HYBRID MODELING APPROACH FOR MELT POOL PREDICTION IN LASER POWDER BED FUSION ADDITIVE MANUFACTURING

Tesfaye Moges¹, Zhuo Yang², Kevontrez Jones³, Shaw Feng⁴, Paul Witherell⁴, Yan Lu⁴

¹ Guest Researcher at National Institute of Standards and Technology (NIST) and Affiliated with Indian Institute of Technology Delhi, New Delhi, India
² Guest Researcher at NIST and Affiliated with University of Massachusetts, Amherst, MA, USA
³ Northwestern University, Evanston, IL, USA
⁴ National Institute of Standards and Technology (NIST), Gaithersburg, MD, USA

ABSTRACT
Multi-scale multi-physics computational models are a promising tool to provide detailed insights to understand the process-structure-property-performance relationships in additive manufacturing (AM) processes. To take advantage of the strengths of both physics-based and data-driven models, we propose a novel hybrid modeling framework for laser powder bed fusion (L-PBF) processes. Our unbiased model integration method combines physics-based data and measurement data for approaching more accurate prediction of melt pool width. Both a high-fidelity computational fluid dynamics (CFD) model and experiments utilizing optical images are used to generate a combined dataset of melt pool widths. From this aggregated dataset, a hybrid model is developed using data-driven modeling techniques, including polynomial regression and Kriging methods. The performance of the hybrid model is evaluated by computing the average relative error and compared with the results of the simulations and surrogate models constructed from the original CFD model and experimental measurements. It is found that the proposed hybrid model performs better in terms of prediction accuracy and computational time. Future work includes a conceptual introduction on the use of an AM ontology to support improved model and data selection when constructing hybrid models. This study can be viewed as a significant step towards the use of hybrid models as predictive models with improves accuracy without the sacrifice of speed.

Keywords: Additive manufacturing, Laser powder bed fusion, Hybrid model, Melt pool width, Gaussian process/Kriging, Data-driven surrogate model, Ontology

1. INTRODUCTION
Metal additive manufacturing (AM) produces metallic parts by fusing materials in a layer-by-layer fashion directly from a 3D CAD model [1]. Laser powder bed fusion (L-PBF) is the most common AM process used for the fabrication of metallic components. In the L-PBF process, a thin powder layer is spread on a substrate and a laser beam selectively melts and fuses powder with neighboring particles and previous layer. This process is repeated until the final part is formed. L-PBF has tremendous potential of producing metallic parts with complex geometry, internal structures, and conformal heating/cooling channels for a wide range of applications including aerospace, automotive, and biomedical implants [2–4]. Compared to conventional manufacturing techniques, L-PBF has many advantages as it allows to locally control microstructures by varying process parameters so that parts with desirable mechanical properties can be produced [5,6]. It also minimizes material wastes and reduces lead time [7]. While offering many advantages, the L-PBF process also faces challenges such as inconsistent part quality and defects in terms of porosity, poor surface finish, delamination, crack formation, and residual stress [8]. Without proper control mechanisms, these challenges can lead to unstable mechanical properties and poor dimensional accuracy on produced parts.

To overcome these challenges and help develop process control mechanisms, there have been many research efforts that have aimed to understand the influence of different process parameters and material properties on part quality [9,10]. The research efforts can be broadly categorized as experimental-based and physics-based investigations. The experimental-based investigation is more realistic as it directly captures observed physical phenomena occurring during the process. However, this
approach can be time consuming and costly while lending itself to inherent process variabilities and sensor calibration errors that cause large measurement uncertainty [11]. Additionally, not every process parameter can be measured. For these reasons, amongst others, interest in computational models for AM continues to rise.

For the L-PBF process, computational models have been widely used to simulate the heat transfer, fluid flow, and phase transformations in L-PBF process and estimate the temperature fields, flow velocities, melt pool characteristics, solidification rate, and residual stress [12,13]. Although these computational models are promising tools to understand the physics of the process, their prediction accuracy can be potentially affected by the assumptions used during model formulation [14]. In addition to modeling assumptions, inaccurate selection of model parameters, such as absorption coefficient and material properties, also leads to significant discrepancy between computationally predicted and experimental results [15]. Furthermore, even though high fidelity physics-based models describe the L-PBF process in detail, solving these models is time consuming due to the complex physics.

Overall, physics-based models suffer from prediction accuracy caused by uncertainties that are associated with the modeling assumptions made about the L-PBF process. The prediction accuracy of the experimental-based investigations is highly affected by sensor noise and measurement uncertainties due to the inherent complexity and stochastic nature of the process. Alterations to each approach have sought to compensate for their drawbacks. For instance, reduced-order physics-based models have been sought to reduce computational demands, though the tradeoff is they can be less accurate. Highly complex, multi-variable experiments of processes like L-PBF are also difficult to implement due to their high cost and time requirements. We propose that the most effective way to overcome the challenges associated with each approach is to develop a model that embraces the advantages of each approach.

In this study, we propose a hybrid model by intertwining physics-based simulation data and experiment-based empirical data for the prediction of output quantities of interest (QoIs) in L-PBF process. To develop an unbiased model which integrates computational and experimental data, data-driven modeling techniques including polynomial regression and Gaussian process/Kriging methods, which are suitable for multi-dimensional problems having limited data, are employed. Melt pool width is chosen as our primary output QoI since the melt pool plays a significant role in determining the microstructure, residual stresses, and mechanical properties of a part fabricated by L-PBF process [16]. A CFD model is used to predict melt pool widths for various combinations of laser power and scan speed. Empirical data is obtained from ex-situ melt pool width measurement data taken from single-track experiments with similar process parameters.

In developing the hybrid model, a polynomial regression method is first applied to construct the initial simulation-based surrogate model using data from a CFD model. Then, the Kriging method is applied to model the residual error between experimental and computational results. An adaptive modeling method is used to iteratively update the Kriging model to improve the predictive errors of the surrogate model. To improve the model, measurement data are iteratively selected by updating sample points using the maximum average relative error as a determining factor. The performance of the hybrid model is evaluated by comparing the relative error to the individual simulation-based and experiment-based surrogate models. It is found that the hybrid model performs better in terms of prediction accuracy and computational time. Once the hybrid model is developed, making predictions on a new set of input variables is straightforward.

This paper is organized as follows: in Section 2, we briefly review the prior research efforts on the physics-based computational models focusing on thermal models and data-driven modeling approaches used in L-PBF AM. Then we provide an overview of the CFD model, the data-driven modeling techniques used in this study, and the hybrid model development methodology in Section 3. In Section 4, we demonstrate the proposed methodology to develop a hybrid model for melt pool width prediction and analyze its performance using CFD simulation data and measurement data. A conceptual use of AM ontology for model selection is introduced in Section 5. Section 6 presents the concluding remarks of the current study and future work. We view this work to be an essential step to developing fast and accurate hybrid models that compliment both physics-based and experimental approaches to improve L-PBF part quality.

2. BACKGROUND

Though L-PBF has shown to be a capable technology for producing complex parts, challenges remain, e.g., inconsistent part properties with defects. Part defects are often attributed to porosity in the fabricated parts, surface roughness, anisotropy in microstructure, residual stresses, delamination, and cracks [17]. Physics-based modeling techniques are based on cause-effect principles in physics, such as fluid dynamics, thermodynamics, heat transfer, and kinetics. While such techniques provide invaluable insight into the general behavior of a process, fall short in predicting defect generation and propagation and final part quality for a specific process implementation. There are at least three reasons for that: (a) a complete set of process variables may not be obtainable due to a lack of knowledge regarding the process, (b) unknown complex energy and material behaviors in the process, such as non-linear interactions, and (c) the physics-based models have not been validated due to limitation in available measurement techniques.

AM data analytics includes measured data, data fusion, data analysis, statistical methods, and machine learning [18]. Real data are acquired from in-situ or ex-situ sensors. To model the process, the relations between inputs and outputs are created by data-driven modeling techniques. Data-driven models can be employed to model the non-linear relationship between sensor outputs with process parameters and predict output quantities of interest for different sets of input variables. However, as stated previously, the measurement data used to build the data-driven
models are affected by uncertainties related to error in sensor calibration and noise, imprecise measurement methods, and variations in the measurements. The following subsections review relevant works associated with each of these modeling approaches.

2.1. Physics-based thermal models
There have been tremendous efforts in the last decade in developing computational models to simulate powder layer deposition, powder-laser interactions, melt pool formation, solidification and grain growth, and residual stress and deformation with different levels of fidelity in the L-PBF process [12–14,19]. The term fidelity used in this context is based on the different physical phenomena captured by the computational models. For instance, a model that captures a larger number of phenomena refers as high-fidelity, whereas, low fidelity is the one with least number of phenomena.

Physics-based computational modeling has been crucial to understanding process-structure-property relations in metal AM [20], and these models have come in different varieties. With regard to the transient manufacturing process, thermal models based on the semi-analytical, the finite element, and finite volume methods have been developed. The semi-analytical Rosenthal-based low-fidelity thermal models can solve the heat conduction equation for temperature profile and melt pool geometry [21,22]. However, they neglect physical phenomena such as the effective powder layer, laser spot diameter, and other phenomena related to heat transfer and fluid flow. Typically, the finite element medium-fidelity models can provide a thermal history of an entire part being built [23–26], but they do so by purely considering heat conduction and neglecting the fluid flow behavior within the melt pool. This simplification can lead to predictions of inaccurate temperature fields. For example, Manvatkar et al. [27] showed that by ignoring Marangoni convection in the molten melt pool, cooling rates in laser assisted AM may be overestimated by as much as double the correct values. Conversely, Gan et al. [28] demonstrated that incorporation of fluid-flow and vaporization can significantly enhance a model ability to accurately predict melt pool geometry, peak temperature and surface topology. For these reasons, several high-fidelity models based on the finite volume method and CFD have been developed to account for additional physics within the melt pool [29–34].

Ensuring the accuracy of high-fidelity models requires an extensive use of well-designed and highly controlled experiments for validation. Due to extremely high temperatures, violent metallic powder spattering, and highly complex physics occurring at multiple length scales within very short time scales, it is difficult to conduct in-situ measurements in the L-PBF process. Typically, as is the case for Ghosh et al. [35], ex-situ measurements are used for numerical validation. It is apparent that assumptions and simplifications made about the process can significantly affect a computational model’s predictive capabilities.

2.2. Data-driven models
Data-driven modeling methods, as a more stochastic approach, have been deployed to analyze AM data. Different data-driven approaches have been implemented to construct surrogate models from experiment-based empirical data and physics-based simulation data. Data-driven surrogate models that rely on empirical data can help estimate data points in a new design space and evaluate correlations between input parameters and output QoIs [36,37]. In addition, simulation-based data-driven surrogate models aim support a black-box approach to reduce the high computational cost of high-fidelity simulations [38]. Data-driven machine learning techniques have been used in AM for different applications throughout the AM lifecycle [39].

Using experiment-based empirical data, Fathi and Mozaﬁfraz [40] developed a data-driven framework to relate process parameters including laser power, scan speed, and powder flow rate to melt pool depth and deposition height for the laser-based direct energy deposition (DED) process. Similarly, Lu et al. [41] used a neural network to map process parameters to deposition height for the DED process. Since obtaining measurement data for such complex processes is difﬁcult and time consuming, in lieu of empirical data, researchers have used physics-based simulation data to develop data-driven surrogate models. From the semi-analytical Rosenthal-based thermal model, Yang et al. [42] developed a Dynamic Variance-Covariance Matrix (DVM) method to investigate the influence of input parameters such as laser power, scan speed, absorption coefficient, and thermal diffusivity on melt pool width for L-PBF process. Kamath [43] built data-driven surrogate models from the Eager-Tsai thermal model using regression trees and Gaussian process regression to predict melt pool depth for the L-PBF process. Recently, Tran and Lo [44] developed an approach to optimize process parameters such as laser power, scan speed, and layer thickness using artiﬁcial neural network (ANN) for L-PBF process.

The previous works focused on developing data-driven surrogate models based on either experiment-based empirical data or physics-based simulation data to predict output QoIs. As measurement data contain errors and uncertainties associated with sensor calibration and noise, imprecise measurement methods, and variation in the process; and physics-based simulation data have uncertainty associated with modeling assumptions, numerical approximation, and variability in input parameters, data-driven surrogate models that are developed solely from empirical data or simulation data may exhibit variable predictive capabilities. Therefore, hybrid models that embraces the advantages of experiment-based and simulation-based approaches are needed.

In general, physics-based models typically do an acceptable job at predicting output QoIs in a wide range of input variables. However, there are multiple sources of uncertainty in these computational models that can cause significant prediction errors [15]. On the other hand, data-driven modeling techniques often rely on a specific system to help find the relationships between inputs and outputs, and without the explicit knowledge of the physical behavior of the system [45]. In such scenarios,
due to data sparsity, extrapolation is limited and careful selection of training data, training algorithm, and model complexity is required [46]. The proposed hybrid model offers an approach where extrapolation can be done for a wide input of parameter ranges while capturing the complex and highly non-linear behavior of the L-PBF process [46].

3. HYBRID MODEL DEVELOPMENT

As mentioned above, researchers have developed various approaches by utilizing different scales, disciplines, and perspectives to overcome the complexity of AM processes. All these efforts can significantly improve knowledge mining in AM. However, these efforts also raise more challenges in data filtering, algorithm selection, and model integration. Different simulations, for example, may only work under specific conditions if they were developed based on different physical phenomena. It is also hard to guarantee consistency between different datasets from various sources. Another typical issue is the stochastic error observed between computational and experimental data. This error may initiate from the fundamental hypothesis or numerical approach of the simulation. However, physical experiments cannot benefit from selective simplifying assumptions like simulations. Hence, experimental results tend to include more complicated physical phenomena than those obtained from a simulation. Thus, even high-fidelity simulations cannot truly match the real experimental data based on different AM machines, labs, and material.

Modifications applied to physics-based models can potentially address the above issues. This section presents the development of an unbiased model integration method to combine computational and experimental data to provide accurate predictions regardless of the sample size and fidelity of the data. The high-fidelity physics-based model, though requires higher computational cost, has the freedom to generate a vast amount of data. On the other hand, the experimental data is assumed to be ground truth but usually limited and expensive to sample.

Figure 1 shows the proposed hybrid modeling framework for the L-PBF process. Various combinations of process parameters, physics-based models and experimental measurement techniques provide datasets with inherent uncertainties [47–49]. Then, multiple sampling methods filter through each dataset to systematically create subsets of the data to be used for training and validation. Once these data subsets are determined, they are used in conjunction with data-driven models to build simulation-based and experiment-based surrogate models. To improve the accuracy of these surrogate models, a hybrid model that combines the physics-based data and experimentally measured data using an unbiased model integration method is created. With this framework in mind, a brief overview of the physics-based CFD model and data-driven approaches used in the current study is provided, and the workflow and algorithm of the proposed hybrid model are discussed in detail in this section.

3.1. Brief details on CFD model

A well tested, three dimensional, transient, thermal-fluid flow model for L-PBF [28,29] is adapted to compute temperature and velocity fields to generate the data referenced in this work. The thermal-fluid flow model solves for conservation of mass (Eqn. 1), momentum (Eqn. 2), and energy (Eqn. 3) to consider liquid flow within the melt pool driven by Marangoni convection.

\[
\frac{\partial p}{\partial t} + \frac{\partial \rho u_i}{\partial x_i} = 0
\]
\[
\frac{\partial \rho u_i}{\partial t} + \frac{\partial \rho u_i u_j}{\partial x_j} = \frac{\partial}{\partial x_j} \left( \mu \left( \frac{\partial u_j}{\partial x_j} \right) \right) - \frac{\partial p}{\partial x_i} + 180 \mu (1 - f_j)^2 \frac{\partial}{\partial x_i} \left( \frac{\partial u_j}{\partial x_j} \right) - \frac{\partial^2}{\partial x_i^2} \frac{1}{\rho} \left( \frac{\partial^2 u_i}{\partial x_j^2} + B \frac{\partial u_j}{\partial x_j} \right) + \rho_{\text{ref}} g_i \beta (T - T_{\text{ref}})
\]

where \( t \) is the time, \( u_i \) is the \( i^{th} \) component of the velocity, \( \mu \) is the viscosity, \( p \) is the pressure, \( T \) is the temperature, \( \rho \) is the density, \( k \) is the thermal conductivity, and \( \beta \) is the thermal expansion coefficient. In this work, \( B \) is a small parameter with a value of \( 10^{-3} \) to avoid division by zero and \( \delta \) is the approximate primary dendritic spacing, which was set to 1\( \mu \)m. Additionally, enthalpy is related to temperature according to the following:

\[
\phi_h = \int_0^T \rho c_p dT + \rho L f_i
\]

where \( c_p \) is the specific heat capacity, \( L \) is the latent enthalpy of fusion, and \( f_i \) is the volume fraction of liquid phase.

The thermal boundary condition including the heat source model at the metal-gas interface is specified as:

\[
q_{\text{ener}} = \frac{2Q \eta}{\pi r_b^2} \exp\left( \frac{-2((x - V_s t)^2 + y^2)}{r_b^2} \right) - h_c (T - T_{\infty}) - \sigma \varepsilon (T^4 - T_{\text{ref}}^4)
\]

where \( Q \) is the laser power, \( \eta \) is the absorptivity, \( r_b \) is the laser beam radius, \( V_s \) is the scanning speed, \( h_c \) is the convective heat transfer coefficient, \( T_{\infty} \) is the ambient temperature, \( \sigma \) is the Stefan-Boltzmann constant, \( \varepsilon \) is the emissivity, and \( T_{\text{ref}} \) is the reference temperature.

The momentum boundary condition at the liquid-gas interface is:

\[
F_{L/G} = \gamma \kappa + \nabla_T^2 \frac{d\gamma}{dT}
\]

where \( \gamma \) is the surface tension coefficient, \( \kappa \) is the outward pointing normal of the surface, and \( \kappa \) is the curvature of the surface. The thermophysical properties of Inconel 625 and approximated processing conditions used in the simulations are summarized in Table 1. Densities taken from [50] for ambient and liquidus temperatures are used for the solid and liquid densities, respectively. Additionally, viscosity at the liquidus temperature taken from [51] was assumed to maintain a constant value within the melt pool. Temperature-dependent polynomial functions were fitted to experimental measurements [52] of thermal conductivity and specific heat capacity for the solid phase.

### Table 1: Thermophysical properties and processing conditions used for the thermal-fluid flow model

<table>
<thead>
<tr>
<th>Physical Property</th>
<th>Value</th>
<th>Reference</th>
</tr>
</thead>
<tbody>
<tr>
<td>Solid density (kg·m⁻³)</td>
<td>8440</td>
<td>[50]</td>
</tr>
<tr>
<td>Liquid density (kg·m⁻³)</td>
<td>7640</td>
<td>[50]</td>
</tr>
<tr>
<td>Solidus temperature (K)</td>
<td>1563</td>
<td>[51]</td>
</tr>
<tr>
<td>Liquidus temperature (K)</td>
<td>1623</td>
<td>[51]</td>
</tr>
<tr>
<td>Solid specific heat capacity (J·kg⁻¹·K⁻¹)</td>
<td>0.2437T</td>
<td>[52]</td>
</tr>
<tr>
<td>Liquid specific heat capacity (J·kg⁻¹·K⁻¹)</td>
<td>709.25</td>
<td>[50]</td>
</tr>
<tr>
<td>Solid thermal conductivity (W·m⁻¹·K⁻¹)</td>
<td>0.01530T</td>
<td>[52]</td>
</tr>
<tr>
<td>Liquid thermal conductivity (W·m⁻¹·K⁻¹)</td>
<td>3.078</td>
<td>[50]</td>
</tr>
<tr>
<td>Latent heat of fusion (KJ·kg⁻¹·K⁻¹)</td>
<td>29.0</td>
<td>[50]</td>
</tr>
<tr>
<td>Dynamic viscosity (Pa·s)</td>
<td>7 × 10⁻³</td>
<td>[51]</td>
</tr>
<tr>
<td>Coefficient of thermal expansion (K⁻¹)</td>
<td>5 × 10⁻⁵</td>
<td>[51]</td>
</tr>
<tr>
<td>Preheat temperature (K)</td>
<td>353</td>
<td>-</td>
</tr>
<tr>
<td>Laser spot radius (( \mu )m)</td>
<td>45</td>
<td>-</td>
</tr>
</tbody>
</table>

### 3.2. Data-driven modeling techniques

This section aims to introduce the data-driven modeling techniques used in this study, namely the polynomial regression (PR) and Kriging methods. In general, the PR model uses the training data to estimate the optimal parameters of the polynomial formulation. On the other hand, the Kriging method offers an interpolation approach from which the prediction is derived based on correlations to existing data. The mathematical formulation of the predictive model can be expressed as:

\[
y(\tilde{x}) = f(\tilde{x}) + \varepsilon
\]

where \( y(\tilde{x}) \) represents the exact solution for new point \( \tilde{x} \), \( f(\tilde{x}) \) is a hypothetical function derived statistically from data that produces the model estimate, \( \varepsilon \) is error, and \( \tilde{x} \) represents a set of input variables. For different modeling approaches, the composition of each of these elements could be different.

#### 3.2.1. Polynomial regression (PR) method

Similar to linear regression, PR formulates the relationship between the input variables \( x \) and the outcome \( y \) with higher-order variation [53]. The efficiency and accuracy of PR make it popular in various engineering domains. The quadratic polynomial function can be presented as:

\[
\hat{y} = \beta_0 + \sum_{i=1}^{k} \beta_i x_i + \sum_{i=1}^{k} \beta_{ii} x_i^2 + \sum_{i=1}^{k} \sum_{j=1}^{k} \beta_{ij} x_i x_j
\]
3.2.2. Kriging method
Unlike traditional parametric modeling methods that can derive specific formulations, the Kriging method predicts a result based on the spatial correlation between an estimated data point and existing data points [54,55]. The general form of the Kriging approach can be presented as:

$$Z_E = \bar{Z} + \sum_{i=1}^{n} \lambda_i (Z_i - \bar{Z})$$  \hspace{1cm} (9)

where $\bar{Z}$ represents the regional mean value of the response and $\lambda_i$ is the distance-correlated weight value, which is determined by computing the spatial correlation.

To calculate the weight factor, $\lambda_i$, one should first compute the spatial correlation $R$ between data points. The value of the spatial correlation can be derived from:

$$R(\theta, x_i, x_j) = \prod_{l=1}^{n} \exp(-\theta_l (x_{il} - x_{jl})^2)$$  \hspace{1cm} (10)

where $x_{il}$ is the $l^{th}$ component of the $i^{th}$ vector $x_i$ [56]. $R(\theta, x_i, x_j)$ depends on the location of points $x_i$ and $x_j$ and the correlation parameter, $\theta$.

3.3. Hybrid modeling approach
This section discusses development of the hybrid model based on a combination of computational and experimental data. The hybrid model applies a polynomial regression method to construct the initial simulation-based surrogate model using computational data obtained from the aforementioned CFD model (Section 3.1.). The Kriging method is then applied to model the residual error between computational and experimental results, using an adaptive modeling method to iteratively update the model and reduce the predictive errors of the hybrid model. Figure 2 outlines our workflow for the hybrid AM model construction.

3.3.1. Workflow of hybrid modeling approach
The proposed method uses physics-based computational data from a CFD model to construct the initial surrogate model by applying the polynomial regression method. This model can generally represent trends, but may include significant error when compared to the experimental data. The next step is to model the residual error between computational and experimental results. An adaptive modeling method can iteratively update the Kriging model to reduce the predictive errors of the hybrid model [57,58,59]. The process adaptively adjusts the training and validation datasets to approach higher predictive accuracy. To evaluate the hybrid model performance, the Average Relative Error Magnitude (AREM) is deployed for individual and global validation [58]. The formulation of AREM can be expressed as:

$$AREM = \frac{1}{m} \left( \sum_{t=1}^{m} |y_t - \hat{y}_t| / y_t \right) (y_t \neq 0)$$  \hspace{1cm} (11)

where $y_t$ is the observed value from given data, $\hat{y}$ is the value predicted by the surrogate model of the data points that were not selected to construct the surrogate model, and $m$ is the number of data points.

3.3.2. Algorithm for hybrid model development
This section provides the algorithmic approach to build the proposed AM hybrid model. We first generate $N_{sim}$ simulation data to construct the initial surrogate model. Similarly, $M_{meas}$ experimental data are measured to improve and validate the hybrid model. The adaptive Kriging model starts with $M_{in}$ initial experimental data points and $M_{up}$ additional points for updating. To validate the final hybrid model, $M_{val}$ experimental data points are used. Therefore, the training-validation data ratio is $(M_{in} + M_{up}) : M_{val}$. The case study in Section 4 uses a total of 72 simulation data points and 21 experimental data points. Here, it is important to note that, in the limited number of experimental data, the predetermined number of validation data can be reached before the error value is less than the threshold. In this case, all
the remaining validation datapoints will be used to validate the hybrid model. This situation can be avoided by providing a greater number of experimental data.

### Algorithm for hybrid modeling approach

<table>
<thead>
<tr>
<th>Step</th>
<th>Task</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Generate experimental and computational data</td>
</tr>
<tr>
<td>2</td>
<td>Construct initial surrogate model</td>
</tr>
<tr>
<td>2.1</td>
<td>Construct initial surrogate model using $N_{sim}$ by PR method</td>
</tr>
<tr>
<td>2.2</td>
<td>Set target threshold for error comparison</td>
</tr>
<tr>
<td>2.3</td>
<td>Validate the surrogate model using additional simulation data</td>
</tr>
<tr>
<td></td>
<td><strong>if</strong> error &gt; threshold <strong>do</strong></td>
</tr>
<tr>
<td></td>
<td>Back to Step 1 and generate more simulation data</td>
</tr>
<tr>
<td></td>
<td><strong>else</strong></td>
</tr>
<tr>
<td></td>
<td>Simulation-based surrogate model is ready and proceed to Step 3</td>
</tr>
<tr>
<td>3</td>
<td>Prepare the training and validation datasets of $M_{meas}$ experimental data</td>
</tr>
<tr>
<td>3.1</td>
<td>Compare the results of the surrogate model against $M_{meas}$ experimental data</td>
</tr>
<tr>
<td>3.2</td>
<td>Select $M_{in}$ experimental data points with largest error to construct the initial training dataset</td>
</tr>
<tr>
<td>3.3</td>
<td>Store the rest of the data that would be used as validation dataset</td>
</tr>
<tr>
<td>4</td>
<td>Construct the hybrid model</td>
</tr>
<tr>
<td>4.1</td>
<td>Calculate the residual error between results from surrogate model and experimental training dataset</td>
</tr>
<tr>
<td>4.2</td>
<td>Build the surrogate model for residual error by Kriging method</td>
</tr>
<tr>
<td>4.3</td>
<td>Combine the PR and Kriging methods to construct the hybrid model</td>
</tr>
<tr>
<td>5</td>
<td>Hybrid model validation</td>
</tr>
<tr>
<td>5.1</td>
<td>Use the validation dataset to validate the hybrid model</td>
</tr>
<tr>
<td>5.2</td>
<td>Set target threshold for error comparison</td>
</tr>
<tr>
<td></td>
<td><strong>if</strong> error &gt; threshold <strong>do</strong></td>
</tr>
<tr>
<td></td>
<td><strong>if</strong> data &gt; $M_{val}$ <strong>do</strong></td>
</tr>
<tr>
<td></td>
<td>Add an $M_{up}$ additional data point with largest error to the training dataset</td>
</tr>
<tr>
<td></td>
<td><strong>end if</strong></td>
</tr>
<tr>
<td></td>
<td><strong>else</strong></td>
</tr>
<tr>
<td></td>
<td>Validate the hybrid model using validation dataset and proceed to Step 6</td>
</tr>
<tr>
<td>5.3</td>
<td>Eliminate the selected data point from the $M_{val}$ validation dataset</td>
</tr>
<tr>
<td>5.4</td>
<td>Go back to Step 4 to validate the hybrid model</td>
</tr>
<tr>
<td>6</td>
<td>Approach the final hybrid model</td>
</tr>
</tbody>
</table>

### 4. Case Study: Hybrid Modeling of Melt Pool Width

This section demonstrates the proposed hybrid model presented in Section 3.3 and discusses its performance by comparing it against computational and experimental data. As previously stated, melt pool widths obtained from CFD model and measurement data are used as output QoIs to demonstrate the hybrid model’s capabilities. The primary input variables used for the simulations and experiments are laser power and scan speed. The range of laser power used for the simulations and experiments is from 49W – 285W and 100W – 195W, respectively. Whereas, the scan speed used for both the simulations and experiments ranges from 100mm/s – 1400mm/s.

#### 4.1 Data for hybrid model

The experiments were performed on Inconel 625 bare plates and ex-situ measurements were conducted to record melt pool widths using optical microscope [60]. The melt pool widths were measured from the optical image of a 1mm long scan track by manually tracing the edges of the track and averaging the distance between the traces at different locations as shown in Figure 3(a). The average and standard deviation of the measured melt pool widths at different combinations of laser power and scan speed are given in Table 2. Similarly, the thermal CFD model is simulated as shown in Figure 3(b), and melt pool widths are extracted for a given input variables. The simulation results for the corresponding experimental data along with the relative percentage error are also given in Table 2.

The average relative percentage error of the simulations against measurement results is 20.78%. This highlights that the physics-based computational model by itself induces major discrepancy against the experimental data. This discrepancy may be due to the different assumptions taken during model formulation including powder particle distribution, spattering of molten metal, gas-liquid-solid interaction, mass loss due to chemical reactions, and others. This discrepancy is due to model uncertainty and also the uncertainty associated with measurements, including sensor error and imprecise measurement methods [15,22].
Figure 3. Experimental-based melt pool width from optical image [60] (a) and melt pool region from the CFD simulation model (b)

Table 2. Melt pool widths: measurement, simulation, and % relative error

<table>
<thead>
<tr>
<th>No. of Run</th>
<th>Laser power (W)</th>
<th>Scan speed (mm/s)</th>
<th>Melt pool width (µm)</th>
<th>% relative error</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td>Measurement with St. dev</td>
<td>Simulation</td>
</tr>
<tr>
<td>1</td>
<td>100</td>
<td>100</td>
<td>259.77 (8.22)</td>
<td>175</td>
</tr>
<tr>
<td>2</td>
<td>100</td>
<td>200</td>
<td>177.26 (4.73)</td>
<td>146</td>
</tr>
<tr>
<td>3</td>
<td>100</td>
<td>400</td>
<td>123.39 (7.73)</td>
<td>120</td>
</tr>
<tr>
<td>4</td>
<td>100</td>
<td>600</td>
<td>86.79 (3.43)</td>
<td>110</td>
</tr>
<tr>
<td>5</td>
<td>100</td>
<td>1000</td>
<td>73.918 (2.47)</td>
<td>94.1</td>
</tr>
<tr>
<td>6</td>
<td>100</td>
<td>1200</td>
<td>72.123 (2.87)</td>
<td>88.2</td>
</tr>
<tr>
<td>7</td>
<td>100</td>
<td>1400</td>
<td>68.932 (2.19)</td>
<td>90</td>
</tr>
<tr>
<td>8</td>
<td>150</td>
<td>200</td>
<td>225.87 (11.16)</td>
<td>161</td>
</tr>
<tr>
<td>9</td>
<td>150</td>
<td>400</td>
<td>166.96 (9.90)</td>
<td>132</td>
</tr>
<tr>
<td>10</td>
<td>150</td>
<td>600</td>
<td>146.89 (8.70)</td>
<td>121</td>
</tr>
<tr>
<td>11</td>
<td>150</td>
<td>800</td>
<td>105.73 (5.15)</td>
<td>110</td>
</tr>
<tr>
<td>12</td>
<td>150</td>
<td>1000</td>
<td>86.922 (3.26)</td>
<td>110</td>
</tr>
<tr>
<td>13</td>
<td>150</td>
<td>1200</td>
<td>93.365 (4.62)</td>
<td>105</td>
</tr>
<tr>
<td>14</td>
<td>150</td>
<td>1400</td>
<td>85.735 (5.34)</td>
<td>101</td>
</tr>
<tr>
<td>15</td>
<td>195</td>
<td>100</td>
<td>362.25 (13.33)</td>
<td>206</td>
</tr>
<tr>
<td>16</td>
<td>195</td>
<td>200</td>
<td>256.05 (16.98)</td>
<td>167</td>
</tr>
<tr>
<td>17</td>
<td>195</td>
<td>400</td>
<td>188.7 (10.69)</td>
<td>142</td>
</tr>
<tr>
<td>18</td>
<td>195</td>
<td>600</td>
<td>149.62 (5.28)</td>
<td>130</td>
</tr>
<tr>
<td>19</td>
<td>195</td>
<td>800</td>
<td>126.3 (4.33)</td>
<td>120</td>
</tr>
<tr>
<td>20</td>
<td>195</td>
<td>1000</td>
<td>107.17 (4.55)</td>
<td>113</td>
</tr>
<tr>
<td>21</td>
<td>195</td>
<td>1200</td>
<td>90.15 (4.97)</td>
<td>110</td>
</tr>
</tbody>
</table>

4.2. Implementation of the proposed hybrid model

Figure 4 shows the schematic framework for the case study implementation of the hybrid model for melt pool width prediction. First, the CFD model is simulated and 72 data points of melt pool widths were extracted for the given combinations of laser power and scan speed, as shown in PV map (left) in Figure 4. Similarly, experimental data provided 21 data points of melt pool widths are measured for the given combinations of laser power and scan speed, as shown in the PV map (right). All simulation data are used to construct the initial surrogate model using a polynomial regression method. The adaptive Kriging model starts with 5 initial data points and 10 additional points for updating. As a result, 6 experimental data points are excluded from the model construction and are used to validate the model. The training-validation data ratio is 15:6. The hybrid model is then used to predict melt pool widths for other possible combinations of laser power and scan speed in the design space with good accuracy and computational time.

The accuracy of the proposed hybrid model mainly depends on the accuracies and the number of computational and experimental data used and the data-driven techniques applied. Obtaining more experimental data used for training, updating, and validating the proposed hybrid model is crucial for improving the accuracy.
4.3. Verification of Surrogate models
To develop and evaluate the hybrid model, both computational-based and experimental-based surrogate models are constructed using corresponding data. To evaluate the accuracy of these surrogate models and compare them with the original simulation and measured data, model verification is conducted for the same processing parameters. The computational-based surrogate model is compared to the original CFD model as shown in Figure 5(a) and (b), and the average error is about 5%. The cause of this discrepancy may be due to the variabilities in data-driven modeling approaches. Similarly, the experimental-based surrogate model has around 10% difference when compared to the original measurement data as shown in Figure 5(c) and (c). This error is mainly attributed to the limited number of measurement data points (only 15) used to build the surrogate model.

Figure 4 Schematic implementation framework of hybrid model of melt pool width
4.4. Performance evaluation of hybrid model

In order to evaluate the performance of the proposed model, we compared the average relative error of hybrid model to the experimental-based surrogate model and computational-based surrogate model as shown in Figure 6 and Table 3. The blue and red dots in the graph represent the measured and predicted results with the validation parameters, respectively.

The results of melt pool widths predicted at the 6 validation data points (that are not used in the model construction) using the experimental-based surrogate model has an average relative error of 13.45%, as shown in Figure 6(a). Similarly, the melt pool widths predicted using the simulation-based surrogate model has an average error of 12.89% as shown in Figure 6(b). Figure 6(c) depicts the prediction of melt pool widths using the proposed hybrid model. The hybrid model has an average error of 7.58%. Table 3 provides the input parameters used for validation, the measured results, simulated results, predictive results from experimental-based and simulation-based surrogate models, and their deviations from the measured results using percentage relative error in parentheses. It can be observed that the hybrid model predicted melt pool width with better accuracy than both simulation-based and experimental-based surrogate models.

<table>
<thead>
<tr>
<th>No. of Run</th>
<th>Laser power (W)</th>
<th>Scan speed (mm/s)</th>
<th>Measurement results</th>
<th>CFD simulation results</th>
<th>Experiment-based surrogate model</th>
<th>Simulation-based surrogate model</th>
<th>Hybrid model</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>100</td>
<td>200</td>
<td>177.26</td>
<td>146 (17.64%)</td>
<td>204.92 (15.60%)</td>
<td>150.24 (15.24%)</td>
<td>187.71 (5.90%)</td>
</tr>
<tr>
<td>2</td>
<td>150</td>
<td>400</td>
<td>166.96</td>
<td>132 (20.94%)</td>
<td>183.63 (9.98%)</td>
<td>151.91 (15.00%)</td>
<td>148.26 (11.20%)</td>
</tr>
<tr>
<td>3</td>
<td>150</td>
<td>1000</td>
<td>86.922</td>
<td>110 (26.55%)</td>
<td>84.13 (3.21%)</td>
<td>100.94 (16.13%)</td>
<td>95.21 (9.54%)</td>
</tr>
<tr>
<td>4</td>
<td>150</td>
<td>1400</td>
<td>85.735</td>
<td>101 (17.80%)</td>
<td>108.64 (26.72%)</td>
<td>104.52 (21.92%)</td>
<td>92.20 (7.54%)</td>
</tr>
<tr>
<td>5</td>
<td>195</td>
<td>800</td>
<td>126.3</td>
<td>120 (4.99%)</td>
<td>111.82 (11.46%)</td>
<td>118.23 (6.39%)</td>
<td>120.34 (4.72%)</td>
</tr>
<tr>
<td>6</td>
<td>195</td>
<td>1000</td>
<td>107.17</td>
<td>110 (2.64%)</td>
<td>92.53 (13.66%)</td>
<td>110.05 (2.69%)</td>
<td>114.23 (6.59%)</td>
</tr>
</tbody>
</table>

Average % relative error: 
- Experiment-based surrogate model: 13.45% 
- Simulation-based surrogate model: 12.89% 
- Hybrid model: 7.58%
Therefore, by integrating the physics-based and experimentally obtained data using the proposed hybrid modeling approach, melt pool widths can be predicted with better accuracy. It can be concluded that with a few experimental data, the overall prediction accuracy of physics-based numerical models can be improved through the proposed approach. Since the hybrid model is a data-driven approach, for new sets of input parameters, the melt pool widths can be predicted in a few seconds. Thus, the hybrid model is computationally efficient compared to pure physics-based models. Due to improved accuracy and computational efficient, the proposed hybrid model better suited for real-time monitoring, control and optimization.

5. FUTURE WORK: AM ONTOLOGY FOR MODEL SELECTION
Our case study demonstrated the ability to create a hybrid model for a specific application, with comparable parameter sets. However, the context under which models are run and experiments are conducted is not always as straight forward. To help resolve potential discrepancies in the data sources, we propose there is a role for ontologies. Ontology can be used to capture the rapidly evolving knowledge of the AM process, computational models, uncertainty sources, and design for AM in an organized structure to help users to interoperate and reuse information [47,49,61]. As stated previously, there are numerous physics-based models ranging from low-fidelity to medium-fidelity to high-fidelity. A well-founded AM ontology can be used to capture the complex interconnections between these models as well as data-driven models and provide useful information for model composition towards developing a more accurate and reliable predictive metamodel [62,63]. In this section, we introduce the conceptual use of AM ontology for providing structured information of the different AM models to help develop a more accurate hybrid model.

An ontology consists of various concepts and correlations among entities. The AM ontology offered in Moges et al. [47]
captures the different features of AM models including assumptions, considered and neglected phenomena, model inputs and outputs, and uncertainty sources. The ontology attempted to capture these features at the five stages of the process: powder layer formation, laser-heat source interaction, melt pool formation, solidification, and residual stress formation. Different physics-based models have been developed to simulate each of these stages of the process. For instance, to simulate the melt pool behavior in L-PBF, there are Rosenthal-based analytical models, FEM thermal models, path-level thermal simulation model [64], CFD models, and Lattice Boltzmann method (LBM) models [14]. Similarly, to determine the amount of absorbed energy by the powder particles, there are different methods including radiation transfer, ray tracing, and Beer-Lambert approach [14]. Since these models are developed based on different methods and formulations, they have different pros and cons in terms of prediction accuracy and computational time. Even some models predict more accurately at a specific region of a PV map than the other regions.

The hybrid model developed in this paper uses computationally predicted data obtained from one physics-based model namely CFD model. It was shown in Section 4 that this hybrid model improves the predictive accuracy of the melt pool width compared to the original physics-based model. However, in order to further improve the accuracy, instead of using simulation data obtained from a single physics-based model, a hybrid model that uses different simulation data obtained from different physics-based models could be more reliable. For this, a systematic approach that leverages the AM ontology is needed to select the simulation data obtained from different physics-based models and integrate them with experimental data to build a more precise hybrid model.

In the future, a more reliable approach need to be developed for (a) investigating the capability of different physics-based models and data-driven techniques, (b) developing strategic approaches to select the different simulation data, (c) integrating the different simulation data with the experimental data using various data-driven methods. Hence, ontology can be an essential tool for strategically selecting models based on their inherent key features. Furthermore, incorporating different data-driven techniques into the existing AM ontology will enable us to select and apply a more suitable technique for developing a hybrid model to predict other output QoIs.

6. CONCLUSION
In this study we proposed a methodology to integrate physics-based data and experimentally measured data into a hybrid model that enables fast and accurate predictions in the L-PBF process. We proposed a hybrid model which comprised of data generated from both numerically predicted and experimentally recorded melt pool widths for various combinations of laser power and scan speed. The numerical results were obtained from a thermal CFD model and ex-situ cross-sectioning was used to gather the experimental results.

In our hybrid modeling approach, we first constructed an initial simulation-based surrogate model using simulation data by applying the polynomial regression method. Then, we applied the Kriging method to model the residual error between the simulated and experimental results using an adaptive modeling method to reduce predictive errors of the hybrid model. The performance of the proposed model is evaluated by comparing predicted results of melt pool widths from a CFD model, a simulation-based surrogate model, an experimental-based surrogate model, and the hybrid model against experimentally measured data. The results showed that on average, the hybrid model had the highest accuracy out of all the models. Additionally, due to improved accuracy and computational efficiency, the proposed hybrid model better suited for real-time process control.

In order to further improve the accuracy of the hybrid model, instead of using data obtained from a single physics-based model, integrating data obtained from multiple physics-based models with experimentally measured data could be more reliable. To address this, AM ontology can be used as an essential tool to help understand the capability of the different physics-based models and data-driven techniques and select the most accurate model at a specific region in the PV map. In the future, the knowledge captured in AM ontology can be leveraged for developing a more accurate hybrid modeling approach by providing predictive capabilities of multiple models. We view this work as a step towards effectively and efficiently utilizing physics-based models, data-driven approaches, experimental-based measurement data, and AM ontology to build reliable and robust predictive models that can be applied for L-PBF process control.

Acknowledgment
The authors gratefully acknowledge Jason Fox from the Engineering Laboratory at National Institute of Standards and Technology (NIST) for conducting experiments and providing measurement data.

DISCLAIMER
Certain commercial systems are identified in this paper. Such identification does not imply recommendation or endorsement by National Institute of Standards and Technology (NIST); nor does it imply that the products identified are necessarily the best available for the purpose. Further, any opinions, findings, conclusions, or recommendations expressed in this material are those of the authors and do not necessarily reflect the views of NIST or any other supporting U.S. government or corporate organizations.

References

12 Copyright © 2020 by ASME


