Data-Driven Part Quality Analysis and Prediction in Cooperative 3D Printing

Anuj Swaminathan*, Zhenghui Sha*[1]
*Walker Department of Mechanical Engineering, The University of Texas at Austin, TX 78712

Abstract

Cooperative 3D printing (C3DP) is a primitive form of swarm manufacturing that utilizes multiple 3D printing robots to overcome the scalability issues present in traditional fused filament fabrication. Our prior work has established various aspects of the operational framework of C3DP, but full autonomy, reliability, and adaptability of the system cannot be achieved without advanced process monitoring and prediction frameworks. To this end, we present a methodology designed for C3DP that utilizes computer vision to assess part quality in situ and a machine learning model to identify trends and make predictions based on the collected datasets. To collect data, a full factorial design of experiment was performed by sweeping parameter combinations of bed temperature, nozzle temperature, print speed, and extrusion multiplier on standardized test coupons, while quantifying warping and top-down quality (the geometric similarity between the printed part's top surface and its intended design) as response variables. Parameter interactions and physics-informed features were also incorporated, and the final models demonstrated improved prediction capability during empirical validation studies.

Keywords: Autonomous Robotic Systems, Hybrid Manufacturing, Swarm Manufacturing, Data-Driven Design, Computer Vision.

1. Introduction

Additive Manufacturing (AM) has evolved not only in its applications but also in its process methodologies. Traditional single-printer fused filament fabrication (FFF) is limited by deposition rate and fixed build volume, constraining workflow efficiency and the tradeoff between part quality and cost. Cooperative 3D Printing (C3DP), an emerging form of Swarm Manufacturing (SM), addresses these limitations by deploying multiple mobile robotic printers on a modular, open factory floor. Figure 1 shows the SCARA printing robots (I) as well as the mobile transporter robots (II) [1].

The C3DP platform consists of a grid of tiles with slots for printers and other robots. Each printer features a Selective Compliance Articulated Robot Arm (SCARA) design, drawing power from nearby nodes and depositing polymers like poly-lactic acid (PLA) onto configurable GeckoTek build surfaces. The system incorporates established strategies for geometric partitioning, scheduling, placement, and path planning, as illustrated in Figure 2 [1–3].

C3DP operates by dividing large objects into smaller, manageable sections, a process known as "chunking", which includes both Z and XY chunking [1–3]. Each chunk is allocated a specific location on the factory floor, and an optimized schedule is generated to maximize printer efficiency [4]. Advanced path planning and motion algorithms prevent collisions as printers move, all coordinated through a central control hub that manages commands and collects operational data such as nozzle temperature, print speed, and extrusion rates from the printers [5, 6].

Despite the establishment of standard parameters such as nozzle temperature and print speed, the C3DP platform continues to face persistent challenges, including extrusion inconsistencies,

[1]Corresponding author: zsha@austin.utexas.edu

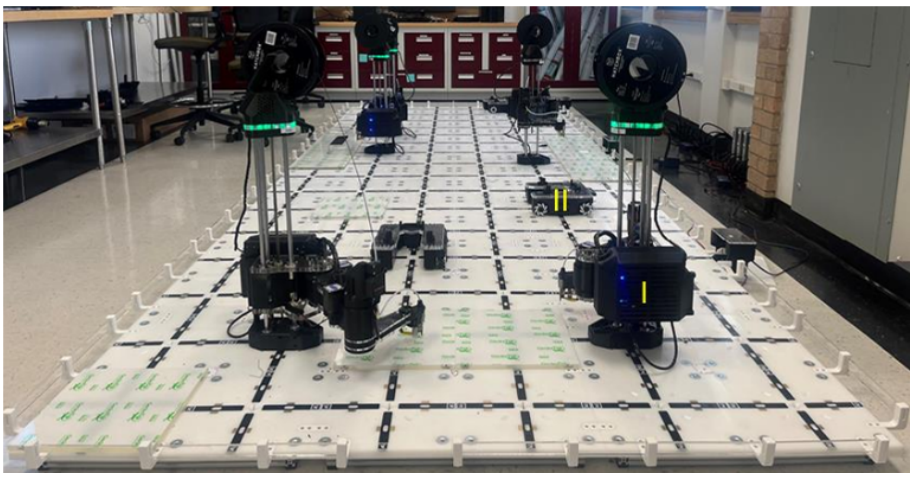


Figure 1: C3DP Platform. (I) SCARA printing robot and (II) mobile transporter robot.

bed miscalibration, warping, and stringing, which collectively impact the reliability and quality of printed parts. As illustrated in Figure 3, common FFF defects include warping (I), layer separation (II), and interfacial gaps (III) and these defects remain prevalent on the current platform. While these defects are manageable in low-cost FFF builds, the integration of advanced manufacturing processes in C3DP, such as direct metal laser sintering (DMLS), cold spray (CS), wire-arc additive manufacturing (WAAM), and subtractive techniques, will require a robust, data-driven approach to fully understand and control parameter-induced variability in part quality.

To address these challenges, our prior work introduced machine-learning (ML) and computer-vision (CV) based methodologies for in-situ print monitoring and closed-loop feedback control, to move towards real-time error mitigation without interrupting builds [7, 8]. The addition of a functional heated print bed, activated automatically upon detection of warping, further enhanced the system's reliability. The current closed-loop feedback platform, shown in Figure 4, is equipped with two SCARA robots in opposing configuration (I and II), a heated bed build surface (III), the heated bed controller (IV), and two Intel RealSense depth cameras (V). These advancements establish the need for continual development of in-situ monitoring methodologies as C3DP and broader swarm manufacturing systems become more autonomous.

A comprehensive analysis of machine performance during builds is essential for identifying and eliminating the root causes of printing errors. This study extends data-driven approaches by quantifying part quality variability through CV and correlating it with machine input parameters using statistical methods. This enables the optimization of future builds via tailored, data-backed parameter selection and physics-informed modeling, while ML integration provides predictive insights for final part quality.

The paper proceeds as follows. In Section 2, we present our literature review and identify research gaps in in-situ monitoring and data-driven additive manufacturing, along with a discussion of the current swarm platform architecture and our recent enhancements. Section 3 presents a detailed methodology that covers experimental design and data analysis. In Section 4, we discuss the results and discuss the performance of the ML algorithm. Finally, in Section 5, we conclude this paper with our closing thoughts on the limitations of this study and the potential directions for

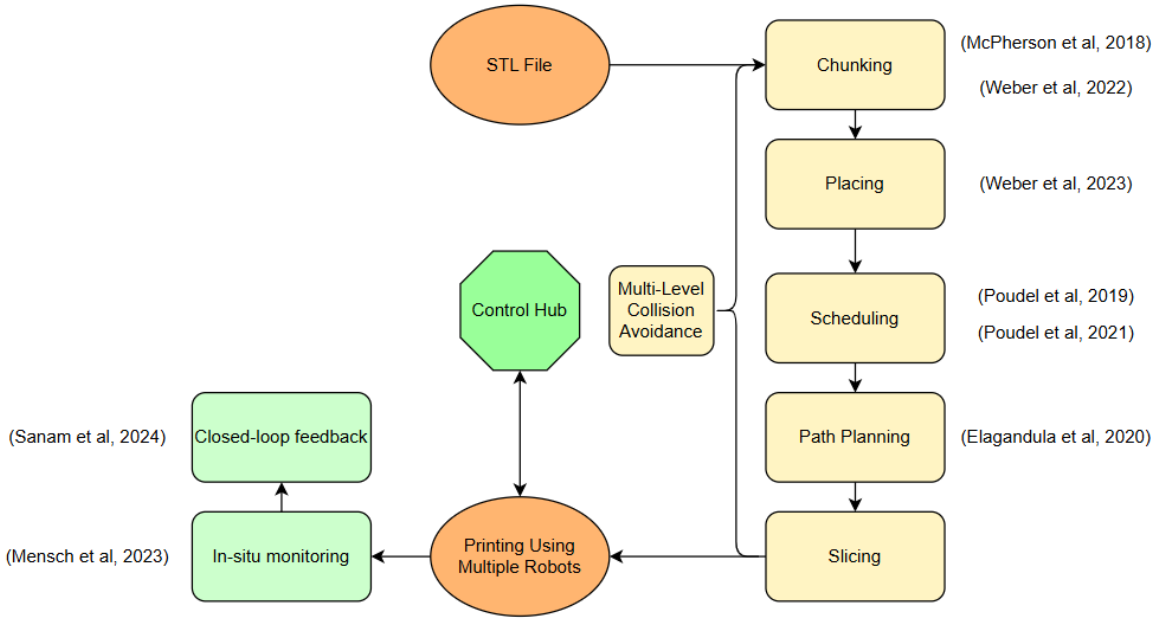


Figure 2: C3DP process framework

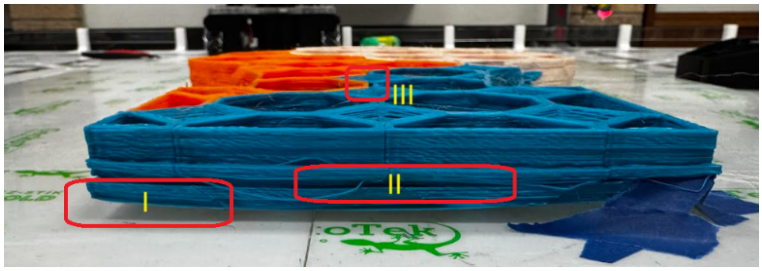


Figure 3: Common issues in C3DP. (I) Warping, (II) layer separation, and (III) interfacial gaps.

future research.

2. Literature Review

Process monitoring in additive manufacturing (AM) remains a critical research focus due to inherent variability in part quality, particularly in sectors like metal AM, where aerospace applications demand stringent quality standards [9].Even consumer-focused fused deposition modeling (FFF) exhibits a 20% failure rate among hobbyist users [10], highlighting the need to develop accessible monitoring solutions across various AM platforms. Vision-based methods dominate this space, with 2D imaging widely adopted for defect detection [11–13], and deep learning enabling real-time error correction by comparing live camera feeds to STL/G-code data [14]. External cameras paired with machine learning can also detect warping and stringing during printing [15–18].

Despite advancements, key gaps persist in AM monitoring: 1) limited accuracy in defect detection, 2) insufficient impact of in-situ systems on final part quality, and 3) underutilization of data-

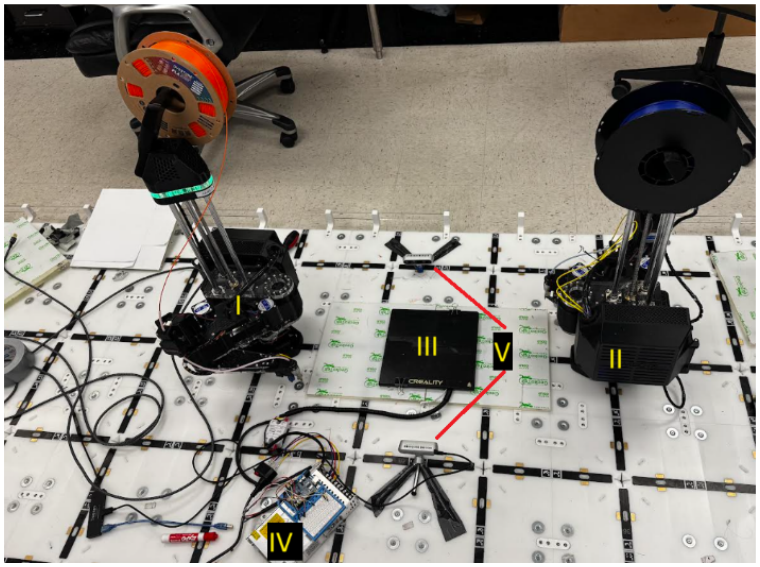


Figure 4: C3DP heated bed configuration. (I) SCARA robot 1, (II) SCARA robot 2, (III) heated bed build surface, (IV) heated bed control system, and (V) real sense depth cameras.

driven insights for process optimization. Previous work in the C3DP platform addressed these through enhanced image processing, ML-driven defect detection, and custom hardware such as heated beds [7, 8]. However, quantitative correlations between machine parameters and part quality in C3DP systems remain underexplored.

Machine learning is increasingly recognized as transformative for AM, enabling design optimization, material tuning, and computational efficiency. While Goh et al. [19] highlight ML's potential to reduce reliance on physics-based modeling, challenges in data acquisition and real-time analysis persist [20–22]. Earlier studies demonstrate ML's utility in predicting mechanical properties, such as compressive strength, tensile strength, and dimensional accuracy [23, 24]. These predictions are often based on parameters like layer thickness and print orientation; however, these approaches often rely on post-build measurements rather than in-situ data [15, 25, 26].

Recent efforts integrate computer vision (CV) for in-situ defect detection [7, 15], yet most lack integration with design of experiments (DoE) workflows or quantitative defect quantification [26]. This study bridges these gaps by combining multi-camera CV for real-time quality assessment, full-factorial DoE for systematic data collection, and ML for predictive parameter optimization, all implemented on a C3DP platform. This integrated approach advances automation in hybrid manufacturing, offering a scalable framework for industrial applications.

3. Methodology

A full-factorial DoE was selected to comprehensively understand how input parameters affect part quality on the C3DP platform. This approach was chosen over fractional-factorial designs to avoid missing critical multi-parameter interactions and ensure robust machine learning model development. Drawing from established FFF literature and the platform's SCARA robot similarities to traditional Cartesian printers [27–29], four key control variables were identified: (1) print speed, (2) nozzle temperature, (3) bed temperature, and (4) extrusion multiplier (material flow rate as percentage) [30–32].

Print speed, nozzle temperature, and extrusion multiplier were directly controlled through slicer configuration, with extrusion multipliers expressed as percentages (e.g., 1.0X = 100% flow, 1.1X = 110% flow). Since the flow rate of the material would be based on the print speed, layer height (fixed at 0.2 mm), nozzle diameter (fixed at 0.4 mm), and PLA density (1.24 g/cm³), the experimental extrusion multipliers of 1.0, 1.1, and 1.2 corresponded to mass flow rates ranging from 10.7-17.9 g/hr at 100% flow, 11.8-19.7 g/hr at 110% flow, and 12.9-21.4 g/hr at 120% flow across the tested print speeds of 30-50 mm/s. Bed temperature, newly integrated through the addition of a hotbed to the C3DP platform, allowed the inclusion of this parameter in the study. The total number of test cases (n) was determined by Equation 1, where y represents the number of control variables and x denotes the number of levels per variable. Level ranges were established considering time efficiency, computational cost, and data resolution requirements.

$$n = x^y \tag{1}$$

With four control variables, three levels per parameter yielded 81 total test cases ($n = 3^4$), balancing data resolution with computational efficiency. This configuration was optimal for identifying non-linear parameter relationships while avoiding the inefficiencies of fewer cases ($n = 27$) or excessive resource demands of higher-level designs ($n = 243$). Parameter values, particularly bed temperature, nozzle temperature, and print speed, were selected based on established FFF operating ranges for PLA filament (see Table 1) [20, 33]. Extrusion multiplier test ranges were selected to be above 100% (over-extrusion cases) due to prior machine performance studies that demonstrated excessive warping with higher material flow rates [7]. Thus, this parameter range was determined to yield more observable and characteristic defects.

Table 1: Control variable levels and values.

Control Variables	**Level Values**		
	1	2	3
Print Speed (mm/s)	30	40	50
Bed Temperature (°C)	50	60	70
Nozzle Temperature (°C)	230	240	250
Extrusion Multiplier	1.0	1.1	1.2

Two response variables quantified part quality in-situ: (1) warping quality score and (2) top-down quality score. The warping quality score measures how much a printed part deviates from its intended geometry due to warping, using side-view images and computer vision analysis. Higher

scores indicate minimal warping and better dimensional accuracy, while lower scores reflect greater distortion. The top-down quality score evaluates the geometric accuracy and surface finish of the part's top surface, based on top-down images compared to the reference design. A higher score means the printed surface closely matches the intended shape and features, while a lower score points to surface defects or missing features. Both scores are scaled from 1 to 10, with 10 representing ideal quality and lower values indicating increasing deviation from the target geometry. Thus, the higher scores indicate better part quality for both metrics, with "ideal" meaning a perfect score of 10. Ideal quality corresponds to a perfect score of 10, representing printed parts that match the ground-truth reference model from the slicer with no detectable deviations in geometry or surface features according to the established grading criteria. These metrics address gaps in existing literature where CV methods detect defects but lack numerical quantification or simultaneous data acquisition capability [7, 8, 27–30, 32]. Unlike traditional post-build metrology approaches focusing on dimensional tolerance and mechanical properties, this study prioritized real-time quality assessment during printing.

A custom test coupon was designed featuring representative geometries to evaluate printer capability (see Figure 5). The geometry incorporates a 20×20 mm square base for initial layer assessment and a 7 mm diameter circular column for continuous curvature deposition, with an overall height of 23 mm. The test coupon's shape was chosen to balance experimental efficiency and robust quality assessment. The square base ensures consistent initial layer adhesion and exaggerates warping at the corners, making defects easier to detect, while the central circular column introduces continuous curvature to test the printer's ability to maintain dimensional accuracy in both planar and curved features. This hybrid geometry enables simultaneous evaluation of warping and vertical fidelity, mirrors common real-world part features, and supports rapid, high-throughput testing with optimized build times. Standardizing this shape ensures results are comparable across printers and parameter sets, making it an effective benchmark for process monitoring, algorithm calibration, and quality prediction in 3D printing. Cura 5.10.0 served as the slicer with standardized parameters of 0.2 mm layer height and 30% grid infill density.

To capture part quality as a function of real-time machine performance across varying cross-sections, multiple response variable scores were obtained throughout each build. Rather than single-point measurements that would miss developing defects due to the ever-changing Z-direction geometry, six top-down and six warping scores were captured per test case at predetermined heights: 3 mm, 7 mm, 11 mm, 15 mm, 19 mm, and 23 mm. This yielded 486 individual scores per response variable across all 81 test cases, which were subsequently averaged to produce final scores for statistical analysis. The capture heights follow Equation 2, where layer height (l) was determined by tracking variable (k). This function enabled systematic data collection at equally spaced intervals from build start to completion, facilitating script indexing and post-processing.

$$l = 4k + 3 \qquad (\text{where } k = 0{:}5) \tag{2}$$

Test cases were organized into three batches by bed temperature for practical implementation: Batch 1 ($50\,^{\circ}$C), Batch 2 ($60\,^{\circ}$C), and Batch 3 ($70\,^{\circ}$C), with 27 unique coupons per batch. This grouping simplified experimental setup, since bed temperature required separate programming from slicer settings. While batching facilitated coupon sorting and preliminary validation of

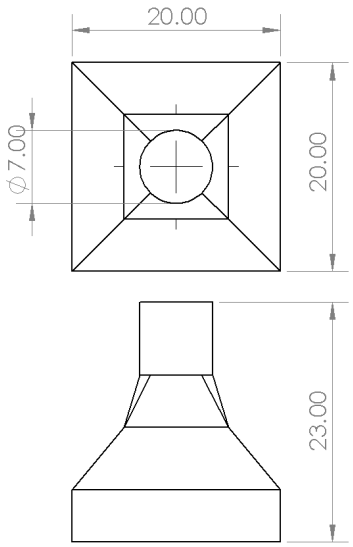


Figure 5: Test coupon geometry (wireframe). All dimensions in mm.

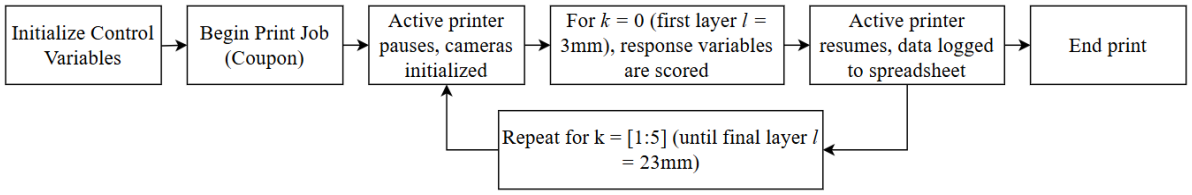


Figure 6: Experimental workflow and data collection process for each test coupon in full factorial DoE.

scoring systems, subsequent statistical analyses regrouped data by all input parameters for comprehensive evaluation. The complete experimental workflow is shown in Figure 6, encompassing platform automation, data capture, scoring, in-situ image processing, and numerical post-processing. Data logging and statistical analysis utilized Microsoft Excel (Data Analysis ToolPak) and Python (Scikit-Learn) for regression modeling.

Several hardware upgrades and software pipelines were implemented to enable robust experimental control from job initiation through data logging. The primary upgrade was a custom-built heated bed system using off-the-shelf components for easy integration with existing GeckoTek build plates. An Arduino Uno provides pulse-width-modulation (PWM) temperature control via thermistor feedback, achieving precise closed-loop control within ± 2 °C. The system accommodates bed mesh leveling adjustments and modified G-code Z-homing for the elevated build surface, while updated Cura profiles support the expanded 300×600 mm build area. This discretized control enables material-specific temperature optimization (PLA, CPE) and improved energy efficiency, which are critical for platform scalability. The heated bed architecture and upgraded controller are shown in Appendix A.

Our previous top-mounted camera configurations proved inadequate in terms of image quality and usability [8]. They required cameras positioned within the optimal $15-50$ mm operating depth

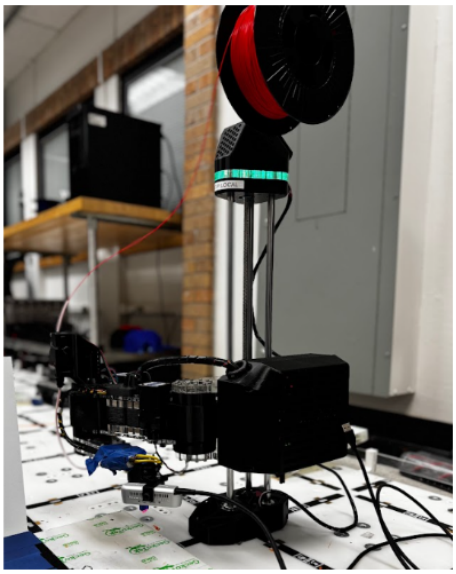


Figure 7: Fully integrated passive imaging robot.

field while maintaining constant distance from the build's topmost feature as parts grew vertically, which is impossible with fixed mounting solutions [7]. In this study, the solution employed a dedicated SCARA robot as a cameraman, enabling close-up, aligned imaging with software-driven dynamic movement. A custom 3D-printed fixture mounted the RealSense camera as an adjustable end-effector, creating hybrid manufacturing agents: an active printing robot and passive imaging robot. The fully integrated camera robot is shown in Figure 7.

To minimize background interference, a white backdrop was positioned behind the side-view camera's capture field. The heated bed's reflective glass surface was covered with blue masking tape to reduce glare from overhead fluorescent lighting. Previous work identified glare from various sources, including BLTouch sensor emissions, as problematic for image capture, necessitating conservative background noise control measures. The final full opposing-configuration, hybrid manufacturing build setup is shown in Figure 8.

3.1. Machine Automation, Data Collection, and In-Situ Image Processing

The experimental process employed four coordinated scripts: (1) printer control and synchronization, (2) top-down image quality analysis, (3) side-view warping analysis, and (4) data logging. Script 1 served as the master controller, triggering scripts 2–4 at predetermined capture points, enabling independent script modification while maintaining system efficiency.

Script 1 coordinated dual SCARA printers using HTTP commands via the Duet API to monitor position, layer height, and G-code execution. At specified capture heights (Equation (2)), the active printer paused and moved away from the build plate, allowing the passive camera robot to position above the part for top-down imaging while simultaneously triggering side-view analysis. The workflow alternated between active build and passive imaging states (Figure 9) until job completion.

Building on prior similarity characterization work, Script 2 used PyRealSense and OpenCV

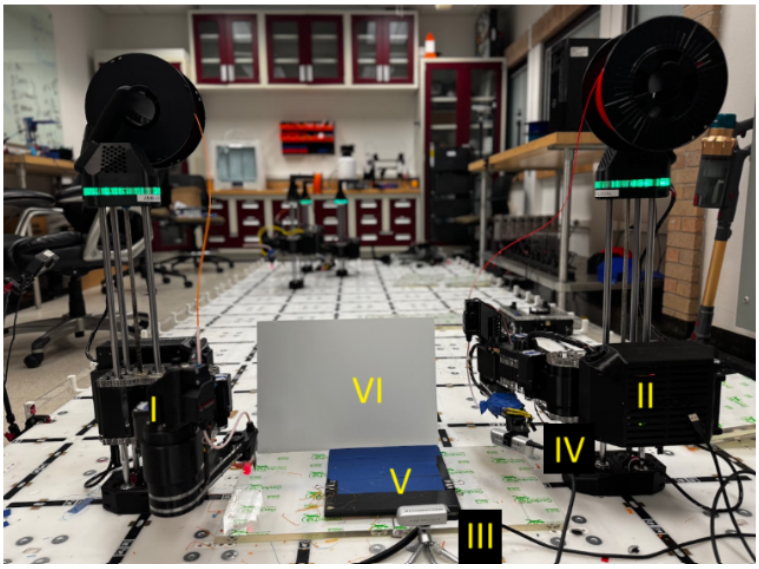


Figure 8: Labeled components of the full testbed setup. (I) Active printing robot, (II) passive imaging robot, (III) side view camera, (IV) top view camera, (V) heated bed, (VI) backdrop.

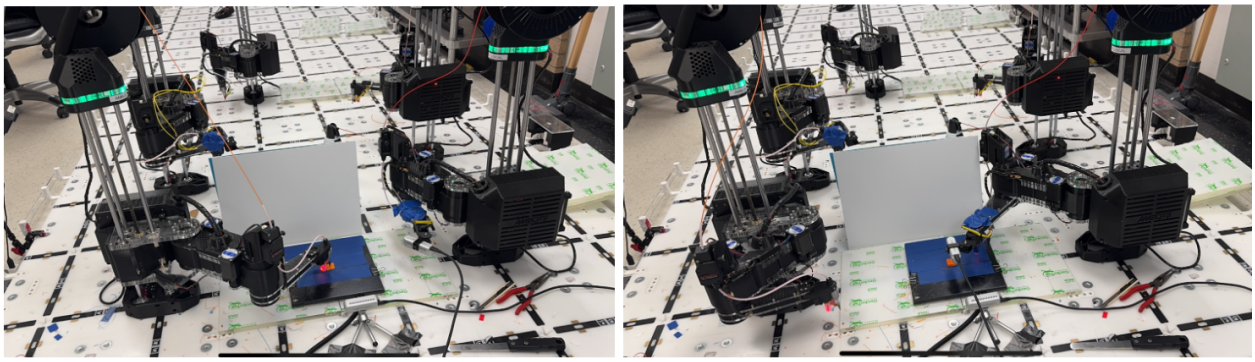


Figure 9: Active (left) and passive (right) state of C3DP platform.

libraries to compare printed layers against reference slicer images (i.e., the ground truth) [8]. The process involved HSV color masking ($H \in [5, 25]$ for orange filament), Gaussian blur preprocessing, and contour extraction with sub-pixel approximation. A three-metric scoring system (1-10 scale) weighted shape similarity (40%), area coverage (40%), and contour presence (20%), with non-linear scaling. Figure 10 demonstrates the script's output over a complete build.

All scoring algorithms were calibrated using standardized test coupons printed on an Ultimaker S3 (assumed ideal quality, score = 10). Variants of the standardized test coupons with deliberately induced defects (also produced by the Ultimaker) were used in a qualitative trial-and-error process to fine-tune the weights of the scoring system parameters by manually adjusting the percentage ranges in Python code, while cross-referencing output scores against human observations. Three additional calibration models of different geometries were also used to fine-tune and verify the scoring capability of the system in a similar fashion. The calibration models can be seen in Appendix A.

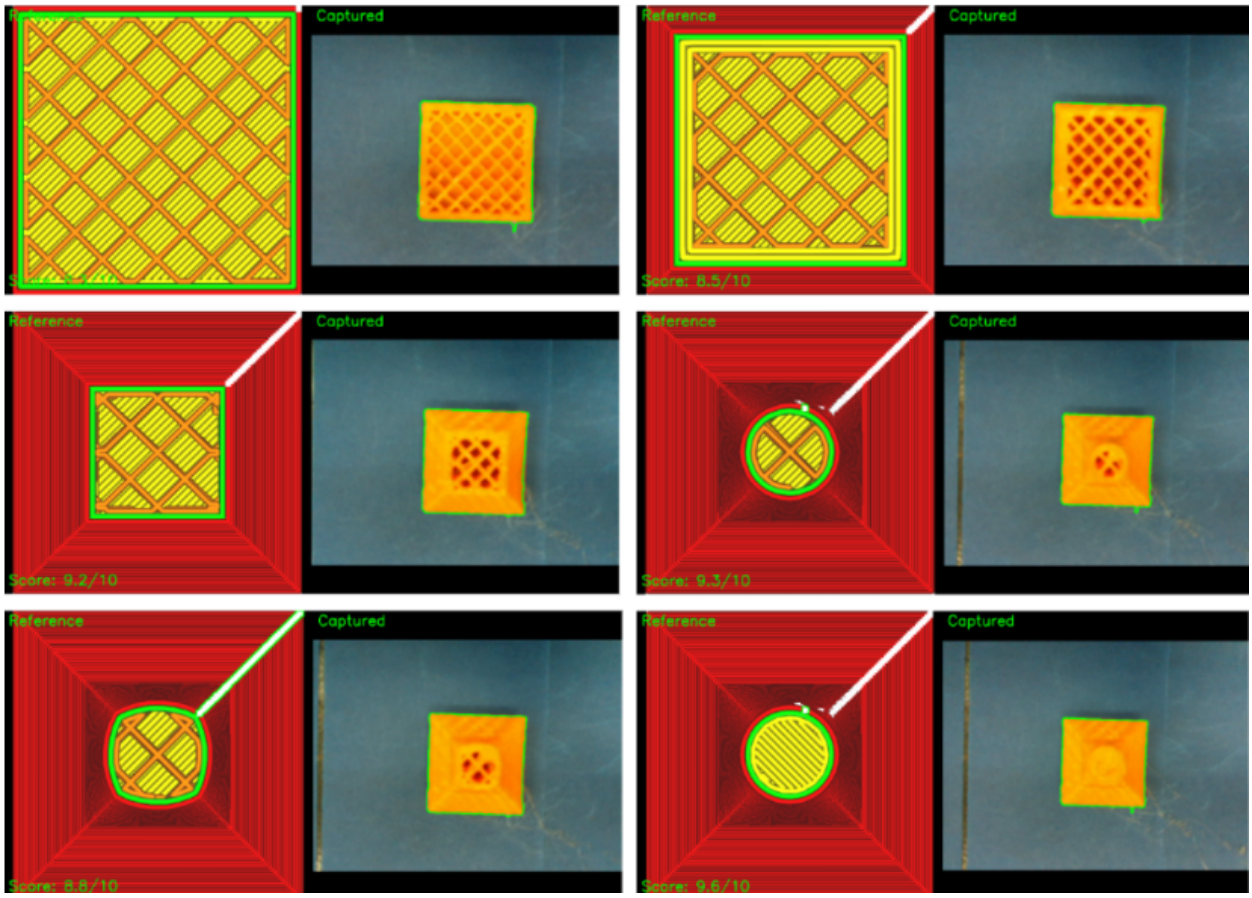


Figure 10: Sample output of the top-down analysis script over a full build (images progress from left to right, then top to bottom).

Table 2: Warping score assignment criteria.

Score	Quality	Contour distance (pixels)	Print impact
9–10	Excellent	<15	No visible defects
8–8.9	Good	16–20	Minor cosmetic issues
6–7.9	Acceptable	21–30	Functional but not ideal
4.5–5.9	Concerning	31–45	Risk of print failure
2–3.9	Poor	46–60	Significant deformation
0–2	Critical (failure)	>60	Print failure

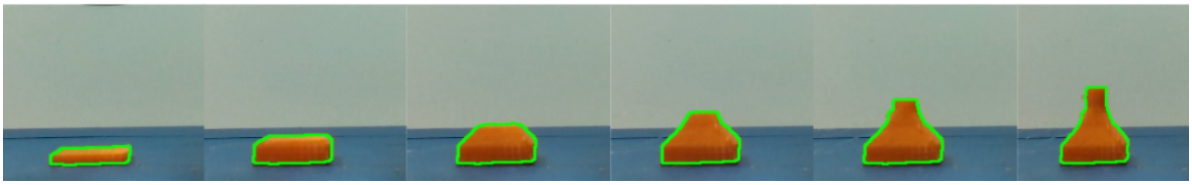


Figure 11: Sample image output of the warping script over a full build.

Script 3 employed contour-based processing with Canny edge detection and morphological operations to assess warping severity through three methods: contour distance sampling (horizontal pixel gaps), bounding box comparison (geometric feature differences), and point cloud analysis (maximum nearest-neighbor distances) [17]. The system applied nonlinear 1–10 scoring criteria detailed in Table 2, capturing six warping assessments per test case (Figure 11). Contour distances are measured in pixels, as determined by the RealSense D435i camera (with 1 px $\approx$ 0.12 mm for this setup). Based on the camera resolution and the distance from the part, a pixel range was established in order to flag early signs of warping at the smallest observable detection range. While thresholds and descriptors were validated for typical FFF test coupons, the definitions in Table 2 can be tuned for specific applications or stricter industry standards. Because regional warping scores are retained for each capture height, users may perform localized analysis in addition to reporting averaged composite metrics, facilitating detection of parameter drifts or localized failures during a build. Images of calibration parts are provided in Appendix A.

3.2. Data Analysis and Regression Models

After logging all quality scores to the master spreadsheet, the six top-down and six warping scores for each test case were averaged, yielding a single composite score for each quality metric per 81 test cases. For both metrics, higher values indicate better part quality (i.e., less warping or higher top-down fidelity). The original regional scores at all six sampled heights for each part are retained, enabling further localized or region-specific analysis if needed.

The data were then grouped by each input parameter, and a range of statistical analyses, including trend visualization, interaction plots, and correlation studies (using Python/Scikit-Learn and Excel), identified key parameter-quality relationships and influential variable combinations. A four-factor Analysis of Variance (ANOVA) and sensitivity analysis informed the development of regression models to statistically predict part quality, resulting in two distinct models for top-down and warping quality.

To further test predictive robustness, a Gaussian Process Regression (GPR) model with a radial basis function (RBF) kernel was implemented for each quality metric, incorporating physics-informed features and providing both predictions and uncertainty estimates. The models were trained on 80% of the dataset (65 cases) and validated on the remaining 20% (16 cases), enabling inverse optimization to identify optimal or worst-case parameter combinations [31].

To validate the robustness and predictive power of both the statistical and GPR models, an empirical validation study was conducted using the same standardized test coupon as the original DoE. Eight validation runs were performed, testing four scenarios: "interpolated" (within the DoE parameter space), "extrapolated" (beyond the original tested range), "GPR-optimal" (model-predicted maximum), and "GPR-worst-case" (model-predicted minimum). Model predictions were compared to actual measured quality scores from real prints. Additionally, three more test cases were evaluated on different geometries: (1) a Benchy model, (2) a latticed rocket engine, and (3) a compressor wheel, using the same workflow with results analyzed in the following section. Images of the coupons and test pieces can be found in Appendix A.

4. Results and Discussion

To qualitatively assess the scoring algorithm, coupons with the maximum, minimum, and average top-down and warping scores from each batch, grouped by bed temperature, were visually compared. For each batch, representative images, found in Appendix B, illustrate the range of print quality: Figures B1.1-1.4 display Batch 1's image data, average, minimum, and maximum scores for both top-down and warping quality; Figures B2.1-2.4 do the same for Batch 2; and Figures B3.1-3.4 for Batch 3. These comparisons demonstrate clear visual distinctions between low, average, and high-quality prints, validating that the scoring system effectively captures meaningful differences in both surface finish and geometric fidelity. Table 3 summarizes the coupon numbers corresponding to each batch's minimum, average, and maximum scores.

Table 3: Batch values for average, maximum, and minimum quality scores (scale: 1–10; higher values indicate better quality). The number in parentheses indicates the coupon number associated with that score. Data is displayed from $n = 27$ test cases per batch, for 81 total cases.

Batch Number	Top-down Quality Score			Warping Quality Score		
	Min.	Avg.	Max.	Min.	Avg.	Max.
1 (50 °C)	7.50 (46)	8.31	8.47 (40)	6.00 (46)	7.77	8.83 (19)
2 (60 °C)	8.48 (50)	8.85	8.97 (5)	6.33 (26)	7.41	8.50 (14)
3 (70 °C)	8.35 (51)	8.74	8.87 (6)	5.50 (6)	6.92	7.83 (51)

As outlined in the methodology, coupon scores were regrouped by input parameter to enable detailed statistical analysis of parameter-quality relationships. For each control variable (bed temperature, nozzle temperature, print speed, and extrusion multiplier), data were organized and visualized to reveal individual parameter effects on both top-down and warping quality scores. Figure 12 presents results for bed temperature effects, with four-panel visualization: the top row shows (left) a box plot and (right) a scatter plot with trendline for warping quality, and the bottom row shows the corresponding plots for top-down quality. The scatterplots reveal that top-down quality remains consistently high (8–9 range) and increases with bed temperature, while warping quality demonstrates greater sensitivity and a declining trend with increasing bed temperature ($R^2 = 0.27$). All

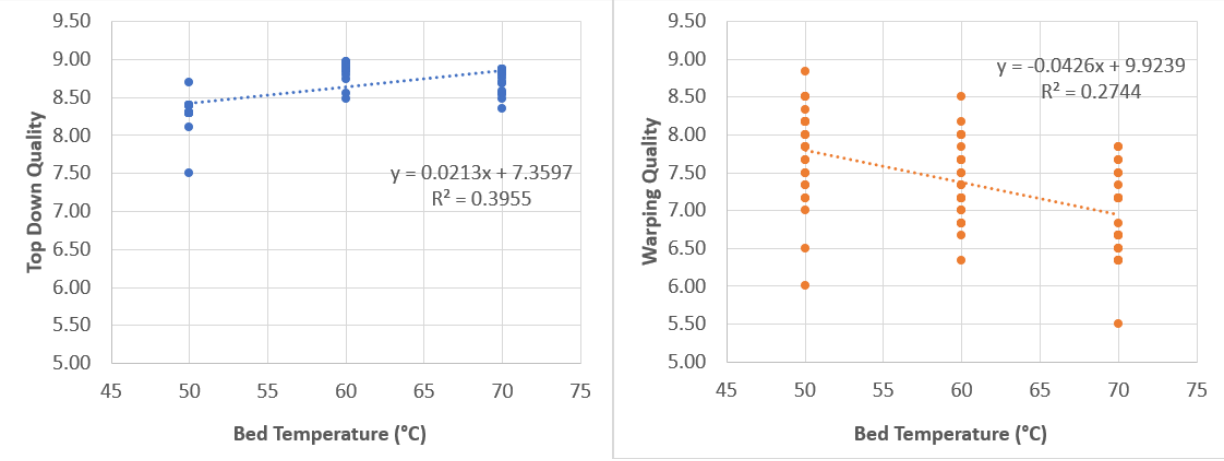


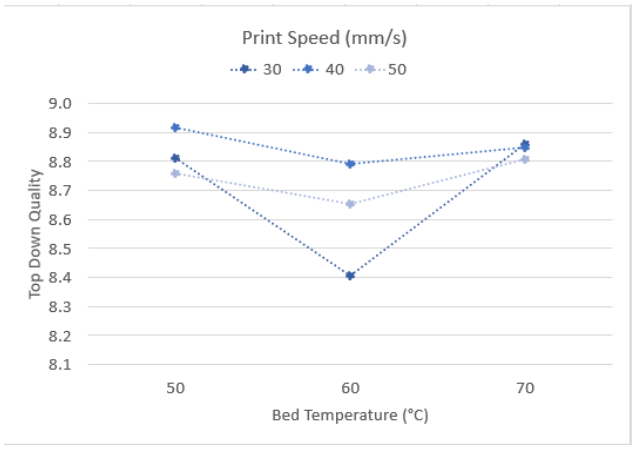
Figure 12: Effect of bed temperature on part quality scores ($n = 81$ test cases). The left panel displays the top-down quality scatterplot with the trendline, while the right panel displays the warping quality scatterplot with the trendline. All scores are on a 1–10 scale, with higher values indicating better quality.

quality score values are on a 1–10 scale, with higher values indicating better part quality.

To analyze parameter interactions, interaction charts were generated by plotting average quality scores against one parameter while representing varying levels of another parameter as separate lines. This approach reveals both main effects and interaction effects, enabling identification of optimal parameter combinations and highlighting the sensitivity of each quality metric to specific variables. Four interaction plots were examined: (1) bed/nozzle temperature, (2) bed temperature/print speed (shown in Figure 13), (3) print speed/extrusion multiplier, and (4) nozzle temperature/extrusion multiplier (Appendix C).

Four-factor ANOVA and sensitivity analyses determined statistical significance and relative parameter importance for each quality metric. ANOVA decomposed total variance into parameter effects and unexplained error, computing F-statistics and p-values to assess significance. Sensitivity analysis calculated each parameter's contribution percentage by dividing the individual sum of squares by total sum of squares. For top-down quality, bed temperature was the most prominent parameter with 13.49% contribution ($p = 0.002$), but other parameters still showed statistically relevant contributions (6.04% for nozzle temperature and 5.67% for print speed). For warping quality, print speed was most influential at 27.07% contribution ($p<0.001$), followed by bed temperature at 4.32%, confirming that warping control is more single-parameter driven than top-down quality maintenance. This result suggests that for minimizing warping, optimizing print speed alone may yield the largest marginal returns, whereas achieving high top-down quality may require simultaneous tuning of multiple parameters, especially bed and nozzle temperature.

Based on these findings, it was revealed that parameter interactions and effects had to be strongly considered for any predictive modeling capability. Linear regression models were tested, but failed to capture the full breadth of multi-parameter effects. The Gaussian Process Regression (GPR) model was designed to capture these multi-parameter interactions and non-linear relationships that standard regression approaches fail to address, as indicated by ANOVA/sensitivity stud-



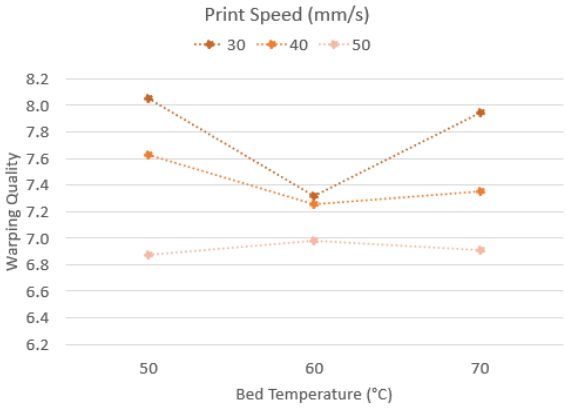


Figure 13: Bed temperature and print speed interactions on part quality ($n = 81$ test cases). Left: top-down quality score; Right: warping quality score. Each line corresponds to a print speed, with points showing average score at each bed temperature. All quality scores are reported on a 1–10 scale, where higher values indicate better quality.

ies and interaction plots. Unlike linear models that isolate individual parameter effects, the GPR framework employs a composite radial basis function (RBF) kernel structure to explicitly model parameter interactions and provide probabilistic predictions with quantified uncertainty (±2σ). The ability of GPR to estimate prediction intervals (±2σ) can be leveraged for risk-aware parameter selection, although in this study, these intervals were not directly used for process control.

Physics-informed features were incorporated into the kernel based on known FFF relationships, including cooling rate (bed temperature × print speed) and mass flow rate (print speed × extrusion multiplier), enabling the GPR to account for complex physical interactions that linear regression could not capture. For instance, including the product of bed temperature and print speed as a feature (representing cooling rate) allows the model to capture latent heat transfer effects impacting part quality, beyond what individual parameters alone would convey.

The RBF kernel measures the similarity between two input points, assigning higher values when the points are close and lower values as they move apart; this enables the Gaussian Process to learn smooth, non-linear relationships in the data by adjusting the kernel's amplitude and length scale coefficients to best fit the training data. The coefficients are learned automatically during model training by maximizing the likelihood of the observed data; thus, GPR does not yield explicit closed-form equations for predictions. Rather, it produces predictions based on the learned kernel function and its optimized parameters, which define the probability distributions over possible functions that fit the data.

The warping GPR model (Figure 14) achieved $R^2 = 0.272$ and RMSE $= 0.581$, capturing the dominant print-speed effect (RBF_1) as well as interactions between bed temperature and print speed ($\text{RBF}_2 \times \text{RBF}_3$). The top-down quality GPR model (Figure 15) yielded $R^2 = 0.134$ and RMSE $= 0.429$, incorporating interactions between bed temperature (RBF_1), as well as extrusion multiplier and nozzle temperature ($\text{RBF}_2 \times \text{RBF}_3$). The learned functions for these models can be seen in Equation 3 (warping quality prediction kernel) and Equation 4 (top down quality prediction kernel). While both GPR models show modest R^2 values, they demonstrate a slight improvement in

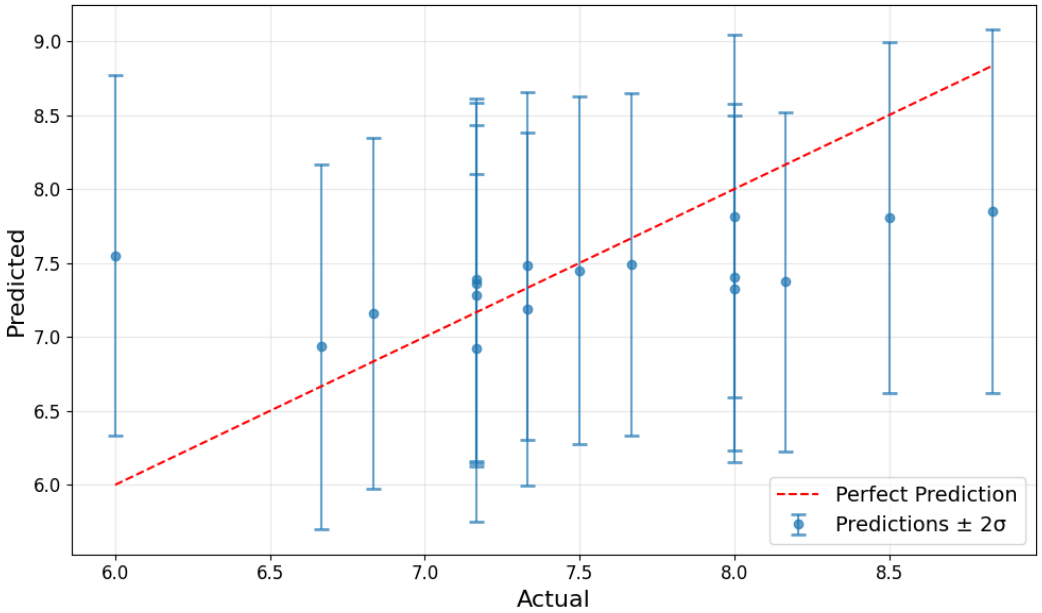


Figure 14: Warping score GPR model. The plot shows average model predictions versus actual values. The two-sigma ($\pm 2\sigma$) confidence interval is shown as range bars for each surveyed data point.

modeling the multi-parameter relationships governing print quality compared to single-parameter regression approaches.

$$\begin{aligned} k(x, x') = \; & 37.45\,\mathrm{RBF}_1(x, x'; 1000) \\ & + 1.66\,[\mathrm{RBF}_2(x, x'; 96.9) \times \mathrm{RBF}_3(x, x'; 96.9)] \\ & + \mathrm{WhiteKernel}(0.314) \end{aligned} \tag{3}$$

$$\begin{aligned} k(x, x') = \; & 36.48\,\mathrm{RBF}_1(x, x'; 1000) \\ & + 0.000042\big[\mathrm{RBF}_2(x, x'; 0.1) \times \mathrm{RBF}_3(x, x'; 1)\big] \\ & + \mathrm{WhiteKernel}(0.0384) \end{aligned} \tag{4}$$

To compare predictive capability between statistical linear regression (SLR) and GPR models, an empirical validation study was conducted using identical test coupons from the original DoE. Four test scenarios were evaluated for each quality metric: (1) interpolated parameters (within DoE range), (2) extrapolated parameters (beyond DoE bounds), (3) GPR-generated best parameter combinations, and (4) GPR-generated worst parameter combinations. The results for warping and top-down quality predictions are summarized in Table 4 and Table 5, respectively.

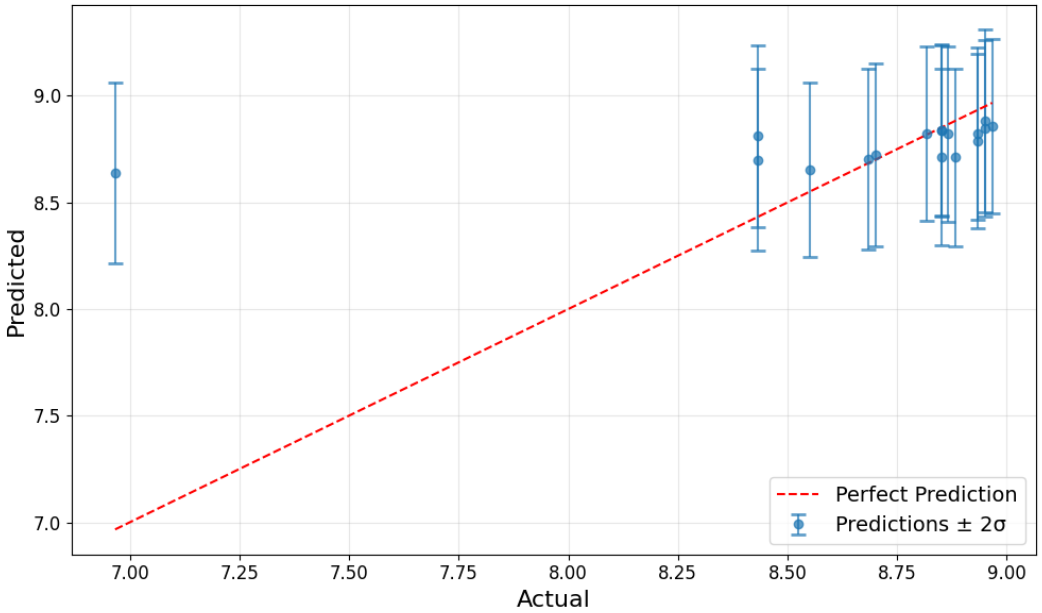


Figure 15: Top-down score GPR model. Model predictions plotted against actual values. The two-sigma ($\pm 2\sigma$) confidence interval is shown as range bars for each surveyed data point. The point at the top left was flagged as an outlier after data review.

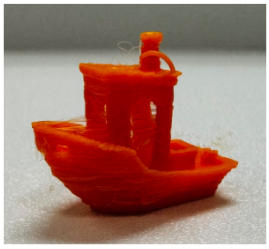
Table 4: GPR and SLR performance comparison for warping quality prediction ($n = 8$ validation runs). GPR error bounds are not shown in this table.

Case Description	**GPR Pred.**	**SLR Pred.**	**Actual Score**	**% Error: GPR**	**% Error: SLR**
Interpolated Parameters	7.6	7.3	8.0	5.96%	9.14%
Extrapolated Parameters	5.7	7.5	7.7	34.98%	2.77%
Best Parameter Combination	8.1	7.3	7.8	3.29%	7.31%
Worst Parameter Combination	6.5	7.3	6.8	5.13%	6.39%

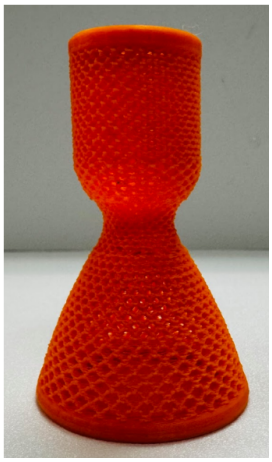
Table 5: GPR and SLR performance comparison for top-down quality prediction ($n = 8$ validation runs). GPR error bounds are not shown in this table.

Case Description	**GPR Pred.**	**SLR Pred.**	**Actual Score**	**% Error: GPR**	**% Error: SLR**
Interpolated Parameters	8.8	9.1	8.9	0.76%	2.13%
Extrapolated Parameters	6.0	7.9	9.0	48.18%	13.43%
Best Parameter Combination	8.8	8.4	9.0	2.08%	6.94%
Worst Parameter Combination	8.4	9.0	8.5	0.79%	5.93%

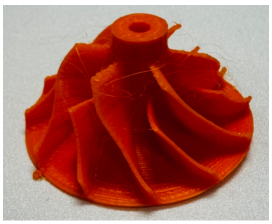
The empirical validation reveals distinct performance characteristics between the two model-



(a) Benchy model [34].



(b) Latticed rocket engine (custom).



(c) Compressor wheel [35].

Figure 16: Non-standard validation geometries: (a) 3DBenchy [34], (b) latticed rocket engine (custom nTop CAD), and (c) compressor wheel [35].

ing approaches. For interpolated cases within the training data range, GPR models demonstrate superior accuracy with errors of 0.76% (top-down) and 5.96% (warping), compared to SLR errors of 2.13% and 9.14%, respectively. However, extrapolation beyond the training data reveals a limitation of the GPR models, particularly for top-down quality prediction where GPR error reaches 48.18% compared to only 13.43% for statistical regression. This pattern suggests that the setup of this GPR model is slightly better at capturing complex parameter interactions within the experimental space, but their uncertainty quantification and RBF kernel structure may lead to poor extrapolation performance.

In contrast, the statistical regression models exhibit more consistent prediction accuracy across both interpolated and extrapolated scenarios. This suggests that GPR models are optimal for parameter optimization within the established experimental bounds, while statistical regression provides more reliable predictions for unexplored parameter combinations. While the GPR also proved more accurate for the best and worst parameter combinations, the generated parameters did not yield scores that exceeded the true maximum and minimum values observed in the DoE, suggesting that further optimizations need to be done to this aspect of the system.

As a final stress test, both models were evaluated on three non-standard geometries using identical input parameters typical for C3DP platform operation: a Benchy model [34] (Figure 16a), a latticed rocket engine designed in nTop by the author (Figure 16b), and a compressor wheel [35] (Figure 16c). The results of the empirical validation for each part are presented in Table 6 below.

Table 6: GPR and SLR performance comparison for validation geometry quality prediction.

Geometry	Quality Parameter	GPR Pred.	SLR Pred.	Actual Score	% Error: GPR	% Error: SLR
Benchy	Top Down	8.9	8.4	8.6	3.18%	2.58%
	Warping	7.9	7.3	8.7	9.70%	18.74%
Rocket Engine	Top Down	8.9	8.4	8.2	8.12%	2.44%
	Warping	7.9	7.3	7.2	10.83%	1.83%
Compressor	Top Down	8.9	8.4	10.0	12.17%	18.37%
	Warping	7.9	7.3	9.8	21.94%	26.78%

The stress tests reveal certain limitations in both modeling approaches when applied to geometries substantially different from the original test coupon. Statistical regression demonstrates more consistent performance across geometries, with top-down quality errors ranging from 2.58% to 6.94%, while GPR errors range from 3.18% to 12.17%. For warping quality, performance varies depending on geometric complexity. Regression achieves exceptional accuracy for the rocket engine (1.83% error) but struggles with the compressor wheel (15.53% error), while GPR shows the opposite trend with better Benchy prediction (9.70% error) but poor compressor performance (21.94% error). Notably, both models tend to predict similar quality scores regardless of geometry (GPR: 8.9 top-down, 7.9 warping; Regression: 8.4 top-down, 7.3 warping), suggesting limited adaptability to different part complexities. While GPR offered a conceptually robust and flexible modeling framework, the relatively weak coefficient of determination in both quality metrics indicates that either finer parameter resolution or additional features may be necessary for predictive precision across diverse print scenarios. This pattern indicates that while both approaches capture parameter-quality relationships for the specific test coupon geometry, their generalization remains constrained, highlighting the need for geometry-specific training data or more robust feature engineering to achieve truly predictive manufacturing quality models.

In terms of study limitations, several challenges span hardware, data processing pipelines, statistical methodology, and modeling fidelity, each of which impacts the current and future performance of the C3DP platform. From a hardware perspective, confounding variables included the use of blue masking tape. While effective at reducing camera glare, altered the thermal profile of the build surface, potentially influencing both adhesion and warping behavior. Temperature instability or drift in both the heated bed and nozzle represents another source of discrepancy between nominal and actual process conditions; these effects warrant closer monitoring through integrated sensing in future work. Parameter ranges, particularly for material flow rate, should have been further analyzed and validated prior to final selection so that under-extrusion defects, a common hardware issue with many FFF platforms, could be captured and characterized in addition to over-extrusion defects. Additionally, the experimental design did not account for potential variability between individual SCARA robots operating on the C3DP platform, raising the possibility of machine-specific biases in scoring. This highlights the necessity for device calibration procedures and cross-validation in large-scale or multi-printer studies.

The design of the current image processing and data capture workflow also imposes constraints on true in-situ process monitoring, as printing must be paused to capture certain images. These

pauses not only disrupt continuous manufacturing workflows but may also affect the thermal history of subsequent layers, thus complicating the causal interpretation of observed print quality trends. The calibration process used for quality scoring further introduces subjectivity, especially for the warping script's default 10/10 first layer score, which may inadvertently skew data for geometries susceptible to early-layer defects. Also, the pixel-dependent warping scoring system introduces the potential for subjective bias based on human assessment, and can drastically change based on camera type and placement. Generally speaking, the calibration and weighting methods for the scoring scripts are based on a qualitative human assessment, which may impact the fidelity and validity of the scoring model for different geometries, and even score extrapolation within the same, standardized test geometries. Further investigation into objective, improved calibration methods is required.

From a data analysis perspective, the use of a full-factorial three-level DoE provided efficient parameter coverage, but resulted in relatively weak correlations ($R^2 < 0.3$) between most process variables and response metrics. This suggests higher-resolution DoEs or alternative experimental strategies might be needed to reveal stronger, statistically significant relationships. The lack of a non-heated-bed control group also limits the ability to draw firm conclusions about the bed heating system's efficacy, as well as the previously mentioned lack of parameter range for under-extrusion cases.

With respect to predictive modeling, GPR models in this study achieved only modest improvements in capturing parameter interactions (as reflected in low R^2 values), and they notably underperformed in extrapolation scenarios (e.g., producing a 48.18% error for top-down quality). The GPR model's performance in predicting optimal and worst-case parameter combinations can also be significantly improved through several approaches. Advanced kernel optimization involves systematically tuning the model's internal parameters (such as the RBF length scales and amplitude values) which effectively control how the model learns relationships between different printing parameters. This, in essence, teaches the model to better recognize patterns in the data by optimizing how it weighs similarities between different parameter sets. Active learning strategies can also employ a more intelligent experimental approach where, instead of testing parameter combinations randomly to find maxima and minima, the model iteratively identifies the most informative experiments to run next based on its current uncertainty; for example, if the model predicts that certain temperature-speed combinations might yield optimal results but is uncertain about this prediction, active learning would prioritize testing those specific combinations to refine the model's accuracy in that region [36].

Nonetheless, GPR models offer unique value within the studied design space by providing uncertainty quantification and by capturing non-linear and multi-parameter effects that linear models inherently miss. The practical implication is that for guiding process selection within validated parameter bounds, particularly in applications where estimation uncertainty is critical, GPR can be advantageous. Conversely, for extrapolation, rapid prototyping, or situations favoring computational efficiency, standard linear regression may still be preferable. These trade-offs underscore the need for model selection to be matched explicitly to use case and data regime, and they should motivate the exploration of deep learning or other advanced methods in future work. In addition to technical limitations, limited generalizability remains a critical issue. The empirical validation on non-standard part geometries demonstrated that neither approach generalized robustly beyond the

test coupon geometry.

5. Conclusion

Future efforts in data-driven part quality analysis for C3DP should focus on developing a more robust image processing pipeline to standardize quality assessment, reducing reliance on subjective calibration. In addition, utilizing alternate prediction models, including deep learning architectures, could enhance accuracy and predictive capability beyond the current statistical and GPR frameworks. Finally, training models on raw image datasets from DoEs in conjunction with quality scores would enable richer feature extraction and potentially improve the system's ability to detect subtle defects and adapt to new geometries. Paired with an improved physics-informed feature set, part quality analysis in this platform could be improved significantly.

Despite the many limitations of this study, a number of unique opportunities emerged. The interactions of the "active" and "passive" robots were unique in the context of C3DP, and present the possibility of utilizing more hybrid systems in tandem. The focus on utilizing data to understand and improve performance also lays a framework for a more integrated, connected platform, perhaps contained within a digital twin model or Robot Operating System (ROS) architecture. By addressing such fundamental research questions on this testbed platform, meaningful advancements can be made toward constructing fully autonomous and adaptive industrial manufacturing systems.

The parameter-quality relationships and computer vision techniques developed in this study offer a direct pathway to improving print reliability, streamlining error detection, and minimizing manual intervention within multi-robot manufacturing environments. These results also provide a broad foundation for evaluating new predictive models or hardware systems, enabling both academic and industrial efforts to benchmark progress in automated additive manufacturing. By deploying and refining the presented methodology, organizations can more rapidly adopt autonomous quality assurance systems and realize the benefits of scalable, intelligent, and robust production workflows.

Acknowledgments

This work was partially supported by the Army Research Lab (ARL) via grant number W911NF-24-2-0007. Any opinions, findings, conclusions, or recommendations expressed in this material are those of the authors and do not necessarily reflect the views of ARL.

References

[1] L. Poudel, W. Zhou, and Z. Sha, "Toward swarm manufacturing: Architecting a cooperative 3D printing system," *Journal of Manufacturing Science and Engineering*, vol. 144, no. 8, pp. 1–14, 2022.

[2] D. Weber, W. Zhou, and Z. Sha, "Z-chunking for cooperative 3D printing of large and tall objects," in *Proceedings of the 33rd Annual International Solid Freeform Fabrication Symposium*, 2022. [Online]. Available: http://dx.doi.org/10.26153/tsw/44190

[3] ——, "Job placement for cooperative 3D printing," in *Manufacturing Science and Engineering Conference (MSEC)*, 2023. [Online]. Available: https://doi.org/10.1115/MSEC2023-104613

[4] L. Poudel, W. Zhou, and Z. Sha, "Resource constrained scheduling for multi-robot cooperative three dimensional printing," *Journal of Mechanical Design*, vol. 143, no. 7, 2021.

[5] S. Elagandula, L. Poudel, Z. Sha, and W. Zhou, "Multi-robot path planning for cooperative 3D printing," in *International Manufacturing Science and Engineering Conference (MSEC)*, 2020.

[6] R. F. P. Stone, W. Zhou, and Z. Sha, "Time-optimal path planning for heterogeneous robots in swarm manufacturing," *ArXiv preprint*, 2023. [Online]. Available: https://doi.org/10.26153/TSW/51043

[7] H. Sanam *et al.*, "Computer vision-based in-situ monitoring of cooperative 3D printing in a closed-loop system," in *ASME 2025 International Design Engineering Technical Conference*, 2025, iDETC2025-168835.

[8] C. Mensch *et al.*, "A real-time monitoring framework for cooperative 3D printing," in *Solid Freeform Fabrication 2024: Proceedings of the 35th Annual International Solid Freeform Fabrication Symposium*, 2024.

[9] S. Usha, "In situ monitoring of metal additive manufacturing process: A review," *Additive Manufacturing*, pp. 275–299, 2021.

[10] Y. Fu, A. Downey, L. Yuan, A. Pratt, and Y. Balogun, "In situ monitoring for fused filament fabrication process: A review," *Additive Manufacturing*, vol. 38, p. 101749, 2021.

[11] Z. Jin, Z. Zhang, and G. X. Gu, "Autonomous in-situ correction of fused deposition modeling printers using computer vision and deep learning," *Manufacturing Letters*, vol. 22, pp. 11–15, 2019.

[12] J. Straub, "Initial work on the characterization of additive manufacturing ({3D} printing) using software image analysis," *Machines*, vol. 3, no. 2, pp. 55–71, 2015.

[13] A. L. Petsiuk and J. M. Pearce, "Open source computer vision-based layer-wise 3D printing analysis," *Additive Manufacturing*, vol. 36, p. 101473, 2020.

[14] D. A. J. Brion, M. Shen, and S. W. Pattinson, "Automated recognition and correction of warp deformation in extrusion additive manufacturing," *Additive Manufacturing*, vol. 56, p. 102838, 2022.

[15] K. Paraskevoudis, P. Karayannis, and E. P. Koumoulos, "Real-time 3D printing remote defect detection (stringing) with computer vision and artificial intelligence," *Processes*, vol. 8, no. 11, p. 1464, 2020.

[16] Q. Wang *et al.*, "Research on process control of 3D printing technology based on multi-sensor technology," in *2020 IEEE 3rd International Conference of Safe Production and Informatization (IICSPI)*, 2020, pp. 675–679.

[17] C. Liu *et al.*, "Image analysis-based closed loop quality control for additive manufacturing with fused filament fabrication," *Journal of Manufacturing Systems*, vol. 51, pp. 75–86, 2019.

[18] F. Mumali and J. Kałkowska, "Intelligent support in manufacturing process selection based on artificial neural networks, fuzzy logic, and genetic algorithms: Current state and future perspectives," *Computers & Industrial Engineering*, vol. 193, p. 110272, 2024.

[19] G. D. Goh *et al.*, "A review on machine learning in 3D printing: Applications, potential, and challenges," *Artificial Intelligence Review*, vol. 54, no. 1, pp. 63–94, 2021.

[20] A. K. Sood *et al.*, "Experimental investigation and empirical modelling of FDM process for compressive strength improvement," *Journal of Advanced Research*, vol. 3, no. 1, pp. 81–90, 2012.

[21] O. Bayraktar *et al.*, "Experimental study on the 3D-printed plastic parts and predicting the mechanical properties using artificial neural networks," *Polymers for Advanced Technologies*, vol. 28, no. 8, pp. 1044–1051, 2017.

[22] S. Kwon and D. Hwang, "Understanding and resolving 3D printing challenges: A systematic literature review," *Processes*, vol. 13, no. 6, p. 1772, 2025.

[23] L. Li *et al.*, "Precision enhancement of 3D printing via in situ metrology," University of Texas at Austin, Tech. Rep., 2018.

[24] R. J. M. Wolfs *et al.*, "A real-time height measurement and feedback system for 3D concrete printing," in *High Tech Concrete: Where Technology and Engineering Meet*, D. A. Hordijk and M. Luković, Eds. Springer International Publishing, 2018, pp. 2474–2483.

[25] O. Holzmond and X. Li, "In situ real time defect detection of 3D printed parts," *Additive Manufacturing*, vol. 17, pp. 135–142, 2017.

[26] R. Nascimento *et al.*, "Computer vision based quality control for additive manufacturing parts," *The International Journal of Advanced Manufacturing Technology*, vol. 124, no. 10, pp. 3241–3256, 2023.

[27] M. F. Khan *et al.*, "Real-time defect detection in 3D printing using machine learning," *Materials Today: Proceedings*, vol. 42, pp. 521–528, 2021.

[28] C. Arnold *et al.*, "Surface quality of 3D-printed models as a function of various printing parameters," *Materials*, vol. 12, no. 12, p. 1970, 2019.

[29] T. S. Tamir *et al.*, "Machine-learning-based monitoring and optimization of processing parameters in 3D printing," *International Journal of Computer Integrated Manufacturing*, vol. 36, no. 9, pp. 1362–1378, 2023.

[30] J. Zhu *et al.*, "Surface quality prediction and quantitative evaluation of process parameter effects for 3D printing with transfer learning-enhanced gradient-boosting decision trees," *Expert Systems with Applications*, vol. 237, p. 121478, 2024.

[31] A. B. Tatar, "Predicting three-dimensional (3D) printing product quality with machine learning-based regression methods," *Firat University Journal of Experimental and Computational Engineering*, vol. 4, no. 1, pp. 206–225, 2025.

[32] U. Delli and S. Chang, "Automated process monitoring in 3D printing using supervised machine learning," *Procedia Manufacturing*, vol. 26, pp. 865–870, 2018.

[33] G. A. R. Sampedro *et al.*, "Design of an in-process quality monitoring strategy for FDM-type 3D printer using deep learning," *Applied Sciences*, vol. 12, no. 17, p. 8753, 2022.

[34] CreativeTools.se, "3dbenchy - the jolly 3d printing torture-test," https://www.thingiverse.com/thing:763622, 2015.

[35] Tkorniev, "Compressor wheel," https://www.thingiverse.com/thing:4683538, 2021.

[36] H. Li, J. Zhang, X. Yang, M. Ye, W. Jiang, J. Gong, Y. Tian, L. Zhao, W. Wang, and Z. Xu, "Bayesian optimization based extreme gradient boosting and GPR time-frequency features for the recognition of moisture damage in asphalt pavement," *Construction and Building Materials*, vol. 434, p. 136675, 2024.

Appendices

A Methodology

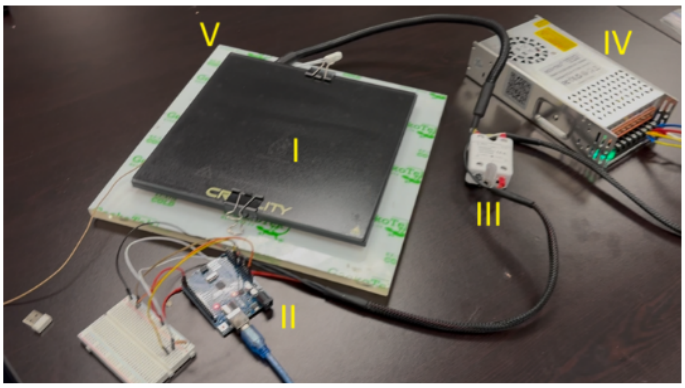


Figure A1: Labeled components of original heated bed. (I) Heated bed with glass surface, (II) Arduino Uno and primary circuit, (III) solid state relay, (IV) power supply unit (PSU), and (V) modified GeckoTek build platform.

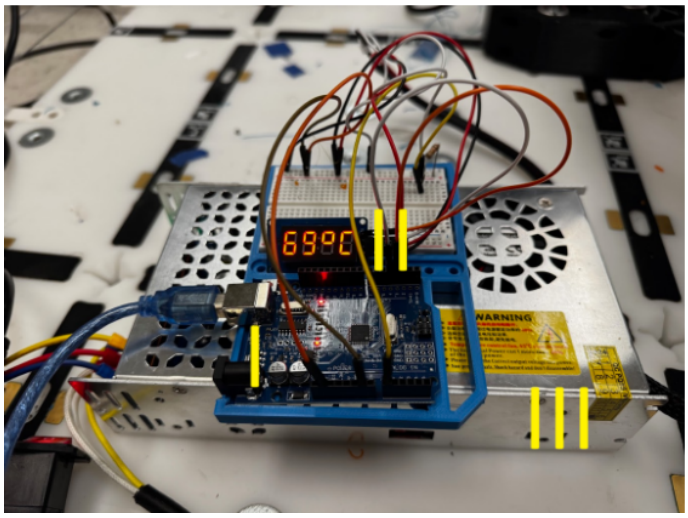


Figure A2: Upgraded hotbed primary controller with Arduino Uno (I), temperature display readout (II), and PSU (III).

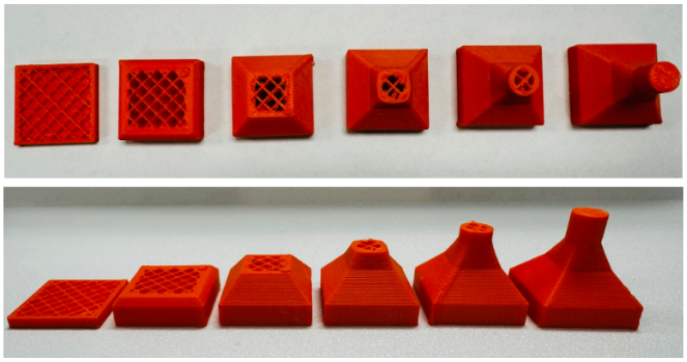


Figure A3: Top: top view of ideal test coupon. Bottom: side view of ideal test coupon.

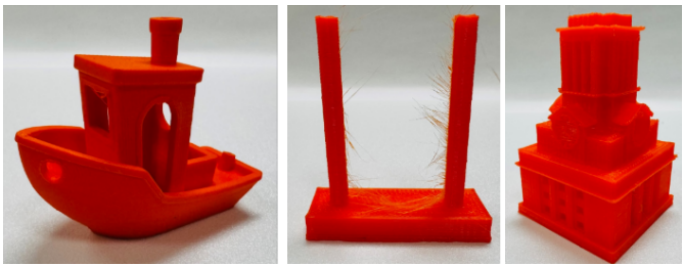


Figure A4: Left: Benchy calibration model. Middle: stringing tower calibration model. Right: UT Tower calibration model.

B DoE Images

B.1 Batch 1

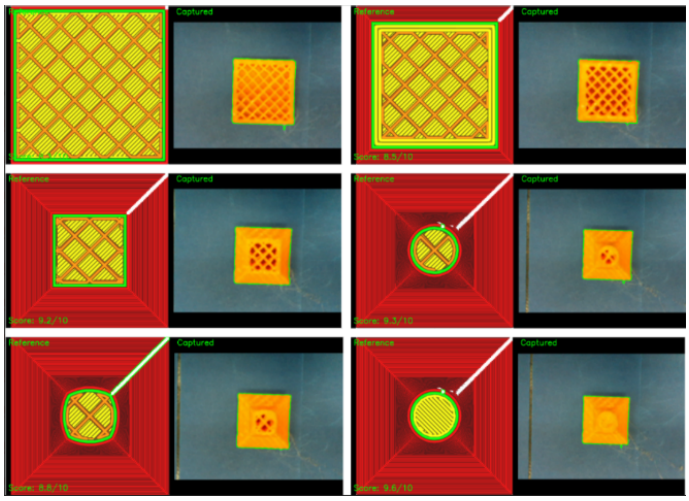


Figure B1.1: Batch 1 average top down quality coupon image data (coupon 1)

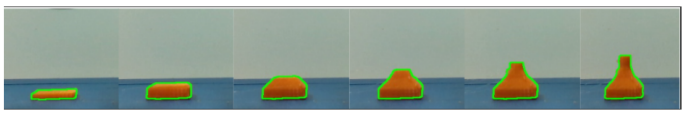


Figure B1.2: Batch 1 average warping quality coupon image data (coupon 4)

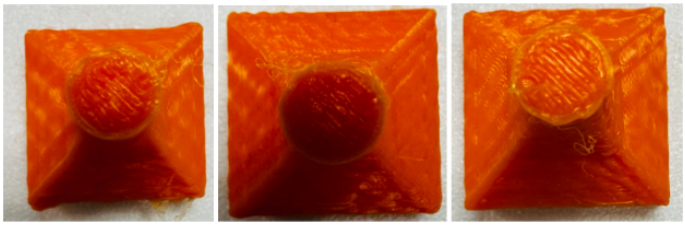


Figure B1.3: Batch 1 minimum (left; coupon 46), average representative (middle; coupon 1), and maximum (right; coupon 40) scores for top down quality

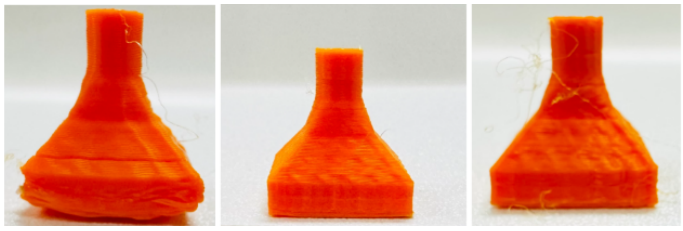


Figure B1.4: Batch 1 minimum (left; coupon 46), average representative (middle ; coupon 4), and maximum (right; coupon 19) scores for warping

B.2 Batch 2

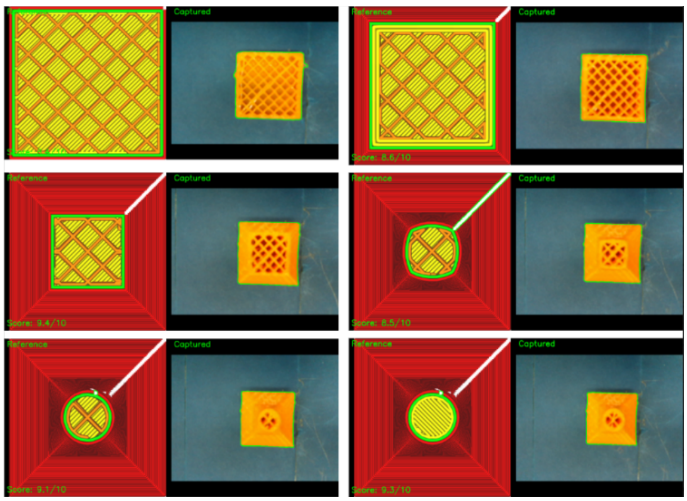


Figure B2.1: Batch 2 average top down quality coupon image data (coupon 41)

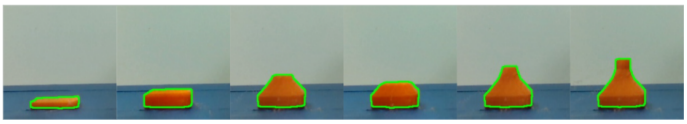


Figure B2.2: Batch 2 average warping quality coupon image data (coupon 41)

Figure B2.3: Batch 2 minimum (left; coupon 50), average representative (middle; coupon 41), and maximum (right; coupon 5) scores for top down quality

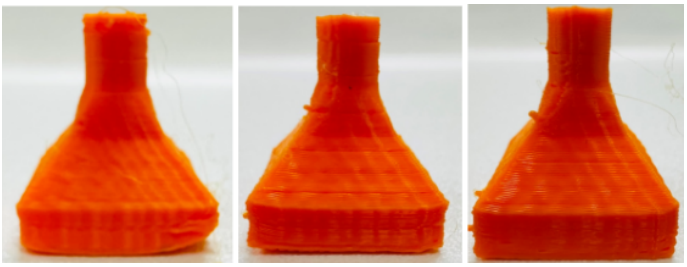


Figure B2.4: Batch 2 minimum (left; coupon 26), average representative (middle; coupon 41), and maximum (right; coupon 14) scores for warping

B.3 Batch 3

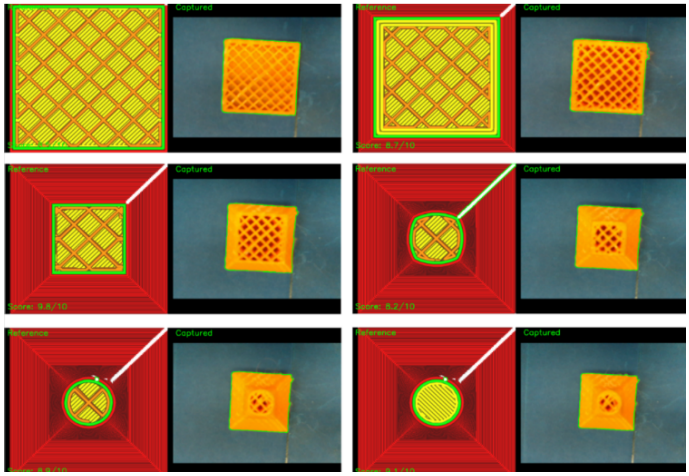


Figure B3.1: Batch 3 average top down quality coupon image data (coupon 12)

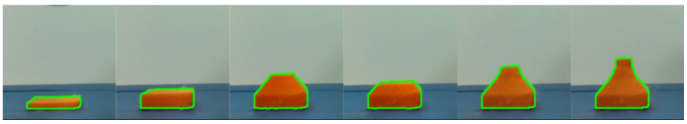


Figure B3.2: Batch 3 average warping quality coupon image data (coupon 12)

Figure B3.3: Batch 3 minimum (left; coupon 51), average representative (middle; coupon 12), and maximum (right; coupon 6) scores for top down quality

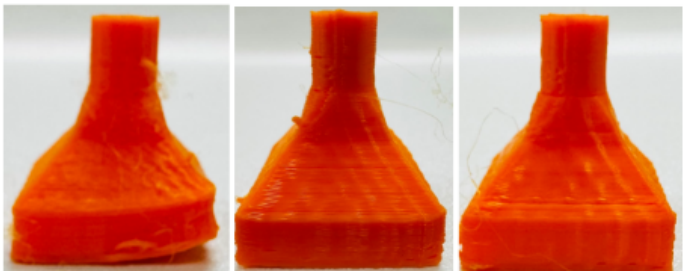


Figure B3.4: Batch 3 minimum (left; coupon 6), average representative (middle; coupon 12), and maximum (right; coupon 51) scores for warping

C Expanded Figures

C.1 Parameter Scatterplots

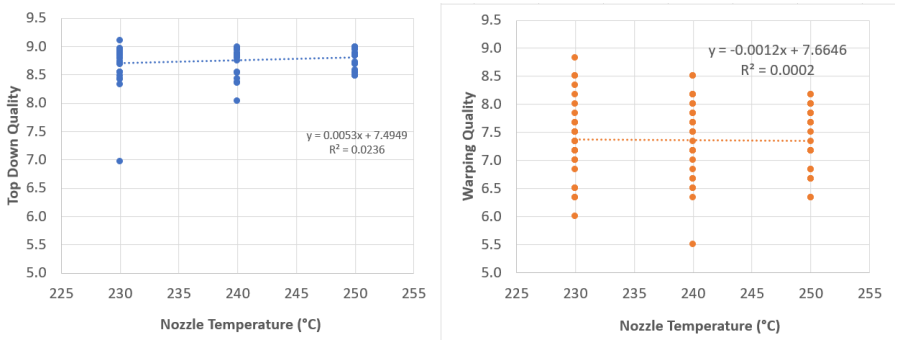


Figure C1.1: Effect of nozzle temperature on part quality scores ($n = 81$ test cases). The left panel displays the top-down quality scatterplot with the trendline, while the right panel displays the warping quality scatterplot with the trendline. All scores are on a 1–10 scale, with higher values indicating better quality.

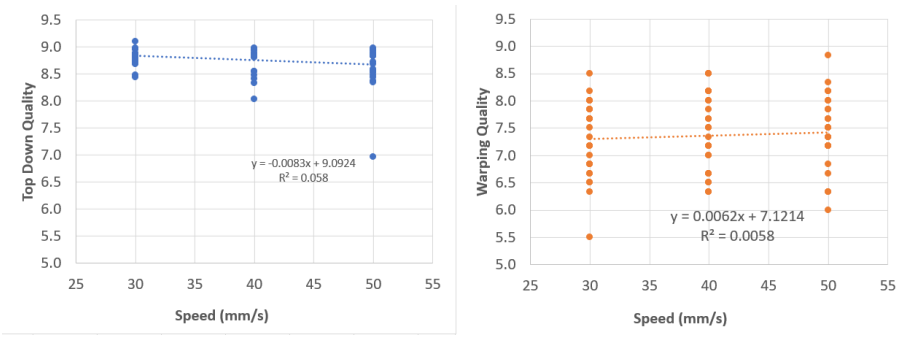


Figure C1.2: Effect of print speed on part quality scores ($n = 81$ test cases). The left panel displays the top-down quality scatterplot with the trendline, while the right panel displays the warping quality scatterplot with the trendline. All scores are on a 1–10 scale, with higher values indicating better quality.

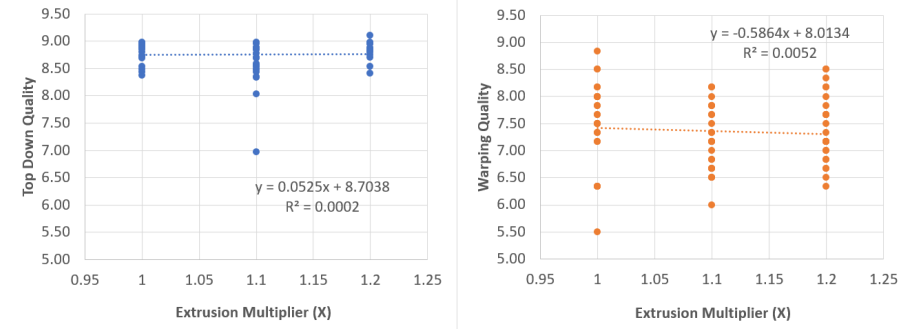


Figure C1.3: Effect of extrusion multiplier on part quality scores ($n = 81$ test cases). The left panel displays the top-down quality scatterplot with the trendline, while the right panel displays the warping quality scatterplot with the trendline. All scores are on a 1–10 scale, with higher values indicating better quality.

C.2 Parameter Interaction Plots

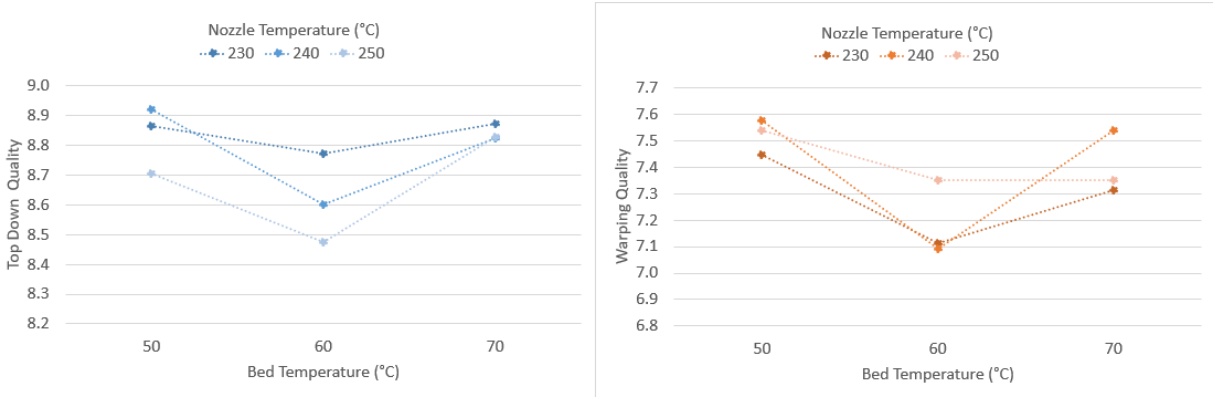


Figure C2.1: Bed temperature and nozzle temperature interactions on part quality ($n = 81$ test cases). Left: top-down quality score; Right: warping quality score. All quality scores are reported on a 1–10 scale, where higher values indicate better quality.

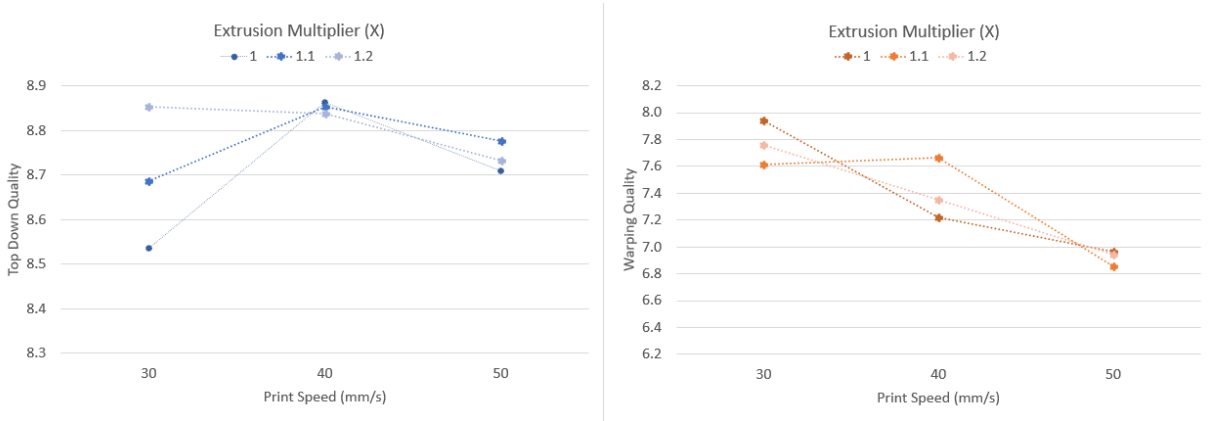


Figure C2.2: Print speed and extrusion multiplier interactions on part quality ($n = 81$ test cases). Left: top-down quality score; Right: warping quality score. All quality scores are reported on a 1–10 scale, where higher values indicate better quality.

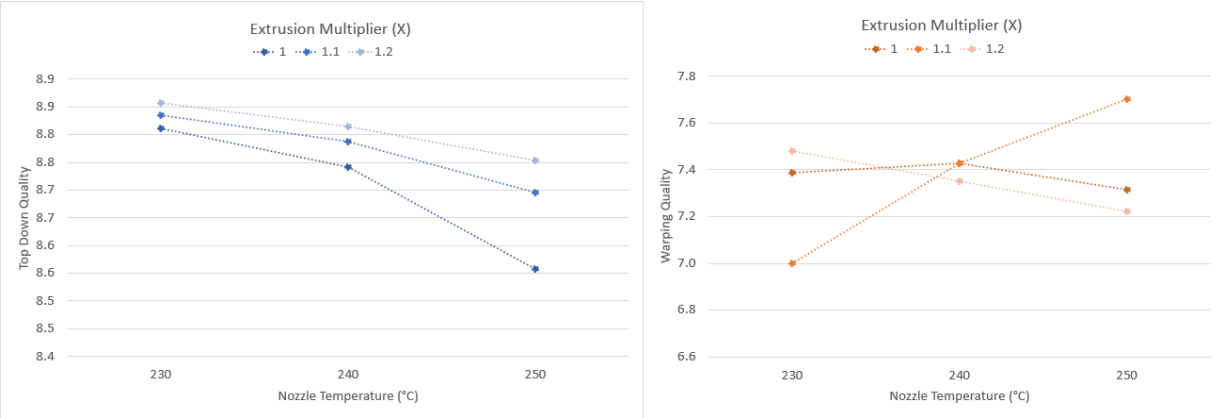


Figure C2.3: Nozzle temperature and extrusion multiplier interactions on part quality ($n = 81$ test cases). Left: top-down quality score; Right: warping quality score. All quality scores are reported on a 1–10 scale, where higher values indicate better quality.