IISE Transactions

ISSN: (Print) (Online) Journal homepage: https://www.tandfonline.com/loi/uiie21

Co-learning of extrusion deposition quality for supporting interconnected additive manufacturing systems

An-Tsun Wei, Hui Wang, Tarik Dickens & Hongmei Chi

To cite this article: An-Tsun Wei, Hui Wang, Tarik Dickens & Hongmei Chi (2022): Co-learning of extrusion deposition quality for supporting interconnected additive manufacturing systems, IISE Transactions, DOI: 10.1080/24725854.2022.2080306

To link to this article: https://doi.org/10.1080/24725854.2022.2080306

IISE TRANSACTIONS
https://doi.org/10.1080/24725854.2022.2080306

Co-learning of extrusion deposition quality for supporting interconnected additive manufacturing systems

An-Tsun Wei[a], Hui Wang[a], Tarik Dickens[a], and Hongmei Chi[b]

[a]Department of Industrial and Manufacturing Engineering, FAMU-FSU College of Engineering, Tallahassee, FL, USA; [b]Department of Computer and Information Sciences, Florida A&M University, Tallahassee, FL, USA

ABSTRACT

Additive manufacturing systems are being deployed on a cloud platform to provide networked manufacturing services. This article explores the value of interconnected printing systems that share process data on the cloud in improving quality control. We employed an example of quality learning for cloud printers by understanding how printing conditions impact printing errors. Traditionally, extensive experiments are necessary to collect data and estimate the relationship between printing conditions vs. quality. This research establishes a multi-printer co-learning methodology to obtain the relationship between the printing conditions and quality using limited data from each printer. Based on multiple interconnected extrusion-based printing systems, the methodology is demonstrated by learning the printing line variations and resultant infill defects induced by extruder kinematics. The method leverages the common covariance structures among printers for the co-learning of kinematics-quality models. This article further proposes a sampling-refined hybrid metaheuristic to reduce the search space for solutions. The results showed significant improvements in quality prediction by leveraging data from data-limited printers, an advantage over traditional transfer learning that transfers knowledge from a data-rich source to a data-limited target. The research establishes algorithms to support quality control for reconfigurable additive manufacturing systems on the cloud.

ARTICLE HISTORY
Received 13 April 2021
Accepted 6 May 2022

KEYWORDS
Cloud additive manufacturing; multi-printer co-learning; quality control; extrusion 3D printing

1. Background

With the advancement of Internet-of-Things (IoT) technology, networked printing services are emerging to flexibly meet customers' diverse demands. The customers can upload their printing models, which are processed by different online printers. One characteristic of such networked printers compared with the conventional printing process is that quality/process data can be stored and shared on a cloud platform (Baumann and Roller, 2017), creating an interconnected environment for multiple printers. State-of-the-art research has demonstrated the applications of such interconnected Additive Manufacturing (AM) in offering networked manufacturing services for personalized demands or multi-printer collaboration to co-create the same structure with reduced time. Relatively limited research has explored the values of interconnected AM and the shared process data in improving quality control. This article addresses the limitation by establishing a co-learning methodology for multiple cloud printers to jointly learn printing variations and defects based on limited data. The methodology will be demonstrated by learning infill defects induced by extruder kinematics in extrusion-based AM as a case study, such as fused deposition modeling or fused filament fabrication.

Extrusion-based AM usually has the limitations of low printing precision as reflected in the shape deviation from the nominal CAD model. The printing errors can be attributed to material shrinkage and printing process setups, such as extruder movement speed, acceleration, extrusion rate, and temperature. These errors include not only the inaccuracy of printed contours, but also infill defects at each layer, such as overfill and underfill problems (Figure 1). It is essential to quantitatively understand the impact of these setup parameters on printing quality in order to calibrate printing processes, thereby improving printing quality and reducing potential post-processing efforts. A major challenge to learning process–quality relationships is acquiring process data under different printing conditions, which can be time-consuming. A considerable number of tests should be conducted given various combinations of printing speeds, accelerations, and/or temperatures to learn a data-driven quality prediction model. The cost of experiments and time to understand the process–quality relationship can significantly delay the production and new printer setup to deal with scalable printing tasks. New printers are frequently engaged in a reconfigurable printing system to deal with larger printing tasks, and limited time and budget are available to understand the printers' variations.

CONTACT Hui Wang hwang10@fsu.edu
Supplemental data for this article is available online at https://doi.org/10.1080/24725854.2022.2080306

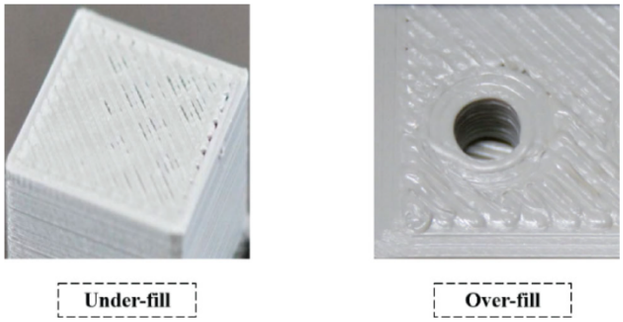


Figure 1. The kinematic-induced defects of infill pattern learned by cloud printers. The left picture shows the gap between adjacent infill lines (Underfill defect), and the right one shows excessive material deposition (Overfill defect).

This research envisions a solution to the data limitation challenge by using multiple interconnected printers to co-learn the process–quality relationship given limited data from each printer. Based on the learning performance in the case study, the co-learning strategy can effectively utilize the data from multiple printers shared on the cloud and reduce the experimentation necessary on each printer to obtain sufficient data to learn process–quality relationship and printing calibration.

2. State-of-the-art and research gaps

This section reviews recent research on (i) the quantitative relationship between process setup and quality and (ii) the learning strategy to obtain such a relationship based on limited data. Quality prediction and modeling for interconnected printing processes are also reviewed. Finally, methodological gaps are summarized for this research to address.

2.1. Printing process–quality relationship

Compared with other printing technologies, relatively limited research has been conducted to predict the dimensional inaccuracy in extrusion-based AM. Wang *et al.* (2017) modeled the extruder's position error and material deformation for the compensation of the CAD design. Noriega *et al.* (2013) adopted artificial neural networks to optimize the distance between parallel faces in the prismatic part to improve dimensional accuracy by adjusting the input values in the CAD model. More importantly, most literature has discussed the quality effect of the process parameter on printing quality. Cheng *et al.* (2018) proposed a two-step scheme to integrate the Gaussian process and a kernel smoothing method to capture the nonlinearity between shape accuracy and the different combinations of the process parameters. Lyu and Manoochehri (2018) proposed the integrated error model to simulate the effects of the extruder's temperature, infill density, and layer thickness. In extrusion-based AM, the printing quality of the part can be affected by many uncertainties. Equbal *et al.* (2017) analyzed the quality relationship on three-factor inputs, i.e., raster angle, raster width, and air gap, and minimized the dimensional error using a response surface method with a composite desirability function.

A dearth of research was found that modeled the geometric variation of the printing line. Agarwala *et al.* (1996) addressed the dimensional error in the start and stop position of the single printing line, which might partially result from the material overflow and underflow from the extruder at the two ends. Bouha (1999) developed a look-ahead trajectory algorithm and integrated a position-and-deposition system to reduce start/stop errors. A strategy has been studied by Bellini *et al.* (2004) to improve the printing consistency using the dynamics of the liquefier and established flow control strategies during the printing acceleration and deceleration (ACC/DCC) phases. Qin *et al.* (2019) proposed a real-time speed control strategy to improve the position error resulting from the speed fluctuation (ACC/DCC) on the sharp corners and transition points from the toolpath. Ravi *et al.* (2017) discovered the nozzle-bed distance and the extruder's temperature could impact the printing linewidth. Comminal *et al.* (2018) developed a model to analyze the deposition flow by changing the parameter combinations.

Most research has focused on the effects of printer mechanisms or part design on geometric errors; however, limited research quantitatively models the quality of the inner structure or infill defects inside a part, such as the underfill and overfill defects shown in Figure 1. In Ren *et al.* (2021), the printing line variation was estimated by speed, acceleration, and jerk using the G-code. The training still encounters challenges in obtaining sufficient quality data under different printing conditions.

2.2. Small-sample modeling and learning

One popular methodology to learn quality prediction models based on limited data is sequential learning over multiple stages, including Bayesian approaches (Lin and Wang, 2012), the Markov Decision Process (Alagoz *et al.*, 2010), and model-free reinforcement learning (van Hasselt *et al.*, 2016). These methods can update the distribution of the learning parameters over time and incrementally increase the small sample size. However, this line of methods does not address the challenge of learning at a time when limited samples are available.

Small-sample learning methodologies have emerged in recent decades. For example, Data Augmentation (DA) is one popular method that mutates the available limited data to enlarge training datasets. The size of the input data for learning can be increased by applying crop, flip, scale, rotate, translation, and Gaussian noises (Takahashi *et al.*, 2019) to the limited data. The extra augmented samples from small samples can provide more features and avoid overfitting for the model. The limitation of DA is that it does not supplement new information about the process for accuracy improvement. Semi-supervised Learning (SML) is another line of popular methods that learn the latent structure between labeled and unlabeled data (van Engelen and Hoos, 2020) to supplement information for training. The limitation is that it can become expensive to collect a sufficiently

labeled dataset. The performance of the method deteriorates when the total sample size that is labeled is considerably limited.

In AM applications, modeling the deviation for various geometric shapes is challenging, since the operational cost may not be affordable for users to produce sufficient training data. Sabbaghi *et al.* (2018) utilized a Bayesian discrepancy measure to cluster the distributions of the local features based on the small distinct printed shapes for the newly observed shape, and further built an adaptive Bayesian hierarchical model to estimate the model parameters for local geometric deviations. The method considers the learning contribution from the disparate printed shapes. These small-sample modeling methods primarily focused on shape variation and did not consider the process–quality relationship.

2.3. Quality prediction and learning for interconnected printing process control

Research on quality control for IoT has emerged and has focused on real-time fault alarm monitoring for the cloud printing system. Data-driven algorithms, e.g., artificial neural networks (Wang *et al.*, 2019), have been proposed to distinguish abnormal images taken by the top-mounted camera during the printing operation. When the printing alert is triggered, the printing process is terminated to reduce material waste. Chan *et al.* (2018), Majeed *et al.* (2019) and Majeed *et al.* (2021) estimated the effects of parameters on improving the printing accuracy and reducing the time cost and energy consumption under a big-data-driven framework.

In recent years, Transfer Learning (TL) has gaineds increasing attention in dealing with quality prediction, given limited data from the target process of interest, by leveraging shared data from different printing processes. Unlike DA and SML, TL can take advantage of the relations between the target and source domains and allow the knowledge to be transferred from the target to the source. The TL is considered suitable for quality learning in an interconnected environment, where each printer can benefit from the data shared by other printers. Cheng *et al.* (2017) categorized the deviation errors of the 2D printing shape into the Shape-Independent Error (SIE) and Shape-Specific Error (SSE),; with the SIE modeling parameters being shared with the data-limited printer. However, its performance in learning more complex shapes is very limited. Cheng *et al.* (2021) extended their initial research by establishing feature-based TL to enable the data-limited printer to co-learn a common pattern for SSE on the local shape with diverse printing shapes. Sabbaghi and Huang (2018) and Francis *et al.* (2020) proposed an effect-equivalent framework to address the challenge of model transfer between different processes and material systems. This line of research primarily focused on the modeling of geometric error.

In contrast, the impacts of printing process conditions, such as speed, acceleration, and jerk specified in G-code, on quality have not been modeled. In a recent study (Ren *et al.*, 2021), a between-printer TL algorithm (Figure 2) was developed to supplement the information from one source printer to one target printer to improve the learning of a kinematics–quality model for the target printer with limited samples. Nevertheless, this method is primarily focused on transferring the knowledge from one data-rich to one data-limited printer. The performance of the aforementioned AM TL degrades when the data from the source printer is limited, and the methods do not address the problem of co-learning for process–quality relationships from multiple data-limited printers.

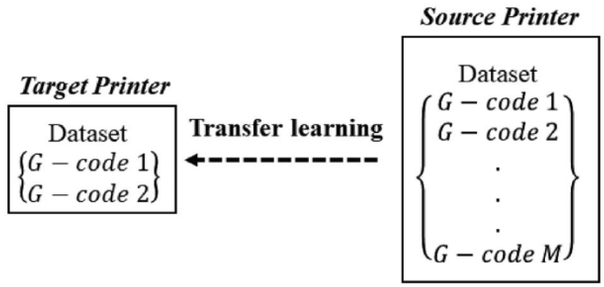


Figure 2. TL between two printers, as addressed by Ren *et al.* (2021).

Relatively less research has been conducted on multi-source learning problems. Yao and Doretto (2010) proposed an upgraded boost learning method that can integrate the information from multiple sources in TL and improve learning accuracy by updating the target's parameters. Shao *et al.* (2017) developed multi-task learning of a surface model to improve the prediction of the machined surface shape by leveraging the spatial data from a number of similar, but not-identical, processes. To tackle the issue of small-sample modeling for each task, Huang *et al.* (2012) introduced Bayesian hierarchical modeling to construct a common pattern for multiple related graphical models for medical image data. A novel hybrid approach, multiple task learning, was proposed by Saha *et al.* (2016) to transfer the data knowledge between the related source and target and online update the learning tasks' parameters upon the arrival of the new data using the Bayesian method. These methods focused on different applications from AM. A recent study by Liu *et al.* (2021) proposed the Naïve Bayes model to transfer the knowledge from multiple metal printers to a new printer by updating the prior knowledge of the relationship between process parameters (e.g., laser energy and power) and mechanical property (e.g., density, microhardness). The methodology is purely data-driven and did not incorporate process insights to simplify the model structures to be learned. As such, its performance in dealing with co-learning among multiple small-data processes can be limited.

Based on literature reviews, this article envisions that leveraging the data from multiple existing printers can be an effective solution to the data scarcity problem on new printers. The interconnected printers built on a cloud platform can leverage the shared data to reduce experiments for model training and expedite each printer's learning process. As such, a bidirectional and unidirectional co-learning

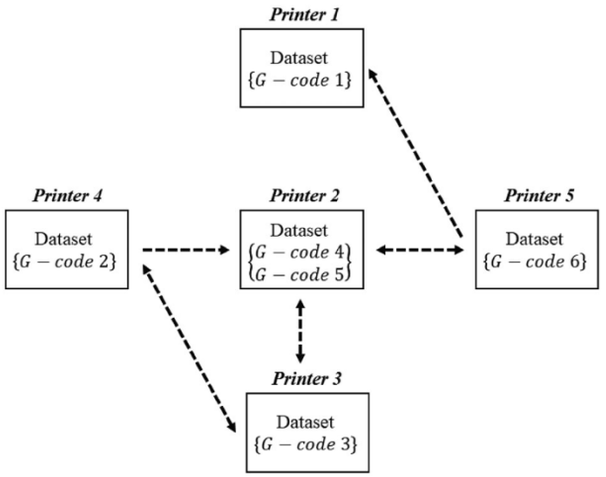


Figure 3. Envisioned co-learning network from multiple data-limited printers. Each printer may only collect the data under limited printing conditions (e.g., one or two G-code setup(s) per printer).

network can be established among multiple printers. As shown in Figure 3, the arrow between two printers indicates that the knowledge from a printer (at one end of the arrow) can partially contribute to the learning of another printer (the other end of the arrow) in the network. Multiple interconnected printers can jointly contribute to the learning of a data-limited printer, whereas some printers possessing less relatedness provide insignificant contributions. This envisioned network can allow multiple printers to co-learn the process–quality relationship even though each printer has very limited measurements.

From the analysis of the literature survey, the research challenges can be summarized as follows:

1. There is a lack of research that addressed the co-learning problem among multiple data-limited printers under the inter-connected structure, as shown in Figure 3. Prior TL in AM falls short of learning process–quality relationship based on limited data from each printer.
2. How to capture interpretable similarity or relatedness between different printing processes remains a challenge. Physical insights into the between-process relatedness can help simplify the process–quality modeling structure and potential computations compared with data-driven learning methodologies.
3. How to effectively explore the large search space of solutions for the co-learning problem, which involves joint estimation of common parameters of process—quality models and printer-specific parameters for each printer.

To deal with challenge 1, this article provides an interpretable between-printer relatedness model to develop the co-learning algorithm among multiple printers. Specifically, the model allows the common covariance structure for all the printers involved in the co-learning to share the information among printers. For challenge 2, this article utilizes the kinematics–quality relationship developed in Ren *et al.* (2021) as an example to capture the similar impacts of extruder kinematics on printed line variations, aiming to provide an interpretable model for multi-printer co-learning. This article further develops a hybrid metaheuristic approach to reduce the search space and solve the co-learning problem 3. This article also discusses the printing process selection to identify the printers that significantly contribute to improving co-learning accuracy. A case study demonstrates the proposed methodology for co-learning among four data-limited printers from different manufacturers.

The remainder of this article is organized as follows. Following Sections 1 and 2, Section 3 focuses on the methodology development for co-learning based on the modeling of between-printer relatedness guided by engineering insights into the kinematics–quality relationship. Section 4 develops a hybrid metaheuristic algorithm for solving the formulated co-learning problem. Section 5 demonstrates the co-learning between a pair of data-limited printers, and Section 6 discusses the printer selection among multiple printers to build co-learning networks. Experimental data from four different printers validate the results. Section 7 summarizes the results and future work. The detailed additional information of the appendices mentioned in the paragraphs is demonstrated in *Supplemental Online Materials*.

3. Between-printer relatedness modeling for co-learning

Based on Ren *et al.* (2021), this section proposes a more generalized model to capture the between-printer relatedness considering physical insights into how printing conditions impact printed quality. Without losing generality, the study will utilize the effect of extruder kinematics on the printing line variation as an example to demonstrate the development of the co-learning model.

3.1. Review of kinematics–quality modeling for extrusion-based printing

Ren *et al.* (2021) dealt with the between-printer transfer of knowledge on the impact of printing kinematics on printing quality. This research indicates that even though the feed rate has been fixed, there are still three stages, including acceleration, constant speed, and the deceleration phases along the printing nozzle's movement direction. Observations from a single printed line indicate that the kinematics can be strongly correlated to printed line variation in five phases, as labeled by (p1, p2 … p5) in Figure 4. p1 denotes the phase when the nozzle is moving from the starting point and line width decreases until it stops. p2 represents the phase when the line width starts increasing until it stops at the nominal value as specified. p3 is when the line width keeps increasing from the specified nominal value until it reaches the constancy. p4 is the phase when both the nominal feed rate and line width remain constant. In the final phase, p5 denotes when the feed rate starts to decelerate, and the line width begins to increase until the extruder stops or moves on to the next printing task.

To characterize the variation of the line caused by the kinematics-induced parameter, Ren *et al.* (2021) also

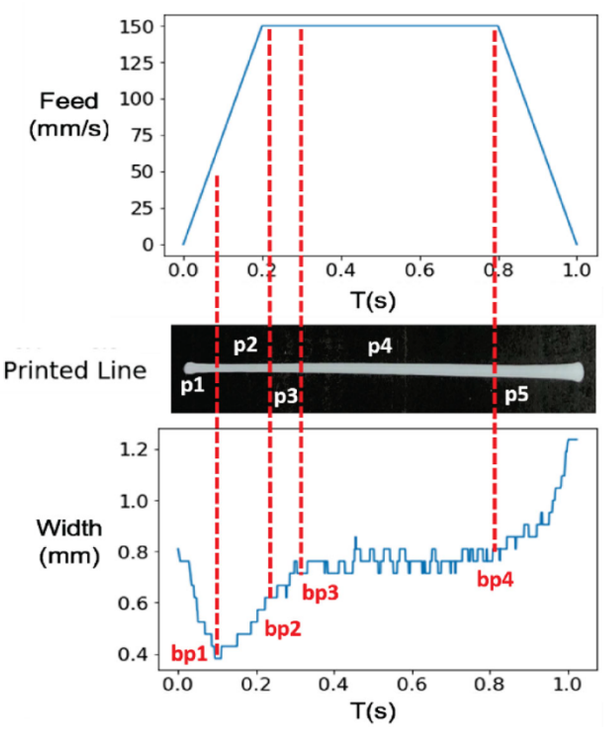


Figure 4. Five phases of straight-line printing.

modeled the kinematic–quality relationship as a piecewise function with five segments, which links the printing line variation to various kinematic parameters, i.e., speed, acceleration, and jerk as shown in (1) as follows:

$$w(t) = \begin{cases} \dfrac{1}{(\theta_1 v(t) + \theta_2)} + e_1 & \text{if } 0 \le t \le bp1 \\ \theta_3\sqrt{(t - bp1)v(t)a} + w(bp1) + e_2 & \text{if } bp1 < t < bp2 \\ \theta_4\sqrt{(t - bp2)v(t)} + w(bp2) + e_3 & \text{if } bp2 < t < bp3 \\ \theta_5 + e_4 & \text{if } bp3 < t < bp4 \\ \theta_6 \log\left(\dfrac{F}{v(t)}\right) + w(bp4) + e_5 & \text{if } bp4 < t \end{cases} \quad (1)$$

where $w(t)$ is the line width at time t. $v(t)$ denotes the actual nozzle travel speed at time t. a and F denote the actual print acceleration and print speed. This model can predict the line width under different kinematic settings as specified in G-code. (The validation of the kinematic-piecewise model (Ren *et al.*, 2021) and the simulation for the non-uniformity of infill pattern are given in Appendices 1.1 and 1.2.)

3.2. Co-learning problem formulation based on the engineering-driven relatedness model

This study explores similar kinematic variation patterns in process variation data between multiple similar-but-non-identical printers. A prior study (Ren *et al.*, 2021) revealed the correlation between model parameters in (1) shares a similar covariance structure (Appendix 1.3 shows an example of the statistical testing on the correlation between model parameters based on experimental data). Six model parameters, denoted as $\vec{\theta} = (\theta_1, \theta_2 \cdots \theta_6)^T$ in process-quality model (1), are utilized to characterize the kinematic variation on straight-line printing. The covariance structure $\boldsymbol{\Sigma}^{(\boldsymbol{\kappa})}$ for the model parameters $\vec{\theta}$ quantifies the kinematic–quality relationship between different line segments. The structure also allows for selecting multiple interconnected printers $\boldsymbol{\kappa}$ which contain useful data to co-learn the common kinematic variation pattern and improve the learning of parameters for each data-limited printer. The covariance structure is a 6×6 matrix $\boldsymbol{\Sigma}^{(\boldsymbol{\kappa})}$, where each entry is $cov(\theta_i, \theta_j)$.

Assume the printing quality of segment i for a target printer can be estimated by a piecewise linear function (e.g., $Y_i = X_i\theta_i + \varepsilon_i$) from (1), where X_i represents the terms consisting of kinematic inputs (e.g., speed, acceleration) in (2) for the estimation of segment i. Take p3 (Figure 4) as an example, X_3 can be expressed as $\sqrt{(t - bp2)v(t)}$. Y_i is the printed linewidth of segment i, which can be expressed as $w(t) - w(bp(i-1))$. The variations of different printing line segments from the same printer are statistically the same (Ren *et al.* 2021). As such $\varepsilon_i \sim N(0, \sigma^2)$. In addition, the learning problem is affected by the appropriate selection of cloud printer combinations $\boldsymbol{\kappa}$. Thus, the co-learning problem considering common covariance structure and printer selection can be represented in (2), where $\|\|_\gamma^2$ denotes the weighted norm squared, e.g., $\|x\|_\gamma^2 = x^T\gamma\ x$. λ, in this article, is tuned based on the range of the lower-bound and upper-bound from sample variance for different printed line segments due to the small samples (one or two G-codes) from each data-limited printer.

Denote Y as the printing quality data from the target printer. The model parameters $\vec{\theta} \sim MVN(\mu, \Sigma)$ is assumed in the kinematics–quality model for the target printer and can be estimated leveraging other cloud printers by solving a least-squares estimation problem or maximizing a posterior $P(\vec{\theta}\,|\mathbf{data})$, e.g., by

$$\max_{\vec{\theta}} P(\vec{\theta}\,|Y) \propto P(Y|\vec{\theta})\ P(\vec{\theta}).$$

The assumption of multivariate normal prior $\vec{\theta} \sim MVN(\mu, \Sigma)$ was made based on the numerical results from Ren *et al.* (2021), which conducted the transfer learning between two printers. It was also motivated by past experience that most coefficients fitted from experimental data are centered around a value under the same printing condition, and model coefficients from different phases in the piecewise model are not independent. Assume $\boldsymbol{\kappa}$ is the cloud printer combination selected for co-learning and the target printer $\in \boldsymbol{\kappa} \subset \boldsymbol{K}$. $\boldsymbol{\Sigma}^{(\boldsymbol{\kappa})}$ is a common covariance structure among model parameters jointly shared with all the printers from $\boldsymbol{\kappa}$. The estimation is also represented by

$$\max_{\vec{\theta}} \log P\left(Y\vec{\theta}\right) + \log P\left(\vec{\theta}\right),$$

then the objective can be formulated as

$$\min_{\boldsymbol{\kappa}} \min_{\vec{\theta}, \boldsymbol{\Sigma}^{(\boldsymbol{\kappa})}} \sum_{i=1}^{6} \left(\|\vec{Y}_i - \vec{X}_i\theta_i\|_2^2 \right) + \lambda\|\vec{\theta}\|_{\boldsymbol{\Sigma}^{(\boldsymbol{\kappa})-1}}^2$$
$$subject\ to\ det\left(\boldsymbol{\Sigma}_n^{(\boldsymbol{\kappa})}\right) > 0, \forall n \in \{1, 2 \ldots 6\} \quad (2)$$

The inner minimization (first-stage optimization) learns the model parameters $\vec{\theta}$ along with a common covariance structure $\mathbf{\Sigma}^{(\boldsymbol{\kappa})}$, given the cloud printer combinations $\boldsymbol{\kappa}$ and the outer minimization (second-stage optimization) identifies $\boldsymbol{\kappa}$. Also, a constraint is added to verify the semi-positive definiteness of each candidate $\mathbf{\Sigma}^{(\boldsymbol{\kappa})}$ generated in the optimization problem. Specifically, let $\mathbf{\Sigma}_n^{(\boldsymbol{\kappa})}$ be as an upper left $n \times n$ submatrix, where $1 \leq n \leq 6$. If all the $det\left(\mathbf{\Sigma}_n^{(\boldsymbol{\kappa})}\right)$s are positive, then $\mathbf{\Sigma}_n^{(\boldsymbol{\kappa})}$ is positive definite. (The derivation of (2) is given in Appendix. 1.4)

Ren *et al.* (2021) learned the coefficients in (1) by leveraging a data-rich printer to a data-limited printer. The common covariance structure $\mathbf{\Sigma}^{(\boldsymbol{\kappa})}$ of the model parameter in (1) is estimated by sample covariance based on the data from a data-rich printer. The limitation is that this formulation does not facilitate the co-learning among multiple data-limited printers. The formulation in this study has two major considerations beyond Ren *et al.* (2021): (i) the common covariance structure of $\vec{\theta}$ is jointly learned by multiple printers, and the covariance structure $\mathbf{\Sigma}^{(\boldsymbol{\kappa})}$ is unknown in the optimization problem; and (ii) selection of appropriate cloud printer combinations $\boldsymbol{\kappa}$ to facilitate the co-learning, which is not completely addressed in Ren *et al.* (2021). The challenges of solving the co-learning problem in (2) are outlined as follows:

1. The unknown covariance structure among different segments of model parameters is commonly shared among multiple printers. A simple aggregation of printers' data to estimate a pooled sample covariance is not accurate, since the data from each printer may be insufficient to estimate the model coefficients and their correlation. The common covariance structure should be jointly learned with model parameters associated with each printer and the selection of cloud printer combinations.
2. The joint learning of covariance structure $\mathbf{\Sigma}^{(\boldsymbol{\kappa})}$ along with the selection of cloud printer combinations $\boldsymbol{\kappa}$ significantly enlarge the search space for the joint optimization problem.
3. Traditional gradient-based approaches also encounter challenges in the complexity of ensuring positive definiteness. In addition, the printer selection makes the combinatorial search NP-hard. There is a strong need to develop an efficient metaheuristic customized to the joint optimization formulated as a two-stage decision problem.

To estimate variables $\left[\vec{\theta}, \mathbf{\Sigma}^{(\boldsymbol{\kappa})}, \boldsymbol{\kappa}\right]$ in the co-learning problem, this article proposes a hybrid metaheuristic method to solve the problem by first reducing the feasible search space and then combining two metaheuristics to explore different search spaces. By contrast, the solution to the problem in Ren *et al.* (2021) only involves a least-square estimation with a closed-form representation.

4. Hybrid metaheuristic to solve the multi-printer co-learning problem

Denote $\mathbf{Y} = \left[\vec{Y}_1, \vec{Y}_2 \cdots \vec{Y}_6\right]$ and $\mathbf{X} = \left[\vec{X}_1^T, \vec{X}_2^T \cdots \vec{X}_3^T\right]$. Taking the first derivative of f for $\vec{\theta}$ yields

$$\vec{\theta} = \left(\mathbf{X}^T\mathbf{X} + \sigma^2\mathbf{\Sigma}^{(\boldsymbol{\kappa})^{-1}}\right)^{-1}\mathbf{X}^T\mathbf{Y} \tag{3}$$

Therefore, the search for model parameters $\vec{\theta}$ for the target printer requires the solution to covariance structure $\mathbf{\Sigma}^{(\boldsymbol{\kappa})}$ and the selection of cloud printers $\boldsymbol{\kappa}$.

This section first proposes a posterior-distribution-based sampling estimation to deal with the challenge of estimating $\mathbf{\Sigma}^{(\boldsymbol{\kappa})}$ from all the printers involved in co-learning. To narrow the search space, this study first models the distribution of model parameters $\vec{\theta}$ based on multivariate statistics. A Bayesian Framework is first established to update the probability of the common covariance structure by using the data from multiple data-limited printers. Once the posterior distribution of $\mathbf{\Sigma}^{(\boldsymbol{\kappa})}$ is obtained, the statistical sampling can be implemented to construct the most likely intervals that parameters should fall within, based on which a metaheuristic algorithm can be developed to estimate $\mathbf{\Sigma}^{(\boldsymbol{\kappa})}$ with the minimized intervals. The accomplishment of the proposed method could also lead to a new opportunity of leveraging the cloud data uploaded from printers on the users' ends to improve the co-learning of quality control strategies.

4.1. Posterior distribution for the common covariance structure

In Bayesian statistics, one common conjugate prior to the unknown covariance and unknown mean of the multivariate normal distribution is a Normal Inverse Wishart (NIW) Distribution, which is a four-parameter distribution with $\vec{\mu}_0$, $\mathbf{\Lambda}_0$, ν_0, K_0. $\vec{\mu}_0$ denotes the parametric mean vector of the dataset, $\mathbf{\Lambda}_0$ denotes the scaling matrix and denotes a $d \times d$ positive definite scaling matrix. Since the prior distribution of the common covariance is unknown, it is common to set a non-informative prior distribution to start with and minimize the effect on the posterior from the prior. (For completeness of the discussion, Appendix. 2.1 presents a weakly informative prior for $\mathbf{\Sigma}^{(\boldsymbol{\kappa})}$). By following Schuurman *et al.* (2016), μ_0 is chosen to be a zero vector and $\mathbf{\Lambda}_0$ as an identity matrix to maintain the constraint of its positive definiteness. d is the number of the model parameters and is equal to six for the kinematics–quality model. ν_0 denotes degrees for freedom and $\nu_0 > d - 1$, and $K_0 > 0$ denotes the scaling factor. Based on Murphy (2007), the prior for the covariance can be estimated by $\mathbf{\Sigma}_{prior} \sim IW_{\nu_0}(\mathbf{\Lambda}_0)$ and mean $\vec{\mu}_{prior}|\mathbf{\Sigma}_{prior} \sim N\left(\vec{\mu}_0, \frac{\mathbf{\Sigma}_{prior}}{K_0}\right)$. Thus, $(\vec{\mu}_{prior}, \mathbf{\Sigma}_{prior})$ follows a NIW Distribution $(\vec{\mu}_{prior}, \mathbf{\Sigma}_{prior}) \sim NIW(\vec{\mu}_0, \mathbf{\Lambda}_0, \nu_0, K_0)$. The probability density function for $(\vec{\mu}_{prior}, \mathbf{\Sigma}_{prior})$ is determined by

$$\mathrm{P}\left(\vec{\mu}_{prior},\ \boldsymbol{\Sigma}_{prior}\right) = \frac{1}{\psi}|\boldsymbol{\Sigma}_{prior}|^{-\left(\frac{\nu_0+d}{2}+1\right)}\exp\left(\frac{1}{2}tr(\boldsymbol{\Lambda}_0^{-1}\boldsymbol{\Sigma}^{-1}) - \frac{K_0}{2}(\vec{\mu}_{prior}-\vec{\mu}_0)^T\boldsymbol{\Sigma}^{-1}\left(\vec{\mu}_{prior}-\vec{\mu}_0\right)\right),$$
$$\text{where } \psi = \frac{2^{\left(\frac{\nu_0 d}{2}\right)}\Gamma_d\left(\frac{\nu_0}{2}\right)\left(\frac{2\pi}{K_0}\right)^{\frac{d}{2}}}{|S|^{\frac{\nu_0}{2}}} \tag{4}$$

After specifying the prior, the posterior distribution can be obtained from the data from all printers. The posterior distribution would also follow a NIW Distribution, and the initial parameters can be updated as $NIW(\vec{\mu}^{\,*}, K^*, \boldsymbol{\Lambda}^*, \nu^*|data, \vec{\mu}_0, \boldsymbol{\Lambda}_0, \nu_0, K_0, \vec{\mu}_{prior}, \boldsymbol{\Sigma}_{prior})$, where

$$\vec{\mu}^{\,*} = \frac{K_0}{K_0+n}\vec{\mu}_0 + \frac{n}{K_0+n}\bar{\vec{x}},$$
$$K^* = K_0 + n,\ \nu^* = \nu_0 + n,$$
$$\boldsymbol{\Lambda}^{*-1} = \boldsymbol{\Lambda}_0^{-1} + \boldsymbol{S} + \frac{K_0 n}{K_0+n}\left(\bar{\vec{x}}-\vec{\mu}_0\right)\left(\bar{\vec{x}}-\vec{\mu}_0\right)^T,$$
$$\boldsymbol{S} = \sum_{i=1}^{n}(\vec{x}_i-\bar{\vec{x}})(\vec{x}_i-\bar{\vec{x}})^T,$$

where n denotes the total number of experimental runs, $\vec{x}_i$ denotes the vector of all parameters in the ith experiment run, and $\bar{\vec{x}}$ is the sample mean $\frac{\sum_{i=1}^{n}\vec{x}_i}{n}$ based on the data from all printers.

4.2. Metaheuristic over MCMC-sampling-refined search space for co-learning

Based on the distribution for covariance structure, this section proposes a Markov Chain Monte Carlo (MCMC) sample-refined metaheuristic to learn the model parameters (given printer combination selection) in two steps. The first step narrows the search space again after specifying the posterior distribution for $\boldsymbol{\Sigma}^{(\boldsymbol{\kappa})}$ by sampling. The second step utilizes a solution-solution-based metaheuristic to seek the optimal $\boldsymbol{\Sigma}^{(\boldsymbol{\kappa})}$ in the reduced search space.

Step 1. Reduce search space by MCMC sampling

The posterior distribution for the marginal covariance matrix $\boldsymbol{\Sigma}_{post}$ given data from all printers can be derived by $\boldsymbol{\Sigma}_{post}|data \sim IW(\boldsymbol{\Lambda}^*, \nu^*)$. The estimation of the covariance requires the multi-dimensional integration that involves posterior distributions, posing a major challenge to computation. Therefore, MCMC sampling can be implemented to simulate a myriad of samples to approximate the random variables (Yildirim, 2012). In this case, MCMC can be employed to sample each entry $cov(\theta_i,\ \theta_j)$ from $\boldsymbol{\Sigma}_{post}$ following the marginal distribution $IW(\boldsymbol{\Lambda}^*,\ \nu^*)$ and estimate the optimal covariance $\boldsymbol{\Sigma}^{(\boldsymbol{\kappa})}$ based on the Metropolis–Hastings algorithm. The steps of the algorithm are shown in **Algorithm I.**

To narrow the sampling range, the credible intervals of estimated parameters need to be constructed on the marginal distribution of the covariance. A credible interval represents the probability that the unobserved parameter value would fall in this range. Since the credible ranges are not unique, one challenge is which interval is more justifiable to be sampled since the interval could vary with the posterior distribution. In this case, the narrowest credible interval with a 95% probability (95%-NCI) can be selected to include the most plausible value of every entry $cov(\theta_i,\ \theta_j)$ from the posterior $\boldsymbol{\Sigma}^{(\boldsymbol{\kappa})}$. (The visualization of $cov(\theta_i,\ \theta_j)$ with a 95%-NCI is shown in Appendix. 2.2)

Algorithm I. Sampling from $\boldsymbol{\Sigma}_{post}|data \sim IW(\boldsymbol{\Lambda}^*, \nu^*)$ via Metropolis-Hastings algorithm

Input

TT	Total time,
B	Burn-in period,
M	Collection of MCMC samples for each entry of $\boldsymbol{\Sigma}_{post}$
$\boldsymbol{\Sigma}_{post}\|data$	Posterior target distribution by given data,
$\boldsymbol{\Sigma}_{post}^t$	Latest updated posterior covariance $\boldsymbol{\Sigma}_{post}$ in sampling.
$Q(x\,\|\,x^t)$	Proposal distribution given the mean value x^t.
$q(x^t\,\|\,x^{t-1})$	Proposal probability density function of x^t, where $x^t \sim \mathrm{Q}(x\mid x^{t-1})$.
$f(y)$	Posterior target probability density function of y, where $\mathrm{y} \sim \mathrm{IW}(\boldsymbol{\Lambda}^*, \nu^*)$.

Output

M in TT without B for $\boldsymbol{\Sigma}_{post}|data$

Initialize the $\boldsymbol{\Sigma}_{post}^0$ from $\boldsymbol{\Sigma}_{post}|data$
for $\boldsymbol{t} = 1:\ TT$ do
  Generate a candidate $\boldsymbol{\Sigma}_{post}^t \sim \mathrm{Q}(\boldsymbol{\Sigma}_{post}^t|\boldsymbol{\Sigma}_{post}^{t-1})$
  Generate the random value $\gamma \sim$ Uniform(0, 1)
  If $\gamma < \mathrm{Min}\left\{1, \frac{f(\boldsymbol{\Sigma}_{post}^t)q(\boldsymbol{\Sigma}_{post}^{t-1}|\boldsymbol{\Sigma}_{post}^t)}{f(\boldsymbol{\Sigma}_{post}^{t-1})q(\boldsymbol{\Sigma}_{post}^t|\boldsymbol{\Sigma}_{post}^{t-1})}\right\}$
    Append $\boldsymbol{\Sigma}_{post}^t$ to M
  else
    $\boldsymbol{\Sigma}_{post}^t = \boldsymbol{\Sigma}_{post}^{t-1}$
    Append $\boldsymbol{\Sigma}_{post}^{t-1}$ to M
Return M$[B+1:\ TT]$

Step 2. Optimize $\Sigma^{(\kappa)}$ and $\vec{\theta}$ via a metaheuristic refined by MCMC

The model parameter $\vec{\theta}$ can be estimated by (3) given the sampled values of $\boldsymbol{\Sigma}^{(\boldsymbol{\kappa})}$. Metaheuristic approaches can be implemented to approximate the optimal value of the parameters. Finding the global optimum from a large search space while avoiding local optima would be extremely challenging. Metaheuristic algorithms can mitigate solutions being trapped within the local optima while preserving the process of the hill-climbing approach for optimization. Some representative methods, such as Simulated Annealing (SA), are probabilistic techniques to deal with the optimization problem based on the analogy of the Metropolis–Hasting algorithm (Henderson

et al., 2006). Without losing generality, this article utilizes the SA as an example in **Algorithm II** to maximize the improvement percentage of the Root-Mean-Square Error (RMSE), denoted as $\hat{R}$, by the co-learning methodology. The calculation of $\hat{R}$ between single printer prediction RMSE_{Single} and $\mathrm{RMSE}_{Co-learn}$ can help evaluate the performance of the model prediction,

i.e., $\hat{R} = (\mathrm{RMSE}_{Single} - \mathrm{RMSE}_{Co-learn})/\mathrm{RMSE}_{Single}$.

Algorithm II. SA over MCMC-refined search space

Input

T and T_{min}	Temperature and a minimum temperature, t Number of times of exploring the new solutions in T,
$f(x)$	The cost function for estimating $\hat{R}$, given the co-learned covariance x,
α	Cooling factor,
M(95%-NCI)	The 95%-NCI by MCMC for each entry of posterior covariance $\boldsymbol{\Sigma}^{(\boldsymbol{\kappa})}$,
PD	Boolean operator for indicating the positive definiteness of $\boldsymbol{\Sigma}^{(\boldsymbol{\kappa})}$,
n	Size of the upper sub-matrix of $\boldsymbol{\Sigma}^{(\boldsymbol{\kappa})}$ with a size $n \times n$,
$\boldsymbol{\Sigma}^0$	The latest updated optimal covariance as input for $f(x)$.

Output

The point estimation of each entry of $\boldsymbol{\Sigma}^{(\boldsymbol{\kappa})}$ and model parameters $\vec{\theta}$

Initialize the $\boldsymbol{\Sigma}^0$

while $T > T_{min}$ **do**

  for 1: t **do**

    Generate the $\boldsymbol{\Sigma}^t$ from *M(95%-NCI)*.

  PD = True

  for 1: n **do**

    If $det(\boldsymbol{\Sigma}_n^t) < 0$

      PD = False

  If PD is True

    If $f(\boldsymbol{\Sigma}^t) > f(\boldsymbol{\Sigma}^0)$

      $\boldsymbol{\Sigma}^0 = \boldsymbol{\Sigma}^t$

  else

    Generate the random value γ from [0, 1] for each entry of $\boldsymbol{\Sigma}^0$

    If $\gamma < e^{\left(\frac{f(\boldsymbol{\Sigma}^t)-f(\boldsymbol{\Sigma}^0)}{T}\right)}$

      $\boldsymbol{\Sigma}^0 = \boldsymbol{\Sigma}^t$

  $T = T*\alpha$

Estimate $\vec{\theta}$ by Eqn. (3)

Return $\boldsymbol{\Sigma}^0$, $\vec{\theta}$

4.3. Selection of printer combinations for co-learning by a population-based metaheuristic

An appropriate selection of printers can ensure efficient learning accuracy, since dissimilar printers may introduce data that mislead learning and increase computation. However, there may exist non-unique selections of printers. As such, extra constraints and practical heuristics can be applied to identify the feasible selection of printers. A population-based metaheuristic, such as the Evolutionary Algorithm (EA, Bozorg-Haddad *et al.*, 2017), can provide a higher chance of not being trapped by local optima within a reasonable computational time. The chromosome can be defined as a set **Y** to make different selections of printers by changing the values 1 or 0 for each component y. The fitness value can be chosen to be the RMSE improvement percentage over the single printer learning. The parent chromosomes are generated from the current population by using the roulette wheel selection method. The chromosome with a larger fitness value in the current population has a higher chance to be selected as a parent for producing offspring. The crossover and mutation operators perform the strategy of exploration and exploitation to maintain the competitiveness and diversity of the population in each generation. The EA iteration stops when a practical computational time is reached. The EA procedure is also provided in **Algorithm III.** (The EA operators are illustrated in Appendix. 2.3).

Algorithm III. EA searching for printer combination $\boldsymbol{\kappa}$ hybrid with MCMC-refined metaheuristic

Input

$\boldsymbol{P}$	Population consists of n chromosomes,
$\boldsymbol{PP}$	Population of parent chromosomes,
$\boldsymbol{PO}$	Population of offspring chromosomes,
m	The number of printers in a chromosome,
ii	th generation,
jj	th chromosomes in $\boldsymbol{P}_{\mathrm{i}}$,
kk	th printers in $\boldsymbol{P}_{i,j}$,
λ	The number of current chromosomes passed to the next generation,
η	$j - \lambda$,
C_r	Crossover rate,
M_r	Mutation rate,
$f(\boldsymbol{P}_{ij})$	The cost function for estimating $\hat{R}$, given $\boldsymbol{\Sigma}^{(\boldsymbol{\kappa})}$ and $\vec{\theta}$ optimized by MCMC-refined SA metaheuristic from $\boldsymbol{P}_{i,j}$,

Output

The chromosome with the best $\hat{R}$ in $\boldsymbol{P}_i$

i = 1

for 1: n do

  $j = 1$

  for 1: m do

    $k = 1$

    $\boldsymbol{P}_{i,j,k} = randint(0, 1)$

    $k = k + 1$

  Optimize $\boldsymbol{\Sigma}^{(\boldsymbol{\kappa})}$ and $\vec{\theta}$ for $(\boldsymbol{P}_{i,j})$ through Eqn. (2) and MCMC-refined SA metaheuristic

  $j = j + 1$

If η %2! = 0

  $\lambda = \lambda + 1$

while *NotTerminated()* **do**

  for 1: λ do

    $j = 1$

    $\boldsymbol{P}_{i+1,\ j}$ = selectParent($\boldsymbol{P}_i$)

If $rand(0,1) < C_r$
Append $\boldsymbol{P}_{i+1,\ j}$ to $\boldsymbol{PP}_i$
$j = j + 1$
for 1: $\eta/2$ **do**
$\boldsymbol{PP}_i =$ Two chromosomes randomly selected from $\boldsymbol{PP}_i$
$\boldsymbol{PO}_{i,1},\ \boldsymbol{PO}_{i,2} = \text{crossover}(\boldsymbol{PP}_{i,1},\ \boldsymbol{PP}_{i,2})$
for 1: m **do**
$k = 1$
If $rand(0,1) < M_r$
$\text{PO}_{\text{i},1,\text{k}} = \text{mutate}(\text{PO}_{\text{i},1,\text{k}})$
If $rand(0,1)_{i2k} < M_r$
$\text{PO}_{\text{i},2,\text{k}} = \text{mutate}(\text{PO}_{\text{i},2,\text{k}})$
$k = k + 1$
Append PO_i to $\mathbf{P}_{\text{i}+1}$
$i = i + 1$
for 1: n do
$j = 1$
Update the optimized $\Sigma^{(\kappa)}$ and $\vec{\theta}$ for $(\text{P}_{\text{i,j}})$ through Eqn. (2) and MCMC-refined SA
$j = j + 1$

Return The chromosome with the best $\hat{R}$ in $\boldsymbol{P}_i$

4.4. A summary of hybrid metaheuristic refined by MCMC for two-stage optimization

By combining the procedures in Sections 4.1 to 4.3, the proposed hybrid metaheuristic (SA + EA) on the search space refined by MCMC sampling can be summarized as follows:

1. Determine the prior model parameters that follow a NIW Distribution.
2. Obtain the posterior distribution of model parameters by considering the likelihood of all the printers' data (training data) on the prior distribution and derive the marginal distribution of the covariance matrix $\boldsymbol{\Sigma}^{(\boldsymbol{\kappa})}$ that follows an Inverse Wishart Distribution.
3. Estimate $\boldsymbol{\Sigma}^{(\boldsymbol{\kappa})}$ by using MCMC sampling and building up the 95%-NCI for each entry of $\boldsymbol{\Sigma}^{(\boldsymbol{\kappa})}$, e.g., $cov(\theta_i,\ \theta_j) = L_{0.95} \le cov(\theta_i,\ \theta_j) \le U_{0.95}$
4. Obtain the optimal $\boldsymbol{\Sigma}^{(\boldsymbol{\kappa})}$ from the 95%-NCI leading to the largest $\hat{R}$ by single solution-based metaheuristic, i.e., SA.
5. Re-optimize the model parameters $\vec{\theta}$ by identifying/updating the combination $\boldsymbol{\kappa}$ of cloud printers in the network over a regular time interval (over hours or days) by a population-based metaheuristic, such as EA)

5. Case study 1: Co-learning between two data-limited printers

Based on experiments, this case study focuses on the co-learning between two printer models, each of which has limited data collected under different G-code setups. This section also compares the learning performance between the TL (Ren *et al.*, 2021) and the proposed co-learning method.

5.1. Experimental condition and setup

Two extrusion-based printer models are employed in the experiment, including LulzBot Taz 5 and LulzBot Taz 6. Taz 6 is the printer of interest, whereas Taz 5 is a printer that can share data with it over a common cloud platform. The nozzle diameter for both printers is 0.5 mm. Taz 6 has the data obtained under only one G-code setup, and Taz 5 also has limited samples. The data collected under one G-code setup were chosen as the training data for Taz 6, and the data under other G-code setups in set $\boldsymbol{T}_1$, where $\boldsymbol{T}_1 = \left(\{\text{speed (mm/s)},\ \text{acceleration (mm/s}^2)\}\right) = \{\{150,\ 800\},\ \{150,\ 600\},\ \{50,\ 600\},\ \{150,\ 400\},\ \{50,\ 400\}\}$, are for testing. Both printers are data-limited, presenting a challenge to estimate the common covariance as adopted by Ren *et al.* (2021).

5.2. Results

To estimate $\boldsymbol{\Sigma}^{(\boldsymbol{\kappa})}$, the search space needs to be narrowed down by deriving the marginal distribution of the covariance structure under the Bayesian Framework. The four parameters μ_0, Λ_0, ν_0, K_0 for the prior NIW Distribution should be defined for each entry $cov(\theta_i,\ \theta_j)$ of $\boldsymbol{\Sigma}^{(\boldsymbol{\kappa})}$. The diagonal value of $\boldsymbol{\Lambda}_0$ was tuned to be 10 in this case. The value of ν_0 is six. K_0 is chosen to be one to minimize the scalar effect and $\mu_{prior} = (0,\ 0)'$. After specifying the parameters for the prior NIW, the posterior NIW could be updated as $NIW(\mu_{post},\ \boldsymbol{\Sigma}_{post}|\ \vec{\mu}^{\,*},\ \boldsymbol{\Lambda}^*,\ \nu^*,\ K^*)$ by considering $\vec{\theta}$. The posterior marginal distribution $\boldsymbol{\Sigma}_{post}$ is also derived from NIW as $\boldsymbol{\Sigma}_{post}|data \sim IW(\boldsymbol{\Lambda}^*, \nu^*)$ for MCMC sampling.

Given $\boldsymbol{\Sigma}_{post}$ for $cov(\theta_i,\ \theta_j)$, the MCMC sampling method was run to reduce the search space of $\boldsymbol{\Sigma}^{(\boldsymbol{\kappa})}$. This case runs 12,000 iterations and burn-in constitutes the first 2000 iterations. The normal distribution will be selected as a proposal distribution to sample a candidate $cov(\theta_i,\ \theta_j)^t \sim N(cov(\theta_i,\ \theta_j)^{t-1}, \sigma^2)$ and $cov(\theta_i,\ \theta_j)^t = cov(\theta_j,\ \theta_i)^t$. σ is treated as a moving step and tuned by evaluating the autocorrelation convergence plot. The samples were collected from the remaining 10,000 to build a 95%-NCI for $cov(\theta_i,\ \theta_j)$ of $\boldsymbol{\Sigma}$ in Fig. 5(a)-(b).

To test the stability of the MCMC, we first examine the trace plot (Figure 5(a)) and burn in the MCMC sample in the early iterations. Different starting values are tested to ensure that the trace plots are mostly overlapped. Then we plot the autocorrelation (Figure 5(c)) of the samples between different lags to ensure that the Markov chain is stationary (high correlation leads to a slow convergence pattern). As can be seen from one example $cov(\theta_i,\ \theta_j)$ sampling in Figure 5, the high correlations only appear among the short lags, and the algorithm can reach convergence within a few iterations. This NCI preliminarily narrows the range for plausible values of $\boldsymbol{\Sigma}^{(\boldsymbol{\kappa})}$, although the potential search space is still large. SA is then implemented to optimize each $cov(\theta_i,\ \theta_j)$ with a lower and acceptable value of the RMSE. After hyperparameter tuning, the temperature (T) of the algorithm is set to be one, and the minimum temperature (T_{min}) is set to a small value of 0.00001. The cooling

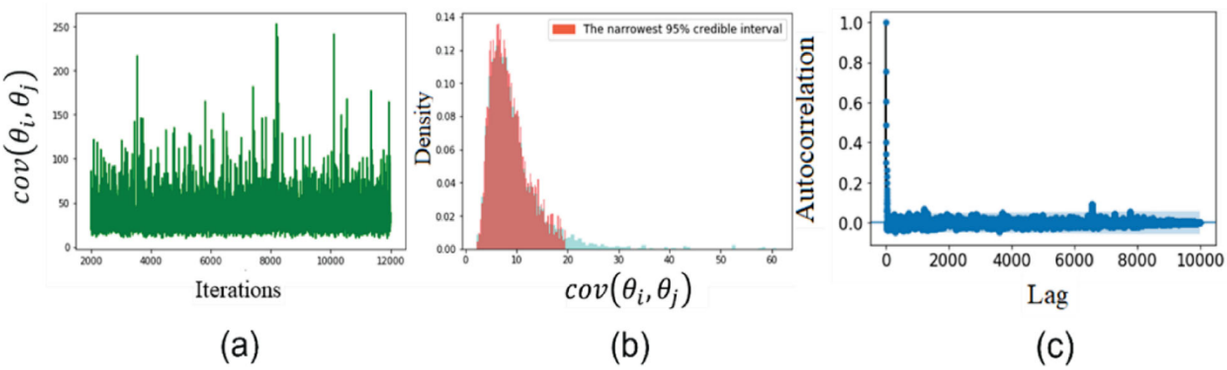


Figure 5. The example of (a) the trace plot; (b) the 95%-NCI; (c) the autocorrelation plot for checking the stability in the MCMC sampling of $\boldsymbol{\Sigma}^{(\kappa)}$.

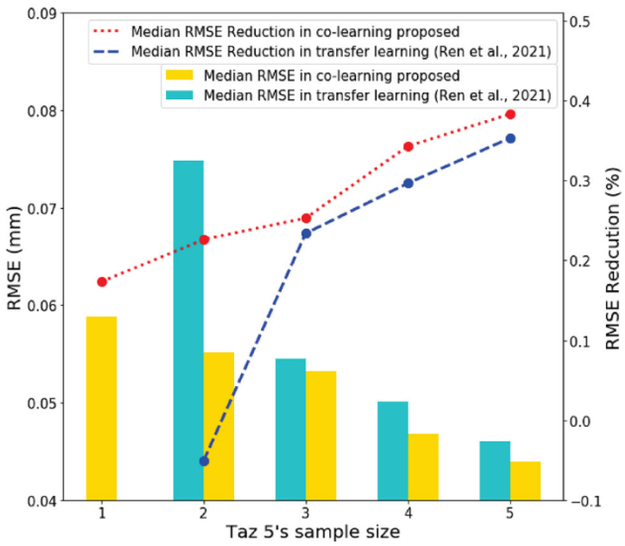


Figure 6. TL (Ren *et al.*, 2021) vs. proposed co-learning approach given different sample sizes from Taz 5. The left axis refers to the RMSE for prediction, as shown by the bar chart. The right axis is the median RMSE reduction (%) compared with the single-printer learning represented by dash lines.

factor α in a cooling step is chosen to be 0.9 for annealing. An identity matrix is chosen to be the initial covariance to start computing. The cost function is the percentage of the median RMSE reduction between the median RMSE obtained from the single-printer learning and proposed co-learning approach. As such, the learning problem is to find the combination of $cov(\theta_i,\ \theta_j)$ leading to the largest percentage of the median RMSE reduction. The results in Figure 6 show the improvement of the proposed co-learning methodology for two data-limited printers. Specifically, when each printer has only one or two samples, the co-learning methodology can significantly reduce the RMSE in quality prediction for Taz 6 compared with TL by Ren *et al.* (2021). (The detailed data in a tabular format for Figure 6 and an example of linewidth prediction by co-learning on Taz 6 are documented in Appendices 3.1 and 3.2)

6. Case study 2: Mulitple data-limited printers co-learning network

The incorporation of more extrusion printers into the co-learning can potentially contribute to the improvement in learning. This case study considers the co-learning among multiple printers to simulate the scenario when the users received quality and G-code data from more networked printers over the cloud. It is worth noting that the TL proposed by Ren *et al.* (2021) is limited to the knowledge transfer between a data-rich to a data-limited printer and does not apply to this case study.

6.1. Selection of cloud printer combinations

As in case study 1, the printer LulzBot Taz 6 is still the printer of interest, and the other one LulzBot Taz 5, and two Creality Ender's are the real printers sharing the data on the cloud. For simplicity, the two Creality Ender's are labeled as Ender 1 and 2, respectively. The nozzle diameter for two newly involved printers is 0.4 mm. The experiment compares the impacts of kinematics setup in G-codes on printing quality between Taz 6 and the other three networked printers. The printed line quality generated by the five G-codes as shown in $\boldsymbol{T}_1$ is evaluated for each printer involved in the co-learning. All printers exhibit similar variation patterns in phases p3 and p5, and their correlation information can be shared across printers. Due to the different characteristics of printer models, it is not guaranteed that more data from different printers can contribute to the co-learning process. In this study, all possible combinations of printers for co-learning include Taz 5, Ender 1, Ender 2, Taz 5+ Ender 1, Ender 1+ Ender 2, Taz 5 + Ender 2, and Taz 5 + Ender 1 + Ender 2.

The data in this case study were constructed so that only one sample is available for each printer to train the quality prediction model. Since the experiment tested five G-codes on each printer, one G-code from $\boldsymbol{T}_1$ is selected as training data at a time for Taz 6 to co-learn the printed quality with various G-codes from other printers. The RMSEs are collected by cross-validation on different one-selected training and four non-selected testing G-codes in $\boldsymbol{T}_1$. Then, the InterQuartile Rrange (IQR) and the median RMSE are constructed to evaluate the learning performance of each printer combination. Figure 7 compares different printer selections and shows the most significant improvement in the IQR and the median RMSE achieved by the co-learning. The results show that the selection of printers for co-learning with Taz 6 should include Taz 5 and Ender 1. This study

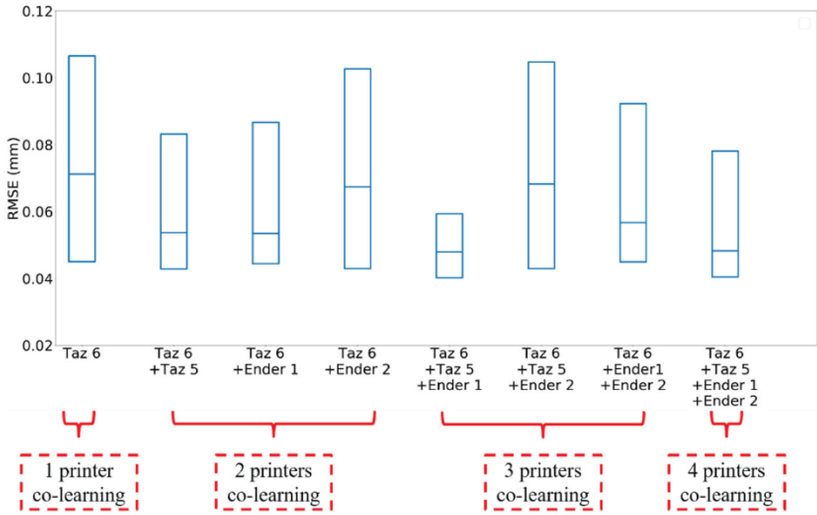


Figure 7. Examples of IQR of the RMSEs in co-learning with some printer selections.

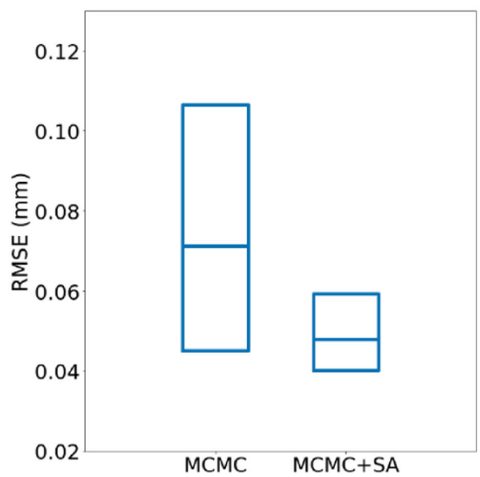


Figure 8. A comparison between the IQR of the method with and without applying SA on posterior-based sampling.

also compared the learning performance of employing MCMC only vs. MCMC+SA based on the best printer combination identified (Taz 6+Taz 5+Ender 1) in Figure. 8. The results of the IQRs show that the median RMSE from MCMC+SA significantly outperforms MCMC's values by more than 32%. Appendix. 4.1 further shows the autocorrelation convergence plot for MCMC diagnostics.

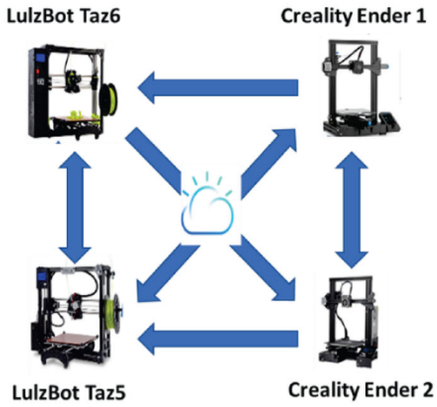
The co-learning can be implemented on all printers on the same cloud. The target printer of interest can be selected as Taz 5, Ender 1, and 2, and each printer still contains only one G-code. Figure 9 shows the co-learning network among four printers, where the arrows indicate the information from the printer at the arrow's origin can help co-learn the quality from the printer at the arrow's end. Taz 5 and Ender 1 have been tested to be the best printer combination for co-learning of printing quality in Taz 6. In addition, Taz 5 co-learns the printing quality with the other three printers involved and can maximize the learning performance up to around 36%. The printer Ender 1 can be co-learned with the combination of Taz 5 and Ender 2 with the best accuracy improvement of 22.56%. Ender 2 appears to exhibit more different behaviors from other printers and has the least improved accuracy through co-learning by up to 4.65%. It is noticeable that after the printer selection, the co-learning is not bi-directional between a pair of printers, indicating that the information sharing from the cloud is not equivalent for all printers.

6.2. Simulation of selection of cloud printer network

In this section, 20 printers are simulated from four groups of printers (Taz 5, Taz 6, Ender 1 & 2) for the co-learning network. Each group has five simulated printers. For group Taz 5, Ender 1 & 2, each printer contains only one G-code quality data generated by adding the Gaussian noise on the real data from $\boldsymbol{T}_1$. For group Taz 6, each of the four simulated printers contains one of the extra four untested G-codes $\boldsymbol{T}_2 = \left(\{\text{speed (mm/s)}, \text{ acceleration (mm/s}^2)\}\right) = \{\{100, 800\}, \{50, 800\}, \{100, 600\}, \{100, 400\}\}$ with Gaussian noise, and only one printer contains the real quality data (without noise) from $\boldsymbol{T}_1$. The printer with the real G-code data serves as the printer of interest. Also, the kinematic–quality relationship between line segments p3 and p5 is learned across different printer combinations.

In this study, the first-stage optimization is the same as Section 5.2 except for the cooling factor α. By considering the time limit, the co-learning of the first-stage optimization process in each printer combination is set not to exceed 120 seconds. Therefore, α is tuned to be 0.5 to accelerate the learning process. For the second-stage optimization, the population size n is five. The number of chromosomes chosen from P_i to P_{i+1} is three. The crossover and mutation rates are mostly set in a range [0.8-0.95] and [0.001-0.05] as suggested by Yang *et al.* (2015). In this article, the crossover rate (C_r) and mutation rate (M_r) are also defined as 0.8 and 0.01 within the above-suggested range. The stopping criterion of the two-stage optimization is set to the CPU running time of 3600 seconds for practical implementation.

To evaluate the proposed two-stage optimization, the study cross-validated the performance of a method by selecting one G-code data in $\boldsymbol{T}_1$ from a non-simulated printer as training data, and the rest of the non-selected four G-codes



Printer	Best Improvement in RMSE reduction by Co-learning
LulzBot Taz6	32.67%
LulzBot Taz5	36.62%
Creality Ender 1	22.56%
Creality Ender 2	4.65%

Figure 9. An established co-learning network from four data-limited printers (e.g., one or two G-code setup(s) per printer). For each printer, the arrows indicate the combination of printers that can jointly contribute to the co-learning accuracy with the best performance. The table to the right highlights the best improvement in RMSE reduction for quality prediction on each printer.

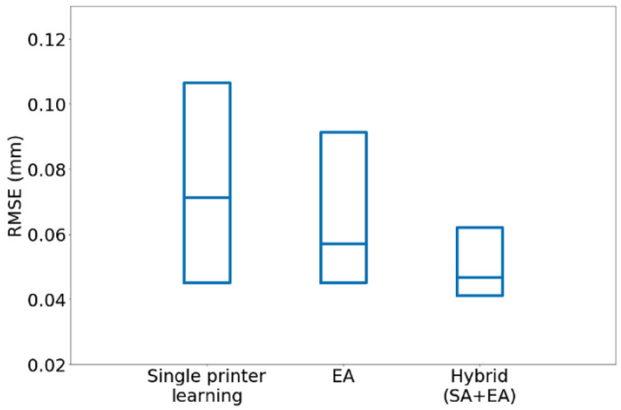


Figure 10. A comparison among three methods (single-printer learning, metaheuristic using EA only, and hybrid metaheuristic using SA for the first stage and EA for the second stage) for IQRs of the RMSEs from the cross-validation results.

in $\boldsymbol{T}_1$ are testing data. Then, the IQR and median RMSE are measured based on the collection of all the testing results (RMSEs). For each training G-code, three more runs are conducted, and the results with the largest median RMSE reduction are selected. In this section, the learning performances of the single-printer learning and two-stage co-learning solved by an EA metaheuristic are also compared with the proposed hybrid metaheuristic (single-solution-based SA + population-based EA). Figure 10 compares he IQR and median RMSE for three different methods.

It can be seen that the proposed two-stage co-learning improves the learning accuracy over single-printer learning by 34.33% and outperforms EA by 14.51% within the computation time. Given the results, it is seen that the single-solution-based metaheuristic (SA) is more efficient compared with the population-based method (EA) in dealing with the continuous space search in the first-stage optimization. The autocorrelation convergence plot for MCMC diagnostics is shown in Appendix. 4.2.

7. Conclusions and future work

Cloud AM has emerged over recent decades as a way to effectively utilize scattered manufacturing resources to deliver on-demand customized or personalized products. Interconnected printers on the cloud can share printing data for collaborative production state-of-the-art research mainly focuses on system integration and production scheduling for cloud printing; however, limited research is available on the quality control issues by taking advantage of data sharing on the cloud.

This article proposes a co-learning methodology that allows multiple interconnected printers to learn the process–quality relationship, especially multiple data-limited printers, potentially helping expedite process calibration for quality control. The methodology is established using an example of extrusion-based printers that have low precision. Traditionally, quality improvement requires the calibration of the printer setup that involves extensive experimentation to estimate how printing conditions in the G-code affect quality. The established co-learning methodology leverages process physics insights, a kinematics–quality model, to address the challenge. The contributions can be summarized as follows:

1. A two-stage optimization problem is formulated to jointly learn the printer selection, common covariance structure of the model coefficients, and the model coefficient for the target printer.
2. An MCMC-refined hybrid metaheuristic algorithm is developed to solve the two-stage optimization, which involves the challenge of ensuring positive definiteness of the covariance structure and the combinatorial problem. The MCMC based on the distribution of covariance structure first identifies the credible interval to narrow the search space for $\boldsymbol{\Sigma}^{(\kappa)}$. Based on the narrowed search space, a hybrid metaheuristic is developed, including a single-solution-based search to find the model parameters and $\boldsymbol{\Sigma}^{(\kappa)}$ within feasible computational timeframe at Stage 1 and a population-based search to explore the cloud printer combination at Stage 2.
3. The method to obtain the kinematics–quality relationship by co-learning the data from cloud printers provide a quantitative model for a new research direction, i.e., adjusting a printer's kinematics parameters, including

speed and acceleration, to compensate for the infill errors.
4. The proposed work can also be applied in (i) building reconfigurable and scalable manufacturing with various personalized products in future manufacturing, and (ii) lowering the entry barrier for new manufacturers such as small business owners to contribute their manufacturing resources, which is especially helpful for mobilizing local resources to manufacture products with high priority under a national emergency, and (iii) other applications such as IoT services and power control for a smart grid.

Future research may include:

1. Extension of the kinematics–quality model leveraging the data from cloud printers has the potential to compensate for printing infill errors by adjusting the extruder's moving speed and acceleration in the printer's G-code along printing paths. Such a compensation strategy is expected to improve infill uniformity and structural performance.
2. Extension of the error estimation to 2D and 3D geometric errors and infill patterns. Training the model based on the part with different dimensions will be another study to be conducted.
3. Consideration of the impact of machine degradation and retooling in the printer. The model parameters concerning the time sequence will also be developed in the model upgrade.
4. Extension of the methodology to other printing processes, such as stereolithography or metal printing, by co-learning how printing conditions impact quality.
5. Exploration of practical security models for cloud-based AM platform; cloud-based AM has potential risks and vulnerabilities for hackers or adversary to exploit it. This security access control model will help AM to be more reliable and secure.

Funding

This research is partially supported by the National Science Foundation grants CMMI-1901109 and HRD-1646897.

Notes on contributors

An-Tsun Wei received his BBA in logistics management from National Kaohsiung University of Science and Technology, Taiwan in 2015, and his MS in industrial engineering from Florida State University in 2020. He is a PhD candidate in industrial engineering at Florida State University. His research focuses on machine learning and cloud data analytics with application to quality control for additive manufacturing.

Hui Wang is an associate professor of industrial engineering at the Florida A&M University-Florida State University College of Engineering and a member of the High-Performance Materials Institute (HPMI). His research has been focused on (i) data modeling and analytics to support quality control for manufacturing processes, including small-sample learning under an interconnected environment, and (ii) optimization of manufacturing system design and supply chain. He received his PhD in industrial engineering from the University of South Florida and an MSE in mechanical engineering from the University of Michigan.

Tarik Dickens is an associate professor of industrial engineering at the Florida A&M University-Florida State University College of Engineering and a member of HPMI, a flagship program in nanocomposite materials. His research interests are in the areas of composite structure design, multifunctional composite processing, intelligent processing, additive and automated processing, and mechanical testing with a focus on failure prognosis (with interconnected sensors). He received his PhD in industrial and manufacturing engineering from Florida State University. Before that, he also earned his BS and MS degrees from Florida State University. His research has received more than $6 million in funding from the National Science Foundation, including the 2017 CREST Center award for Complex Materials Design for Additive Manufacturing, where he is a co-founder.

Hongmei Chi is a professor of computer and information sciences at Florida A&M University. Her research focuses on areas of applied cybersecurity, mobile health privacy, Monte Carlo and quasi-Monte Carlo, and data science. She received her PhD in computer science at Florida State University.

ORCID

An-Tsun Wei http://orcid.org/0000-0003-3346-8577
Hui Wang http://orcid.org/0000-0002-9412-3516

Data availability statement

The data supporting this study's findings are available at https://github.com/FAMU-FSU-IME/IISE-Co-learning.git.

Division of Civil, Mechanical and Manufacturing Innovation; Division of Human Resource Management.

References

Agarwala, M.K., Jamalabad, V.R., Langrana, N.A., Safari, A., Whalen, P.J. and Danforth, S.C. (1996) Structural quality of parts processed by fused deposition. *Rapid Prototyping Journal*, **2**(4), 4–19.

Alagoz, O., Hsu, H., Schaefer, A.J. and Roberts, M.S. (2010) Markov decision processes: A tool for sequential decision making under uncertainty. *Medical Decision Making*, **30**(4), 474–483. https://doi.org/10.1177/0272989X09353194.

Baumann, F. and Roller, D. (2017) Additive manufacturing, cloud-based 3D printing and associated services—overview. *Journal of Manufacturing and Materials Processing*, **1**(2), 15. https://doi.org/10.3390/jmmp1020015.

Bellini, A., Güçeri, S. and Bertoldi, M. (2004) Liquefier dynamics in fused deposition. *Journal of Manufacturing Science and Engineering, Transactions of the ASME*, **126**(2), 247–246.

Bouhal, A. (1999) Tracking control and trajectory planning in layered manufacturing applications. *IEEE Transactions on Industrial Electronics*, **46**(2), 445–451.

Bozorg-Haddad, O., Solgi, M. and Loáiciga, H.A. (2017) Meta-heuristic and evolutionary algorithms for engineering optimization. *Wiley Series in Operations Research and Management Science*, 53–67. https://doi.org/10.1002/9781119387053.

Chan, S.L., Lu, Y. and Wang, Y. (2018) Data-driven cost estimation for additive manufacturing in cybermanufacturing. *Journal of Manufacturing Systems*, **46**, 115–126.

Cheng, L., Tsung, F. and Wang, A. (2017) A statistical transfer learning perspective for modeling shape deviations in additive manufacturing. *IEEE Robotics and Automation Letters*, **2**(4), 1988–1993.

Cheng, L., Wang, A. and Tsung, F. (2018) A prediction and compensation scheme for in-plane shape deviation of additive manufacturing with information on process parameters. *IISE Transactions*, **50**(5), 394–406.

Cheng, L., Wang, K. and Tsung, F. (2021) A hybrid transfer learning framework for in-plane freeform shape accuracy control in additive manufacturing. *IISE Transactions*, **53**(3), 298–312.

Comminal, R., Serdeczny, M.P., Pedersen, D.B. and Spangenberg, J. (2018) Numerical modeling of the strand deposition flow in extrusion-based additive manufacturing. *Additive Manufacturing*, **20**, 68–76.

Equbal, A., Sood, A.K., Ansari, A.R. and Equbal, M.A. (2017) Optimization of process parameters of FDM part for minimizing its dimensional inaccuracy. *International Journal of Mechanical and Production Engineering Research and Development*, **7**(2), 57–66.

Francis, J., Sabbaghi, A., Ravi Shankar, M., Ghasri-Khouzani, M. and Bian, L. (2020) Efficient distortion prediction of additively manufactured parts using Bayesian model transfer between material systems. *Journal of Manufacturing Science and Engineering, Transactions of the ASME*, **142**(5), 051001.

Henderson, D., Jacobson, S.H. and Johnson, A.W. (2006) The theory and practice of simulated annealing, in *Handbook of Metaheuristics*, Springer, Boston, MA, pp. 287–319.

Huang, S., Li, J., Chen, K., Wu, T., Ye, J., Wu, X. and Yao, L. (2012) A transfer learning approach for network modeling. *IIE Transactions*, **44**(11), 915–931.

Lin, J. and Wang, K. (2012) A Bayesian framework for online parameter estimation and process adjustment using categorical observations. *IIE Transactions*, **44**(4), 291–300.

Liu, S., Stebner, A.P., Kappes, B.B. and Zhang, X. (2021) Machine learning for knowledge transfer across multiple metals additive manufacturing printers. *Additive Manufacturing*, **39**. https://doi.org/10.1016/j.addma.2021.101877.

Lyu, J. and Manoochehri, S. (2018) Modeling machine motion and process parameter errors for improving dimensional accuracy of fused deposition modeling machines. *Journal of Manufacturing Science and Engineering, Transactions of the ASME*, **140**(12), 121012.

Majeed, A., Lv, J. and Peng, T. (2019) A framework for big data driven process analysis and optimization for additive manufacturing. *Rapid Prototyping Journal*, **25**(2), 308–321.

Majeed, A., Zhang, Y., Ren, S., Lv, J., Peng, T., Waqar, S. and Yin, E. (2021) A big data-driven framework for sustainable and smart additive manufacturing. *Robotics and Computer-Integrated Manufacturing*, **67**, 102026.

Murphy, K.P. (2007) Conjugate Bayesian analysis of the Gaussian distribution. Technical report. University of British Columbia. http://citeseerx.ist.psu.edu/viewdoc/summary?doi=10.1.1.126.4603.

Noriega, A., Blanco, D., Alvarez, B.J. and Garcia, A. (2013) Dimensional accuracy improvement of FDM square cross-section parts using artificial neural networks and an optimization algorithm. *International Journal of Advanced Manufacturing Technology*, **69**(9–12), 2301–2313.

Qin, Q., Huang, J. and Yao, J. (2019) A real-time adaptive look-ahead speed control algorithm for FDM-based additive manufacturing technology with Hbot kinematic system. *Rapid Prototyping Journal*, **25**(6), 1095–1107.

Ravi, P., Shiakolas, P.S. and Thorat, A.D. (2017) Analyzing the effects of temperature, nozzle-bed distance, and their interactions on the width of fused deposition modeled struts using statistical techniques toward precision scaffold fabrication. *Journal of Manufacturing Science and Engineering, Transactions of the ASME*, **139**(7), 071007.

Ren, J., Wei, A.-T., Jiang, Z., Wang, H. and Wang, X. (2021) Improved modeling of kinematics-induced geometric variations in extrusion-based additive manufacturing through between-printer transfer learning. *IEEE Transactions on Automation Science and Engineering*, 1–12. https://doi.org/10.1109/tase.2021.3063389.

Sabbaghi, A. and Huang, Q. (2018) Model transfer across additive manufacturing processes via mean effect equivalence of lurking variables. *Annals of Applied Statistics*, **12**(4), 2409–2429.

Sabbaghi, A., Huang, Q. and Dasgupta, T. (2018) Bayesian model building from small samples of disparate data for capturing in-plane deviation in additive manufacturing. *Technometrics*, **60**(4), 532–544.

Saha, B., Gupta, S., Phung, D. and Venkatesh, S. (2016) Multiple task transfer learning with small sample sizes. *Knowledge and Information Systems*, **46**(2), 315–342.

Schuurman, N.K., Grasman, R.P.P.P. and Hamaker, E.L. (2016) A comparison of inverse-Wishart prior specifications for covariance matrices in multilevel autoregressive models. *Multivariate Behavioral Research*, **51**(2–3), 185–206.

Shao, C., Ren, J., Wang, H., Jin, J. and Hu, S.J. (2017) Improving machined surface shape prediction by integrating multi-task learning with cutting force variation modeling. *Journal of Manufacturing Science and Engineering, Transactions of the ASME*, **139**(1). https://doi.org/10.1115/1.4034592.

Takahashi, R., Matsubara, T. and Uehara, K. (2019) Data augmentation using random image cropping and patching for deep CNNs. *IEEE Transactions on Circuits and Systems for Video Technology*, **30**(9), 2917—2931.

van Engelen, J.E. and Hoos, H.H. (2020) A survey on semi-supervised learning. *Machine Learning*, **109**(2), 373–440.

van Hasselt, H., Guez, A. and Silver, D. (2016, March) Deep reinforcement learning with double Q-Learning. *Proceedings of the AAAI Conference on Artificial Intelligence*, Phoenix, Arizona, **30**(1). https://doi.org/10.1609/aaai.v30i1.10295.

Wang, A., Song, S., Huang, Q. and Tsung, F. (2017) In-plane shape-deviation modeling and compensation for fused deposition modeling processes. *IEEE Transactions on Automation Science and Engineering*, **14**(2), 968–976.

Wang, Y., Lin, Y., Zhong, R.Y. and Xu, X. (2019) IoT-enabled cloud-based additive manufacturing platform to support rapid product development. *International Journal of Production Research*, **57**(12), 3975–3991.

Yang, X.S., Chien, S.F. and Ting, T.O. (2015) Bio-inspired computation and optimization: An overview, in *Bio-inspired computation in telecommunications*, pp 1–21, Boston, Morgan Kaufmann, MA, USA.

Yao, Y. and Doretto, G. (2010, June) Boosting for transfer learning with multiple sources, in *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, pp 1855–1862, San Francisco, CA. https://doi.org/10.1109/CVPR.2010.5539857.

Yildirim, I. (2012) Bayesian inference: Metropolis-Hastings sampling, Department of Brain and Cognitive Sciences, University of Rochester, Rochester, NY. https://doi.org/10.1177/1475921718794299. [Online]: https://www2.bcs.rochester.edu/sites/jacobslab/cheat_sheet/MetropolisHastingsSampling.pdf. (accessed 19 June 2022).