Proceedings of the ASME 2017 International Design Engineering Technical Conferences & Computers and Information in Engineering Conference
IDETC/CIE 2017
August 6-9, 2017, Cleveland, Ohio, USA

DETC2017-67807

A DOMAIN-DRIVEN APPROACH TO METAMODELING IN ADDITIVE MANUFACTURING

Zhuo Yang, Thomas Hagedorn, Douglas Eddy, Sundar Krishnamurty, Ian Grosse
University of Massachusetts Amherst
Department of Mechanical and Industrial Engineering
Amherst, MA 01003
Email: zhuoyang@engin.umass.edu, thagedorn@engin.umass.edu, dceddy@engin.umass.edu, [skrishna, grosse]@ecs.umass.edu

Peter Denno, Yan Lu, Paul Witherell
National Institute of Standards and Technology
Engineering Laboratory
Gaithersburg, MD 20899
Email: [peter.denno, yan.lu, paul.witherell]@nist.gov

ABSTRACT

Recent studies have shown advantages to utilizing metamodeling techniques to mimic, analyze, and optimize system input-output relationships in Additive Manufacturing (AM). This paper addresses a key challenge in applying such metamodeling methods, namely the selection of the most appropriate metamodel. This challenge is addressed with domain-specific AM information, derived from physics, heuristics and prior knowledge of the process. Domain-specific input/output models and their interrelationships are studied as a basis for a domain-driven metamodeling approach in AM. A metamodel selection process is introduced that evaluates global and local modeling performances, with different AM datasets, for three types of surrogate metamodels (polynomial regression (PR), Kriging, and artificial neural network (ANN)). A salient feature of this approach is its ability to seamlessly integrate domain-specific information in the model selection process. The approach is demonstrated with the aid of a metal powder bed fusion (PBF) case study and the results are discussed.

1. INTRODUCTION

AM techniques have promising applications in such different domains as aerospace, medical devices, and heavy industry [1]. However, these often multi-physical processes are still not fully understood or controlled [2]. For example, in the powder bed fusion (PBF) process, the relative density of AM parts is affected by various user controllable parameters such as laser power, scan speed, or powder density[3]. In some instances, the hardness of AM parts might be different despite being produced by same machine using the same input parameters [4]. The difficult-to-predict performances of AM processes introduce significant uncertainty into quality control and engineering design [5]. To mitigate this uncertainty, researchers are seeking mathematical predictive models that can predict material properties prior to production by the manufacturing process. Pure physics-based models [Pal et al, 2013], numerical simulation models [6], and metamodels [7] have been proposed to predict AM behavior. This paper discusses a novel method that leverages domain knowledge to improve the selection of metamodels for AM predictive modeling problems.

Metamodels, also known as surrogate models, construct a model of a model to understand complex systems [8]. Unlike physics-based or numerical simulation models that often require detailed knowledge of internal processes such as problem physics [5], metamodeling techniques focus on the input/output relationships[9]. Metamodels can significantly reduce the cost of organizing knowledge for a poorly understood system. However, these models can introduce modeling uncertainties due to lack of representative process knowledge [10], a hindrance our domain-driven technique seeks to address. Indeed, some problem-specific knowledge is desired to select an appropriate

sampling strategy and account for system behaviors and sensitivity properties [11, 12], which can fundamentally affect model performance.

The term model performance in this paper refers to predictability and efficiency that is believed to be correlated to data sets such as sample size, data gradient, and kurtosis [13]. Previous modeling selection frameworks have used information about data, called data features, to identify the optimal modeling algorithm for certain types of problem. "Data features" in this context are considered characterizations of how each parameter impacts the responses in a given domain. For example, Rice et al. proposed a model can use extracted characteristics and performance measurements such as normalized root mean square error (NRMSE) [13] and maximum relative error magnitude (MREM) [14] from a given dataset to select the optimal modeling algorithm from a set of candidates [15]. Cui et al.'s energy model recommendation framework uses dual performance evaluation criteria and criteria reduction methods to implement a meta-learning procedure for modeling algorithm selection [13]. While these well-established methods work for some general problems or problems in specific domains, they usually require significant work in data feature characterization.

This paper aims to address AM predictive modeling challenges by constructing a specific-to-AM framework that looks to efficiently and accurately identify optimal metamodeling methods for given problems, prior to deeply investigating data features related to a specific domain. The term "domain" in this paper indicates the topic area to which the parameters apply. For example, laser power is a parameter in the AM thermal domain and powder density resides in the AM material domain, etc. In the following sections, we will first discuss the performance of metamodels with different domain-inspired AM input/output parameters. Unlike Rice's model that uses data features for algorithm recommendation, the method proposed in this paper focuses on using correlations between AM parameters to efficiently identify data features. An AM input/output correlation chart was developed to visually present the nonlinearity of different combinations of parameters. We introduce the domain-driven framework in Section 5. For demonstrative purpose, a case study based on AM datasets is presented in Section 6. The benefits of using this domain-driven approach are explained and further discussed in these sections.

Section 2 provides a brief overview of predictive models in AM. Section 3 discusses parametric correlations in AM models based on a detailed literature review. The remaining sections are built upon the findings from Section 3. Section 4 provides a brief background of metamodeling algorithms. The three candidates presented in this paper are used to demonstrate the framework. The fundamental structure of the framework and an illustrative case study with empirical data from an AM process are introduced in Sections 5 and 6. Section 7 concludes with a comprehensive discussion of the early framework.

2. OVERVIEW OF AM PREDICTIVE MODELING

Porosity, relative density, surface roughness, geometric accuracy, hardness, and tensile strength are examples of critical properties that define AM product quality [1]. While important to physics based models, values of intermediate parameters such as melt pool width, penetration depth, and melting temperature cannot represent critical AM properties directly [16]. Advancement in AM processes and product design quality requires a clear understanding of the relationship between various AM input and output parameters.

Several studies have explored the effect of AM process parameters on part performance [17]. Witherell et al (2014) analyzed the metal PBF process and divided it into four critical categories that summarize the complex inter-relationships between AM parameters [7]. In this approach, the PBF process is modeled as a set of sub-processes, including heat source/absorption model, a melt pool formation model and a solidification model, where each physical process also has multiple sub-processes. For example, Marangoni, capillary, and heat convection/conduction models are sub-sets of a larger melt pool model [18].

Other studies focus on the details of sub-processes such as heat source/absorption and melt pool models. For instance, the isotherm migration method uses theoretical analysis to estimate the surface temperature of the powder bed, which can then be used to study phase changing during 3D printing [19]. Finite element analysis (FEA) can also simulate melt pool information [20, 21]. Ma et al (2015)'s 3D FEA model of melt pool for example used user controllable inputs (laser power, scan speed, powder density, etc.) to predict melt pool width and depth [22]. These studies contribute to improved AM process knowledge, but they do not completely characterize the full set of physical phenomena in an AM processes. As a result, there is interest in using robust and efficient design of experiments (DOEs) and metamodeling approaches to study AM problems.

Metamodeling methods use a statistical approach that treats a complex system as a black box to avoid limitations stemming from lack of knowledge during model construction [8]. The data that is used to construct metamodels can be collected from computer simulations or actual measurements from DOE [23]. In general, one can improve the metamodel by adapting sampling strategies that are well suited to a specific problem [Shao et al, 2008], optimizing modeling parameters [24], and selecting a more appropriate modeling method [15].

Optimal sampling can potentially inform a metamodel with information about the unknown system [25]. For example, Shao and Krishnamurty's model updating method selects sample points that are closer to local optima to improve the predictive ability of a Kriging model [26]. Selection of modeling parameters can also affect the model performance within the same given sample set [27] [Yang et al, 2017]. Both methods assume that a pre-selected metamodeling algorithm is appropriate, which is not necessarily true. For example, it can be expected that a linear model would have difficulty producing accurate predictions for highly nonlinear problems regardless of sampling technique or model parameters. Model selection usually requires significant work to characterize potentially significant data features in a given dataset [13] or uses a complete, multi-stage adaptive sampling process [26].

A rapid model selection procedure that can potentially skip investigation processes, but still provides reliable solutions, is highly desirable. In this paper, an AM metamodeling algorithm selection framework is proposed to accomplish this goal. This paper builds on our recent efforts in metamodeling for additive manufacturing, focusing on efficiency, effectiveness and optimal sampling [14]. Specifically, this work provides a basis for domain-driven enhanced metamodeling by explicating and exploiting the unknown and complicated correlations in the system behavior using a priori knowledge of the domain being studied.

3. ANALYSIS OF THE CORRELATION BETWEEN INPUT/OUTPUT OF PBF PROCESS

This paper investigates a metamodel selection method that leverages pre-existing knowledge of parametric relations instead of pure data analysis. For an identical system under the same conditions, one might perceive that the basic relation between inputs and outputs is unchanged. The following beam bending example is introduced to explain this hypothesis.

In the general case of any beam, when an analysis of beam bending is desired many of its characteristics may be unknown. For example, consider a large beam made of lattice of different orders of magnitude (Figure 1), with a load applied. Measurements of loading conditions can be taken, but not enough information can be obtained to perform a complete analysis due to the presence of too many variables.

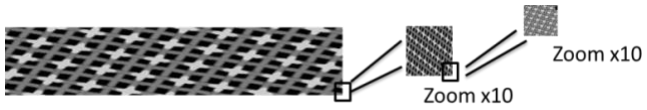

Figure 1. Beam created with lattice.

However, given the loading and constraints, we can observe that the case behaves like a bending beam. While only partial measurements can be taken at micro and meso scales, we can make the hypothesis that it will perform according to beam theory at the meso and macro levels, thus making domain-specific observations. These observations allow us to extrapolate measured values, based on a combination of measured results and expected macro performance.

A similar situation is found in metal PBF processes in that they are very complex and involve a large number of parameters (more than 50) [1, 16]. The input/output relation is difficult to discern using theoretical analysis. If one considers every single PBF parameter, there are millions of combinations of different models. Thus, section 5 will introduce a domain-driven framework that uses past model performance to predict the appropriate modeling method for a new problem. The rest of this section will summarize parametric relations between AM parameters from the literature to construct the prerequisite knowledge for development of a model selection framework.

Figure 2 shows 8 inputs and 6 outputs that have high occurrence in recent PBF literature. The relation between these parameters is currently marked as unknown due to lack of information.

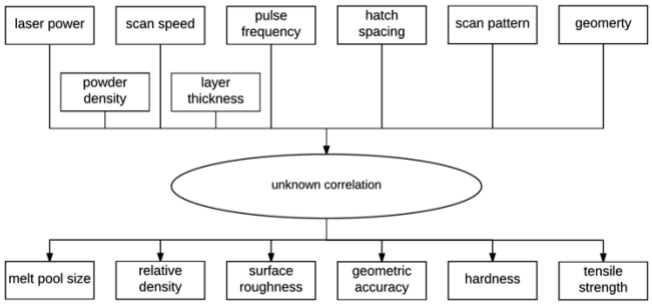

Figure 2. Infrastructure of principal AM inputs/outputs

For single input cases, PBF parameters such as laser power, scan speed, and hatch spacing have highly linear correlations to certain outputs [29, 30]. For example, Tang et al (2003)'s metal laser sintering experiment indicates the surface roughness and tensile strength is linearly increased with laser power [31]. Similarly, tensile strength decreases monotonically with higher scan speed with other parameters held constant. From the same study, however, surface roughness was not linearly related to scan speed. With reduced layer thickness, scan speed and roughness instead had a slight nonlinear relation. Another study, using different materials but the same laser parameters, found scan speed and layer thickness have a linear relation to relative density when varied individually [32]. Such similar results using multiple materials suggests that relative density is linearly related to laser power, scan speed, and layer thickness [33]. Intermediate outputs such as penetration depth (not included in Figure 2) have also been shown to exhibit a linear relation with some parameters. Kruth et al (2003)'s laser sintering experiment found higher scan speeds produced a linearly decreasing layer thickness [34]. Some inputs, such as laser pulse frequency, have been found to have a highly nonlinear relation to relative density [35].

Compared to the single input/output problem, the relations between variables become considerably more complicated when studying multiple input parameters. Tang et al (2003) found that the surface roughness is not linearly related to a combination of laser power and hatch spacing [31]. Similarly, when considering the relation of laser power and layer thickness to surface hardness, the relation is also nonlinear [32]. A similar, nonlinear relation is also observed in Morgan et al (2004)'s empirical result [35]. The simple linear relation between scan speed and relative density becomes significantly more complicated when pulse density is also varied. This evidence seems to imply that more variables generate larger uncertainties in PBF due to an increase in unknown interactions. However, other outputs such as the tensile strength are not that sensitive to the same combinations of inputs. The observed relation remains linear under a combination of factors. Thus, more inputs can increase PBF problem complexity but do not necessarily indicate increasing nonlinearity of the input/output relation.

This paper focuses on the development of a general algorithm recommendation framework based on input/output parameters before deeply investigating physical interactions between PBF parameters. As such, it is necessary to understand the general factors that cause the uncertainty in PBF processes. Beaman et al (1997) first introduced the concept of energy density for AM, which is described by the Equation (1) [36]:

$$E_\rho = \frac{4P}{\pi r^2}\frac{2r}{v}\frac{2r}{s} \quad (1)$$

where E_ρ is Energy density, P is laser power, r is beam radius, v is scan speed, and s is hatch spacing.

Equation (1) indicates that higher laser power, lower scan speed, and closer hatch spacing produce higher energy density. More energy delivered to the powder usually means better melting conditions. Improved melting conditions will result in lower porosity and thus higher relative density. For example, Meier et al (2008)'s experiment with the metal laser sintering process shows the relative density increases from 69% to 99% with a power increase from 30W to 90W with other parameters held constant [37]. Another study [38] concluded higher energy density tends to produce a continuous melting track against irregular melt shape. These findings imply the linear relations may be more likely if the involved input/output parameters can be related to an overall energy density dependency.

In contrast, nonlinear relations were found in studies on outcomes related to part microstructure. Meier et al (2008) found surface roughness is not monotonically increased with scan speed, with the optimal roughness obtained in the middle of the range of scanning speeds tested [37]. Other research suggests that microstructure varies throughout the entire PBF part, and thus can be considered a local rather part-wide property. Wang et al (2012) found the hardness tests at different locations/directions in the same AM part produce different results [4]. Similarly, studies of thermal conditions indicate that variation of thermo-physical properties of AM parts are complicated [39].

Figure 3 summarizes the hypothetical relation of input/output correlations observed in the literature. The thick arrow in the middle of the figure represents the relation from linear to highly nonlinear. The arrow (right to left) on the top represents whether the input parameters can be classified as related to energy density or not. The bottom arrow (left to right) represents whether the outputs are in macro-scale or micro-scale. The observations performed in this section pertain to linearity of input/output relationships. Figure 4 illustrates the more general case to which these techniques may be applied.

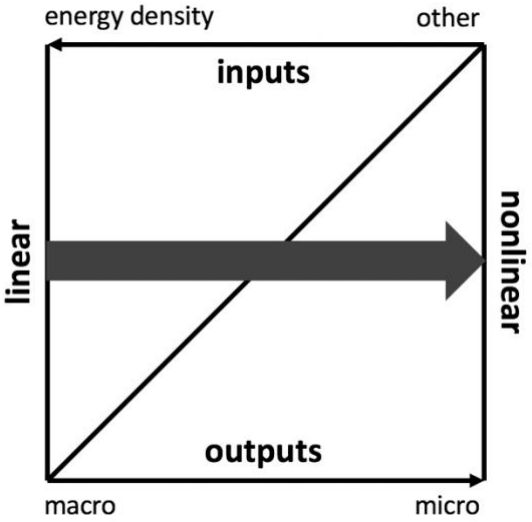

Figure 3. Hypothetical relation of input/output correlations

Figure 4.General case to model input/output relationships

Figure 3 can be used to summarize past literature results for the particular case of PBF. As discussed, laser power, layer thickness, and scan speed are input parameters that relate to energy density. Pulse frequency is located in the upper right corner since it is unrelated to energy density. For outputs, relative density is considered a high macro scale property as it depends on part width rather than local porosity. Surface roughness, however, relates more to AM microstructure. For the problem that involves the parameters in Figure 5(a), the link between inputs and outputs intersects with the bold arrow on the left. Figure 5(b) indicates pulse frequency and surface roughness have a highly nonlinear relation. Figure 5(c) and Figure 5(d) demonstrate the limitations in past research.

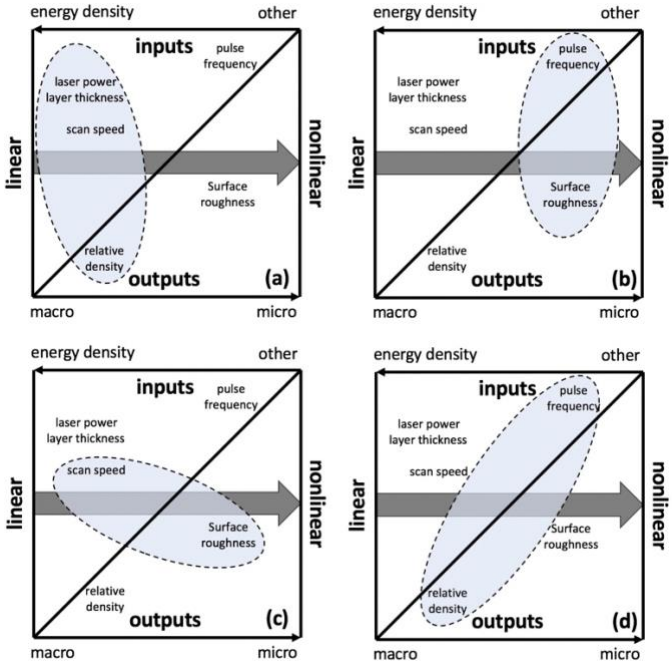

Figure 5. The use of AM input/output correlation chart

Figure 5 visually summarizes the relation between different combinations of PBF parameters observed in the literature. However, it must be stressed that, like the literature, it only summarizes some of the parameters of interest, and might not be sufficient to guide metamodel selection, especially in the case of indeterminate parameter sets. A more rigorous mathematical solution is thus needed. Section 5 introduces a framework to recommend models based on mathematical computation.

4. OVERVIEW OF METAMODELING TECHNIQUES

This section reviews the metamodeling techniques that are applied in Section 5. PR, Kriging, and ANN techniques were used in this study as they cover both parametric and non-parametric techniques [13]. However, the framework should not be limited to only these algorithms. These three were selected in this paper for demonstrative purposes. In general, any metamodeling technique can be considered as a candidate during the actual practice.

A typical metamodel [40] can be expressed as:

$$y(\tilde{x}) = f(\tilde{x}) + \varepsilon \quad (2)$$

where $y(\tilde{x})$ represents an unknown function, $f(\tilde{x})$ is a known function of $\tilde{x}$ derived statistically, and ε is the error part. $\tilde{x}$ represents the set of the system's independent input variables. For different modeling methods, the function represented by each part of the expression is different.

4.1 Polynomial Regression (PR)

The PR technique is a variation of linear regression in which an n^{th} order polynomial is used to model the relationship between the independent variables $\tilde{x}$ and the outcomes $y(\tilde{x})$ [40]. The method is popular in various engineering domains since it is fast and easy to use. A second order quadratic polynomial function has the form of:

$$\hat{y} = \beta_0 + \sum_{i=1}^{k} \beta_i x_i + \sum_{i=1}^{k} \beta_{ii} x_i^2 + \sum_i \sum_j \beta_{ij} x_i x_j \quad (3)$$

where β_0, β_i, and β_{ij} are regression coefficients, and k is the number of design variables.

4.2 Kriging

Unlike the parametric techniques producing an actual formula, the Kriging method as a non-parametric method builds its estimation based on the position of sample data. The underlying assumption of Kriging models is that an unknown point can be estimated from observed (known) points based on spatial correlation [41]. The estimation process is completed by a variogram or so called spatial correlation functions [42, 43]. The general form of a kriging estimation for an unknown predicted value of a point Z_E for a single outcome is [44]:

$$Z_E = \bar{Z} + \sum_{i=1}^{n} \lambda_i (Z_i - \bar{Z}) \quad (4)$$

where $\bar{Z}$ represents the regional mean value of the response and. λ_i is the distance-correlated weight value, which is determined by the computation of spatial correlation. The value of spatial correlation can be derived from:

$$R(\theta, x_i, x_j) = \prod_{l=1}^{n} \exp(-\theta_l (x_{i,l} - x_{j,l})^2) \quad (5)$$

where $x_{i,l}$ is the lth component of the ith vector x_i [44]. $R(\theta, x_i, x_j)$ depends on the location of points x_i and x_j, and the correlation parameter, θ. Multiple Kriging methods exist, including simple Kriging, ordinary Kriging, regression Kriging, etc. [43]. The ordinary Kriging method is used in this paper.

4.3 Artificial Neural Network (ANN)

ANN is a computational algorithm that mimics the central nervous system [45] and has been widely used for solving problems with complicated structures. A typical ANN model consists of an input layer, hidden layers, and an output layer [46]. Each layer consists of "neurons" that are connected across layers to transmit and deduce information. The optimal number of neurons and hidden layers may differ, and depends on the complexity of the problem. The structure of a simple ANN model [47] is shown in Figure 6, where x1 to x3 are input parameters, u1-4 are the neurons in the single hidden layer, and outputs y1 and y2 are located in output layer.

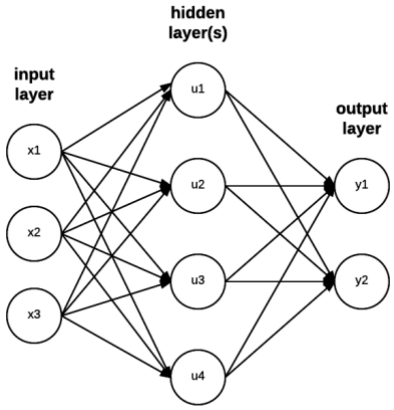

Figure 6. Typical structure of a simple ANN model

5. DOMAIN-DRIVEN MODEL RECOMMENDATION FRAMEWORK FOR AM

An exhaustive search, which is also known as the generate and test method, is the most general problem solving technique for systematically enumerating all possible candidate algorithms and selecting the most appropriate candidate based on a set of criteria [48]. While it is a global optima algorithm, it is also extremely inefficient, especially for those problems with abundant candidates and/or large input datasets. In such cases, it may be more efficient to incorporate prior knowledge into the algorithm selection process.

Many selection or recommendation techniques were developed to improve the efficiency of exhaustive search. Rice's model [15], for example, can recommend the best candidate for a new instance based on previous model selection knowledge. It includes four spaces: the problem space P represents the datasets of learning instances; domain space F contains the characteristics; algorithm space A includes all candidate algorithms; performance space Y is the measured performance of instance P for each algorithm in A [15]. Rice's model compares characteristics of a new instance to all previous examples and then assesses the suitability of each algorithm based on a set of rules or a selection algorithm. The model findings can be used to select the optimal algorithm from a given problem. Once the solution is derived, the performance in the new instance is added to the performance space Y, updating the model with a new point. In this way, a user can avoid exhaustively testing each candidate algorithm for a new instance [13].

The proposed domain-driven method is built upon Rice's algorithm selection method. However, instead of using data-features to characterize the new dataset, this proposed AM framework uses AM knowledge to indicate a possible optimal option from candidate algorithms. This approach is fundamentally different than Rice's method in that the result can be independent of the unfixed data strategy and rely on the relatively fixed knowledge of the physics of the problem. The AM characteristics used are the relations between input/output parameters discussed in section 3. The general workflow of the proposed framework is shown Figure 7. It requires sufficient knowledge to commence the selection process. Knowledge construction consists of collecting existing datasets, classifying the instances, and computing the performance of each candidate algorithm on each dataset. The extracted information is then fed back into the current knowledge model and the system predicts a possible optimal metamodeling algorithm. Before proceeding to model construction with actual data, the newly calculated solution updates the knowledge model. Details of each critical step are discussed in the rest of this section.

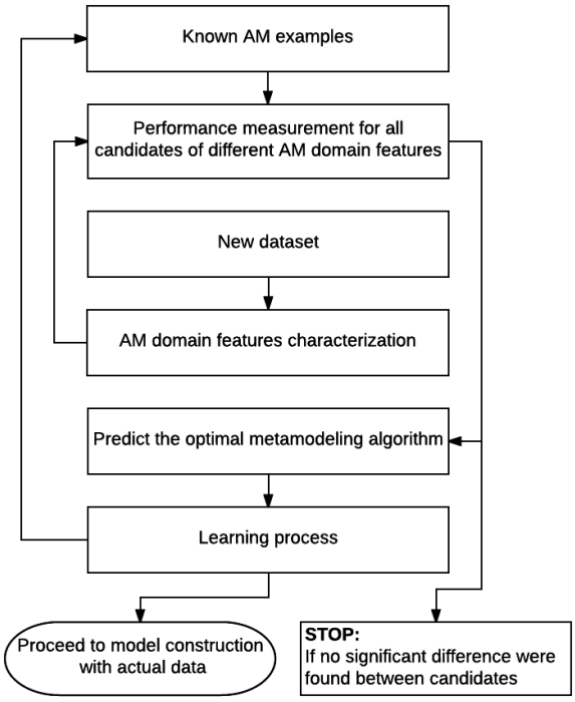

Figure 7. General workflow of the proposed framework of AM domain-driven modeling selection method

5.1 AM characterization

The AM characteristics mentioned at this stage are AM input/output parameters. At this step, all parameters are formed to input vector X of the selection framework:

$$\mathrm{X} = [x_1\ x_2\ x_3 \ldots\ x_n\ y_1\ y_2\ y_3 \ldots\ y_m]^T$$

where the x and y are the inputs and outputs respectively and are equal to 1 or 0. 1 indicates that the problem includes the parameter and 0 indicates the parameter is not considered. These vectors are the inputs to the learning process. For problems that have exactly the same outputs, y can be ignored.

5.2 Performance measurement

Measuring model performance of known datasets is critical to improving model selection accuracy. Two criteria were employed in the case study in section 6 to evaluate the modeling performance by a set of candidate algorithms. For global measurement, normalized root mean square error (NRMSE) [13] is used. Maximum relative error magnitude (MREM), on the

other hand, is used to evaluate the outstanding error of the models [14]. These were formulated as:

$$NRMSE = \sqrt{\frac{\sum_{i=1}^{N}(y_i-\tilde{y}_i)^2}{N}}/(y_{max}-y_{min}) \quad (6)$$

$$MREM = \max(\frac{|y_i-\tilde{y}_i|}{y_i}) \quad (7)$$

Where y_i ($\neq 0$) is the actual observed value, $\tilde{y}_i$ is the estimated value from the metamodel, y_{max}/y_{min} are the maximum/minimum actual observation, and N is the total amount of validation samples. With NRMSE and MREM, the framework can make its recommendation based on both global and local performances of the datasets by assigning appropriate weights to each criteria. In this paper, all case studies consider the NRMSE and MREM criteria equally to not bias either way. However, in some cases, these two objectives could conflict. Under these circumstances, a user could deploy a weighted multi-criteria decision making formulation.

5.3 Prediction process

The prediction process of the proposed recommendation framework could be completed by either model-based or instance-based methods. Model-based methods build predictive models to determine the optimal modeling algorithm. The predictive selection is based on an input vector X (model variables) and the resulting modeling performance (model outcome). Once the model is built, the new instance with PBF-related information in X_{new} would then import to the model and calculate the predictive result. The model-based method is similar to what is discussed in Section 3. For example, the vector X is the input variable set of the recommended predictive model. The values of NRMSE and MREM then become the predictive results. Once the recommended model is built based on existing instances, the model can predict the NRMSE and MREM of candidate algorithms for a new problem. The user can decide which algorithm would be employed according to these indicators of model performance.

An instance-based method by comparison solves the problem based on existing examples. It assumes an algorithm has similar performance on similar problems, where the similarity is measured by Euclidean distance between instance input and output vectors [Brighton et al, 2002]. The k-nearest neighbor (k-NN) ranking approach as was employed in this study. The k-NN approach ranks the nearby k nearest examples for their similarity. However, the simplest case of k-NN is the closest neighbor example based on the comparison of the Euclidean distance of all examples, which also called as 1-NN. The formulation of 1-NN is:

$$dist(i,j) = \sqrt{(a_i - a_j)^2},\ j = 1, \dots, m \quad (8)$$

Where a_i represents the new instance a_j represents the existing examples, and m is the total number of examples. In the case of the metamodeling algorithm, a set of datasets composed of input and output parameters would serve as the existing points, each of which has been characterized by a set of metamodels. By comparing the input parameters (a_j) to those used in the new dataset, the user can then determine how similar the data is to a known dataset. At this point, users would simply compare the performance of various modeling algorithms in the existing examples. Thus, a likely best predictive modeling algorithm can be chosen without costly characterization of information and data features of the dataset. This saves the cost of testing all candidate algorithms individually, allowing the user to directly proceed to model construction and parameters optimization. A demonstrative example in the following section shows how the method works.

6. DEMONSTRATIVE CASE STUDY

A simple example was constructed from existing AM datasets (2 for constructing knowledge and 1 for verification) to illustrate the proposed AM domain-driven framework. Tang et al (2003)'s and Morgan et al (2004)'s metal PBF datasets were used to construct the knowledge base [31, 35]. The dataset from Chatterjee et al (2003)'s was selected to verify the selection accuracy [49]. The three datasets have the same output parameter, relative density. Because of this, the Y output vector is omitted, and the results obtained may potentially be more accurate since the knowledge is constructed from somewhat similar examples.

The 1-NN method was used to predict the optimal modeling algorithm. The predicted algorithm was then compared to actual modeling results with all candidate algorithms to assess the predictive solution [49].

6.1 Knowledge construction

The knowledge used for model predictions was composed of a small dataset with 15 samples [31] and a large dataset with 105 samples [35]. They were selected to build the knowledge model because their similarities: 1) both are metal PBF process; 2) both use similar experimental conditions; 3) they have the same output as the new dataset (relative density). The differences between them also provide opportunities for future model selection for new instances, namely: 1) they use different input parameters; 2) they use a different DOE strategy and 3) both have different variables that are not considered in DOE such as materials and specific machines. Thus, the knowledge base of these datasets is reasonable and has useful variation.

For initial construction, we consider 6 independent input variables though neither dataset can cover the parameter of layer thickness. The two datasets overlap in individual parameters. The matrix of the inputs is shown in Table 1. Note, the order of input parameters that are in the table and elements in the vector are constantly fixed for a future prediction process.

Table 1. The inputs matrix of given datasets

	Laser power	Scan speed	Powder density	Layer thickness	Pulse frequency	Hatch spacing
Tang (2003)	1	1	1	0	0	1
Morgan (2004)	0	1	0	0	1	1

Writing these inputs as input vecors:

$$X_{tang} = [1\ 1\ 1\ 0\ 0\ 1]^T$$
$$X_{morgan} = [0\ 1\ 0\ 0\ 1\ 1]^T$$

The output vector Y is omitted in this case study since both knowledge and verification datasets have the same target output – relative density.

Three algorithms were used to characterize both input datasets. A PR model was built using the pure quadratic regression method. The Kriging model is built by ordinary Kriging method and the Gaussian correlation function with maximum likelihood approach. The ANN model is defined with 10 hidden layers. It should be noted that the candidate algorithms used in this case study are not meant to be exhaustive, but rather to represent a set of common modeling approaches. To calculate the performance, each original dataset is divided into training and testing sets with fixed ratio 80% and 20% using the Latin Hypercube based Minimum Euclidean Distance method [14]. For Tang et al (2003)'s dataset, PR works the best from three candidates as both NRSME (0.1580) and MREM (0.0220) are the lowest (Table 2). However, in the second dataset the Kriging model tested was found to be the best possible choice among the candidate algorithms. Thus, at a system level the knowledge base indicates: 1) while X=[1 1 1 0 0 1]^T, recommend model=PR; 2) while X=[0 1 0 0 1 1]^T, recommend model=Kriging. The future model selection process is built based on these rules.

Table 2. Model performance of both datasets for three candidates

	NRSME			MREM		
	PR	**Kriging**	**ANN**	**PR**	**Kriging**	**ANN**
Tang (2003)	0.1580	0.1757	0.5603	0.0220	0.0230	0.1642
Morgan (2004)	0.2018	0.1332	0.3917	0.1055	0.0669	0.1866

6.2 Modeling algorithm recommendation

The verification dataset consisted of 13 samples manufactured using the metal PBF process [49]. Compared to the datasets in the knowledge model, the experiment used carbon steel powder instead of stainless steel or a copper alloy, and has the smallest sample size (13) and number of input variables (2). The two input variables were layer thickness and hatch spacing, resulting in an input vector of:

$$X_{new} = [0\ 0\ 0\ 1\ 0\ 1]^T$$

Based on 1-NN approach, the Euclidean distance between the new and former datasets are:

$$\text{dist}(\text{new}, \text{tang}) = 2$$
$$\text{dist}(\text{new}, \text{morgan}) = 1.732$$

The knowledge model at this stage is likely insufficient due to a very limited number of example instances. Though it has these defects, the distance results show that the dataset is closer to Morgan's data than Tang's. Thus, the recommended algorithm would be a Kriging model. Once confirmed, the result can be used to update the current knowledge model with a new instance – while X=[0 0 0 1 0 1]^T, and Y is relative density, the optimal candidate model is Kriging. Once this is done, the updated knowledge model was updated and can cover the aspect of layer thickness.

For verification, the performance of each candidate model with the new dataset is shown in Table 3. All models were constructed using the same methods and model parameters as in the knowledge model datasets. Based on the performance measurement, Kriging model shows small advantages in both NRSME and MREM compared to the PR and ANN models. Thus, the result is consistent with the solution predicted using the framework.

Table 3. Model performance for the new dataset

	NRSME			MREM		
	PR	**Kriging**	**ANN**	**PR**	**Kriging**	**ANN**
Chatterjee (2003)	0.3374	0.3186	1.0448	0.0282	0.0257	0.0681

7. DISCUSSION AND FUTURE WORK

The proposed AM domain-driven metamodeling recommendation framework has the potential to provide an efficient and reliable way to predict the optimal metamodel for a new problem. It is efficient as it can avoid exhaustively testing all possible candidate algorithms once a sufficient knowledge model is constructed. Moreover, the solution can help to direct future model construction when considering data-features that might allow the user to hone in from a broad class of algorithms to a specific one. The general framework was established based on the hypothesis that certain combinations of input/output parameters have consistent behavior. The predictive solution could be made more reliable if it were derived from a larger set of consolidated knowledge. A simple demonstrative case study that included three distinct metal PBF datasets shows the algorithm prediction process.

Though the proposed framework shows a multiple of advantages in AM metamodeling problems, the details of the method need further improvement. The current set of candidate algorithms is limited; including only PR, Kriging, and ANN. While suitable for demonstration, this limited size of candidates potentially restrains higher model performance of new datasets. Furthermore, each model only has the basic modeling configuration without the ability for user modification. It may cause false results due to incomplete consideration of modeling options. For example, the finding that the ordinary Kriging model works better than a pure quadratic regression model does not mean that it also works better than a higher order PR model. Without consideration of the range of available algorithm types, the current framework may mislead the user. To overcome these disadvantages, more detailed metamodeling techniques should be added to the current framework. This work is being undertaken.

Beyond adding more broad classes of algorithms, subclasses of algorithms also need to be considered for a more robust solution. For example, consideration of different Kriging methods might enrich the study. Simple Kriging, stochastic Kriging, and dynamic Kriging may further define the Kriging class in the set of candidates. In addition to adding more

algorithms, it may be useful to bring modeling guidelines into the framework. For example, such guidelines might indicate that ANN may not be well suited for use with small datasets. Such considerations may improve the predictive accuracy of the framework for a larger breadth of datasets.

For the specific case of AM and the PBF process, algorithms for the AM characterization process also need further improvement. There are more than 50 independent variables in metal PBF process [1, 16]. This study has included less than 1/3 of them. Another disadvantage is that the framework can only count categorical input/output vectors, rather than considering broad classes of inputs and the relative similarity between different types of variables. For example, in the review of the literature, variables relating to energy density were found to behave very similarly within a range of outputs. This is knowledge that might improve the model selection process. If included in the knowledge model, the system could possess greater insight when calculating the distance between instances. More robust parameter classification may thus be needed for more accurate prediction. Similarly, as research continues, the vectors can be further detailed and classified in multiple levels based on process knowledge and empirical data. For example, materials could be classified as 0 (single component), 0.5 (multiple components without steel), or 1 (multiple components with steel), or using some other scheme to provide greater insight into problem similarity. However, development of reasonable methods requires a more comprehensive understanding of AM processes.

All of this suggests the need for a hybrid approach that utilizes a combination of process-specific knowledge and experience, algorithmic knowledge and dataset-specific considerations. The process knowledge might consist of empirical input/output relations as in this paper, utilize knowledge of problem physics to assess the similarity of datasets and suggest several candidate classes of algorithm. Algorithmic knowledge might consist of a well defined model of broad algorithm classes and subclasses, and defined model and data attributes that affect their performance. Simple data features such as sample size and the utilized DOE methods could then be used.

The most important challenge in our research currently is a lack of data to construct a more reliable knowledge model. In the case study, the naive knowledge model consists of only two instances. To improve the model, more AM knowledge is needed. In the current knowledge model, only empirical datasets can be used to build the knowledge model. Simulation models might be used to enhance the knowledge model. For example, Ma et al (2015)'s FEA model has 10 independent AM input variables, which may allow the framework to include a broader range of problem physics [22]. Formal information models may also contribute to better knowledge construction. More candidate metamodeling algorithms should be considered into such information models to not limit the overall performance. Recent development of an AM ontology might provide the basis for more effective utilization of process specific knowledge. If this information can be utilized in a future version of the proposed framework, it can potentially boost its predictive ability and accuracy.

ACKNOWLEGEMENT

This material is based upon work supported by the National Science Foundation (NSF) under Grant No. 1439683, the National Institute of Standards and Technology (NIST) under Cooperative Agreement number NIST 70NANB15H320, and industry members of the NSF Center for e-Design.

REFERENCES

[1] Gibson, I., Rosen, D.W., and Stucker, B., 2010, "Additive manufacturing technologies," Springer, .

[2] Tapia, G., and Elwany, A., 2014, "A Review on Process Monitoring and Control in Metal-Based Additive Manufacturing," Journal of Manufacturing Science and Engineering, 136(6) pp. 060801.

[3] Kamath, C., El-dasher, B., Gallegos, G. F., 2014, "Density of Additively-Manufactured, 316L SS Parts using Laser Powder-Bed Fusion at Powers Up to 400 W," The International Journal of Advanced Manufacturing Technology, 74(1-4) pp. 65-78.

[4] Wang, Z., Guan, K., Gao, M., 2012, "The Microstructure and Mechanical Properties of Deposited-IN718 by Selective Laser Melting," Journal of Alloys and Compounds, 513pp. 518-523.

[5] Huang, Y., Leu, M. C., Mazumder, J., 2015, "Additive Manufacturing: Current State, Future Potential, Gaps and Needs, and Recommendations," Journal of Manufacturing Science and Engineering, 137(1) pp. 014001.

[6] Yang, L., Peng, X., and Wang, B., 2001, "Numerical Modeling and Experimental Investigation on the Characteristics of Molten Pool during Laser Processing," International Journal of Heat and Mass Transfer, 44(23) pp. 4465-4473.

[7] Witherell, P., Feng, S., Simpson, T. W., 2014, "Toward Metamodels for Composable and Reusable Additive Manufacturing Process Models," Journal of Manufacturing Science and Engineering, 136(6) pp. 061025.

[8] Wang, G. G., and Shan, S., 2007, "Review of Metamodeling Techniques in Support of Engineering Design Optimization," Journal of Mechanical Design, 129(4) pp. 370-380.

[9] Jin, R., Du, X., and Chen, W., 2003, "The use of Metamodeling Techniques for Optimization Under Uncertainty," Structural and Multidisciplinary Optimization, 25(2) pp. 99-116.

[10] Hoffman, F. O., and Hammonds, J. S., 1994, "Propagation of Uncertainty in Risk Assessments: The Need to Distinguish between Uncertainty due to Lack of Knowledge and Uncertainty due to Variability," Risk Analysis, 14(5) pp. 707-712.

[11] Kothari, C.R., 2004, "Research methodology: Methods and techniques," New Age International, .

[12] Holman, J.P., and Gajda, W.J., 1994, "Experimental methods for engineers," McGraw-Hill New York, .

[13] Cui, C., 2016, Building Energy Modeling: A Data-Driven Approach, .

[14] Yang, Z., Eddy, D., Krishnamurty, S., 2016, "Investigating Predictive Metamodeling for Additive Manufacturing," ASME 2016 International Design Engineering Technical Conferences and Computers and Information in Engineering Conference, Anonymous American Society of Mechanical Engineers, pp. V01AT02A020-V01AT02A020.

[15] Rice, J. R., 1976, "The Algorithm Selection Problem," Advances in Computers, 15pp. 65-118.

[16] Frazier, W. E., 2014, "Metal Additive Manufacturing: A Review," Journal of Materials Engineering and Performance, 23(6) pp. 1917-1928.

[17] Olakanmi, E., Cochrane, R., and Dalgarno, K., 2015, "A Review on Selective Laser Sintering/Melting (SLS/SLM) of Aluminium Alloy Powders: Processing, Microstructure, and Properties," Progress in Materials Science, 74pp. 401-477.

[18] Xiao, B., and Zhang, Y., 2007, "Marangoni and Buoyancy Effects on Direct Metal Laser Sintering with a Moving Laser Beam," Numerical Heat Transfer, Part A: Applications, 51(8) pp. 715-733.

[19] Devesse, W., De Baere, D., and Guillaume, P., 2014, "The Isotherm Migration Method in Spherical Coordinates with a Moving Heat Source," International Journal of Heat and Mass Transfer, 75pp. 726-735.
[20] Roberts, I., Wang, C., Esterlein, R., 2009, "A Three-Dimensional Finite Element Analysis of the Temperature Field during Laser Melting of Metal Powders in Additive Layer Manufacturing," International Journal of Machine Tools and Manufacture, 49(12) pp. 916-923.
[21] Contuzzi, N., Campanelli, S., and Ludovico, A., 2011, "3 D Finite Element Analysis in the Selective Laser Melting Process," International Journal of Simulation Modelling, 10(3) pp. 113-121.
[22] Ma, L., Fong, J., Lane, B., 2015, "Using design of experiments in finite element modeling to identify critical variables for laser powder bed fusion," International Solid Freeform Fabrication Symposium, Anonymous Laboratory for Freeform Fabrication and the University of Texas Austin, TX, USA, .
[23] Palmer, K. D., 1998, Data Collection Plans and Meta Models for Chemical Process Flowsheet Simulators, .
[24] Siah, E. S., Sasena, M., Volakis, J. L., 2004, "Fast Parameter Optimization of Large-Scale Electromagnetic Objects using DIRECT with Kriging Metamodeling," IEEE Transactions on Microwave Theory and Techniques, 52(1) pp. 276-285.
[25] Shao, T., 2007, "Toward a structured approach to simulation-based engineering design under uncertainty," ProQuest, .
[26] Shao, T., and Krishnamurty, S., 2008, "A Clustering-Based Surrogate Model Updating Approach to Simulation-Based Engineering Design," Journal of Mechanical Design, 130(4) pp. 041101.
[27] Zhao, D., and Xue, D., 2011, "A Multi-Surrogate Approximation Method for Metamodeling," Engineering with Computers, 27(2) pp. 139-153.
[28] Bauchau, O., and Craig, J., 2009, "Structural Analysis,"Springer, pp. 173-221.
[29] Gong, H., Rafi, K., Gu, H., 2014, "Analysis of Defect Generation in Ti–6Al–4V Parts made using Powder Bed Fusion Additive Manufacturing Processes," Additive Manufacturing, 1pp. 87-98.
[30] Raghunath, N., and Pandey, P. M., 2007, "Improving Accuracy through Shrinkage Modelling by using Taguchi Method in Selective Laser Sintering," International Journal of Machine Tools and Manufacture, 47(6) pp. 985-995.
[31] Tang, Y., Loh, H., Wong, Y., 2003, "Direct Laser Sintering of a Copper-Based Alloy for Creating Three-Dimensional Metal Parts," Journal of Materials Processing Technology, 140(1) pp. 368-372.
[32] Kempen, K., Yasa, E., Thijs, L., 2011, "Microstructure and Mechanical Properties of Selective Laser Melted 18Ni-300 Steel," Physics Procedia, 12pp. 255-263.
[33] Louvis, E., Fox, P., and Sutcliffe, C. J., 2011, "Selective Laser Melting of Aluminium Components," Journal of Materials Processing Technology, 211(2) pp. 275-284.
[34] Kruth, J., Wang, X., Laoui, T., 2003, "Lasers and Materials in Selective Laser Sintering," Assembly Automation, 23(4) pp. 357-371.
[35] Morgan, R., Sutcliffe, C., and O'neill, W., 2004, "Density Analysis of Direct Metal Laser Re-Melted 316L Stainless Steel Cubic Primitives," Journal of Materials Science, 39(4) pp. 1195-1205.
[36] Beaman, J. J., Barlow, J. W., Bourell, D. L., 1997, "Solid Freeform Fabrication: A New Direction in Manufacturing," Kluwer Academic Publishers, Norwell, MA, 2061pp. 25-49.
[37] Meier, H., and Haberland, C., 2008, "Experimental Studies on Selective Laser Melting of Metallic Parts," Materialwissenschaft Und Werkstofftechnik, 39(9) pp. 665-670.
[38] Ciurana, J., Hernandez, L., and Delgado, J., 2013, "Energy Density Analysis on Single Tracks Formed by Selective Laser Melting with CoCrMo Powder Material," The International Journal of Advanced Manufacturing Technology, 68(5-8) pp. 1103-1110.
[39] Mills, K.C., 2002, "Recommended values of thermophysical properties for selected commercial alloys," Woodhead Publishing, .
[40] Box, G.E., and Draper, N.R., 1987, "Empirical model-building and response surfaces," Wiley New York, .
[41] Simpson, T. W., Booker, A. J., Ghosh, D., 2004, "Approximation Methods in Multidisciplinary Analysis and Optimization: A Panel Discussion," Structural and Multidisciplinary Optimization, 27(5) pp. 302-313.
[42] Shan, S., and Wang, G. G., 2010, "Survey of Modeling and Optimization Strategies to Solve High-Dimensional Design Problems with Computationally-Expensive Black-Box Functions," Structural and Multidisciplinary Optimization, 41(2) pp. 219-241.
[43] Cressie, N., 2015, "Statistics for spatial data," John Wiley & Sons, .
[44] Sacks, J., Welch, W. J., Mitchell, T. J., 1989, "Design and Analysis of Computer Experiments," Statistical Science, pp. 409-423.
[45] Rosenblatt, F., 1958, "The Perceptron: A Probabilistic Model for Information Storage and Organization in the Brain." Psychological Review, 65(6) pp. 386.
[46] Yegnanarayana, B., 2009, "Artificial neural networks," PHI Learning Pvt. Ltd., .
[47] Shanmuganathan, S., and Samarasinghe, S., 2016, "Artificial neural network modelling," Springer, .
[48] Narendra, P. M., and Fukunaga, K., 1977, "A Branch and Bound Algorithm for Feature Subset Selection," IEEE Transactions on Computers, 26(9) pp. 917-922.
[49] Chatterjee, A., Kumar, S., Saha, P., 2003, "An Experimental Design Approach to Selective Laser Sintering of Low Carbon Steel," Journal of Materials Processing Technology, 136(1) pp. 151-157.

