Hybrid Multi-Task Learning-Based Response Surface Modeling in Manufacturing

Yuhang Yang[a], Chenhui Shao[a,*]

[a] *Department of Mechanical Science and Engineering*
University of Illinois at Urbana-Champaign
Urbana, IL 61801, United States

Abstract

Response surface modeling is an essential technique for identifying the optimal input parameters in a process, especially when the physical knowledge about the process is limited. It explores the relationships between the process input variables and the response variables through a sequence of designed experiments. Conventional response surface models typically rely on a large number of experiments to achieve reliable modeling performance, which can be cost prohibitive and time-consuming. Furthermore, nonlinear input-output relationships in some processes may not be sufficiently accounted for by existing modeling methods. To address these challenges, this paper develops a new response surface modeling approach based on hybrid multi-task learning (H-MTL). This approach decomposes the variability in process responses into two components—a global trend and a residual term, which are estimated through self-learning and MTL of Gaussian process (GP), respectively. MTL leverages the similarities between multiple similar-but-not-identical GPs, thus achieving superior modeling performance without increasing experimental cost. The effectiveness of the proposed method is demonstrated by a case study using experimental data collected from real-world ultrasonic metal welding processes with different material combinations. In addition, the hyperparameter selection, the effects of the number of

*Corresponding author
Email addresses: yang221@illinois.edu (Yuhang Yang), chshao@illinois.edu (Chenhui Shao)
URL: https://mechse.illinois.edu/people/profile/chshao (Chenhui Shao)

tasks, and the determination of the stopping criterion are discussed in detail.

Keywords: Multi-task learning, Gaussian process, Response surface modeling, Ultrasonic metal welding, Smart manufacturing, Data-efficient learning, Process optimization

---

1. Introduction

In extensive industrial applications, one overarching goal is to optimize the process response, which can be the performance of a process or the quality of a product, by controlling the operational input factors. Due to the complex physical phenomena involved in some applications, quantifying the relationship between input variables and response variables using physics-based models can be extremely difficult. Response surface modeling is a statistical tool to establish an empirical input-output relationship in a process. It has been investigated and successfully applied in a number of scientific and engineering disciplines, such as analytical chemistry [1], environmental science [2], food industry [3], and manufacturing [4–13].

A general response surface model can be expressed by:

$$y = f(x_1, x_2, \ldots, x_d), \tag{1}$$

where y is the response and $x_1, x_2, \ldots,$ and x_d are the input variables. The procedure of constructing response surface modeling is mathematically equivalent to the model fitting in a regression analysis. Most existing studies of response surface modeling fit the regression model from only a single dataset of the targeted process [14]. Though these studies achieved satisfactory modeling performance, the acquisition of training datasets through experiments is expensive and time-consuming. Typically, production costs associated with process optimization in manufacturing can be divided into three major categories: (1) costs of materials and machine usage, (2) labor costs, and (3) costs induced by delayed decision-making in choosing the process parameters during production launch [15–17]. Furthermore, the costs increase significantly as the design

space increases. For instance, for a process with three input variables and each input variable has five levels, the total number of parametric combinations is $5^3 = 125$. This number increases exponentially with the number of input variables. Therefore, a cost-effective response surface modeling method is critically needed.

A manufacturing process/machine may be tasked with different materials, configurations, tools, etc. An emerging opportunity in cost-effectively obtaining the response surface model of a new process lies in sharing the knowledge or information available from other experiments conducted under similar processing conditions. For example, the influence of welding time, clamping pressure/force, and vibration amplitude, on the joint strength in ultrasonic metal welding has been investigated separately in four different studies conducted by different researchers and corresponding response surfaces were obtained [10–13]. The experiments in these studies were carried out on different metal materials (magnesium to titanium, aluminum to steel, aluminum to copper, and copper to copper). In these modeled response surfaces, although the ranges of the maximum failure load and the obtained optimal welding parameters are different, the trends over the three welding parameters share some similarities. This indicates that one can potentially transfer knowledge and insights across different processes, thereby improving the learning/modeling efficiency and performance. Additionally, it has been demonstrated that tooling design [18], tool conditions [16, 19–22], joint configuration [12], and specimen surface condition [18, 23] substantially influence the joint strength in ultrasonic metal welding. Leveraging the similarities between different welding settings can potentially save a large amount of time and costs.

In this paper, we present a new H-MTL-based approach to cost-effectively model the response surfaces of multiple similar-but-not-identical manufacturing processes. In the proposed approach, the response surface for a given process is decomposed into two parts—a global trend and a local spatial variability, which are estimated by self-learning and MTL, respectively. An iterative procedure is established to find the optimal decomposition and model parameters for each

part.

The remainder of this paper is organized as follows. Section 2 reviews existing response surface modeling and MTL methods. In Section 3, we present the H-MTL-based response surface modeling approach. A simulation-based numerical case study and a real-world case study using experimental data collected from ultrasonic metal welding processes are presented in Section 4. The hyperparameter selection, the effects of the number of tasks, and the determination of the stopping criterion are discussed in Section 5. Finally, Section 6 concludes the paper.

2. Literature Review

Response surface methodology was first introduced by Box and Wilson in 1951 [24]. The modeling techniques used in response surface methodology have been evolving since then. The most popularly used models include polynomial regression and multiple linear regression [14]. For example, Awad and Hassan optimized the environmental and quality cost in machining processes using polynomial regression with selected machining parameters, including cutting speed, feed speed, depth of cut, and tool nose radius [8].

In recent years, machine learning techniques have been increasingly adopted to capture intricate, nonlinear input-output relationships, including artificial neural network, support vector machine (SVM), and Gaussian process (GP) [9–12, 25–29]. For instance, Jurkovic et al. [9] studied the influence of three input parameters, including cutting speed, feed rate, and depth of cut, on three output responses, namely, cutting force, surface roughness, and tool life, in a high-speed turning process. They compared the performances of SVM regression, polynomial regression, and artificial neural network. Zhao et al. built a response surface of tensile strength of the joint in ultrasonic welding of magnesium-titanium dissimilar alloys using artificial neural network and identified the optimal ranges of three welding parameters including clamping force, welding time, and vibration amplitude [10]. Satpathy et al. performed a similar study on ultrasonic spot

welding of aluminum-copper dissimilar metals using linear regression model, artificial neural network, and adaptive neuro-fuzzy inference system. They also analyzed the influence of welding parameters and their interactions [11]. Meng et al. characterized the influence of welding time and welding amplitude on the resulting joint quality in terms of shear and peel strengths using machine learning-based response surface models, and analyzed the differences in two quality indices [12]. They further developed a multi-objective optimization approach to optimize a compound strength indicator that was defined as the average of normalized shear and peel strengths. Sun et al. proposed to predict the welding quality using Bayesian regularized neural network with features generated and selected from welding process signals [29]. The relationship between welding process parameters, weld attributes, and joint performance was also investigated.

Recent studies showed that deep reinforcement learning (DRL) have superior performance for sequential optimization problems with a high-dimensional parametric space when the descriptive mathematical models are not available [27, 30]. He et al. developed a DRL-based framework for efficiently optimizing textile chemical manufacturing process with multi criteria [31]. Leng et al. applied a DRL algorithm to minimize the cost of color changeovers in an automotive painting process [32]. In these examples, deep neural networks were used to approximate the interaction between the process variables and systems/environments, and a large amount of data and computation resources are required. However, such a large dataset may not be accessible in some manufacturing applications due to the high cost of physical experiments.

MTL has emerged as a solution for transferring knowledge among multiple similar-but-not-identical processes. MTL is particularly useful in scenarios where it is expensive or infeasible to recollect sufficient training data and rebuild the models for new learning tasks. In the past few years, MTL was successfully used in multiple engineering/manufacturing applications, e.g., [33–36]. In the automotive industry, an engineering-guided MTL approach was developed to improve the machined surface shape prediction by integrating MTL with cut-

ting force variation modeling [33]. In the energy industry, Shireen et al. [34] proposed an iterative MTL approach for time-series modeling of solar panel photovoltaic outputs, and the proposed method overcomes the challenge in time series data modeling caused by a significant amount of missing or unavailable historical data. In grain quality monitoring, Wang et al. [35] established a mixed-effect framework for thermal field estimation and adopted a spatiotemporal MTL approach to model the local variability. Most recently, Chen et al. [36] developed an MTL method for data-efficient spatiotemporal modeling of tool surface progression in ultrasonic metal welding. A new spatiotemporal kernel was devised based on a squared sine exponential damping function to characterize the periodic trend of tool surfaces. They also analyzed the computational efficiency of their method and developed effective computing acceleration methods. Ramezankhani et al. [37] developed a neural network-based transfer learning framework for robust and reliable modeling for exotherm curve in composites autoclave processing. To the best of our knowledge, the application of MTL in response surface modeling has not been explored.

It should be noted that, although many machine learning models are available, it is challenging to include the results from all machine learning techniques in one study. Our pilot study showed that other machine learning techniques, such as SVM, had performance worse than candidate methods discussed later in this manuscript. In this study, because of the spatial correlation pattern shown by the data set, we identify GP as the most appropriate baseline method for modeling the residual variability, and focus on the comparison between the MTL version and single-task learning (STL) version of GP modeling.

3. H-MTL-Based Response Surface Model

Assume there are m similar-but-not-identical processes, the response surface model for process l can be expressed as

$$Z_l(\boldsymbol{x}) = g_l(\boldsymbol{\beta}_l; \boldsymbol{x}) + f_l(\boldsymbol{x}) + \epsilon, \epsilon \sim \mathcal{N}(0, \sigma_\epsilon^2), \tag{2}$$

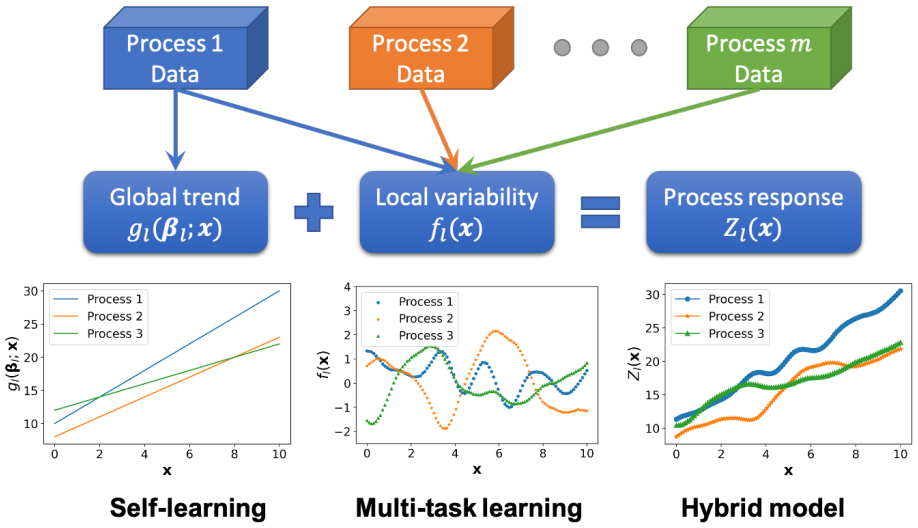


Figure 1: Illustration of the H-MTL learning scheme.

where $\boldsymbol{x}$ is the vector of input parameters, Z_l is the response surface model for process l, $g_l(\cdot)$ is the global trend with parameters $\boldsymbol{\beta}_l$, f_l is the local spatial variability, and ϵ is a zero-mean error term with variance σ_ϵ^2 .

Fig. 1 illustrates the learning scheme of the proposed H-MTL algorithm. Instead of applying the conventional MTL approaches directly, we decompose the response surface of a given process into a global trend that is governed by process input parameters and a local spatial variability term that follows a zero-mean GP. The global trend is estimated individually for each task through self-learning, and the local spatial variability term can be jointly learned across multiple similar-but-not-identical processes. In other words, the H-MTL model is a combination of a self-learning regression model and an MTL-GP model.

This hybrid modeling approach is inspired by the combination of a regression model and a simple GP model in geostatistics. Such a combination has been widely discussed since Matheron first introduced the universal kriging in 1969 [38], and it has been proven that this approach can yield higher prediction accuracy than a plain regression model or GP model in many applications, including mapping pollution in soil layers [39–41], temperature estimation in

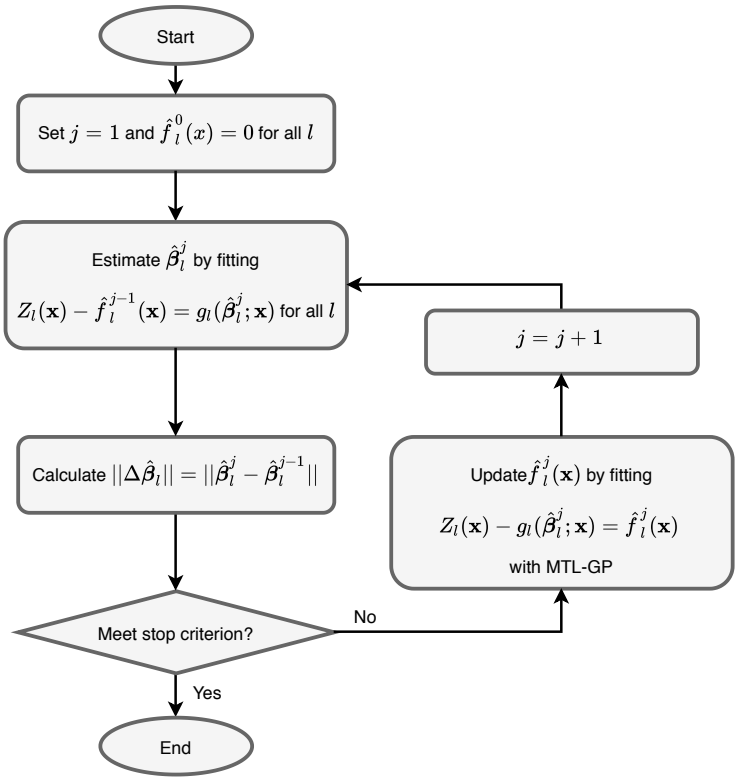


Figure 2: Flowchart of the proposed H-MTL algorithm procedure.

climate monitoring [42], monitoring geometric quality in semiconductor manufacturing [43], and remaining useful life estimation for machining tools [44]. In geostatistics, different terms including kriging with external drift and regression-kriging have been used to describe the combination of a linear regression model and a zero-mean GP [39–42]. This technique has also been referred to as additive GP or integrated GP in some studies, e.g., [43, 44]. In these methods, the regression and GP models are estimated using the same data set, essentially representing an STL scheme. However, in Model (2), we hypothesize that the local variability term, which is modeled by a zero-mean GP, shares similarities among multiple similar-but-not-identical processes. Consequently, MTL can be applied to learn the GP term to improve the learning efficiency and performance.

Two sets of parameters need to be estimated for each process in the Model (2),

the parameters $\boldsymbol{\beta}_l$ in the self-learning model and the parameters in the zero-mean GP f_l. Because it is very challenging to estimate both sets of parameters analytically, we develop an iterative model estimation procedure, which is shown by Fig. 2. Similar iterative estimation procedures have been developed for regression-kriging methods [41, 45] and MTL methods [33–35]. In the initialization of our procedure, we set the local spatial variability component as 0 for all processes. In the first iteration, the initial parameters of the global trend model are estimated individually for each process, and the residual is modeled by MTL-GP. In iteration j, the parameters of the global trend model, $\hat{\boldsymbol{\beta}}_l^j$, are estimated by fitting the model, $g(\hat{\boldsymbol{\beta}}_l^j; \boldsymbol{x})$, to the response subtracting the MTL-GP modeled local variability component from previous iteration, $Z_l(\boldsymbol{x}) - \hat{f}_l^{j-1}(\boldsymbol{x})$. The model of the local variability component, $\hat{f}_l^j(\boldsymbol{x})$, is updated by fitting the new residual, $Z_l(\boldsymbol{x}) - g(\hat{\boldsymbol{\beta}}_l^j; \boldsymbol{x})$, with MTL-GP. The process of model updating is repeated until a stopping criterion is satisfied. In our method, the stopping criterion is defined as follows: if the average change of the parameters in the global trend models for each process, $\left\|\Delta\hat{\boldsymbol{\beta}}\right\| = \frac{1}{m}\sum_{l=1}^{m}\left\|\Delta\hat{\boldsymbol{\beta}}_l\right\|$, is below a predetermined threshold η, the procedure will stop. It is worth noting that other stopping criteria, e.g., $\forall\Delta\hat{\boldsymbol{\beta}}_l < \eta$, can also be used. The discussion of threshold selection can be found in Section 5.3.

As reviewed in Appendix A, the model parameters of the MTL-GP model are estimated through an expectation–maximization (EM) algorithm, and the computational complexity of MTL-GP model is $\mathcal{O}(k_1 m n^3)$, where k_1 is the number of EM steps, m is the number of tasks, and n is the training sample size. In the H-MTL model, the computational complexity of MTL-GP model is dominant, while the computational cost of the self-learning regression can be neglected. The computational complexity of the H-MTL model is $\mathcal{O}(k_1 k_2 m n^3)$, where k_2 is the number of iterations needed to reach stopping criterion.

4. Case Studies

In this section, the effectiveness of the proposed H-MTL-based response surface modeling is demonstrated by a simulation-based numerical case study and a real-world case study using experimental data collected from ultrasonic metal welding.

The H-MTL-based response surface modeling method is compared with four representative state-of-the-art response surface modeling techniques: GP, linear regression (LR), GP with LR, and MTL-GP. A summary of these candidate models is given in Table 1. GP and MTL-GP model response surfaces using a GP with a constant mean. LR models the global trend only. In GP with LR, the response surface is decomposed into a global trend and a local variability term, which are modeled by LR and GP, respectively. This approach is sometimes referred to as universal kriging, kriging with external drift, and regression-kriging [39–42]. GP, LR, and GP with LR are considered as STL approaches because they do not transfer knowledge across processes. The H-MTL model and MTL-GP model, described in Sections 3 and Appendix A, respectively, are both MTL methods.

Table 1: Summary of candidate response surface modeling techniques

Method	Combining GP with regression model?	Transferring knowledge from other processes?
H-MTL	Yes	Yes
MTL-GP	No	Yes
GP w/ LR	Yes	No
LR	No	No
GP	No	No

The prediction error for data point i in the dataset is given by

$$e(\boldsymbol{x}_i) = \hat{Z}(\boldsymbol{x}_i) - Z(\boldsymbol{x}_i), \tag{3}$$

where $\hat{Z}(\boldsymbol{x}_i)$ and $Z(\boldsymbol{x}_i)$ are the predicted value and ground truth of the response, respectively. Root mean squared error (RMSE) and mean absolute error (MAE), which are defined by Eqs. (4) and (5), respectively, are used as the metrics to evaluate the prediction performance. A smaller RMSE or MAE indicates a better prediction performance.

$$\text{RMSE} = \sqrt{\frac{1}{n}\sum_{i=1}^{n} e(\boldsymbol{x}_i)^2}, \tag{4}$$

$$\text{MAE} = \frac{1}{n}\sum_{i=1}^{n} \left|e(\boldsymbol{x}_i)\right|, \tag{5}$$

where n is the total number of settings in testing dataset. Additionally, standard deviation (SD), defined by Eq. (6), is used to characterize the variation of prediction errors. A smaller SD of the prediction errors implies a smaller prediction variability.

$$\text{SD} = \sqrt{\frac{1}{n}\sum_{i=1}^{n} \left(e(\boldsymbol{x}_i) - \bar{e}\right)^2}, \quad \text{where} \quad \bar{e} = \frac{1}{n}\sum_{i=1}^{n} e(\boldsymbol{x}_i). \tag{6}$$

4.1. Numerical Case Study

In this section, we use a numerical example to illustrate the underlying learnt model in the H-MTL-based response surface modeling approach. We simulate a dataset with 5 processes over 100 input values, where the GP priors are inductively correlated. The joint GP prior distributions can be expressed as

$$f \sim \mathcal{N}(0, \Sigma), \quad \Sigma = K^f \otimes K^x, \tag{7}$$

where $\otimes$ denotes the Kronecker product, K^x is the base covariance matrix over inputs, K^f is a $l \times l$ positive semi-definite matrix that specifies the inter-task correlations. The response $\boldsymbol{y}_l$ for process l is a combination of a randomly generated linear trend and the corresponding GP:

$$\boldsymbol{y}_l \sim a_l \boldsymbol{x} + b_l + \mathcal{N}(0, \Sigma_l), \quad \Sigma_l = K^f_{ll} \otimes K^x, \tag{8}$$

Table 2: The averages of RMSE, MAE, and SD of the candidate response surface modeling techniques. The smallest values for each metric are bolded.

	H-MTL	MTL-GP	GP w/ LR	LR	GP
RMSE	**0.502**	0.621	0.724	1.053	1.236
MAE	**0.382**	0.521	0.460	0.821	0.867
SD	**0.441**	0.519	0.714	1.007	1.199

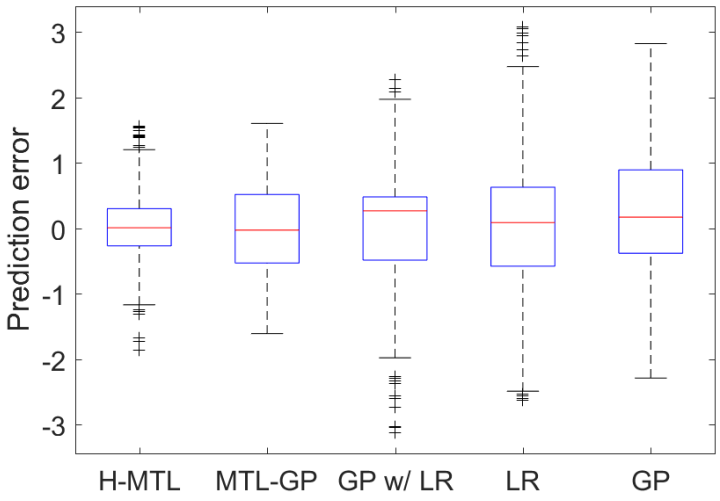


Figure 3: The boxplots of prediction errors in the testing dataset with five candidate methods.

where a_l and b_l are randomly sampled constants. For each task, responses of 100 input values are generated. 20 of them are randomly selected as the training data and the remaining 80 responses are used as the test data.

Table 2 lists the averages of RMSE, MAE, and SD of the candidate response surface modeling techniques. H-MTL model achieves a superior prediction performance with the lowest averages of RMSE and MAE. The average of SD in H-MTL model is also smallest. Fig. 3 shows the distributions of prediction errors from all tasks. H-MTL model has a mean of prediction errors close to 0 and the smallest ranges and the interquartile ranges of the prediction errors. These

imply that the H-MTL model can reduce the bias and variability in the prediction results. This numerical example shows that H-MTL model can improve the performance of the response surface modeling when the residual variability from multiple similar-but-not-identical processes that follow joint GP priors. When the true GP priors are unknown from a real-world dataset, a pilot study can be used to check the validity of the assumption of the task similarity.

4.2. Real-World Case Study

Ultrasonic metal welding is a solid-state joining technique. A bonding between thin metal sheets clamped under pressure is created with oscillating shears generated by ultrasonic vibration. It has wide industrial applications such as electric vehicle battery assembly, electronic packaging, and automotive body construction [10–13, 19–22]. The joint strength of ultrasonic metal welding is influenced by a set of process parameters, including welding pressure, welding time, and vibration amplitude. The relationship between process parameters and joint strength in ultrasonic metal welding has been actively investigated in the literature, e.g., [10–13]. In the rest of this section, we will first introduce the experimental setup and then compare the modeling performances of the H-MTL method and existing response surface modeling approaches.

4.2.1. Experimental Setup

The experiments are conducted on ultrasonic metal welder (Branson Ultraweld L20 Spot Welder) with a universal controller (Branson VersaGraphix controller). 0.254 mm-thick 110 copper sheet and 0.254 mm-thick 1100 aluminum sheets are used in this study. The specimens are 50.8 mm long and 25.4 mm wide. The ethanol-moistened cleaning wipes are used to clean the surfaces of the specimens and to get rid of contaminants before welding. The welding is conducted at a fixed frequency of 20 kHz. Welding pressure, welding time, and vibration amplitude are considered as the process input variables. Table 3 lists the details of input variables and their levels. Welding pressure varies between 25 psi and 70 psi at 15 psi intervals, welding time are chosen from 0.2 s to 1.0

s with increments of 0.2 s, and three levels of vibration amplitude are selected as 30 μm, 35 μm, and 40 μm. A full factorial design of experiments is chosen and the total number of factor level combinations of the three parameters is therefore $4 \times 5 \times 3 = 60$. The maximum load strength obtained through the T-peel test is selected as the response variable to represent the weld quality. The T-peel test is performed following ASTM D1876 with a constant strain rate of 0.5 mm/second at room condition.

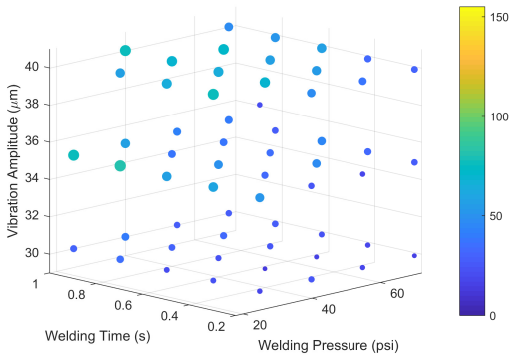
Table 3: Input variables and their levels in the experimental design.

Variable	Unit	Levels
Welding pressure	psi	25, 40, 55, 70
Welding time	second	0.2, 0.4, 0.6, 0.8, 1.0
Vibration amplitude	μm	30, 35, 40

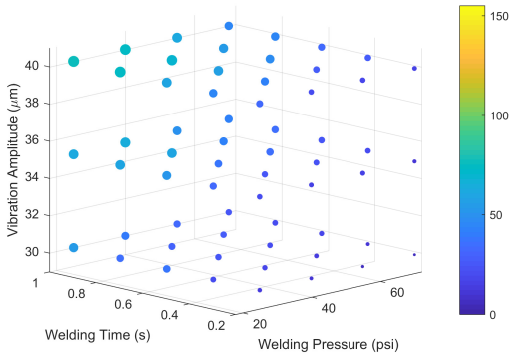
Three combinations of metals, aluminum to aluminum (Al-Al), copper to aluminum (Cu-Al), and copper to copper (Cu-Cu), are used to represent similar-but-not-identical welding processes. Because the ultrasonic welding process can be affected by some external factors, such as the material properties of specimens [18, 23] and the tool health status of the welding machine [20, 21], and these factors cannot be precisely monitored or controlled, the welding process may present a certain level of instability even under the same welding conditions [10, 18, 23, 46, 47]. To ensure the accuracy and reliability of the data, the experiments for each combination of materials at each factor level are repeated three times in a random order and the average maximum loads are used as the responses of joint strength. A total of $60 \times 3 \times 3 = 540$ experiments have been conducted.

4.2.2. Analysis of Results

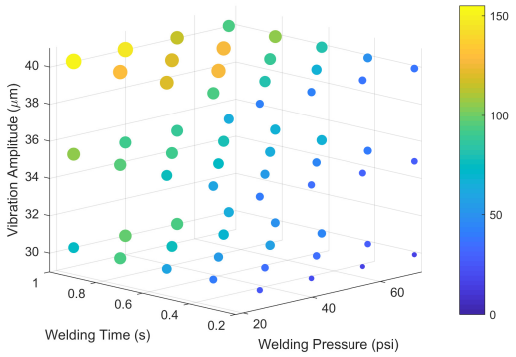
The three-dimensional (3D) distributions of joint strength over three welding parameters are displayed in Fig. 4. In Al-Al welding, two data points at (25 psi, 40 μm, 1 second) and (55 psi, 40 μm, 1 second) are not available, be-



(a) Al-Al



(b) Cu-Al



(c) Cu-Cu

Figure 4: The 3D distributions of joint strengths over three welding parameters. The size and color of the markers represent the values of joint strength. Two data points in Al-Al are not available due to the welding failure.

cause excessive welding energy is generated with the high values of welding time and vibration amplitude, and the top Al plates are cracked or collapsed for all samples at these two settings. The monotonic trends of the welding strength over the welding parameters can be observed. In this study, LR is utilized to model the global trends. The highest maximum load strength of Cu-Cu among all settings is higher than of Al-Al and Cu-Al. Although the ranges of the maximum load strengths in the three response surfaces are different, similar overall trends over the welding parameters can be observed. As such, it is reasonable to consider them as similar-but-not-identical processes.

In order to simulate the process of building a complete response surface model with a reduced number of actual experiments, the following procedure is taken: for each task, 20 out of 60 parametric combinations are randomly selected as the training data, and the remaining 40 combinations are used as the test data (38 settings in testing data for Al-Al).

The comparative results are presented by Table 4. It is clear that the H-MTL model outperforms other approaches in terms of the averages of RMSE, MAE, and SD. In all tasks, the MTL-based approaches demonstrate a significant improvement compared with STL approaches. The distributions of prediction errors from each task are shown in Fig. 5. It can be noticed that, in all three tasks, the means of prediction errors of the H-MTL model are close to 0 and the ranges and the interquartile ranges of the prediction errors of the H-MTL model are the smallest, indicating that the H-MTL model leads to the lowest bias and variability. In conclusion, with the same number of experiments, the H-MTL-based response surface model achieves a superior prediction performance compared to the state-of-the-art methods.

5. Discussion

In this section, we discuss three practical issues for implementing the proposed H-MTL method in industrial applications, including hyperparameter selection, the effects of the number of tasks, and the determination of the stopping

Table 4: The performance comparison of the response surface modeling techniques with all three tasks. The smallest values for each task are bolded. H-MTL model achieves the best overall performance while MTL-GP is the second to best method.

(a) RMSE

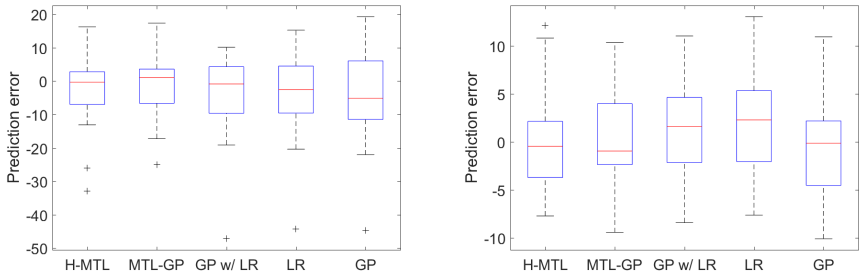
	H-MTL	MTL-GP	GP w/ LR	LR	GP
Al-Al	9.603	**8.826**	11.269	11.373	12.755
Cu-Al	**4.202**	4.664	4.824	5.669	4.873
Cu-Cu	**8.166**	10.857	11.010	11.725	14.079
Mean	**7.323**	8.116	9.034	9.589	10.569

(b) MAE

	H-MTL	MTL-GP	GP w/ LR	LR	GP
Al-Al	**6.654**	6.961	7.727	8.457	10.063
Cu-Al	**3.312**	3.702	3.962	4.583	3.924
Cu-Cu	**6.066**	8.422	8.192	9.241	10.424
Mean	**5.344**	6.362	6.627	7.427	8.137

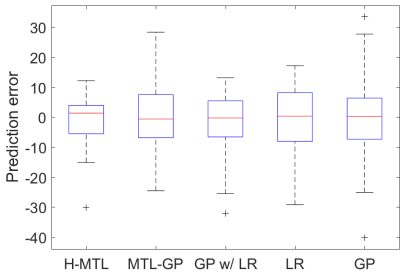
(c) SD

	H-MTL	MTL-GP	GP w/ LR	LR	STL-GP
Al-Al	9.041	**8.993**	10.811	10.572	12.284
Cu-Al	**3.752**	4.299	4.678	5.214	4.822
Cu-Cu	**7.700**	9.415	10.809	11.828	14.078
Mean	**6.831**	7.569	8.766	9.205	10.395



(a) Al-Al

(b) Cu-Al



(c) Cu-Cu

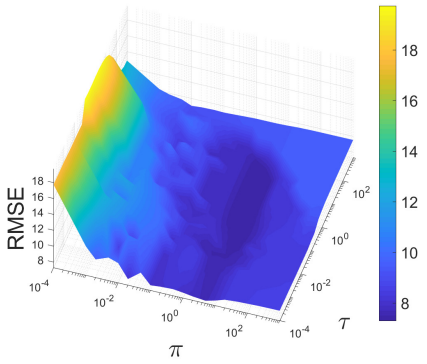
Figure 5: The boxplots of prediction errors from each testing dataset with five candidate methods.

criterion.

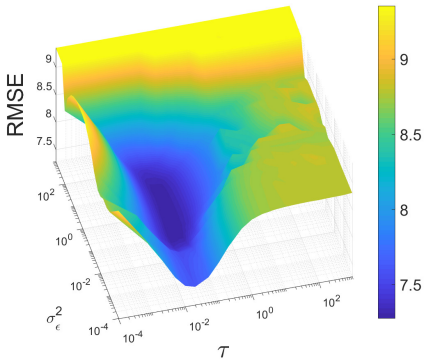
5.1. Hyperparameter Selection

The performance of the H-MTL model can be affected by three hyperparameters, i.e., τ and π that specify the normal-inverse-Wishart distribution, as well as the initial value of the error variance σ_ϵ^2, and these hyperparameters need to be determined prior to the modeling procedure. In this section, experiments are conducted to investigate the effects of the hyper parameters. In every run, each hyperparameter is assigned with one of 18 candidate values, varying from 0.0001 to 200. All $18 \times 18 \times 18 = 5832$ possible combinations are tested to identify the appropriate ranges of the hyperparameters.

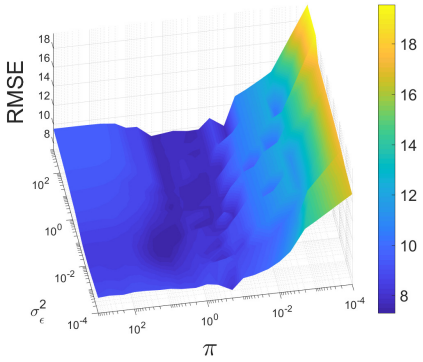
Fig. 6 displays the trends of RMSE when a pair of hyperparameters change while another hyperparameter is kept at a fixed value. It can be observed that



(a) RMSE vs. τ and π while σ_ϵ^2 is fixed at 0.02



(b) RMSE vs. τ and σ_ϵ^2 while π is fixed at 5



(c) RMSE vs. σ_ϵ^2 and π while τ is fixed at 0.05

Figure 6: Slices of the effects of hyperparameter pairs on RMSE.

low RMSEs are obtained with $0.02 < \tau < 0.1$, $2 < \pi < 10$, $0.01 < \sigma_\epsilon^2 < 0.2$. In other words, satisfying prediction results can be achieved when the hyperparameters are selected from appropriate ranges. The optimal ranges of the hyperparameters may vary case by case, and a cross-validation study can be conducted to find the appropriate values using the training data.

5.2. Effects of the Number of Tasks

The effects of the number of tasks have been discussed in a number of MTL-related studies [33–35]. In general, the prediction performance will improve when the number of similar tasks increases because more information is trans-

ferred. A similar conclusion could be drawn from this study.

In Section 4.2.2, the training data from all three tasks has been utilized. In this section, three experiments are conducted for each pair of tasks under the same experimental setup as described in Section 4.2.1. The results are summarized in Table 5 and Table 6. In all cases, the RMSEs are reduced in response surfaces built with the H-MTL model. It proves that even the information from only one external task is shared, the H-MTL approach can still improve the prediction performances of the response surface for the target task, though the improvement rate is smaller, as shown by Table 6. This implies that the MTL-based response surface modeling can benefit from more similar tasks.

Table 5: The RMSE comparison for GP w/ LR model and H-MTL model for each pair of tasks.

(a) Al-Al and Cu-Al

	GP w/ LR	H-MTL
Al-Al	11.269	10.932
Cu-Al	4.824	4.664

(b) Al-Al and Cu-Cu

	GP w/ LR	H-MTL
Al-Al	11.269	9.729
Cu-Cu	11.010	10.317

(c) Cu-Al and Cu-Cu

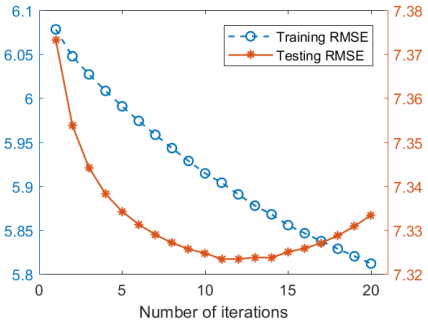
	GP w/ LR	H-MTL
Cu-Al	4.824	4.382
Cu-Cu	11.010	8.969

5.3. Effects of the Number of Iterations and the Stopping Criterion

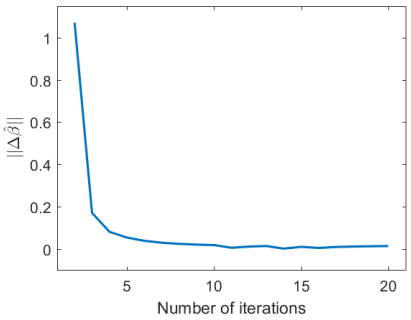
The proposed H-MTL model could estimate the optimal decomposition of individual trends and local variability components iteratively, and it is important to choose an appropriate stopping criterion to prevent overfitting. In this section, we first study the effects of the number of iterations in the parameter estimation process. Fig. 7 shows the changes of the training and testing RMSE as well as the difference between estimated parameters between iterations, $\left\|\Delta\hat{\boldsymbol{\beta}}\right\|$

Table 6: The improvement rate for each pair of tasks. The row name represents the target task to be predicted and the column name represents the external task that is included in learning.

	Al-Al	Cu-Al	Cu-Cu	Two external tasks
Al-Al	N/A	2.99%	10.59%	14.78%
Cu-Al	3.31%	N/A	9.17%	12.91%
Cu-Cu	4.78%	18.58%	N/A	25.83%



(a) The training and testing RMSEs vs. the number of iterations.



(b) $\left\|\Delta\hat{\beta}\right\|$ vs. the number of iterations.

Figure 7: The effects of the number of iterations.

vs. the number of iterations. It is noticed that, as the number of iterations increases, the training RMSE keeps decreasing, while the testing RMSE first decreases and then increases, and the change of estimated parameters between two consecutive iterations reduces continuously. This implies that the algorithm can be considered as converged when the estimated parameters do not change significantly and we should stop the iteration process to prevent overfitting. In this example, $\left\|\Delta\hat{\boldsymbol{\beta}}\right\|$ is smaller than 0.01 after 10-th iteration, and the smallest testing RMSE can be obtained from 11-th to 14-th iteration. Hence, it is reasonable to choose $\left\|\Delta\hat{\boldsymbol{\beta}}\right\| < 0.01$ as the stopping criterion in this case study.

It should be noted that a similar iterative procedure has been widely applied and discussed for the combination of a regression model and GP model in the

literature [45, 48]. Although it has been proven that a satisfactory prediction performance can be achieved with a single iteration in practice [48], fitting each component iteratively until convergence is required for obtaining the optimal result [45]. Our study agrees with these opinions. The training RMSE with one single iteration deviates from the optimal RMSE at 11-th iteration by 3% and the testing RMSE deviates by 1%. This performance is acceptable for practitioners who prefer computational efficiency and responsive decision-making. On the other hand, convergence is indeed required for achieving the optimal performance at the cost of more iterations. In practice, users should choose single-iteration or convergence approaches based on their applications.

The convergence of the iterative procedure may not be guaranteed and slow convergence or non-convergence may happen and lead to undesirable prediction results. A practical solution is to monitor the changes in estimated parameters and re-initialize the parameters randomly when fluctuation occurs or convergence is not achieved within a certain number of iterations.

5.4. Future Work

The proposed H-MTL model can be potentially applied to a number of other processes and disciplines where cost-effective response surface modeling is needed. However, it should be noted that some difficulties may prevent successful applications. First, the modeling performance relies heavily on the task similarity. If a process is not related/similar to other processes, MTL models could be misled and give even worse performances than STL. Another potential challenge is the selection of self-learning models. The similarity in residual variability can be affected by an improper self-learning model. However, there is a lack of quantitative evaluation for task similarity, and a pilot study on a validation dataset may provide valuable insights.

Sampling design, or design of experiments, in MTL is still an open question. Different sampling schemes may lead to largely varying prediction performance. The goal of sampling design is to optimally allocate the experimental efforts such that the parametric space can be more effectively explored. In this study,

the training datasets are randomly sampled from all combinations of parameters. An intelligent sampling design could further reduce the cost of actual experiments and improve the learning performance.

6. Conclusion

In this paper, an H-MTL-based response surface modeling approach has been developed for the first time. In this approach, the response surface of a given process is decomposed into two components—a global trend and a residual variability components, which are fitted through self-learning and MTL, respectively. By leveraging the similarities between multiple similar-but-not-identical processes, the H-MTL-based response surface modeling can achieve modeling performance superior to conventional STL-based response surface modeling methods. A simulation-based numerical case study and a real-world case study with the data collected from ultrasonic metal welding processes demonstrate that the proposed method can improve the prediction performance of response surface modeling without increasing the cost of experiments. Furthermore, the hyperparameter selection, the effects of number of tasks, and the determination of the stopping criterion are discussed in detail. Finally, future research is suggested to focus on the quantification of task similarity and sampling design for MTL.

Acknowledgment

This research has been supported by the National Science Foundation under Grant No. 1944345.

Appendix A. Review of MTL-GP model

There are a number of MTL algorithms for GP available with different ways of defining the commonality among different tasks. Without loss of generality, we adopt the inductive approach of MTL-GP developed by Yu et al. [49] in this study to illustrate the concept of MTL. In MTL, datasets from multiple

processes are taken in the training stage. Assume there are m similar-but-not-identical processes with the same number of input parameters, d. For each process l, the observation data can be represented as $\mathcal{D}_l = \{(\mathbf{X}_l, \boldsymbol{y}_l)\}$, where $\mathbf{X}_l \in \mathbb{R}^{n_l \times d}$, $\boldsymbol{y}_l \in \mathbb{R}^{n_l}$, n_l, is the size of observation data for process l. We further assume that there are in total n distinct data points in $\{\mathcal{D}_l\}$ with $\min(\{n_l\})$ $\leq n \leq \sum_l n_l$ and denote $\mathbf{X} = \cup \mathbf{X}_l, l = 1, \ldots, n$ as the set of distinguished $\boldsymbol{x}$ in $\{\mathcal{D}_l\}$. The goal of MTL is to estimate m related functions f_l, $l = 1, \ldots, m$, and predict the responses with unobserved parameters for a certain process based on observation data from all m processes.

Given the mean $\boldsymbol{\mu}_f$ and covariance $\mathbf{K}$ of function values $f = [f(\boldsymbol{x}_1),\ldots,$ $f(\boldsymbol{x}_n)]^\intercal$, there exist unique $\boldsymbol{\mu}_\alpha \in \mathbb{R}^n$ and $\mathbf{C}_\alpha \in \mathbb{R}^{n \times n}$ such that $\boldsymbol{\mu}_f = \boldsymbol{\kappa}\mu_\alpha$, $\mathbf{K} = \boldsymbol{\kappa}\mathbf{C}_\alpha\boldsymbol{\kappa}$. $\boldsymbol{\mu}_\alpha$ and $\mathbf{C}_\alpha$ are sampled from the hyper prior specified by a normal-inverse-Wishart distribution

$$p\left(\boldsymbol{\mu}_\alpha, \mathbf{C}_\alpha\right) = \mathcal{N}\left(\boldsymbol{\mu}_\alpha | 0, \frac{1}{\pi}\mathbf{C}_\alpha\right) \mathcal{IW}\left(\mathbf{C}_\alpha | \tau, \boldsymbol{\kappa}^{-1}\right), \tag{A.1}$$

where π and τ are the hyper parameters that specify the distribution. For each function f_l in process l, there exists a unique $\alpha^l \in \mathbb{R}^n$ such that $f_l = \boldsymbol{\alpha}^l\boldsymbol{\kappa}$ and $\boldsymbol{\alpha}^l \sim \mathcal{N}(\boldsymbol{\mu}_\alpha, \mathbf{C}_\alpha)$. Given a parameter vector, $\boldsymbol{x} \in \mathbf{X}_l$, the response can be expressed as

$$y = \sum_{i=1}^{n} \alpha_i^l \boldsymbol{\kappa}(\boldsymbol{x}_i, \boldsymbol{x}) + \epsilon, \epsilon \sim \mathcal{N}(0, \sigma_\epsilon^2), \tag{A.2}$$

where $\boldsymbol{x}_i \in \mathbf{X}$.

The model parameters $\theta = \{\boldsymbol{\mu}_\alpha, \mathbf{C}_\alpha, \sigma_\epsilon^2\}$ and $\boldsymbol{\alpha}^l$ can be estimated through an expectation–maximization (EM) algorithm. The EM algorithm is an iterative method to find maximum likelihood estimates of parameters by alternating between an expectation step and a maximization step, which are shown below.

- *E-step*: Given the current $\theta = \{\boldsymbol{\mu}_\alpha, \mathbf{C}_\alpha, \sigma_\epsilon^2\}$, this step estimate the expectation and covariance of $\boldsymbol{\alpha}^l$:

$$\hat{\boldsymbol{\alpha}}^l = \left(\frac{1}{\sigma_\epsilon^2}\boldsymbol{\kappa}_l^\intercal\boldsymbol{\kappa}_l + \mathbf{C}_\alpha^{-1}\right)^{-1}\left(\frac{1}{\sigma_\epsilon^2}\boldsymbol{\kappa}_l^\intercal\boldsymbol{y}_l + \mathbf{C}_\alpha^{-1}\boldsymbol{\mu}_\alpha\right), \tag{A.3}$$

$$\mathbf{C}_{\alpha^l} = \left(\frac{1}{\sigma_\epsilon^2}\boldsymbol{\kappa}_l^\intercal\boldsymbol{\kappa}_l + \mathbf{C}_\alpha^{-1}\right)^{-1}, \tag{A.4}$$

where $\boldsymbol{\kappa}_l^\intercal \in \mathbb{R}^{n_l \times n}$ is the base kernel $\kappa(\cdot,\cdot)$ evaluated between $\mathbf{X}_l$ and $\mathbf{X}$.

- *M-step*: This step optimizes the model parameters θ based on the estimations in *E-step*:

$$\boldsymbol{\mu}_\alpha = \frac{1}{\pi + m} \sum_l \hat{\boldsymbol{\alpha}}^l, \tag{A.5}$$

$$\mathbf{C}_\alpha = \frac{1}{\tau + m} \left\{ \pi \boldsymbol{\mu}_\alpha \boldsymbol{\mu}_\alpha^\intercal + \tau \boldsymbol{\kappa}^{-1} + \sum_l \mathbf{C}_{\alpha^l} + \sum_l \left[\hat{\boldsymbol{\alpha}}^l - \boldsymbol{\mu}_\alpha\right] \left[\hat{\boldsymbol{\alpha}}^l - \boldsymbol{\mu}_\alpha\right]^\intercal \right\}, \tag{A.6}$$

$$\sigma_\epsilon^2 = \frac{1}{\sum_l n_l} \sum_l \left\| \boldsymbol{y}_l - \boldsymbol{\kappa}_l \hat{\boldsymbol{\alpha}}^l \right\|^2 + \mathrm{tr}[\boldsymbol{\kappa}_l \mathbf{C}_{\alpha^l} \boldsymbol{\kappa}_l^\intercal]. \tag{A.7}$$

Once the model parameters θ and $\hat{\boldsymbol{\alpha}}^l$ are obtained with the EM algorithm, the predictor of an unobserved response in process l corresponding to an input parameter vector $\boldsymbol{x}_*$ can be expressed as

$$\hat{\boldsymbol{f}}_l(\boldsymbol{x}_*) = \sum_{i=1}^{n} \hat{\alpha}_i^l \kappa(\boldsymbol{x}_i, \boldsymbol{x}_*). \tag{A.8}$$

References

[1] M. A. Bezerra, R. E. Santelli, E. P. Oliveira, L. S. Villar, L. A. Escaleira, Response surface methodology (rsm) as a tool for optimization in analytical chemistry, Talanta 76 (5) (2008) 965–977.

[2] E. Lima, V. Lima, C. Almeida, K. Justi, Application of response surface methodology and machine learning combined with data simulation to metal determination of freshwater sediment, Water, Air, & Soil Pollution 228 (9) (2017) 370.

[3] M. Yolmeh, S. M. Jafari, Applications of response surface methodology in the food industry processes, Food and Bioprocess Technology 10 (3) (2017) 413–433.

[4] W. Cai, J. Wang, P. Jiang, L. Cao, G. Mi, Q. Zhou, Application of sensing techniques and artificial intelligence-based methods to laser welding real-time monitoring: A critical review of recent literature, Journal of Manufacturing Systems 57 (2020) 1–18.

[5] T. H. Kim, J. Yum, S. J. Hu, J. Spicer, J. A. Abell, Process robustness of single lap ultrasonic welding of thin, dissimilar materials, CIRP annals 60 (1) (2011) 17–20.

[6] M. K. Gupta, P. Sood, V. S. Sharma, Machining parameters optimization of titanium alloy using response surface methodology and particle swarm optimization under minimum-quantity lubrication environment, Materials and Manufacturing Processes 31 (13) (2016) 1671–1682.

[7] L. G. de Oliveira, A. P. de Paiva, P. P. Balestrassi, J. R. Ferreira, S. C. da Costa, P. H. da Silva Campos, Response surface methodology for advanced manufacturing technology optimization: theoretical fundamentals, practical guidelines, and survey literature review, The International Journal of Advanced Manufacturing Technology 104 (5-8) (2019) 1785–1837.

[8] M. I. Awad, N. M. Hassan, Joint decisions of machining process parameters setting and lot-size determination with environmental and quality cost consideration, Journal of manufacturing systems 46 (2018) 79–92.

[9] Z. Jurkovic, G. Cukor, M. Brezocnik, T. Brajkovic, A comparison of machine learning methods for cutting parameters prediction in high speed turning process, Journal of Intelligent Manufacturing 29 (8) (2018) 1683–1693.

[10] D. Zhao, D. Ren, K. Zhao, S. Pan, X. Guo, Effect of welding parameters on tensile strength of ultrasonic spot welded joints of aluminum to steel–by experimentation and artificial neural network, Journal of Manufacturing Processes 30 (2017) 63–74.

[11] M. P. Satpathy, S. B. Mishra, S. K. Sahoo, Ultrasonic spot welding of aluminum-copper dissimilar metals: A study on joint strength by experimentation and machine learning techniques, Journal of Manufacturing Processes 33 (2018) 96–110.

[12] Y. Meng, M. Rajagopal, G. Kuntumalla, R. Toro, H. Zhao, H. C. Chang, S. Sundar, S. Salapaka, N. Miljkovic, P. Ferreira, S. Sinha, C. Shao, Multi-objective optimization of peel and shear strengths in ultrasonic metal welding using machine learning-based response surface methodology, Mathematical Biosciences and Engineering 17 (6) (2020) 7411.

[13] D. Zhao, K. Zhao, D. Ren, X. Guo, Ultrasonic welding of magnesium–titanium dissimilar metals: A study on influences of welding parameters on mechanical property by experimentation and artificial neural network, Journal of Manufacturing Science and Engineering 139 (3).

[14] R. H. Myers, D. C. Montgomery, C. M. Anderson-Cook, Response surface methodology: process and product optimization using designed experiments, John Wiley & Sons, 2016.

[15] E. Ruschel, E. A. P. Santos, E. d. F. R. Loures, Industrial maintenance decision-making: A systematic literature review, Journal of Manufacturing Systems 45 (2017) 180–194.

[16] Y. Yang, Y. Zhang, Y. D. Cai, Q. Lu, S. Koric, C. Shao, Hierarchical measurement strategy for cost-effective interpolation of spatiotemporal data in manufacturing, Journal of Manufacturing Systems 53 (2019) 159–168.

[17] M. Celen, D. Djurdjanovic, Integrated maintenance and operations decision making with imperfect degradation state observations, Journal of Manufacturing Systems 55 (2020) 302–316.

[18] L. Nong, C. Shao, T. H. Kim, S. J. Hu, Improving process robustness in ultrasonic metal welding of lithium-ion batteries, Journal of Manufacturing Systems 48 (2018) 45–54.

[19] C. Shao, W. Guo, T. H. Kim, J. J. Jin, S. J. Hu, J. P. Spicer, J. A. Abell, Characterization and monitoring of tool wear in ultrasonic metal welding, in: 9th International Workshop on Microfactories, 2014, pp. 161–169.

[20] C. Shao, T. H. Kim, S. J. Hu, J. J. Jin, J. A. Abell, J. P. Spicer, Tool wear monitoring for ultrasonic metal welding of lithium-ion batteries, Journal of Manufacturing Science and Engineering 138 (5) (2016) 051005.

[21] Y. Zerehsaz, C. Shao, J. Jin, Tool wear monitoring in ultrasonic welding using high-order decomposition, Journal of Intelligent Manufacturing 30 (2) (2019) 657–669.

[22] Q. Nazir, C. Shao, Online tool condition monitoring for ultrasonic metal welding via sensor fusion and machine learning, Journal of Manufacturing Processes 62 (2021) 806–816.

[23] S. Shawn Lee, C. Shao, T. Hyung Kim, S. Jack Hu, E. Kannatey-Asibu, W. W. Cai, J. Patrick Spicer, J. A. Abell, Characterization of ultrasonic metal welding by correlating online sensor signals with weld attributes, Journal of Manufacturing Science and Engineering 136 (5).

[24] G. E. Box, K. B. Wilson, On the experimental attainment of optimum conditions, Journal of the royal statistical society: Series b (Methodological) 13 (1) (1951) 1–38.

[25] M. Sharp, R. Ak, T. Hedberg Jr, A survey of the advancing use and development of machine learning in smart manufacturing, Journal of manufacturing systems 48 (2018) 170–179.

[26] O. A. Wahab, J. Bentahar, H. Otrok, A. Mourad, Resource-aware detection and defense system against multi-type attacks in the cloud: Repeated bayesian stackelberg game, IEEE Transactions on Dependable and Secure Computing.

[27] B. Wang, S. J. Hu, L. Sun, T. Freiheit, Intelligent welding system technologies: State-of-the-art review and perspectives, Journal of Manufacturing Systems 56 (2020) 373–391.

[28] D. Nainwal, P. K. Kankar, P. K. Jain, Condition monitoring in additive manufacturing using support vector machine, in: Recent Advances in Mechanical Engineering, Springer, 2021, pp. 119–126.

[29] L. Sun, S. J. Hu, T. Freiheit, Feature-based quality classification for ultrasonic welding of carbon fiber reinforced polymer through bayesian regularized neural network, Journal of Manufacturing Systems 58 (2021) 335–347.

[30] G. Rjoub, J. Bentahar, O. A. Wahab, A. Bataineh, Deep smart scheduling: A deep learning approach for automated big data scheduling over the cloud, in: 2019 7th International Conference on Future Internet of Things and Cloud (FiCloud), IEEE, 2019, pp. 189–196.

[31] Z. He, K.-P. Tran, S. Thomassey, X. Zeng, J. Xu, C. Yi, A deep reinforcement learning based multi-criteria decision support system for optimizing textile chemical process, Computers in Industry 125 (2021) 103373.

[32] J. Leng, C. Jin, A. Vogl, H. Liu, Deep reinforcement learning for a color-batching resequencing problem, Journal of Manufacturing Systems 56 (2020) 175–187.

[33] C. Shao, J. Ren, H. Wang, J. J. Jin, S. J. Hu, Improving machined surface shape prediction by integrating multi-task learning with cutting force variation modeling, Journal of Manufacturing Science and Engineering 139 (1) (2017) 011014.

[34] T. Shireen, C. Shao, H. Wang, J. Li, X. Zhang, M. Li, Iterative multi-task learning for time-series modeling of solar panel pv outputs, Applied energy 212 (2018) 654–662.

[35] D. Wang, K. Liu, X. Zhang, H. Wang, Spatiotemporal multitask learning for 3-d dynamic field modeling, IEEE Transactions on Automation Science and Engineering.

[36] H. Chen, Y. Yang, C. Shao, Multi-task learning for data-efficient spatiotemporal modeling of tool surface progression in ultrasonic metal welding, Journal of Manufacturing Systems 58 (2021) 306–315.

[37] M. Ramezankhani, B. Crawford, A. Narayan, H. Voggenreiter, R. Seethaler, A. S. Milani, Making costly manufacturing smart with transfer learning under limited data: A case study on composites autoclave processing, Journal of Manufacturing Systems 59 (2021) 345–354.

[38] G. Matheron, Le krigeage universel, Vol. 1, École nationale supérieure des mines de Paris Paris, 1969.

[39] M. Knotters, D. Brus, J. O. Voshaar, A comparison of kriging, co-kriging and kriging combined with regression for spatial interpolation of horizon depth with censored observations, Geoderma 67 (3-4) (1995) 227–246.

[40] T. Hengl, G. B. Heuvelink, A. Stein, A generic framework for spatial prediction of soil variables based on regression-kriging, Geoderma 120 (1-2) (2004) 75–93.

[41] T. Hengl, G. B. Heuvelink, D. G. Rossiter, About regression-kriging: From equations to case studies, Computers & geosciences 33 (10) (2007) 1301–1315.

[42] G. Hudson, H. Wackernagel, Mapping temperature using kriging with external drift: theory and an example from scotland, International journal of Climatology 14 (1) (1994) 77–91.

[43] L. Zhang, K. Wang, N. Chen, Monitoring wafers' geometric quality using an additive gaussian process model, IIE Transactions 48 (1) (2016) 1–15.

[44] S. Aye, P. Heyns, An integrated gaussian process regression for prediction of remaining useful life of slow speed bearings based on acoustic emission, Mechanical Systems and Signal Processing 84 (2017) 485–498.

[45] J. D. Opsomer, D. Ruppert, M. P. Wand, U. Holst, O. Hössjer, Kriging with nonparametric variance function estimation, Biometrics 55 (3) (1999) 704–710.

[46] W. Guo, C. Shao, T. H. Kim, S. J. Hu, J. J. Jin, J. P. Spicer, H. Wang, Online process monitoring with near-zero misdetection for ultrasonic welding of lithium-ion batteries: An integration of univariate and multivariate methods, Journal of Manufacturing Systems 38 (2016) 141–150.

[47] C. Shao, K. Paynabar, T. H. Kim, J. J. Jin, S. J. Hu, J. P. Spicer, H. Wang, J. A. Abell, Feature selection for manufacturing process monitoring using cross-validation, Journal of Manufacturing Systems 32 (4) (2013) 550–555.

[48] P. K. Kitanidis, Generalized covariance functions in estimation, Mathematical Geology 25 (5) (1993) 525–540.

[49] K. Yu, V. Tresp, A. Schwaighofer, Learning gaussian processes from multiple tasks, in: Proceedings of the 22nd international conference on Machine learning, 2005, pp. 1012–1019.