 1b
$\delta = D_{orb} / D$	Eq. 1c

The chip volume per tooth can be represented by a simplified descriptor as:

$$MRR^* = a_p a_e f_z \quad \text{Eq. 2}$$

which should be read as a compact indicator of relative cutting intensity rather than a complete volumetric material removal rate, since it omits the feed-rate term explicitly. Nominal spindle speed and feed rate were derived from the selected cutting speed v_c and feed per tooth f_z as $n = 1000v_c/(\pi D)$ and $v_f = f_z \cdot z \cdot n$, with z the number of teeth.

A second useful descriptor is the sweep angle, i.e. the angular range over which a cutting edge remains engaged with the workpiece during a cutting event. In conventional milling this angle is governed mainly by radial immersion and tool geometry; in trochoidal milling it becomes strongly path-dependent, because engagement is shaped by the instantaneous relative motion between tool and workpiece along the loop, and it can vary within a single trochoidal period even when the programmed parameters are constant. Sweep angle and h_{max} therefore provide a compact, complementary description of engagement severity and are used here as physical background for interpreting the measured force and power responses, rather than as separate experimental classes.

2.2 Experimental plan

The experimental campaign was organised as a controlled dataset in which tool diameter and selected cutting/toolpath parameters (a_e , a_p , f_z , D_{orb}) were varied while the machined feature geometry, a straight channel milled into a 260 mm x 260 mm AA7075-T651 plate, was adapted per test so that each channel spanned a comparable number of trochoidal cycles rather than a fixed absolute length (the machined length itself is not a controlled variable); channel width followed the swept tool envelope, $w = D + D_{orb}$ (Section 2.1, Fig. 1). Three tool diameters were tested ($D = 8, 10$ and 12 mm) with 24 tests per diameter, for a total of 72 trials. Within each diameter, a_e , a_p , f_z and D_{orb} were varied within bounds representative of high-performance roughing of AA7075-T651, so that the resulting dimensionless descriptors λ , μ and δ (Eq. 1) spanned a comparable range across tool sizes. This normalisation is what subsequently allows tests performed with different tools to be pooled into a single modelling dataset (Section 4) instead of being treated as separate experimental blocks. Tool diameter is therefore best understood as a blocking factor superimposed on a continuous variation of engagement- and feed-related parameters, rather than as one level of a classical full-factorial array.

Table 1
Ranges of process parameters and resulting normalised descriptors.

D [mm]	a_p [mm]	a_e [mm]	f_z [mm/Z]	D_{orb} [mm]	λ [-]	μ [-]	δ [-]
8	18.80– 23.44	0.34– 1.02	0.016– 0.044	2.49– 5.88	0.783– 0.977	0.043– 0.128	0.311– 0.735
10	23.59– 29.25	0.40– 1.29	0.016– 0.044	3.05– 7.40	0.786– 0.975	0.040– 0.129	0.305– 0.740
12	28.09– 35.09	0.51– 1.55	0.016– 0.044	3.71– 8.98	0.780– 0.975	0.043– 0.129	0.309– 0.748

Bounds were set individually per diameter so that the resulting normalised descriptors λ , μ and δ spanned closely comparable ranges across the three tool sizes (Table 1), the criterion that makes tests performed with different tools directly comparable and poolable into a single modelling dataset.

All tests were performed on AA7075-T651 aluminium alloy, machined as a straight channel feature of constant nominal geometry (width w , cutting length L_c), so that only tool diameter and process/toolpath parameters were varied between tests. Figure 2 shows (a) representative trochoidal channels obtained on the workpiece and (b) the experimental set-up.

Machining was carried out on a Kondia B640 vertical milling centre with a Fagor CNC 8070 controller. Because trochoidal toolpaths are curvature-rich, constant-feed interpolation (G196) with cutter-radius compensation was kept active throughout the trochoidal loop to avoid uncontrolled feed-rate fluctuations that could otherwise confound the comparison between tests. A water-soluble metalworking fluid (Quaker Houghton QH ALTIMAX M 5290) was used throughout, with concentration and pH periodically checked as a consistency control rather than as an independent variable. Three end-mill diameters were tested ($D = 8, 10$ and 12 mm) introduced as a primary input because it simultaneously affects engagement geometry, cutting stiffness, chip formation and dynamic behaviour. Each NC program executed three consecutive tests (a “triplet”) separated by controlled pauses, so that spindle power returned to near-idle between tests and each triplet could be segmented reliably during post-processing.

2.3 Signal processing and feature extraction

Cutting forces were measured with a Kistler 9255B three-component piezoelectric dynamometer mounted under the workpiece, connected to a Kistler 5070 charge amplifier and logged with DynoWare. The in-plane force components F_x , F_y and the axial component F_z were recorded with force ranges of 2000 N (in-plane) and 4000 N (axial) and an effective sampling frequency of ≈ 25 kHz. Spindle power was logged independently from the machine monitoring system at a substantially lower rate (≈ 16 Hz), a controller-computed estimate derived from spindle drive current rather than a directly transduced instantaneous signal and therefore inherently smoother than the force channel, and was used as the primary event signal to segment the three consecutive tests contained in each acquisition file and to identify effective-cutting intervals.

Force and power were acquired on different systems and are therefore not sampled at the same frequency; the workflow uses power as the event indicator for the start/end of cutting segments and then maps the resulting cutting state onto the force timeline. Prior to feature extraction, the raw signals were corrected for isolated timestamp irregularities, time-aligned and resampled onto a common time base.

Triples segmentation was performed by thresholding the power signal relative to an estimated idle level P_{idle} and searching for three contiguous active intervals of sufficient duration; each detected segment

$[t_{start}, t_{end}]$ was then trimmed at both ends by an edge-trimming window Δt_{trim} to remove start/end transients, giving the stable interval $[t_1, t_2]$ used for the subsequent statistics.

Force signals can exhibit offset and slow drift, e.g. due to thermal drift in the measurement chain or pre-load variations. Baseline values were therefore estimated from short idle windows of length T_{base} , separated from each segment boundary by a gap T_{gap} and identified from the idle condition $P(t) \leq P_{idle,thr}$. Denoting the two baseline reference points (t_{ref1}, b_1) and (t_{ref2}, b_2) , an offset-only correction was applied, while an offset-plus-linear-drift correction – subtracting a linear baseline constructed between (t_{ref1}, b_1) and (t_{ref2}, b_2) from the force vector $F(t) = [F_x(t), F_y(t), F_z(t)]$ – was applied. This correction is deliberately simple and transparent: it reduces systematic bias in mean values and prevents slow drift from artificially inflating peak-to-peak metrics. The in-plane force magnitude was computed as $F_{xy}(t) = \sqrt{F_x^2(t) + F_y^2(t)}$. From the cutting-only samples of each test, five output indicators were then extracted: $F_{xy,mean} = \text{mean}(F_{xy}(t))$; $F_{xy,max} = \max(F_{xy}(t))$; $F_{xy,ptp} = P_{90}(F_{xy}) - P_{10}(F_{xy})$; $F_{z,max} = \max(|F_z(t)|)$; $P_{mean} = \text{mean}(P(t))$; i.e. the mean and maximum in-plane force, a robust peak-to-peak in-plane force defined as the difference between the 90th and 10th percentiles of F_{xy} to limit the influence of isolated spikes, the maximum absolute axial force, and the mean cutting-only spindle power.

All results were consolidated into a single spreadsheet ledger in which each row corresponds to one test and links the programmed inputs ($D, a_p, a_e, f_z, D_{orb}$), the derived descriptors ($\lambda, \mu, \delta, h_{max}, \text{MRR}^*$) and the five extracted output features to one NC program instance and one pair of acquisition files. After removing incomplete rows, the final modelling dataset comprised 72 tests, equally distributed across the three tool diameters (24 tests each).

Before pooling the data for modelling, absence of acquisition saturation, correct triplet segmentation and baseline stability were verified for every test; percentile-based force variability ($F_{xy,ptp} = P_{90} - P_{10}$) was used instead of a simple maximum-minus-minimum range to reduce sensitivity to isolated spikes.

The main sources of uncertainty in the resulting dataset are the differences in dynamic response and chip evacuation between tool diameters, possible feed modulation from axis dynamics along the curvature-rich trochoidal motion, the substantially lower sampling frequency of the power signal relative to the force signal or and local transient events not fully represented by global, test-level descriptors. Tool wear, runout and vibration were not measured explicitly or introduced as explicit variables and should be therefore treated as sources of experimental uncertainty similarly.

3. Effect of process parameters on cutting force and spindle power

The measured responses covered a wide range of process severity: mean in-plane force ranged from 15.8 to 294.9 N, maximum in-plane force from 106.6 to 845.3 N, maximum axial force from 25.1 to 193.5 N, and mean spindle power from 0.282 to 2.002 kW. This variability confirms that the campaign covered clearly distinct cutting conditions and provides a meaningful basis for comparing the modelling

structures introduced in Section 4. Four raw process parameters, tool diameter D , feed per tooth f_z , radial engagement a_e and orbital diameter D_{orb} , are examined individually below. Because the experimental plan scales engagement and orbital geometry with tool size (Section 2.2), these four parameters are not statistically independent: D correlates moderately with a_e ($r = 0.47$) and with D_{orb} ($r = 0.55$), a consequence of the experimental design itself (Table 1) rather than an independent physical association, while f_z was varied essentially independently of the other three ($|r| < 0.08$ in every case). Table 2 therefore reports, for every parameter-target pair, both the simple Pearson correlation r and the partial correlation controlling for D , so that an effect genuinely attributable to a given parameter can be distinguished from one that merely reflects the tool diameter at which it happened to be tested.

Table 2
Correlation between raw process parameters and measured responses across all 72 tests (r , Pearson correlation; $r_{|D}$, partial correlation controlling for tool diameter; p , two-sided significance).

Parameter	Target	r	p	$r_{ D}$	$p_{ D}$
D	$F_{xy,mean}$	0.451	< 0.001	—	—
D	$F_{xy,max}$	0.392	0.001	—	—
D	$F_{xy,ptp}$	0.374	0.001	—	—
D	$F_{z,max}$	0.597	< 0.001	—	—
D	P_{mean}	-0.503	< 0.001	—	—
f_z	$F_{xy,mean}$	0.183	0.124	0.206	0.084
f_z	$F_{xy,max}$	0.113	0.344	0.124	0.304
f_z	$F_{xy,ptp}$	0.140	0.239	0.152	0.205
f_z	$F_{z,max}$	0.043	0.721	0.055	0.650
f_z	P_{mean}	0.224	0.059	0.258	0.030
a_e	$F_{xy,mean}$	0.304	0.009	0.119	0.325
a_e	$F_{xy,max}$	0.247	0.036	0.080	0.509
a_e	$F_{xy,ptp}$	0.418	< 0.001	0.297	0.012
a_e	$F_{z,max}$	0.380	0.001	0.143	0.233
a_e	P_{mean}	0.058	0.627	0.382	0.001
D_{orb}	$F_{xy,mean}$	0.290	0.014	0.058	0.630
D_{orb}	$F_{xy,max}$	0.294	0.012	0.105	0.385
D_{orb}	$F_{xy,ptp}$	0.218	0.066	0.018	0.882
D_{orb}	$F_{z,max}$	0.439	< 0.001	0.168	0.162
D_{orb}	P_{mean}	-0.508	< 0.001	-0.323	0.006

3.1 Influence of tool diameter D

Tool diameter shows significant correlations with all five measured responses ($|r| = 0.39\text{--}0.60$, $p \leq 0.001$; Table 2). Table 3 summarises the corresponding response means for each diameter group. The 12 mm tool produced the highest mean and peak force levels: $F_{xy,\text{mean}}$ increased from 65.13 N for $D = 10$ mm to 145.11 N for $D = 12$ mm, while $F_{xy,\text{max}}$ increased from 223.65 to 448.29 N. Mean spindle power followed a different trend, decreasing from 1.140 kW for $D = 8$ mm to 0.674 kW for $D = 12$ mm. This apparently counter-intuitive pattern is explained further in Section 3.4: because the experimental plan scales the orbital diameter D_{orb} with D (Table 1), larger tools also engage over a gentler, less abrupt orbital loop, which lowers average power even as the larger absolute chip load raises force. P_{mean} is the controller-reported spindle power (Section 2.3) averaged over the cutting-only interval of each test (Section 2.4); because segmentation operates at the test level rather than within each trochoidal cycle, this average necessarily includes the brief low- or zero-engagement arcs that are an intrinsic part of every trochoidal loop.

These grouped values should not, however, be interpreted as isolated effects of tool diameter. As shown in Table 3, mean f_z remains essentially constant across diameter groups (0.0299–0.0300 mm/tooth), whereas a_e increases from 0.681 to 1.023 mm and D_{orb} from 4.20 to 6.30 mm between $D = 8$ and 12 mm. This co-variation is also reflected in Table 2, where D correlates with both a_e ($r = 0.47$) and D_{orb} ($r = 0.55$). Figure 3 shows the resulting distributions of $F_{xy,\text{mean}}$ and P_{mean} for the three diameter groups. The substantial within-group dispersion indicates that diameter defines only part of the observed response variability. Partial correlations are therefore considered in Sections 3.3 and 3.4 to assess the associations of a_e and D_{orb} after controlling for D .

Table 3
Mean raw process parameters and output responses grouped by tool diameter.

Tool D [mm]	f_z [mm/tooth]	a_e [mm]	D_{orb} [mm]	$F_{xy,\text{mean}}$ [N]	$F_{xy,\text{max}}$ [N]	$F_{xy,\text{ptp}}$ [N]	$F_{z,\text{max}}$ [N]	P_{mean} [kW]
8	0.030	0.681	4.20	79.26	299.52	168.74	48.32	1.140
10	0.030	0.845	5.24	65.13	223.65	122.60	67.82	0.865
12	0.030	1.023	6.30	145.11	448.29	266.32	95.44	0.674

3.2 Influence of feed per tooth f_z

f_z ranged from 0.016 to 0.044 mm/tooth (mean 0.030 mm/tooth) and, unlike a_e and D_{orb} , was varied essentially independently of tool diameter ($r = -0.002$; Table 3 shows its group mean is the same to three significant figures for all three diameters), so its simple and partial correlations in Table 2 are nearly identical and can be read directly. The result is a genuinely different pattern from diameter: f_z shows weak, non-significant correlations with every in-plane and axial force target ($r = 0.04\text{--}0.18$, $p \geq 0.12$; Fig. 4b), but a positive relationship with mean spindle power that strengthens from borderline ($r = 0.224$, $p = 0.059$) to statistically significant once tool diameter is controlled for ($r|D = 0.258$, $p = 0.030$); grouping tests into low, mid and high feed terciles shows mean power rising monotonically from 0.775

kW to 0.905 kW to 1.003 kW (Fig. 4a). This is physically coherent: f_z sets the material-removal rate directly and therefore the energetic demand on the spindle, whereas instantaneous force is governed more by the engagement geometry (a_e , D) than by feed rate alone over the comparatively narrow range tested here. Practically, this means feed per tooth can be used within this range to trim power demand with little direct effect on force-based tool loading – a genuinely useful, and not obvious a priori, degree of freedom for process planning.

3.3 Influence of radial engagement a_e

a_e ranged from 0.34 to 1.55 mm and is moderately confounded with tool diameter ($r = 0.47$; Table 3 shows its group mean rising from 0.681 to 1.023 mm across the three diameters), so its simple correlations in Table 2 must be read alongside the partial ones. Simple correlations are significant for four of the five targets ($r = 0.25$ – 0.42), but three of these – $F_{xy,mean}$, $F_{xy,max}$ and $F_{z,max}$ – collapse to non-significance once D is controlled for ($r|D = 0.08$ – 0.14 , $p > 0.2$), showing that most of their apparent dependence on a_e is really a diameter effect in disguise. Two results survive, however, and are genuinely informative. First, a_e retains a significant independent association with force variability, $F_{xy,ptp}$ ($r|D = 0.297$, $p = 0.012$; low/mid/high tercile means 137.3, 191.9 and 228.4 N), consistent with radial engagement directly setting the amplitude of the chip-load oscillation within each trochoidal loop. Second, and more strikingly, a_e shows essentially no simple correlation with mean spindle power ($r = 0.058$, $p = 0.627$) but the strongest partial correlation in the entire table once D is controlled for ($r|D = 0.382$, $p = 0.001$). This is a classic suppression effect, made visible in Fig. 5a: D and a_e are positively correlated but pull P_{mean} in opposite directions (Section 3.1), so their simple correlations with power partly cancel when all three diameters are pooled together (the flat, noisy trend across colours), yet within each individual colour/diameter group power rises clearly with a_e . In practice this means that, for a fixed tool diameter, increasing radial engagement is a genuine and fairly strong driver of spindle power, even though this relationship is invisible in a naive diameter-pooled scatter plot.

3.4 Influence of orbital diameter D_{orb}

D_{orb} ranged from 2.49 to 8.98 mm and is the most diameter-confounded of the four raw parameters ($r = 0.55$ with D ; Table 3 shows its group mean rising from 4.20 to 6.30 mm). As with a_e , most of its simple correlations with the force targets ($r = 0.22$ – 0.44) are substantially reduced after controlling for D ($r|D = 0.02$ – 0.17 , all non-significant), indicating that orbital diameter mainly tracks tool size rather than exerting an independent influence on force. Mean spindle power is the exception: D_{orb} shows the strongest simple correlation of any parameter–target pair in Table 2 ($r = -0.508$, $p < 0.001$), and this negative relationship survives controlling for D ($r|D = -0.323$, $p = 0.006$); mean power falls from 1.118 kW in the lowest D_{orb} tercile to 0.697 kW in the highest. Figure 5b shows this relationship directly: within each diameter colour group, power still trends downward as D_{orb} increases, unlike the a_e panel where the within-group trend only appears after conditioning. A plausible physical reading is that a larger orbital loop, for a given tool diameter, spreads material removal over a gentler, less abrupt engagement history, reducing the average energetic severity of the cut even though it does not consistently reduce

instantaneous force. Together with Section 3.3, this identifies orbital and radial engagement geometry – not just tool diameter – as genuine, independent drivers of spindle power once diameter is accounted for.

3.5 Influence of engagement and chip-load descriptors

Beyond the four raw parameters above, the maximum uncut chip thickness h_{max} and the severity descriptor MRR^* were used to interpret trends in the experimental data (Fig. 6). The relationship between h_{max} and $F_{xy,mean}$ is physically plausible but not strongly linear: higher h_{max} values are generally associated with higher mean in-plane force, though with substantial scatter, indicating that chip thickness alone does not fully explain the force response. The relationship between MRR^* and P_{mean} shows the same suppression pattern already identified for radial engagement in Section 3.3: an essentially flat, non-significant simple correlation once diameters are pooled ($r = 0.059$, $p = 0.62$ – the weakest pairing in this section, and visible as the near-flat trend in Fig. 6b), which becomes strong once D is controlled for (partial $r = 0.481$, $p < 0.001$), since $MRR^* = a_e \cdot a_p \cdot f_z$ directly combines the two raw parameters already shown to be independent drivers of power in Sections 3.3–3.4. These trends motivate the inclusion of h_{max} and MRR^* as additional, physically meaningful inputs in the hybrid model structure described in Section 4.

3.6 Combined interpretation of parameter effects

The parameter-by-parameter analysis above resolves a question that grouped, diameter-only statistics (Section 3.1) leave open: which raw variables carry genuinely independent information, and which are largely proxies for tool size? Feed per tooth is the cleanest case – independent of D by design, and a real, if modest, independent driver of spindle power. Radial and orbital engagement are more entangled with diameter, but each retains at least one statistically robust, diameter-independent effect: a_e on force variability and, together with D_{orb} , on spindle power via the suppression effect identified in Section 3.3. No single variable – D , f_z , a_e , D_{orb} , h_{max} or MRR^* – explains the measured response in isolation, and several are only partially independent of one another (Section 4.2 quantifies this further for the descriptors used in modelling). This is precisely why D is retained as an explicit input throughout Section 4 rather than assumed redundant with the normalised ratios, and why a multivariable, data-driven framework – rather than a series of univariate trend lines – is required to represent the process response with reasonable accuracy.

4. Physics-Informed Machine Learning Modelling

4.1 Input structures and modelling dataset

Each of the five output indicators ($F_{xy,mean}$, $F_{xy,max}$, $F_{xy,ptp}$, $F_{z,max}$, P_{mean}) was modelled independently from three progressively enriched input structures (Table 4): a minimal 5-variable model using only D , λ , μ , δ

and f_z ; a hybrid 5 + 2 model that adds h_{max} and MRR*; and an enriched 5 + 4 model that further reintroduces the raw engagement variables a_e and a_p . Comparing the three structures makes it possible to evaluate whether additional physical variables improve predictive performance or mainly introduce redundancy.

The five targets were selected because they represent complementary aspects of the process response: $F_{xy,mean}$ describes the average mechanical load in the plane of motion; $F_{xy,max}$ and $F_{xy,ptp}$ are more sensitive to peaks and fluctuations, potentially associated with local engagement variations and dynamic effects; $F_{z,max}$ captures the axial load component; and P_{mean} represents the energetic demand of the operation. Rows with missing values in the selected inputs or outputs were removed prior to training, so that the same validation procedure could be applied consistently across all targets and input structures.

Table 4
Input structures used for the machine learning models.

Model structure	Input variables	Purpose
Minimal (5)	$D, \lambda, \mu, \delta, f_z$	Compact baseline using only fundamental process descriptors
Hybrid (5 + 2)	$D, \lambda, \mu, \delta, f_z, h_{max}, \text{MRR}^*$	Adds local chip-load severity and global process-severity information
Enriched (5 + 4)	$D, \lambda, \mu, \delta, f_z, h_{max}, \text{MRR}^*, a_e, a_p$	Tests whether additional raw engagement variables improve performance or introduce redundancy

4.2 Correlation analysis

Pairwise correlation was computed for each input set to check redundancy before training (Fig. 7 shows the enriched 5 + 4 structure), using the Pearson correlation coefficient between each pair of candidate input variables, the same measure used for Table 2 in Section 3. The minimal set shows low correlation among its five variables and is therefore a clean, non-redundant baseline. In the hybrid set, h_{max} correlates strongly with f_z ($r = 0.865$) because it is partly derived from feed per tooth, and MRR* correlates with a_e ($r = 0.803$), consistent with its definition; these correlations are expected, since the added variables are physically derived from existing descriptors rather than independent measurements. The enriched set shows the strongest redundancy, notably between D and a_p ($r = 0.926$) and between μ and a_e ($r = 0.874$), confirming that reintroducing raw engagement variables on top of an already normalised parameterisation adds complexity without necessarily adding independent information.

4.3 Learning algorithm and validation

Predictive models were built in MATLAB using least-squares boosting (LSBoost), an ensemble of regression trees trained sequentially on the residuals of the previous trees. This approach was adopted

for three reasons. First, the relationship between trochoidal milling parameters and process outputs is expected to be non-linear, reflecting well-established non-linearities in milling mechanics, such as the sine-type dependence of instantaneous chip thickness on engagement angle (Section 2.1) and threshold-like transitions in chip formation and tool-workpiece contact as engagement conditions change. Second, interaction effects are likely, since the effect of one variable, e.g. feed per tooth, depends on engagement condition, tool diameter and chip thickness. Third, tree-based ensembles can model non-linear tabular data without requiring an explicit analytical equation, which suits an exploratory, physics-informed modelling stage in which the functional form of the force/power response is not assumed a priori.

The modelling workflow followed the same steps for every target and input structure: (i) read the experimental dataset; (ii) select the input structure (Table 4); (iii) remove rows with missing input or output values; (iv) compute the feature correlation matrix; (v) train one boosted-tree model per target under 5-fold cross-validation; (vi) train a final model on the complete dataset; (vii) compute predictor importance for that final model. A separate model was trained for each output target under each input structure, so that the predictive difficulty of each target could be evaluated independently rather than through a single multi-output model.

Model performance was assessed by 5-fold cross-validation: the dataset was split into five subsets, each used once as the test set while the remaining four were used for training. Three metrics were used to quantify prediction performance. The mean absolute error, $MAE = (1/n_s) \sum |y_i - \hat{y}_i|$ provides an intuitive average error magnitude. The root-mean-square error, $RMSE = \sqrt{[(1/n_s) \sum (y_i - \hat{y}_i)^2]}$, penalises larger errors more strongly and is therefore more sensitive to outliers or poorly predicted cases. Finally, the coefficient of determination, $R^2 = 1 - \sum (y_i - \hat{y}_i)^2 / \sum (y_i - \bar{y})^2$ indicates how much of the test-set variance is explained by the model. Here, y_i , $\hat{y}_i$, $\bar{y}$ and n_s are the measured value, predicted value, mean measured value and number of samples in the test fold, respectively. Negative R^2 values indicate that the model performs worse than simply predicting the mean of the test data. A final model was subsequently trained on the complete dataset for each target to compute predictor importance.

5. Results and discussion

5.1 Comparison of the three model structures

Table 5 reports MAE, RMSE and R^2 for all three input structures and five targets. The hybrid 5 + 2 model provides the best overall compromise: for the average in-plane force, R^2 increases from 0.241 (minimal) to 0.364 (hybrid) while MAE decreases from 37.214 to 32.694 N; for mean spindle power, R^2 increases from 0.374 to 0.473 while MAE decreases from 0.210 to 0.191 kW. The enriched 5 + 4 model does not provide a consistent additional gain – for most targets it performs similarly to or worse than the hybrid model (e.g. $R^2 = 0.303$ for $F_{xy,mean}$ and $R^2 = 0.467$ for P_{mean}) – indicating that its extra raw engagement

variables mainly reintroduce the redundancy identified in Section 4.2 rather than genuinely new information. Figure 8 visualises this comparison across all five targets.

Table 5
Predictive performance of the three model structures under 5-fold cross-validation.

Model	Target	MAE	RMSE	R^2
Minimal (5)	$F_{xy,mean}$	37.214	47.346	0.241
Minimal (5)	$F_{xy,max}$	95.597	132.080	0.029
Minimal (5)	$F_{xy,ptp}$	62.138	79.798	0.242
Minimal (5)	$F_{z,max}$	19.759	25.982	-0.151
Minimal (5)	P_{mean}	0.210	0.291	0.374
Hybrid (5 + 2)	$F_{xy,mean}$	32.694	42.923	0.364
Hybrid (5 + 2)	$F_{xy,max}$	98.228	130.840	0.069
Hybrid (5 + 2)	$F_{xy,ptp}$	62.603	79.027	0.263
Hybrid (5 + 2)	$F_{z,max}$	19.852	26.174	-0.190
Hybrid (5 + 2)	P_{mean}	0.191	0.270	0.473
Enriched (5 + 4)	$F_{xy,mean}$	36.315	45.184	0.303
Enriched (5 + 4)	$F_{xy,max}$	102.550	138.620	-0.016
Enriched (5 + 4)	$F_{xy,ptp}$	64.931	82.763	0.168
Enriched (5 + 4)	$F_{z,max}$	20.607	27.917	-0.189
Enriched (5 + 4)	P_{mean}	0.193	0.269	0.467

5.2 Comparison against linear regression and bagged-tree baselines

The choice of boosted regression trees (Section 4.3) was checked against two baselines on the hybrid 5 + 2 input structure: ordinary multiple linear regression, and a bagged-tree ensemble (a random-forest-style model, 100 trees). All three algorithms were re-implemented independently in MATLAB (*fitlm* for linear regression; *fitensemble* with the 'Bag' and 'LSBoost' methods for the two ensembles) and evaluated on the same 72-test dataset using 5-fold cross-validation with a single fixed random partition

(seed = 42) shared identically across all three algorithms, so that differences in Table 6 reflect the learning algorithm alone and not the fold assignment. Following standard boosting practice, the LSBoost weak learners were restricted to shallow trees (maximum 10 splits) with 150 learning cycles at a learning rate of 0.1, so that no single tree could overfit the residuals within a training fold; the bagged ensemble used unrestricted trees, as is standard for bagging. Because this is an independently re-run 5-fold partition rather than the exact fold assignment behind Table 5, the resulting LSBoost figures below differ somewhat from the hybrid-model row of Table 5 – an expected consequence of cross-validation variance at $n = 72$ – but the qualitative pattern is consistent: P_{mean} and $F_{xy,mean}$ remain the most predictable targets, and $F_{xy,max}$ and $F_{z,max}$ remain among the hardest. Notably, $F_{z,max}$ is the target whose sign flips between the two independent runs (from a negative R^2 in Table 5 to a modestly positive one here); rather than weakening the conclusions of Section 5.6, this fold-to-fold instability is itself evidence that $F_{z,max}$ sits at the edge of what any of the tested global descriptors can reliably explain.

Table 6
Comparison of learning algorithms for the hybrid 5 + 2 input structure under 5-fold cross-validation (identical folds across algorithms).

Algorithm	Target	MAE	RMSE	R^2
Linear regression	$F_{xy,mean}$	40.441	51.709	0.247
Bagged trees	$F_{xy,mean}$	35.818	47.312	0.370
LSBoost	$F_{xy,mean}$	34.184	44.144	0.451
Linear regression	$F_{xy,max}$	111.522	149.485	0.070
Bagged trees	$F_{xy,max}$	90.777	131.243	0.283
LSBoost	$F_{xy,max}$	100.088	136.814	0.221
Linear regression	$F_{xy,ptp}$	71.509	89.013	0.302
Bagged trees	$F_{xy,ptp}$	64.001	83.893	0.380
LSBoost	$F_{xy,ptp}$	64.807	83.522	0.386
Linear regression	$F_{z,max}$	20.120	29.175	0.180
Bagged trees	$F_{z,max}$	19.905	28.055	0.242
LSBoost	$F_{z,max}$	19.656	26.729	0.312
Linear regression	P_{mean}	0.188	0.284	0.436
Bagged trees	P_{mean}	0.201	0.285	0.434
LSBoost	P_{mean}	0.189	0.276	0.468

LSBoost outperforms plain linear regression on every target, with the largest gains for $F_{xy,mean}$ (R^2 0.247 $\rightarrow$ 0.451) and $F_{z,max}$ (0.180 $\rightarrow$ 0.312), confirming that the input–output relationship captured by the tree ensemble is genuinely non-linear rather than an artefact of model complexity. LSBoost also matches or exceeds bagging on four of the five targets, and is a close second on $F_{xy,max}$ (0.221 vs. 0.283). The gap between the two tree ensembles is consistently smaller than the gap between either ensemble and linear regression, which indicates that most of the predictive benefit comes from moving to a non-linear, tree-based representation at all; the specific choice between bagging and boosting is of secondary importance for this dataset, though boosting retains a small, fairly consistent edge and was therefore retained as the primary algorithm for Sections 4–5.

5.3 Cross-diameter (leave-one-out) generalisation

Sections 5.1–5.2 evaluate the models with random 5-fold cross-validation, in which every training fold already contains examples from all three tool diameters: the models therefore only ever have to interpolate. A materially harder and practically more relevant question is whether a model trained on only two tool diameters can predict the third, entirely unseen one – the situation a user would actually face when selecting parameters for a new tool size. Three leave-one-diameter-out scenarios were tested on the hybrid 5 + 2 structure: training on $D = 8$ and 10 mm to predict $D = 12$ mm; training on 8 and 12 mm to predict 10 mm; and training on 10 and 12 mm to predict 8 mm, using the same linear regression and LSBoost configurations as Section 5.2 (48 training tests, 24 held-out test tests per scenario). A diameter-agnostic variant of LSBoost, with the raw tool diameter D removed from the input set (leaving only λ , μ , δ , f_z , h_{max} and MRR^*), was also tested to check whether the normalised descriptors alone could substitute for an unseen value of D .

Table 7
Leave-one-diameter-out generalisation for the hybrid 5 + 2 structure: training on two tool diameters to predict the third (summary; full results in Appendix A, Table A1).

Held-out (test)	Algorithm	Mean R^2 (5 targets)	Range across targets (R^2)
$D = 12$ (train 8 + 10)	Linear regression	-1.42	-2.86 to -0.14
$D = 12$ (train 8 + 10)	LSBoost	-1.09	-1.57 to -0.54
$D = 10$ (train 8 + 12)	Linear regression	-2.01	-4.61 to 0.27
$D = 10$ (train 8 + 12)	LSBoost	-5.29	-9.91 to -1.67
$D = 8$ (train 10 + 12)	Linear regression	-4.52	-9.96 to 0.21
$D = 8$ (train 10 + 12)	LSBoost	-1.26	-4.02 to 0.09

Across the three leave-one-diameter-out scenarios, two algorithms and five targets, the mean R^2 was approximately -2.6, compared with the 0.07–0.47 range obtained under random cross-validation (Tables 5 and 6). Only 3 of the 30 scenario $\times$ algorithm $\times$ target combinations produced positive R^2 values. These corresponded to mean spindle power predicted by linear regression for $D = 10$ mm ($R^2 = 0.269$) and $D = 8$ mm ($R^2 = 0.206$), and to $F_{xy,ptp}$ predicted by LSBoost for $D = 8$ mm ($R^2 = 0.085$). The largest errors were obtained for $F_{xy,max}$, with $R^2 = -9.96$ for linear regression when predicting $D = 8$ mm from the 10 and 12 mm datasets, and $R^2 = -9.91$ for LSBoost when predicting $D = 10$ mm from the 8 and 12 mm datasets. These results show that the performance obtained under random cross-validation does not extend to an unseen tool diameter.

Removing the raw tool diameter from the input set did not produce a consistent improvement. For example, for $F_{xy,max}$ with $D = 10$ mm held out, R^2 increased from -9.908 to -6.858, while the remaining targets showed no systematic change. The poor cross-diameter performance therefore cannot be attributed solely to the inability of tree-based models to extrapolate in D . Although the normalised descriptors λ , μ and δ facilitate pooling data from different tool sizes, they do not fully represent diameter-dependent changes in tool stiffness, dynamic response, contact conditions and chip

evacuation. Consequently, the present models are suitable for interpolation when the relevant tool diameters are represented in the training dataset, but they should not be applied to an unseen diameter without additional experimental data.

5.4 Performance of the hybrid input structure

The hybrid model extends the minimal input set by incorporating h_{max} and MRR^* , which provide additional information on local chip-load severity and overall cutting intensity, respectively. Both descriptors are physically motivated by the trends discussed in Section 3. In contrast, the enriched model additionally includes the raw engagement variables ae and ap , which are already partly represented through the normalised descriptors μ and λ . This is reflected in the relatively high correlations reported in Section 4.2, including $r = 0.874$ between μ and ae and the strong correlation between D and ap .

The cross-validation results in Table 5 show that adding h_{max} and MRR^* improves the predictive performance of the minimal model, whereas the subsequent inclusion of ae and ap does not provide a consistent additional gain. The hybrid 5 + 2 structure therefore provides the best balance between predictive performance and input redundancy among the three configurations considered. Its lower dimensionality also facilitates interpretation and reduces the number of partially redundant variables required for model implementation.

5.5 Feature importance

Figure 9 reports predictor importance for the final hybrid 5 + 2 models trained on the complete dataset. Tool diameter D is the dominant predictor for the force-related targets, particularly $F_{z,max}$ (56.1) and $F_{xy,mean}$ (47.6), reflecting its combined influence on engagement geometry, stiffness and contact conditions; the normalised descriptor λ contributes consistently across the in-plane force targets, allowing tests on different diameters to be compared on a common basis. h_{max} contributes most to force variability (28.9 for $F_{xy,ptp}$) and to power (24.2), while MRR^* is the leading predictor of P_{mean} (30.2), consistent with its role as a global cutting-intensity indicator. Because several inputs are correlated (Section 4.2), importance should be read as a description of model behaviour rather than a direct causal ranking of the process variables. As with the algorithm comparison in Section 5.2, this final model was trained independently for this analysis; exact percentages can differ somewhat from a separate LSBoost training run because of the randomness inherent in tree-ensemble fitting, though the broad pattern is consistent: tool diameter and the normalised axial-engagement ratio dominate the in-plane force targets, while maximum chip thickness and the severity descriptor lead for force variability and mean power respectively.

5.6 Target-by-target interpretation

The five targets show markedly different predictability, consistent with their physical meaning: average, system-level responses are captured far better than peak-based, local responses. P_{mean} is the best-

predicted output ($R^2 = 0.473$) because it is an averaged energetic response, less sensitive to short-duration transients than instantaneous force. $F_{xy,mean}$ is the second-best predicted target ($R^2 = 0.364$): the improvement obtained by adding h_{max} and MRR^* confirms that these descriptors add genuine chip-load information beyond the minimal geometric set, although the moderate R^2 also shows that global, test-level descriptors cannot fully capture a quantity that depends on the continuous evolution of engagement along the trochoidal path. $F_{xy,ptp}$ is moderately predicted ($R^2 = 0.263$), reflecting that part – but not all – of the in-cycle force variability is explained by global descriptors. $F_{xy,max}$ is weakly predicted ($R^2 = 0.069$) because a maximum value is dominated by short, local events (a sudden engagement increase, a chip-evacuation issue, a controller-related motion variation) that are not represented by test-level inputs. $F_{z,max}$ is the hardest target: all three model structures return a negative R^2 (best case $R^2 = -0.151$, minimal model), meaning that none of the tested input structures outperforms simply predicting the mean of the test data. Rather than a modelling failure, this is a diagnostic result: it indicates that axial force peaks are governed by local or transient phenomena – tool runout, axial vibration, local contact conditions, chip evacuation – that are not represented by any of the tested global descriptors, however enriched.

5.7 Practical implications and limitations

The results indicate that $F_{xy,mean}$ and P_{mean} , together with the hybrid 5 + 2 descriptor set, are the most suitable outputs for compact screening of trochoidal milling conditions in terms of average mechanical and energetic demand. In contrast, peak-related quantities ($F_{xy,max}$, $F_{xy,ptp}$ and $F_{z,max}$) were less accurately predicted, reflecting their greater sensitivity to local engagement and transient phenomena. These quantities may therefore be more appropriate as monitoring or constraint variables than as direct optimisation targets within the present modelling framework. The hybrid model also offers a favourable balance between predictive performance and complexity, using seven physically meaningful inputs while avoiding the additional redundancy introduced by the enriched 5 + 4 structure. This compact formulation may facilitate its eventual use in process-planning or CNC decision-support applications, although such deployment would require validation under independent machining conditions.

The applicability of the models is limited by both the size and structure of the experimental dataset. The 72 tests distributed across three tool diameters provide sufficient information to compare descriptor structures within the investigated domain, but they do not represent the full range of possible trochoidal milling conditions. Moreover, the input variables are test-level descriptors and therefore do not resolve the instantaneous variation of cutter–workpiece engagement within each trochoidal cycle. This limitation is consistent with the better prediction of averaged responses compared with peak quantities, particularly $F_{z,max}$. Tool wear, runout, vibration and detailed tool geometry were not explicitly included as model inputs and remain part of the experimental variability.

The cross-diameter results further define the range of validity of the proposed models. As shown in Section 5.3, satisfactory interpolation within a dataset containing all three tested diameters does not imply generalisation to an unseen tool size. Changes in diameter also modify stiffness, contact

conditions, dynamic response and chip evacuation, effects that are only partially represented by the normalised descriptors. Consequently, the present models should be restricted to machining conditions represented by the training domain unless additional experimental data are introduced for new tool sizes. Finally, predictor importance should be interpreted as a measure of how the trained models use the available variables rather than as a causal ranking of process effects, particularly when correlated descriptors are present. Similarly, cross-validation provides an appropriate basis for comparing model structures within the available dataset, but independent experimental validation remains necessary before extending the models to new machining conditions.

6. Conclusions and future work

This study investigated high-performance trochoidal milling of AA7075-T651 aluminium alloy through 72 instrumented tests using three end-mill diameters (8, 10 and 12 mm). Cutting-force and spindle-power measurements were combined with a physics-guided machine-learning framework to assess the effect of input representation, learning algorithm and cross-diameter generalisation. The main conclusions are:

- Tool diameter was strongly associated with the measured force responses, with the 12 mm tool producing the highest average and peak force levels. Mean spindle power showed the opposite overall trend, decreasing from 1.140 kW for $D = 8$ mm to 0.674 kW for $D = 12$ mm.
- The hybrid 5 + 2 input structure, which adds maximum uncut chip thickness (h_{max}) and the severity descriptor MRR^* to the five-variable baseline, provided the best overall predictive performance. For $F_{xy,mean}$, R^2 increased from 0.241 to 0.364, while for P_{mean} it increased from 0.374 to 0.473. Adding a_e and a_p in the enriched 5 + 4 structure did not produce consistent further improvement because these variables were strongly correlated with descriptors already included in the model.
- Non-linear tree ensembles generally outperformed multiple linear regression. On identical cross-validation folds, LSBoost provided higher R^2 than linear regression for all five targets, although its performance was similar to that of bagged trees and was lower for $F_{xy,max}$. The main modelling improvement therefore resulted from using a non-linear tree-based representation rather than from the specific choice of boosting over bagging.
- Prediction accuracy depended strongly on the type of response considered. Mean spindle power and mean in-plane force were predicted with moderate accuracy, whereas peak-related quantities were less accurately reproduced. In particular, $F_{z,max}$ was not consistently predicted better than a mean-value reference under the main cross-validation analysis, indicating that global test-level descriptors do not adequately represent the local and transient phenomena governing axial-force peaks.
- Cross-diameter generalisation remains limited. Random cross-validation, with all three diameters represented in the training data, yielded R^2 values of up to 0.47, depending on the target and model. When one diameter was completely excluded from training, predictive accuracy decreased substantially, with a mean R^2 of approximately -2.6 across the evaluated cases. Excluding tool diameter from the input set did not materially improve this behaviour. These results suggest that the

normalised engagement descriptors provide a useful basis for combining data from different diameters within the explored range, while prediction for a completely unseen tool diameter remains more challenging.

- Overall, the proposed descriptor set is suitable for compact modelling of average force and power responses within the investigated experimental domain, but its applicability is limited to conditions represented in the training data. Future work should incorporate additional tool diameters, independent validation data and time-resolved descriptors of cutter–workpiece engagement, together with variables related to runout, vibration, tool geometry and wear, to improve prediction of transient force responses and cross-diameter generalisation.

Nomenclature

Symbol	Description	Unit
D	Tool diameter	mm
D_{orb}	Orbital diameter of the trochoidal path ($= 2R$)	mm
a_e	Radial engagement	mm
a_p	Axial depth of cut	mm
v_c	Cutting speed	m/min
n	Spindle speed	rpm
f_z	Feed per tooth	mm/tooth
v_f	Feed rate	mm/min
z	Number of teeth	–
λ	Normalised axial engagement, $a_p/(3D)$	–
μ	Normalised radial engagement, a_e/D	–
δ	Orbital diameter ratio, D_{orb}/D	–
h_{max}	Maximum uncut chip thickness	mm
MRR*	Simplified process-severity descriptor, $a_e \cdot a_p \cdot f_z$	mm ³ /tooth
$F_{xy}(t)$	In-plane force magnitude	N
$F_{xy,mean}$	Mean in-plane cutting force	N
$F_{xy,max}$	Maximum in-plane cutting force	N
$F_{xy,ptp}$	Robust peak-to-peak in-plane force ($P_{90} - P_{10}$)	N
$F_{z,max}$	Maximum absolute axial force	N
P_{mean}	Mean spindle power during cutting	kW

Declarations

CRedit authorship contribution statement

Gorka Urbikain-Pelayo: Writing – original draft, Validation, Software, Resources, Methodology, Investigation, Data curation, Conceptualization. **Daniel Olvera-Trejo:** Writing – review & editing, Writing – original draft, Validation, Software, Resources, Methodology, Investigation, Data curation,

Conceptualization. **Luis Norberto López de Lacalle:** Writing – review & editing, Supervision, Project administration. **Alex Elías-Zúñiga:** Writing – review & editing, Supervision, Project administration.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

The experimental dataset and MATLAB processing scripts generated and analysed during this study are available from the corresponding author upon reasonable request.

Acknowledgements

This project acknowledges the following projects funded by the Spanish Ministry of Research: MICINN-TED2021-130044B-I00: ‘A new way to assess quality in manufacturing processes by merging process data in high connected production systems in aeroturbines (Quolink-1)’ and INNPACTO CPP2021-008932 ‘Sistemas de producción para control de calidad exhaustivo en componentes mecanizados a partir de nuevos procesos primarios ultra near net shape’. Special thanks also to Project NEOPHYM from Call: Generation of Knowledge (PID2022-137380OB-I00). Thanks are due to Basque Government by Excelent group IT 894-26. This research was funded by Tecnológico de Monterrey through the Institute of Advanced Materials for Sustainable Manufacturing. Besides, we acknowledge the support and advices of Garikoitz Goikoetxea during the work.

References

1. Kusiak A (2018) Smart manufacturing. *Int J Prod Res* 56(1–2):508–517. 10.1080/00207543.2017.1351644
2. Urbikain Pelayo G, Olvera-Trejo D, Luo M, Tang K, López de Lacalle LN, Elías-Zúñiga A (2021) A model-based sustainable productivity concept for the best decision-making in rough milling operations. *Measurement* 186:110120. 10.1016/j.measurement.2021.110120
3. Otkur M, Lazoglu I (2007) Trochoidal milling. *Int J Mach Tools Manuf* 47(9):1324–1332. 10.1016/j.ijmachtools.2006.08.002
4. Wu S, Ma W, Li B, Wang C (2016) Trochoidal machining for the high-speed milling of pockets. *J Mater Process Technol* 233:29–43. 10.1016/j.jmatprotec.2016.01.033
5. Pleta A, Mears L (2016) Cutting force investigation of trochoidal milling in nickel-based superalloy. *Procedia Manuf* 5:1348–1356. 10.1016/j.promfg.2016.08.105

6. Pleta A, Akhavan Niaki F, Mears L (2017) Investigation of chip thickness and force modelling of trochoidal milling. *Procedia Manuf* 10:612–621. 10.1016/j.promfg.2017.07.063
7. Pleta A, Akhavan Niaki F, Mears L (2018) A comparative study on the cutting force coefficient identification between trochoidal and slot milling. *Procedia Manuf* 26:570–579. 10.1016/j.promfg.2018.07.067
8. Ducroux E, Prat D, Viprey F, Fromentin G, d'Acunto A (2019) Analysis and modelling of trochoidal milling in Inconel 718. *Procedia CIRP* 82:473–478. 10.1016/j.procir.2019.04.044
9. Rodriguez-Alabanda O, Guerrero-Vaca G, Molero E, Romero PE (2021) Experimental analysis of deep slot milling in EN AW 2024-T3 alloy by stretched trochoidal toolpath and variable helix angle tool. *CIRP J Manufact Sci Technol* 35:346–360. 10.1016/j.cirpj.2021.07.002
10. Li Z, Chen L, Xu K, Gao Y, Tang K (2020) Five-axis trochoidal flank milling of deep 3D cavities. *Comput Aided Des* 119:102775. 10.1016/j.cad.2019.102775
11. Kozový P, Šajgalík M, Holubják J, Joch R, Drbúl M (2023) Influence of trochoidal milling parameters on tool load. *Transp Res Procedia* 74:709–716. 10.1016/j.trpro.2023.11.201
12. Shah AM, Agarwal A, Mears L (2023) Tool wear area estimation through in-process edge force coefficient in trochoidal milling of Inconel 718. *Manuf Lett* 35:391–398
13. Li Z, Niu J, Sun S, Sun Y (2025) A parameter-variant trochoidal-like tool path planning method for chatter-free and high-efficiency milling. *Chin J Aeronaut* 38:103082. 10.1016/j.cja.2024.05.038
14. Nguyen V-H, Le T-T, Nguyen A-T, Hoang X-T, Nguyen N-T (2025) Comprehensive machine learning and multi-criteria optimization workflow for trochoidal milling parameter selection in AISI 4140 steel. *Int J Adv Manuf Technol* 141(7–8):4071–4094. 10.1007/s00170-025-16916-7
15. Guo Q, Wang W, Jiang Y, Sun C, Liu Z, Yang Z, Ren Q (2026) Trochoidal milling technology in the machining of difficult-to-machine materials. *Int J Adv Manuf Technol* 144(5–6):3039–3059. 10.1007/s00170-026-18075-9
16. Groover MP (2010) *Fundamentals of Modern Manufacturing: Materials, Processes, and Systems*, 4th edn. Wiley, Hoboken, NJ
17. Stephenson DA, Agapiou JS (2016) *Metal Cutting Theory and Practice*, 3rd edn. CRC, Boca Raton, FL
18. Gupta HN, Gupta RC, Mittal A (2009) *Manufacturing Processes*, 2nd edn. New Age International, New Delhi
19. Engin S, Altintas Y (2001) Mechanics and dynamics of general milling cutters. Part I: helical end mills. *Int J Mach Tools Manuf* 41(15):2195–2212. 10.1016/S0890-6955(01)00045-1
20. Engin S, Altintas Y (2001) Mechanics and dynamics of general milling cutters. Part II: inserted cutters. *Int J Mach Tools Manuf* 41(15):2213–2231. 10.1016/S0890-6955(01)00046-3
21. Altintas Y, Budak E (1995) Analytical prediction of stability lobes in milling. *CIRP Ann* 44(1):357–362. 10.1016/S0007-8506(07)62342-7
22. Zhang X, Wang X, Zhang P, Chen K, Cao F (2024) Fast extraction of cutter-workpiece engagement for milling force prediction in multi-axis machining. *Measurement* 231:114490.

- 10.1016/j.measurement.2024.114490
23. Teti R, Jemielniak K, O'Donnell G, Dornfeld D (2010) Advanced monitoring of machining operations. *CIRP Ann* 59(2):717–739. 10.1016/j.cirp.2010.05.010
24. Teti R, Mourtzis D, D'Addona DM, Caggiano A (2022) Process monitoring of machining. *CIRP Ann* 71(2):529–552. 10.1016/j.cirp.2022.05.009
25. Duro JA, Padget JA, Bowen CR, Kim HA, Nassehi A (2016) Multi-sensor data fusion framework for CNC machining monitoring. *Mech Syst Signal Process* 66–67. 10.1016/j.ymssp.2015.04.019
26. Denkena B, Abele E, Brecher C, Dittrich M-A, Kara S, Mori M (2020) Energy efficient machine tools. *CIRP Ann* 69(2):646–667. 10.1016/j.cirp.2020.05.008
27. Pawar SS, Bera TC, Sangwan KS (2022) Modelling of spindle energy consumption in CNC milling. *Procedia CIRP* 105:192–197. 10.1016/j.procir.2022.02.032
28. Wuest T, Weimer D, Irgens C, Thoben K-D (2016) Machine learning in manufacturing: advantages, challenges, and applications. *Prod Manuf Res* 4(1):23–45. 10.1080/21693277.2016.1192517
29. Wu D, Jennings C, Terpenney J, Gao RX, Kumara S (2017) A comparative study on machine learning algorithms for smart manufacturing: tool wear prediction using random forests. *J Manuf Sci Eng* 139(7):071018. 10.1115/1.4036350
30. Xie J, Hu P, Chen J, Han W, Wang R (2023) Deep learning-based instantaneous cutting force modeling of three-axis CNC milling. *Int J Mech Sci* 246:108153. 10.1016/j.ijmecsci.2023.108153
31. Heitz T, He N, Ait-Mlouk A, Bachrathy D, Chen N, Zhao G, Li L (2025) Investigation on eXtreme Gradient Boosting for cutting force prediction in milling. *J Intell Manuf* 36(1):285–301. 10.1007/s10845-023-02243-9
32. Ebrahimi MH, Niknam SA (2026) AI-driven prediction of surface roughness and cutting force in milling aluminum alloy under data-scarce conditions. *Machines* 14(7):756. 10.3390/machines14070756
33. Drew D, Betts J, Sadrafshari S, Mohammadi A, Shokrani A (2025) Application of machine learning for tool condition monitoring using sensor integrated tooling. *Procedia CIRP* 133:66–71. 10.1016/j.procir.2025.02.013
34. Friedman JH (2001) Greedy function approximation: a gradient boosting machine. *Annals Stat* 29(5):1189–1232. 10.1214/aos/1013203451
35. Ra D, Lee J, Lee M, Kwak S, Lee S, Lee S (2026) Physics-informed machine learning across manufacturing processes: recent advances, challenges, and directions. *J Manuf Syst* 85:72–95. 10.1016/j.jmsy.2026.01.002
36. Pashmforoush F, Araghizad AE, Budak E (2025) Tool wear prediction in milling process using physics-informed machine learning and thermo-mechanical force model with monitoring applications. *J Manuf Syst* 82:1192–1212. 10.1016/j.jmsy.2025.08.014
37. Araghizad AE, Pashmforoush F, Tehranizadeh F, Kilic K, Budak E (2024) Improving milling force predictions: a hybrid approach integrating physics-based simulation and machine learning for

- remarkable accuracy across diverse unseen materials and tool types. *J Manuf Process* 114:92–107. 10.1016/j.jmapro.2024.02.001
38. Bergs T, Biermann D, Erkorkmaz K, M'Saoubi R (2023) Digital twins for cutting processes. *CIRP Ann* 72(2):541–567. 10.1016/j.cirp.2023.05.006
39. Kim H, Okwudire CE (2023) Intelligent feedrate optimization using a physics-based and data-driven digital twin. *CIRP Ann* 72(1):325–328. 10.1016/j.cirp.2023.04.063
40. Xie H, Li G, Longstaff AP, Fletcher S, Ding S, Pan W (2025) Study of cutting force predictability, signal complexity of different end milling CWE stages with different modelling methods. *Int J Adv Manuf Technol* 136(3–4):1163–1190. 10.1007/s00170-024-14877-x
41. Danish M, Gupta MK, Irfan SA, Ghazali SM, Rathore MF, Krolczyk GM, Alsaady A (2024) Machine learning models for prediction and classification of tool wear in sustainable milling of additively manufactured 316 stainless steel. *Results Eng* 22:102015. 10.1016/j.rineng.2024.102015
42. Elminir HK, El-Brawany MA, Ibrahim DA, Elattar HM, Ramadan EA (2024) An efficient deep learning prognostic model for remaining useful life estimation of high speed CNC milling machine cutters. *Results Eng* 24:103420. 10.1016/j.rineng.2024.103420

Figures

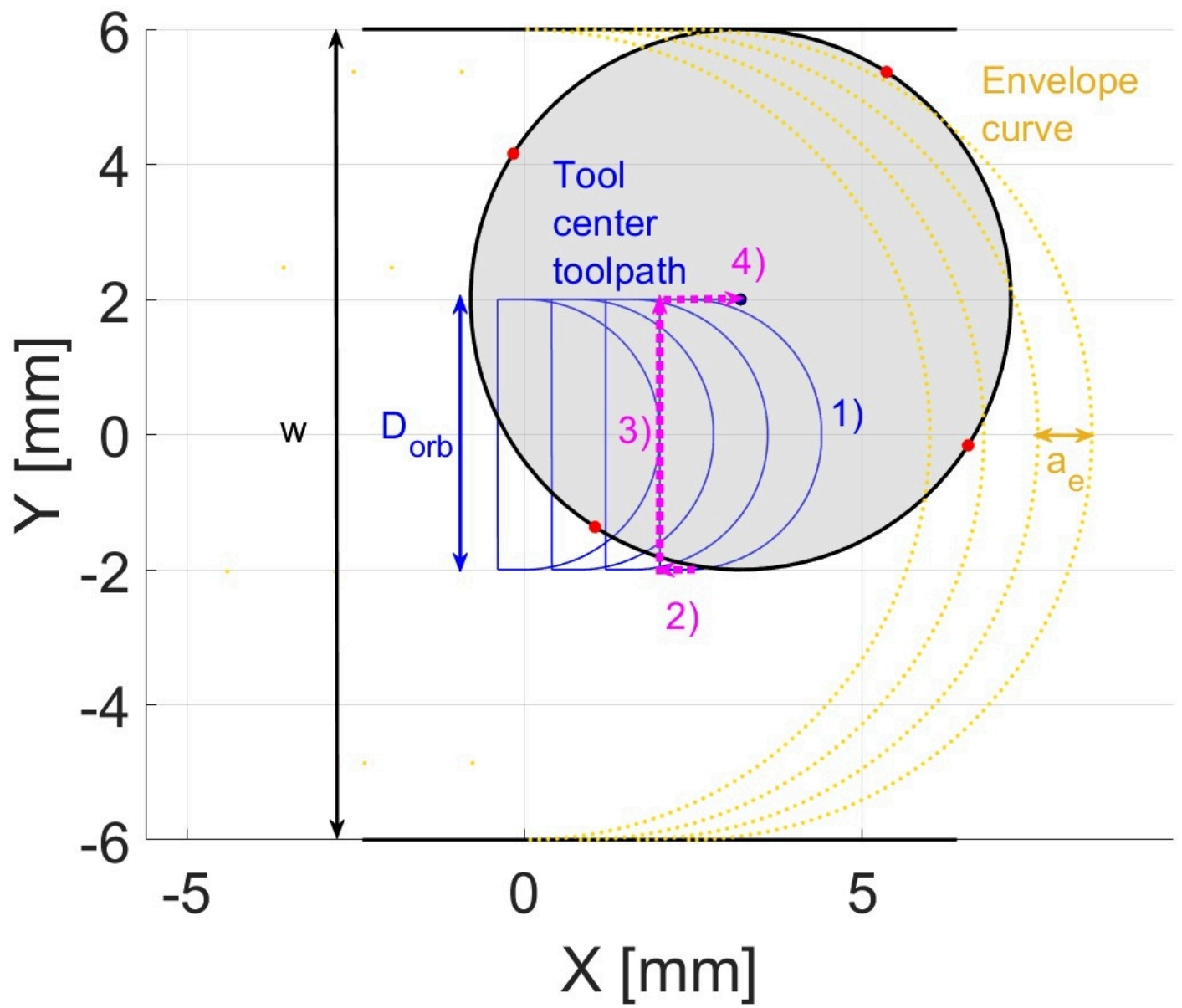
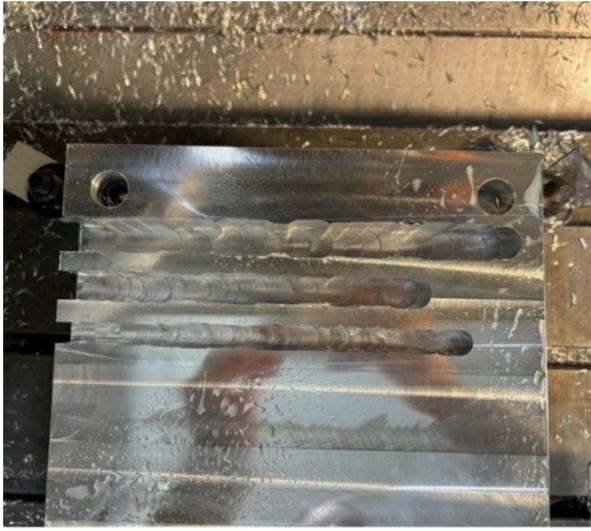


Figure 1

Trochoidal milling process: case $D = 8 \text{ mm}$, $D_{orb} = 4 \text{ mm}$ ($w = 12$).



(a)



(b)

Figure 2

(a) Channels produced during the tests; (b) Experimental set-up.

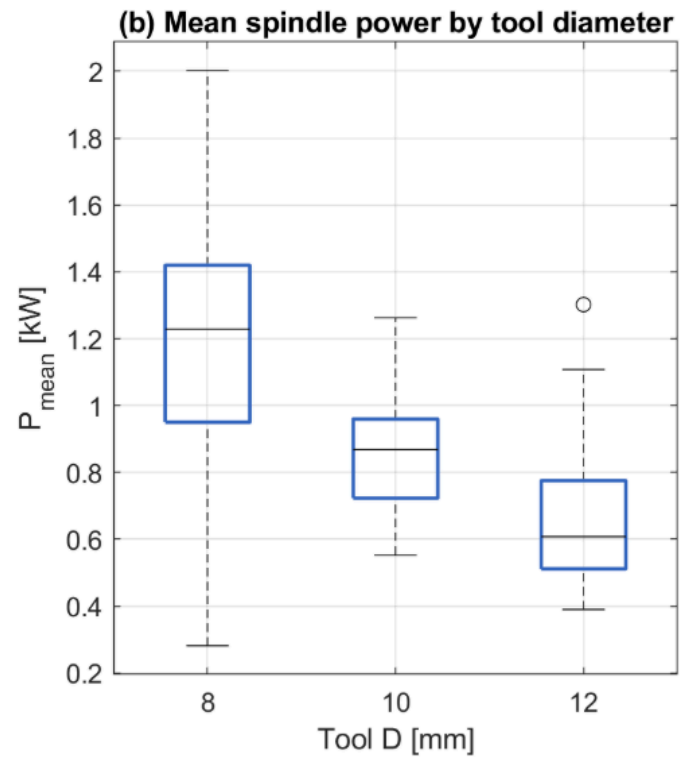
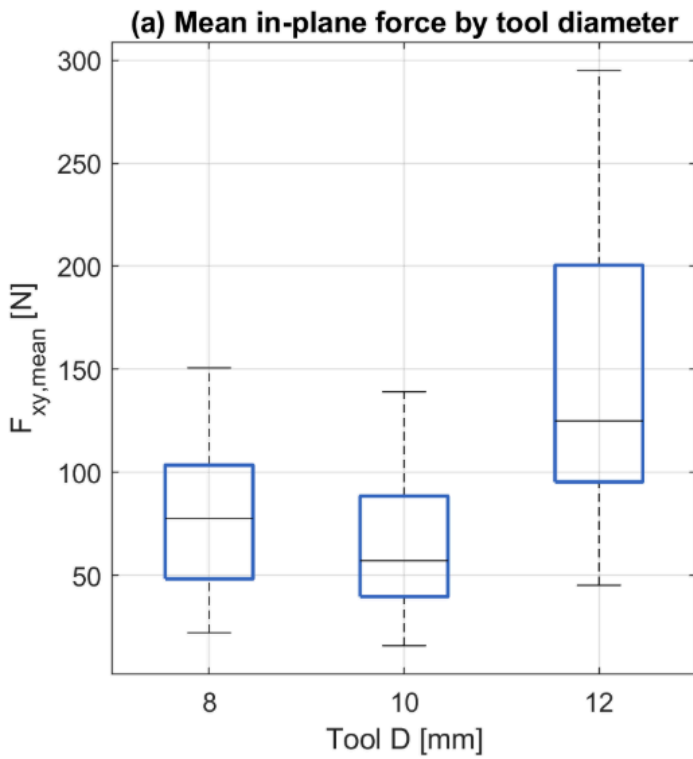


Figure 3

Distribution (median, interquartile range and whiskers) of (a) mean in-plane force and (b) mean spindle power for each of the three tested tool diameters.

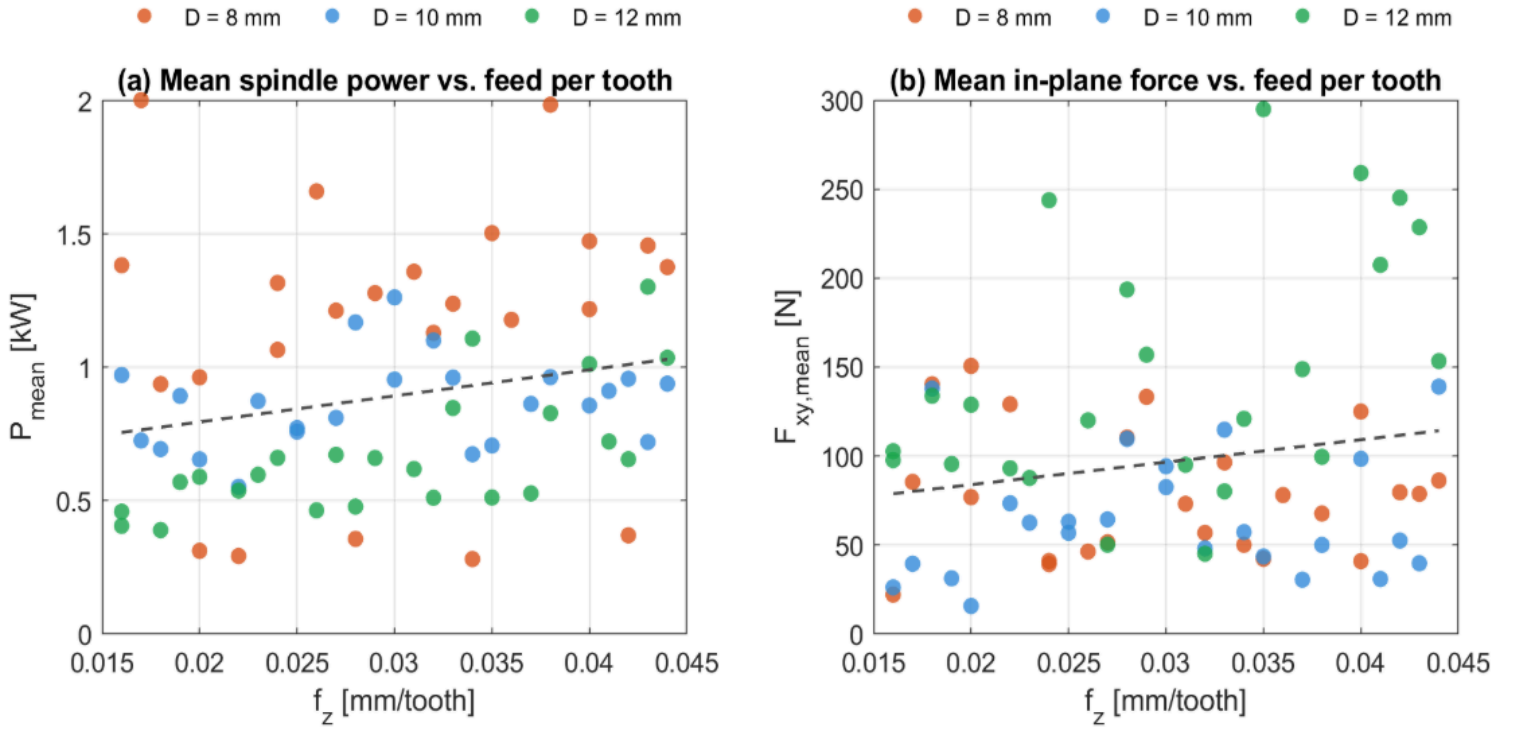


Figure 4

(a) Mean spindle power and (b) mean in-plane force as a function of feed per tooth, with least-squares trend lines (all 72 tests pooled).

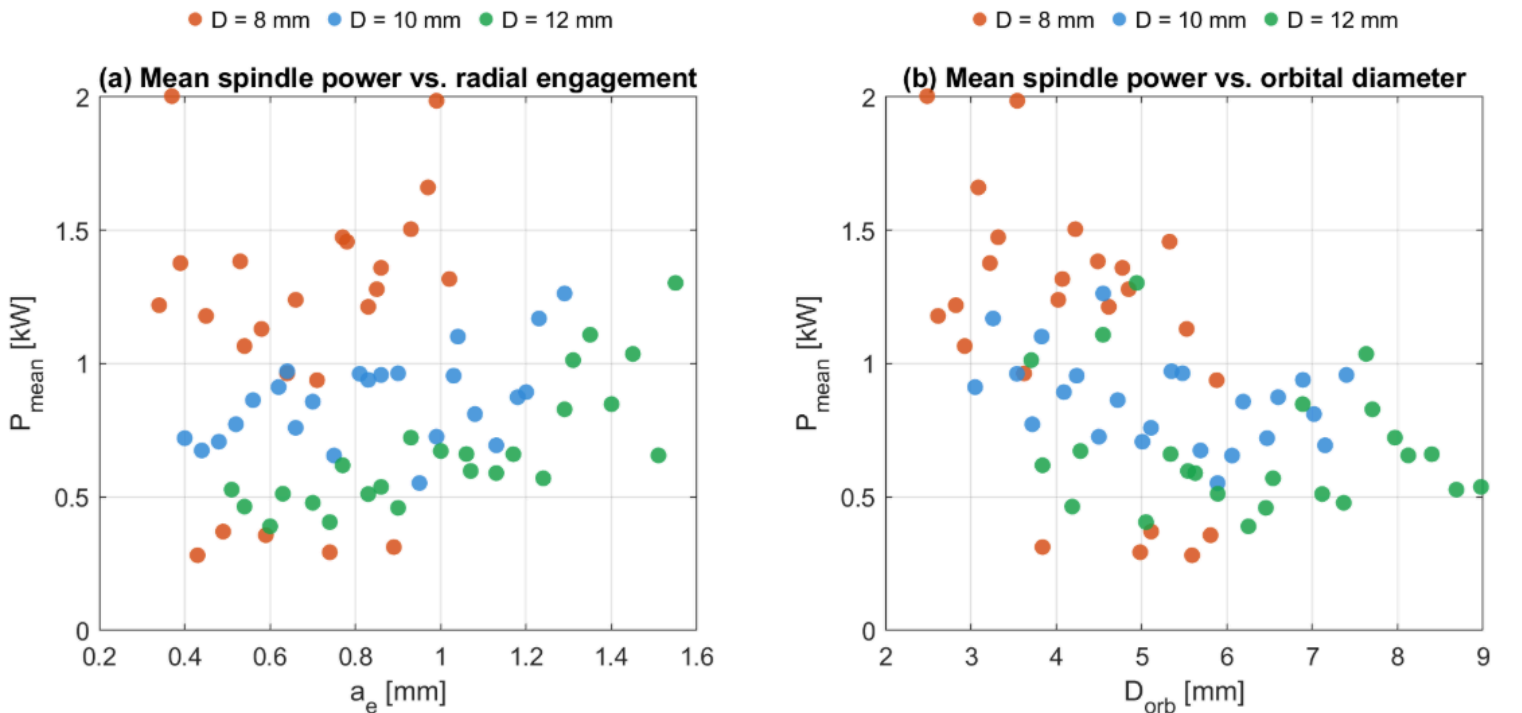


Figure 5

Mean spindle power as a function of (a) radial engagement and (b) orbital diameter, coloured by tool diameter.

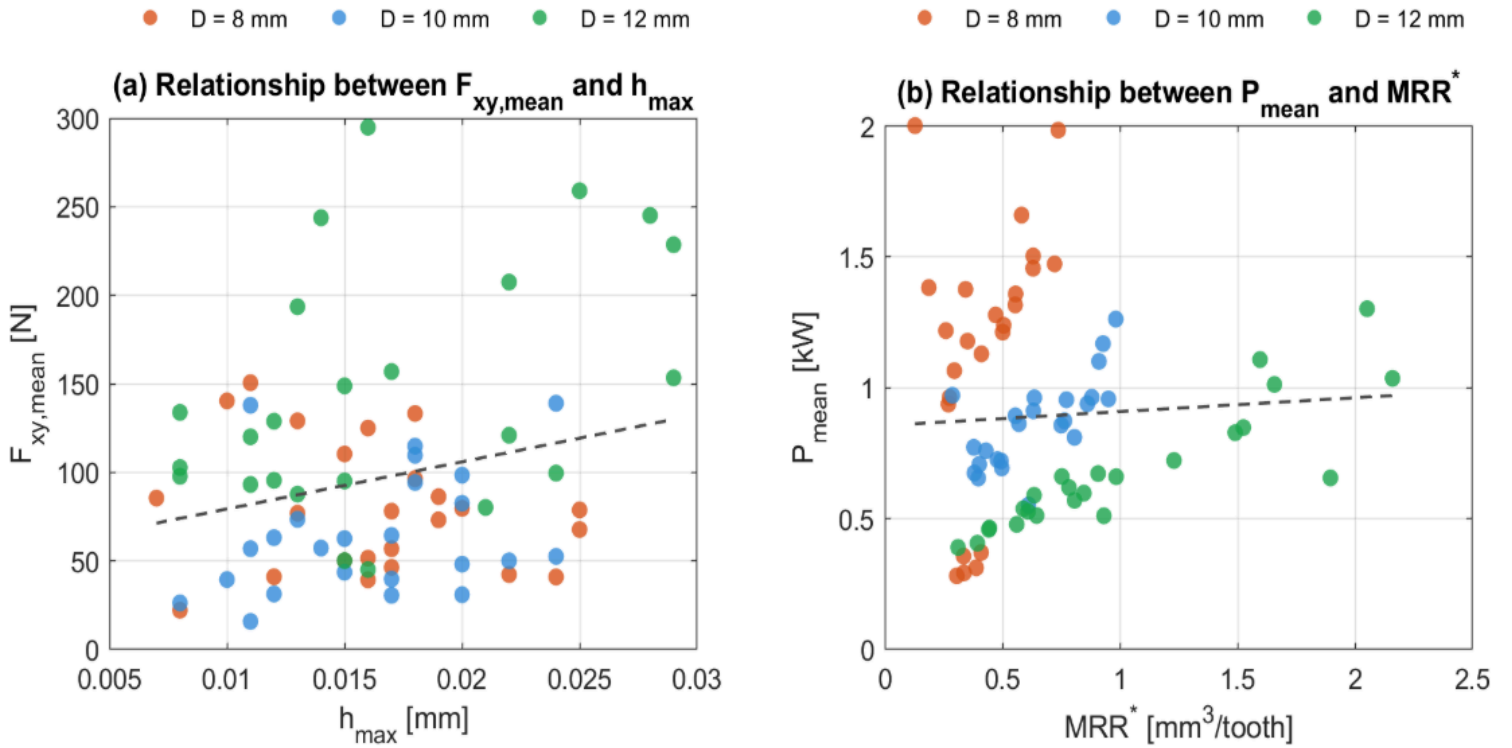


Figure 6

Relationship between selected physics-informed descriptors and measured responses: (a) $F_{xy,mean}$ as a function of h_{max} ; (b) P_{mean} as a function of the simplified process-severity descriptor MRR^* .

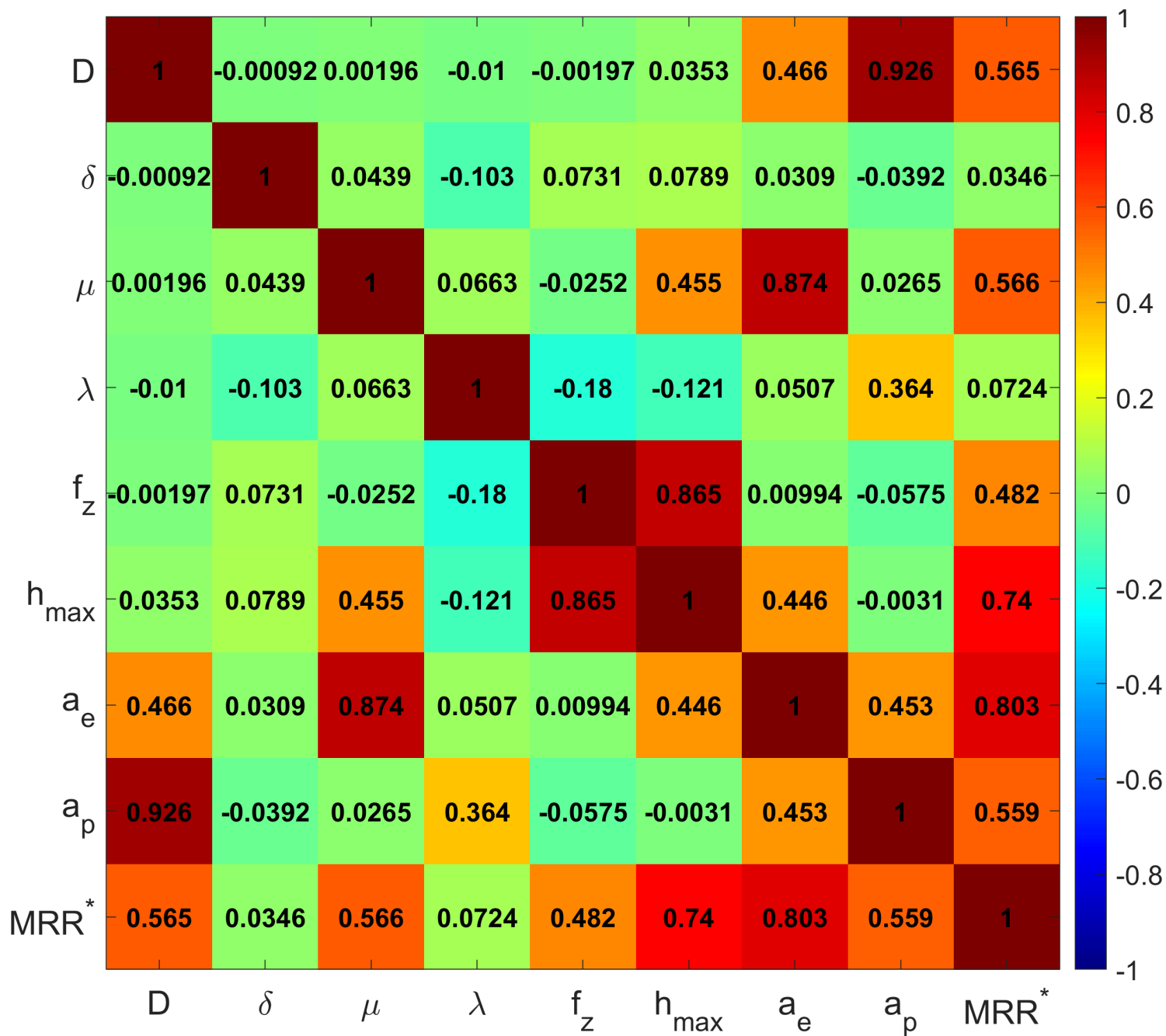


Figure 7

Feature correlation matrix for the enriched 5+4 input structure.

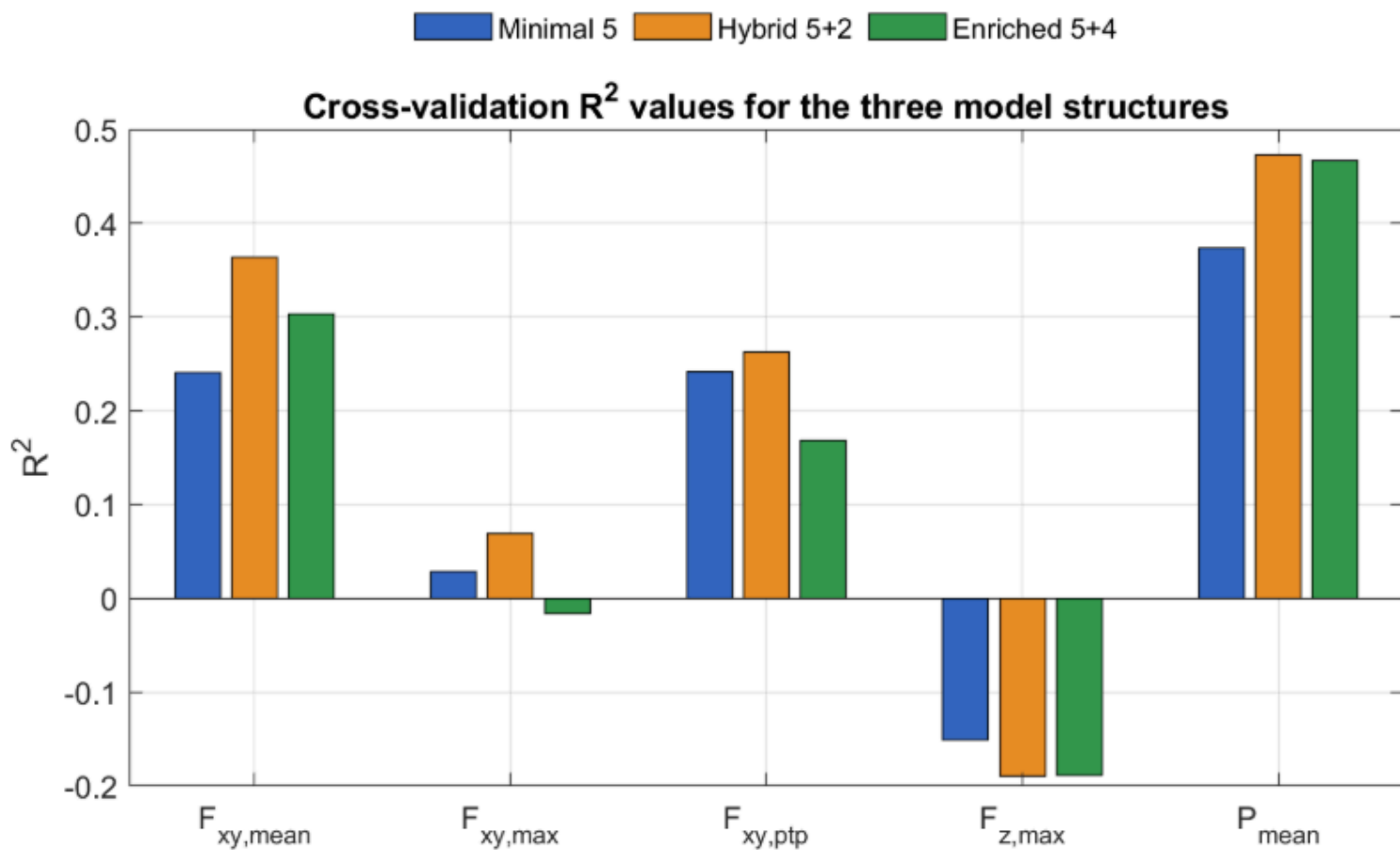


Figure 8

Comparison of cross-validation R^2 values for the three model structures and the five output targets.

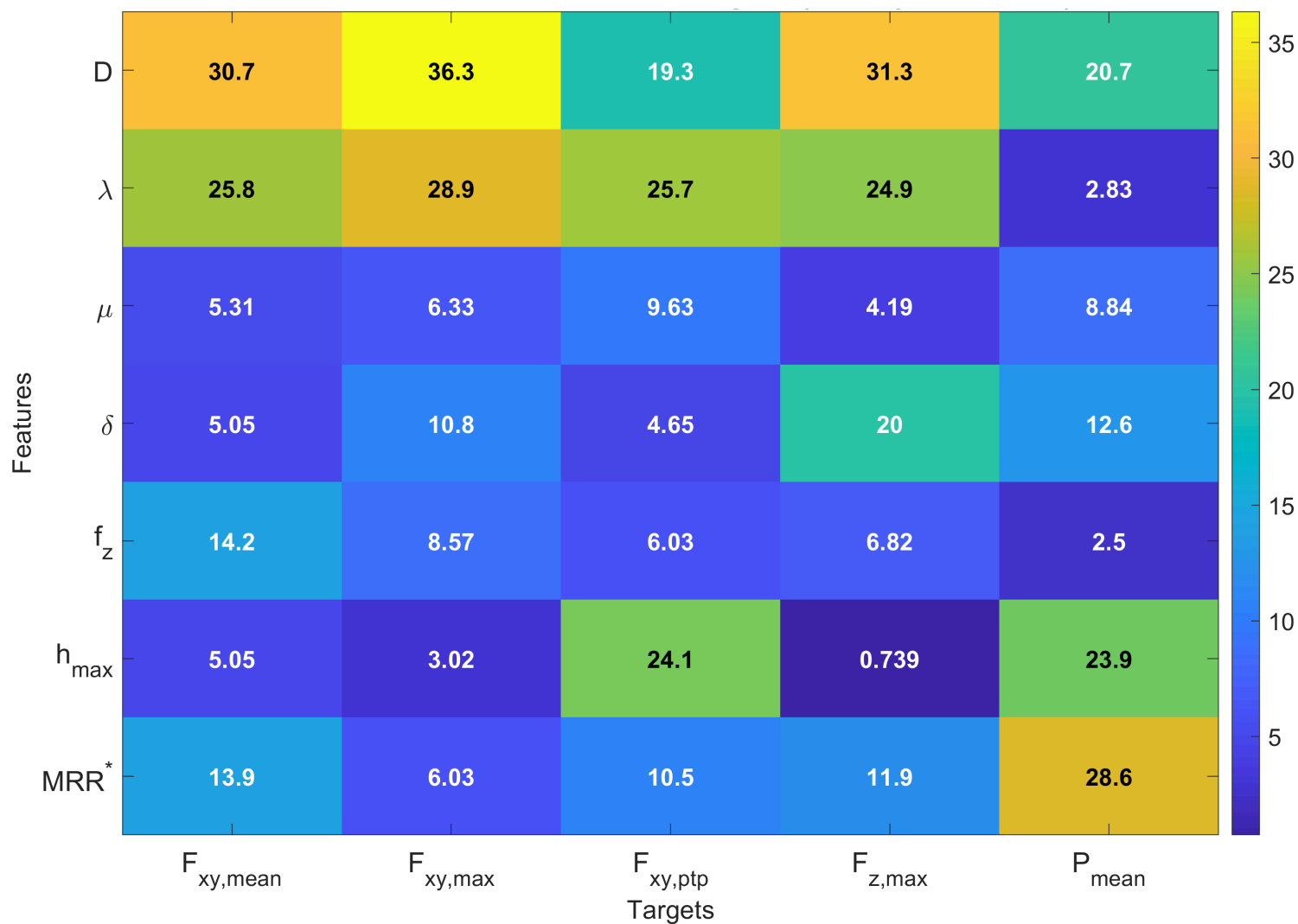


Figure 9

Feature importance across targets for the hybrid 5+2 model.

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- [AppendixAandB.docx](#)