

# Gaussian process-based surrogate modeling framework for process planning in laser powder-bed fusion additive manufacturing of 316L stainless steel

Gustavo Tapia<sup>1</sup> · Saad Khairallah<sup>2</sup> · Manyalibo Matthews<sup>3</sup> · Wayne E. King<sup>3</sup> · Alaa Elwany<sup>1</sup> 

Received: 24 April 2017 / Accepted: 30 August 2017 / Published online: 22 September 2017  
© Springer-Verlag London Ltd. 2017

**Abstract** Laser Powder-Bed Fusion (L-PBF) metal-based additive manufacturing (AM) is complex and not fully understood. Successful processing for one material, might not necessarily apply to a different material. This paper describes a workflow process that aims at creating a material data sheet standard that describes regimes where the process can be expected to be robust. The procedure consists of building a Gaussian process-based surrogate model of the L-PBF process that predicts melt pool depth in single-track experiments given a laser power, scan speed, and laser beam size combination. The predictions are then mapped onto a power versus scan speed diagram delimiting the conduction from the keyhole melting controlled regimes. This statistical framework is shown to be robust even for cases where experimental training data might be suboptimal in quality, if appropriate physics-based filters are applied. Additionally, it is demonstrated that a high-fidelity simulation model of L-PBF can equally be successfully used for building a surrogate model, which is beneficial since simulations are getting more efficient and are more practical to study the response of different materials, than to re-tool an AM machine for new material powder.

**Keywords** Additive manufacturing · Laser powder-bed fusion · 316L stainless steel · Gaussian processes · Bayesian statistics

## 1 Introduction

Additive manufacturing (AM) technologies have evolved over the past two decades from being limited to producing prototypes geared towards accelerating the product development cycle to the production of end-use parts in a range of applications such as aerospace, biomedical, and defense. Among the key drivers of this evolution are the research and technological advances in metal-based AM technologies which enabled the fabrication of direct end-use parts from a variety of metallic alloys such as stainless steels, titanium alloys, and nickel-based super alloys [1, 2]. Seven sub-categories of additive manufacturing technologies have been defined to date, four of which are capable of producing metallic parts [3]. Among these, powder-bed fusion processes tend to be the most common due to their capability of producing parts with improved density, resolution, and surface finish that require less post-processing compared to other processes such as binder jetting [1, 4].

Laser powder-bed fusion (L-PBF) AM processes employ a high-energy laser beam to selectively fuse fine metallic powder particles in a layer-by-layer fashion. Although the mechanism appears to be simple, the challenge lies in the underlying complex physical phenomena involved in the process such as rapid melting, evaporation, solidification, recoil, and re-heating upon successive passes of the laser beam within the same layer or across successive layers. This makes the end part susceptible to defects like porosity, residual stress, and micro-cracks due to the very high thermal gradients and cooling rates. These defects can

---

✉ Alaa Elwany  
elwany@tamu.edu

<sup>1</sup> Department of Industrial and Systems Engineering, Texas A&M University, College Station, TX, USA

<sup>2</sup> Engineering Directorate, Lawrence Livermore National Laboratory, Livermore, CA, USA

<sup>3</sup> Physical and Life Sciences Directorate, Lawrence Livermore National Laboratory, Livermore, CA, USA

be detrimental to the mechanical properties of the parts. Moreover, the current limited understanding of the underlying physical phenomena hinders our ability to predict the microstructure and properties of the parts in an effort to ensure that they meet design specifications. This is further exacerbated because some of these phenomena occur over multiple lengths and time scales.

Due to the complexity of L-PBF processes, the majority of recent road-mapping efforts have strongly emphasized the need for developing modeling and simulation tools to foster our understanding of the process and ultimately serve as predictive tools to optimize the process and mitigate defects [5–8]. Indeed, many research efforts have been conducted to address this technical need, many of which have focused on modeling and predictive simulation using finite element methods (*FEM*). Modeling and simulation efforts using *FEM* are summarized and reviewed in King et al. [9]. Other research efforts followed different modeling approaches, such as [10–14] that have focused on developing high-fidelity mesoscopic simulation models to study the physics of complex melt flow and the mechanisms that drive the formation of pores and spatter in L-PBF. The key challenge in using both *FEM* models and high-fidelity models is the computational burden associated with running the simulations. Although these models are still invaluable for understanding the physics of the process, their direct use in process optimization is impractical and sometimes unfeasible.

Other works followed an experimental approach towards process optimization to circumvent these computational challenges. For example, King et al. [15] present experimental observations of keyhole mode laser melting in L-PBF and identify process conditions under which favorable conduction controlled melting occurs. The underlying challenges with experimental approaches are the fact that L-PBF experiments are time- and cost-intensive. Furthermore, in most cases experimental approaches are not machine- or material-agnostic. Changing the material system or the build platform requires a new set of experiments. Ideally, a combination of physics-based simulation models with experimental data is desired to conduct effective process optimization.

An attractive alternative that can be used to reduce some of the hurdles associated with computationally expensive simulation models and costly experiments is through the use of surrogate modeling. This is a key functional area in the field of uncertainty quantification (*UQ*) that focuses on constructing computationally efficient approximations or surrogates that can be used in lieu of the original simulation model (i.e., a model of the model [16]). In the case of L-PBF, a carefully constructed surrogate model can be used to replace a *FEM* model for instance to generate the sufficiently large number of simulations needed to draw

meaningful insight into the L-PBF process or to conduct process optimization. Moreover, in addition to the original focus of surrogate modeling on approximating physics-based simulation models, it can also be leveraged as an effective tool to construct response surfaces from experimental data as opposed to simulations. These response surfaces can be subsequently used to predict the process output at other unobserved experimental settings. In other words, given a limited number of experimental observations with finite accuracy at particular settings, a surrogate model can be used to predict the output at other settings of interest instead of conducting further experiments.

In this paper, we focus on developing surrogate models to enhance L-PBF process optimization. More specifically, we start by constructing an efficient Gaussian process-based (*GP*) surrogate for a high-fidelity physics-based model proposed in [9, 10, 17] to study the physical mechanisms of AM processes. Next, we use predictive Gaussian process models to construct a surrogate response surface model for experimental data acquired from measuring the depth of the melt pool in L-PBF processes. The melt pool is the region at the laser-powder interface at which metallic powder particles fuse to form a pool of molten metal then solidify after the laser beam moves to another location. The depth of the melt pool ( $d$ ) has been of much interest in several prior studies since it gives an indication of how well successive layers bond to one another [18], as well as a way to prevent keyhole mode [15]. Careful selection of the combination of laser power ( $P$ ), scanning speed ( $v$ ), and beam size ( $\sigma$ ) is needed to achieve the latter [15], which in turn prevents pore defects in fabricated parts. Our Gaussian process-based response surface model leverages single track experimental data to determine processing windows that meet these conditions while reducing the need for large numbers of experimental observations.

A Gaussian process-based surrogate modeling framework is employed in this work, motivated by the attractive properties offered by Gaussian processes such as analysis and quantification of uncertainty of functions, appealing mathematical and computational properties based on the broad in-depth knowledge of multivariate statistics and Gaussian distributions, flexibility and richness in modeling dependence among data observed in space, and the ability to incorporate a wide range of smoothness assumptions [19–21]. It is important to point out that a wide variety of powerful data-driven predictive modeling techniques exists, such as artificial neural networks, support vector machines, and logistic regression. These techniques are more commonly used in the context of the classification problem in machine learning [22], and typically rely on the availability of datasets with large enough sizes, since with small datasets there may exist gaps between samples, or only limited different classification cases may be provided [23].

Furthermore, GP models are characterized by the ability to quantify uncertainty in model predictions, and have been the most commonly used surrogate modeling tool [24].

The paper is organized as follows: we provide a high-level overview of the statistical UQ framework in Section 2. The core of the paper is presented in Sections 3, 4, and 5, where all these models and theories are applied to experimental and simulated L-PBF data to create statistical surrogate models and processing windows. These statistical models, results, and validations are reported before finalizing the paper with conclusions and future work directions in Section 6.

## 2 Statistical surrogate modeling framework

Our surrogate modeling approach is centered on Gaussian processes modeling. GP has been used for surrogate modeling of large-scale simulation models [25–30]. In this section, we briefly present the theory underlying these models and present the details of a Bayesian framework for calculating predictions, confidence intervals, and conducting model validation. Readers interested in the application of GP-based surrogate modeling in L-PBF without the underlying details of the statistical model can proceed directly to Section 3.

### 2.1 Gaussian processes

The GP model is a non-parametric statistical model in which a stochastic process  $f(\cdot)$  is assumed to have all of its finite-dimensional distributions as multivariate normal [31]. In other words, the joint probability distribution of the outputs from the stochastic process at any finite set of inputs  $\mathbb{X} = \{\mathbf{x}_1, \dots, \mathbf{x}_n\}$  is modeled as an  $n$ -dimensional multivariate normal distribution:

$$p(f(\mathbf{x}_1), \dots, f(\mathbf{x}_n)) \sim N_n(\boldsymbol{\mu}, \mathbf{C}), \tag{1}$$

where the mean vector  $\boldsymbol{\mu}$  and the covariance function  $\mathbf{C}$  are defined by a mean function  $\mu(\cdot)$  and covariance function  $C(\cdot, \cdot)$ , respectively, with the following properties:

$$\begin{aligned} \mu(\mathbf{x}_i) &= \boldsymbol{\mu}_i = \mathbb{E}[f(\mathbf{x}_i)] \\ C(\mathbf{x}_i, \mathbf{x}_j) &= \mathbf{C}_{i,j} = \text{cov}[f(\mathbf{x}_i), f(\mathbf{x}_j)]. \end{aligned}$$

Within this context, we will denote a Gaussian process as  $f(\cdot) \sim \text{GP}(\mu, C)$ . A further and detailed explanation on this type of stochastic processes is provided in references [19, 31].

### 2.2 Gaussian process-based statistical model

We start by defining some notation to adequately describe the model. Let  $Y$  be the quantity of interest (QoI); that is the

output of the simulation model or process for which we wish to build a surrogate model. Our task is to find a function  $f$  such that  $f: \mathbf{x} \rightarrow Y$ . Formally,  $Y(\mathbf{x}) \in \mathbb{R}$  is a univariate output of the model or process observed at a given input  $\mathbf{x} \in \mathcal{X} \subseteq \mathbb{R}^q$ , where  $\mathcal{X}$  is a  $q$ -dimensional study region or domain of interest. These inputs  $\mathbf{x}$  can be thought of as *locations* on the  $q$ -dimensional space.

The statistical model is defined as follows,

$$Y(\mathbf{x}) = f(\mathbf{x}) + \epsilon(\mathbf{x}), \tag{2}$$

where  $f(\cdot) \sim \text{GP}(\mu, C)$  is a Gaussian process that captures the dependence of outputs  $Y(\mathbf{x})$  at different locations  $\mathbf{x}$ , and  $\epsilon(\cdot)$  is a measurement error term that captures inherent noise associated with experimental measurements. It must be noted that when building a surrogate for a simulation model, no measurement error is present, and therefore, this term is disregarded.

The choices of mean and covariance functions  $\mu(\cdot)$  and  $C(\cdot, \cdot)$ , respectively, are important since they characterize the probability distribution on the outputs of the stochastic process  $f(\cdot)$  given in Eq. 1. Although providing guides for selecting these functions is outside of the scope of this study, the interested reader should refer to previous works like [19, 31, 32]. In the current study, we employ the approaches described in Higdon et al. [33, 34].

Under these settings, we define an  $n \times q$  input matrix  $\mathbf{X}$  and a corresponding  $n$ -dimensional output vector  $\mathbf{Y}$  as

$$\mathbf{X} = \begin{bmatrix} \mathbf{x}_1^\top \\ \mathbf{x}_2^\top \\ \vdots \\ \mathbf{x}_n^\top \end{bmatrix}, \quad \mathbf{Y} = Y(\mathbf{X}) = \begin{bmatrix} Y(\mathbf{x}_1) \\ Y(\mathbf{x}_2) \\ \vdots \\ Y(\mathbf{x}_n) \end{bmatrix},$$

which represent the dataset that has been acquired through simulation (or observed through experimental observations).

Since the  $q$  columns of the matrix  $\mathbf{X}$  typically have different units, we normalize the elements of the matrix to the unit hypercube  $[0, 1]^q$ . We also standardize the elements of the vector  $\mathbf{Y}$  to have mean 0 and variance 1 for mathematical convenience related to subsequent estimation of parameters [33]. In particular, with this step we establish a constant mean function  $\mu(\mathbf{x}) = 0$ , and only focus our attention to the covariance function and error term.

The role of the covariance function is to capture the spatial dependence between two different locations  $\mathbf{x}_i, \mathbf{x}_j \in \mathcal{X}$ . We employ a re-parametrized form of the well-known power exponential covariance function given by:

$$C(\mathbf{x}_i, \mathbf{x}_j) = \frac{1}{\lambda} \prod_{k=1}^q \delta_k^{4(x_{ik} - x_{jk})^2}, \tag{3}$$

where  $\lambda$  is the precision (also known as inverse variance) of the stochastic process  $f$ , and  $\boldsymbol{\delta} = \{\delta_1, \dots, \delta_q\}$  is a set of

parameters that control the strength or relevance of each of the input space dimensions. The resulting  $n \times n$  covariance matrix is calculated from the input set,

$$\mathbf{C} = [C(\mathbf{x}_i, \mathbf{x}_j)]_{i,j=1:n}.$$

The error terms  $\epsilon(\cdot)$  are commonly assumed to be independent and identically distributed (iid) normal random variables with zero mean and variance  $\frac{1}{\tau}$ ,  $\epsilon(\mathbf{x}) \sim N\left(0, \frac{1}{\tau}\right)$ . These error terms (sometimes known as the white noise component) are more relevant to the case of using the model to construct a response surface for experimental data to capture measurement errors. They are less relevant in the case of building a computationally efficient surrogate for a simulation model because we make the assumption that the simulation model is deterministic. In other words, running the simulation model multiple times using the same inputs will give the same output. Nonetheless, they are still occasionally used when building surrogates for deterministic simulation models in the case of large datasets to provide numerical stability in calculating the covariance matrices and their inverse [35].

### 2.3 Bayesian estimation

The statistical model given in Eq. 2 is fully defined by the set of model parameters  $\boldsymbol{\Omega} = \{\lambda, \boldsymbol{\delta}, \tau\}$ . Our task is to estimate these parameters given the observed dataset using a Bayesian framework, which has been previously used in predictive modeling for L-PBF [20, 21, 36].

Under the Bayesian scheme, model parameters are treated as random variables that follow a joint probability distribution  $p(\boldsymbol{\Omega})$ , thus, we estimate their probability distribution after having observed the data using Bayes rule:

$$p(\boldsymbol{\Omega}|\mathbf{X}, \mathbf{Y}) \propto p(\mathbf{Y}|\mathbf{X}, \boldsymbol{\Omega}) \times p(\boldsymbol{\Omega}), \quad (4)$$

where the left-hand side is termed the *posterior* distribution, the first term on the right-hand side is the *likelihood function*, and the last term is the *prior* distribution of the parameters  $\boldsymbol{\Omega}$ .

**Likelihood function** refers to the probability of observing the dataset  $\mathbf{Y}$  given the set of model parameters  $\boldsymbol{\Omega}$ . From the statistical model in Eq. 2, and the definitions of the mean function, covariance function, and error terms, it can readily be shown that the likelihood is given by

$$p(\mathbf{Y}|\mathbf{X}, \boldsymbol{\Omega}) \sim N_n(\mathbf{0}, \boldsymbol{\Sigma}), \quad (5)$$

with the covariance matrix  $\boldsymbol{\Sigma} = \mathbf{C} + \frac{1}{\tau} \mathbf{I}_n \mathbf{1}_{\{\epsilon(\cdot)\}}$ , where  $\mathbf{I}_n$  is the  $n \times n$  identity matrix and  $\mathbf{1}_{\{\epsilon(\cdot)\}}$  is an indicator function that is equal to 1 if the error term is included (in the case of modeling experimental data) and 0 otherwise (in the case of modeling simulation model output).

**Prior distribution** captures our knowledge or prior belief of the model parameters  $\boldsymbol{\Omega}$  before having observed any data. It is not uncommon to have cases where no previous knowledge is available, in which case the use of uninformative prior distributions is employed.

We first assume statistical independence among the model parameters,

$$\begin{aligned} p(\boldsymbol{\Omega}) &= p(\lambda, \boldsymbol{\delta}, \tau) \\ &= p(\lambda) p(\boldsymbol{\delta}_1) \cdots p(\boldsymbol{\delta}_q) p(\tau) \end{aligned} \quad (6)$$

and set individual distributions to each parameter.

Following Higdon et al. [33], we specify independent *gamma* distributions for the precision parameters  $\{\lambda, \tau\}$  and *beta* distributions for range (also known as scale) parameters  $\boldsymbol{\delta}$ .

$$\begin{aligned} p(\lambda) &\sim \text{Gamma}(a_\lambda, b_\lambda) \propto \lambda^{a_\lambda-1} \exp(-b_\lambda \lambda) \\ p(\tau) &\sim \text{Gamma}(a_\tau, b_\tau) \propto \tau^{a_\tau-1} \exp(-b_\tau \tau) \\ p(\boldsymbol{\delta}_i) &\sim \text{Beta}(a_\delta, b_\delta) \propto \delta_i^{a_\delta-1} (1 - \delta_i)^{b_\delta-1}, \quad i = 1, \dots, q. \end{aligned}$$

These choices of these prior distributions present some features that improve the estimation process. First, the gamma family of distributions is conjugate to the normal family in the likelihood function which provides computational advantages [20, 31]. Second, the variance of the process is expected to be close to 1 due to the standardization of vector  $\mathbf{Y}$ , hence we set  $a_\lambda = b_\lambda = 5$  to give more probability weight to values close to  $\lambda = 1$ . In contrast, we assign an uninformative prior distribution for  $\tau$  by setting  $a_\tau = 1$  and  $b_\tau = 0.0001$ , which will make  $\tau$  tend to have large values, equivalent to modeling low noise [33].

The covariance function in Eq. 3 provides further information beyond spatial dependence. In particular, if covariate  $i$  has no influence on the output, then parameter  $\delta_i$  should be estimated to be equal to 1. Thus, under this parameterization, we are also able to identify the inputs that really influence the process and these that do not (*active* or *inactive* inputs). Consequently, we set  $a_\delta = 1$  and  $b_\delta = 0.1$  in the prior beta distribution such that there is substantial prior mass near 1 [33].

**Posterior distribution** Combining Eqs. 4, 5 and 6, the posterior distribution of the model parameters  $\boldsymbol{\Omega}$  (up to a normalizing constant) can be calculated as follows:

$$p(\boldsymbol{\Omega}|\mathbf{X}, \mathbf{Y}) \propto |\boldsymbol{\Sigma}|^{-1/2} \exp\left(-\frac{1}{2} \mathbf{Y} \boldsymbol{\Sigma}^{-1} \mathbf{Y}\right) p(\lambda) p(\boldsymbol{\delta}_1) \cdots p(\boldsymbol{\delta}_q) p(\tau). \quad (7)$$

Since the probability distribution in Eq. 7 does not have an explicit form, we employ Markov Chain Monte Carlo (MCMC) methods to estimate it numerically.

### 2.4 Prediction

The end goal after building a statistical surrogate model is to make predictions of the QoI at new previously unobserved inputs. Since we use a GP-based model, we utilize the *Kriging* estimator (also known as the Best Linear Unbiased Predictor–BLUP) [37]. One of the benefits of the Kriging estimator is that it is not limited to providing a point estimate, but rather a predictive distribution which is important for constructing confidence intervals on the predictions.

It is well established in the literature that the predictive distribution of the model output (or QoI) at a previously unobserved input  $\mathbf{x}_0$  in a GP framework is normal with the following mean and variance:

$$\mathbb{E}[Y(\mathbf{x}_0) | \mathbf{X}, \mathbf{Y}, \boldsymbol{\Omega}] = \mathbf{h}^\top \boldsymbol{\Sigma}^{-1} \mathbf{Y}, \tag{8}$$

$$\text{var}[Y(\mathbf{x}_0) | \mathbf{X}, \mathbf{Y}, \boldsymbol{\Omega}] = C(\mathbf{x}_0, \mathbf{x}_0) - \mathbf{h}^\top \boldsymbol{\Sigma}^{-1} \mathbf{h} + \frac{1}{\tau} \mathbf{1}_{\{\epsilon(\cdot)\}}, \tag{9}$$

where,

$$\mathbf{h} = [C(\mathbf{x}_1, \mathbf{x}_0), \dots, C(\mathbf{x}_n, \mathbf{x}_0)]^\top.$$

### 3 Surrogate modeling of L-PBF experimental data

The aim of the current work, as stated in Section 1, is to provide an efficient alternative, through surrogate models, to computationally expensive physics-based simulation models or to costly L-PBF experiments. In this section, we will utilize the statistical surrogate model developed in Section 2 to construct a GP-based response surface for experimental L-PBF data.

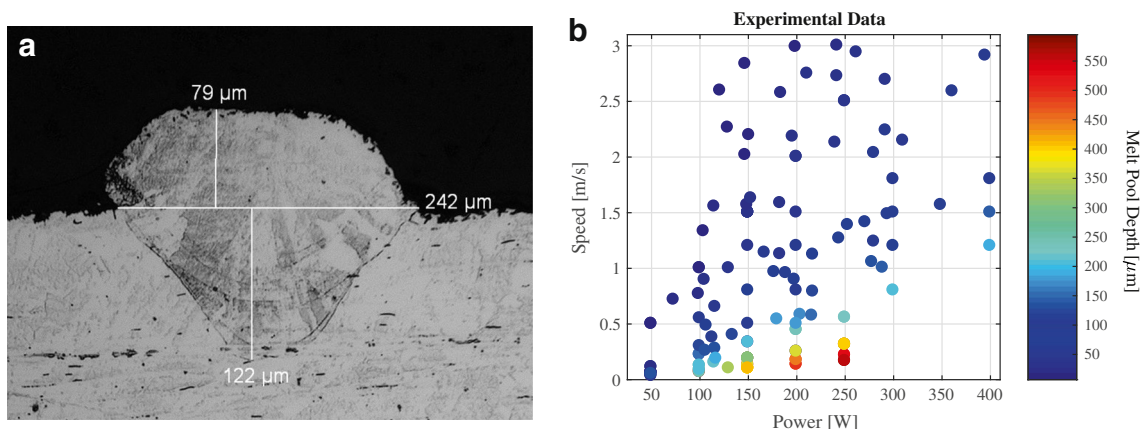
#### 3.1 Experimental datasets

The data used in this study represents experimental measurements of the melt pool depth from L-PBF single track

deposits of 316L stainless steel. In L-PBF, a single track is one single straight laser scan with a specific combination of laser power, scanning speed and laser beam size, on the very first powder-bed layer placed on top of the substrate plate. The melt pool is then defined as the region at the laser-powder interface where powder particles melt and then subsequently solidify to bond to the substrate plate (or to a previous layer when building a complete part). The depth of the melt pool is an important quantity of interest (QoI) that dictates the degree to which the current layer being processed bonds to the previous layer, and has been the focus in many previous studies [38–41]. In our experiments, single tracks of 316L stainless steel are first deposited under different process parameter settings, and then the substrate on which the tracks are deposited is sectioned, etched, and examined under an optical microscope to measure the depth of the melt pool.

We combined three different experimental datasets in this study. The first two datasets were acquired by coauthors of the current study and previously published in King et al. [15] and Kamath [35]. These are henceforth referred to as *King* and *Kamath* datasets respectively. An additional dataset was generated in the current study from similar single-track experiments conducted on a commercial Concept Laser M2 system, and will be distinguished as *Exp1*. Figure 1a shows a sample micrograph of a characterized single track from the experiments. The combined datasets include a total of  $n = 139$  melt pool depth measurements over a wide range of values for laser power  $P$  and scanning speed  $v$ , at laser beam size  $52 \mu\text{m}$  (see note below Table 1 regarding this measure). We have maintained a constant layer thickness of  $50 \mu\text{m}$  in this study since it is the manufacturer recommended value.

The color coded experimental melt pool depth measurements are depicted in Fig. 1b. Table 1 shows a summary of each dataset. An important note is the fact that only dataset *Exp1* included replications over the same combinations



**Fig. 1** Experimental melt pool depth measurements. **a** Optical micrograph of a characterized single track. **b** Data gathered from *King* [15], *Kamath* [35] and additional measurements *Exp1* made in this study

**Table 1** Summary of experimental datasets

Name	Source	$n$	$n_{\text{unique}}$	$P_{\text{min}}$	$P_{\text{max}}$	$v_{\text{min}}$	$v_{\text{max}}$	$\sigma$	$t$
<i>King</i>	[15]	52	52	73	395	0.15	3.0	52	50
<i>Kamath</i>	[35]	14	14	150	400	0.50	1.8	54	50
<i>Expl</i>	Measured	73	31	50	250	0.03	2.5	52	50

$n$  the number of data points in the dataset,  $n_{\text{unique}}$  the number of unique combinations of laser power and speed in the set,  $P$  laser power in W,  $v$  scanning speed in m/s,  $\sigma$  the width in  $\mu\text{m}$  of the laser beam with a Gaussian profile specified at 4 times its standard deviation (often named  $D4\sigma$ ),  $t$  the layer thickness in  $\mu\text{m}$

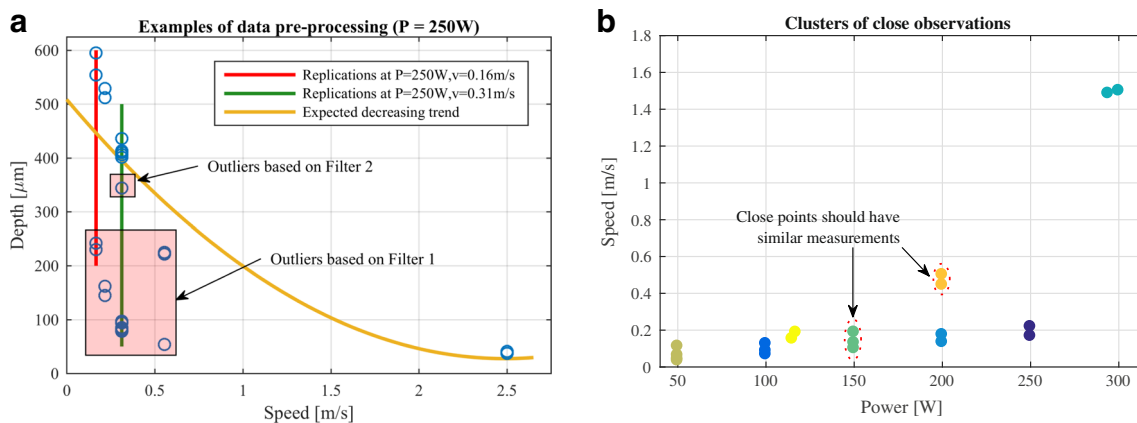
of process parameters, hence some data pre-processing is implemented and presented in the next subsection.

### 3.2 Data pre-processing

Since we combine different datasets with variations in experimental settings, we start by conducting pre-processing of the data before implementing the statistical model in order to account for anomalies. We start with the dataset *Expl*, the only dataset that includes replications, we analyze the data points for each different combination of power and speed, and remove obvious outliers. A point is flagged as outlier if the melt pool depth does not decrease with increasing scanning speed for a given constant laser power and beam size (Filter 1). Furthermore, within the same replication we flag points that lie outside a symmetric 10% range centered around the mean melt pool depth of that replication (Filter 2). It is very important to point out that since the collective data for this study (displayed in Table 1) is compiled from different studies, it was not feasible to revisit and examine the physical samples in order to identify the root cause of these anomalies. Hence, these ad hoc filters were devised based on domain knowledge of the process and expert judgment. More specifically, Filters 1 and 2 imply

that melt pool depth should decrease as scanning speed is increased at a fixed power setting, given the fact that less energy is being input into the powder. It remains true however that upon the availability of the physical samples, the reason for these anomalies and outliers should be carefully studied and identified. The implementation of these two filters are visually depicted in Fig. 2a.

A subsequent pre-processing step is to consider experimental data points with close values of L-PBF processing parameters. We define these as points that are within a circle with a pre-specified threshold radius in the  $P - v$  parameter space. We expect these points to have very similar melt pool depths, otherwise this flags an anomaly. To implement this filter, we first select a subset of data points with similar laser beam sizes. Next, we scale the  $P - v$  axes to the unit hypercube, and then flag those points that are within a 0.01-radius circle from one another. This area corresponds approximately to 7 W and 0.06 m/s. The identified data points are then clustered using a  $k$ -means unsupervised learning algorithm (see Fig. 2b). Finally, after determining these clusters, if the standard deviation of the points within a cluster is relatively low compared to its mean, then no change is needed. Otherwise: (1) if the cluster contains only two data points, the whole cluster is discarded, or (2) if the cluster



**Fig. 2** Demonstration of data pre-processing. **a** Flagging outliers based on Filter 1 and Filter 2 for a constant laser power  $P = 250$  W. **b**  $k$ -means clustering of points with close combinations of processing parameters. Different colors correspond to separate clusters found by the  $k$ -means algorithm

contains more than two data points, we iteratively identify and remove outliers following the same procedure previously explained for the replication case until low standard deviation is obtained or the whole cluster is discarded.

The end result of the pre-processing procedure outlined above consists of  $n = 96$  points with no replications ( $n_{\text{unique}} = 96$ ) after combining all individual datasets.

### 3.3 Constructing the response surface

A response surface is fitted to the pre-processed dataset using the model explained in Section 2. The posterior distribution of the model parameters  $\Omega$  is obtained using an adaptive Metropolis-Hastings MCMC algorithm with 25,000 iterations, burn-in period of 40%, and thinning every 5th iteration. An adaptive Metropolis-Hastings algorithm is used due to its capability of tuning up the parameters of the proposal distribution in order to improve mixing, exploration of the statistical parameters space, and convergence of the chain. After obtaining the posterior distribution, predictions were made over the whole range of the L-PBF processing parameters space as depicted in Fig. 3.

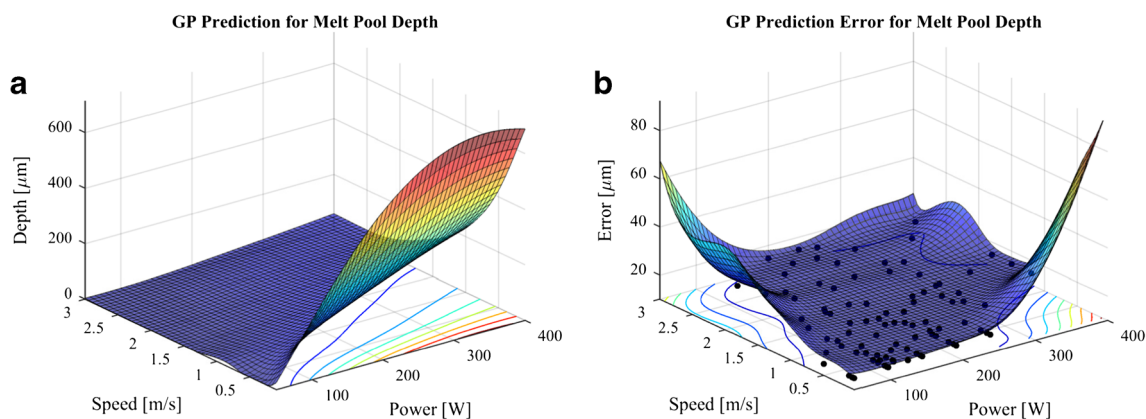
It can be seen that the standard deviation of the predictions is within a value of less than  $20 \mu\text{m}$  for most of the points within the training data points range. It is worth pointing out that areas with larger values of the standard deviation (rightmost corner in Fig. 3b) are attributed to extrapolation errors, which is a typical characteristic of GP predictions. The same holds true for the small peak observed at the  $P = 50 \text{ W}$  and  $v = 1 \text{ m/s}$  combination referring to Fig. 3a, since it can be seen that this combination lies outside the data range, which results in an expected extrapolation error represented by that small peak. It is common practice in GP predictions to overcome this by employing a space-filling sampling design such that observations are uniformly spread over the entire domain of interest.

### 3.4 Cross-validation of the statistical model

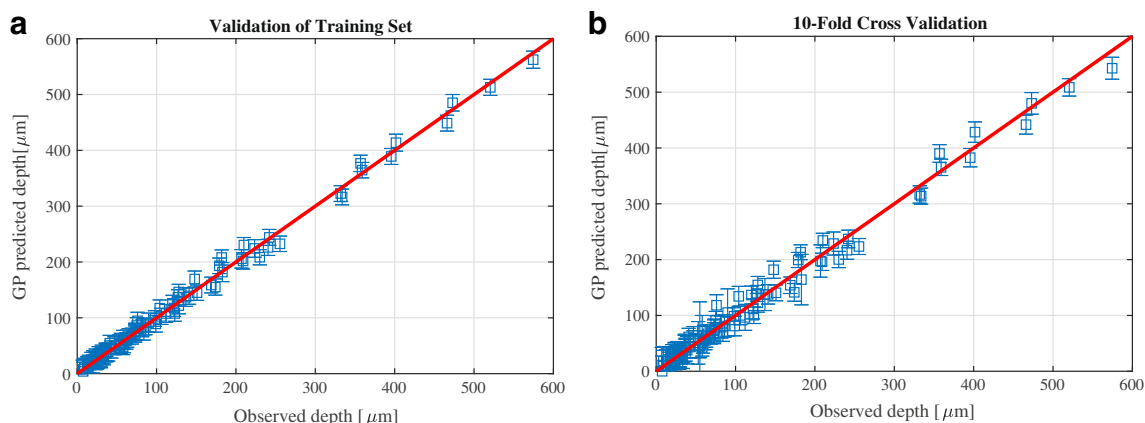
To assess the predictive performance of the model, we carry out cross-validation (CV). We denote the data used to estimate the parameters of the statistical model as the *training* dataset and keep a portion of the data, denoted as the *test* dataset, to test the prediction accuracy of the model by comparing predictions to the actual observed values. We employ *k-fold* cross-validation, which is an iterative procedure whereby the entire dataset is randomly partitioned into  $k$  disjoint subsets of approximately the same size. The model is then trained with all but one subset, which is kept as the *test* dataset. This process is then repeated until all  $k$  subsets have been left out once, and then the resulting prediction errors are evaluated. A special case is when  $k = n$ , called *Leave-One-Out* (LOO) cross-validation, where the model is trained with all but one data point.

Results from the validation of the training set and a 10-fold cross-validation are presented in Fig. 4. The plots represent a comparison between the experimental melt pool depth observations at a specific combination of laser power and speed (abscissa) and the predicted value calculated from the statistical GP model at the same combination of power and speed (ordinate). Ideally, we would want these two values to be equal, which corresponds to the red line on the plots show on Fig. 4a, b. Therefore, a preliminary visual indicator of model performance is how closely the points follow this ideal red line. To quantitatively assess model performance, we define the *mean absolute prediction error* (MAPE) as the average deviation from the red line, given by the following equation,

$$\text{MAPE} = \frac{1}{n} \sum_{i=1}^n |Y_{i,\text{obs}} - Y_{i,\text{pred}}|, \quad (10)$$



**Fig. 3** Predictions calculated from the GP model trained with the pre-processed dataset. **a** Mean value of the predictions over the processing parameter space. **b** Standard deviation of the predictions; black dots represent the locations of the training data points showing very low values of the standard deviation



**Fig. 4** Validation results for the GP model. **a** Training set validation results ( $MAPE_{train}$ ). **b** Tenfold cross-validation results ( $MAPE_{10CV}$ )

where  $n$  is the number of data points used for validation purposes,  $Y_{i,obs}$  is the observed value for data point  $i$ , and  $Y_{i,pred}$  is the GP predicted value for data point  $i$ .

Figure 4a plots the validation results using the whole training dataset of the GP model, and its corresponding predictions. As a first indicator, we want the MAPE for the training set ( $MAPE_{train}$ ) to be as close to zero as possible, indicating adequacy of the statistical model. In contrast, if this error is high a reformulation of the model might be required. In our case,  $MAPE_{train} = 6.023 \mu m$ , implying that the model is adequate and acceptable because this magnitude of error is comparable to errors of our measurement capabilities (i.e. optical microscopy). Furthermore, it can be visually seen that all of the points follow the ideal red line and the error bars are relatively narrow which confirms our result.

Figure 4b shows the results after carrying out a 10-fold cross-validation procedure. Notice that the  $i$ th data point in Eq. 10 is plotted in the graph only when its partition is left out (in other words, it was not used for training the model). Therefore, these results are interpreted as the ability of the model to generalize the process to unobserved combinations of laser power and speed. We see again for this case that the results closely follow the ideal red line, with a low cross-validation  $MAPE_{10CV} = 10.91 \mu m$ .

### 4 Surrogate modeling of a high-fidelity L-PBF simulation model

In Section 3, we presented a framework for using GP-based surrogate modeling to construct a response surface from L-PBF experimental observations. As mentioned earlier, surrogate modeling is also an effective tool for providing a computationally efficient and accurate approximation of computer simulation models with high computational burden. In this section, we employ GP-based framework to construct a surrogate model for a high-fidelity powder-scale simulation model developed in Khairallah et al. [10] to study the physics of complex melt flow in L-PBF processes. This model will be referred to as the *Powder Model* in the remaining of the paper. One simulation dataset was generated, denoted as *Sim1*, and it is summarized in Table 2 and shown in Fig. 5.

We implement the GP-based model, in a similar fashion to what was implemented in Section 3, using the *Sim1* dataset. Plots from Fig. 6 present the results after fitting and validating the GP model for this dataset of simulations.

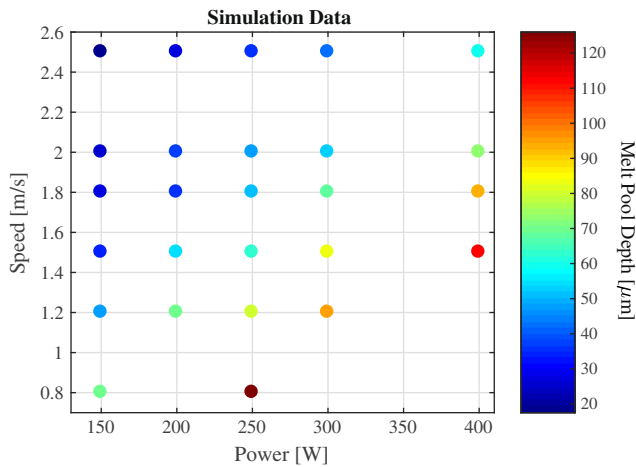
Our results reflect again that the GP predictive model represents a good approximation of the high-fidelity simulation model. An important observation is the behavior of the training set validation plot (Fig. 6c). All points lie directly

**Table 2** Summary of the simulation dataset

Name	Source	$n$	$n_{unique}$	$P_{min}$	$P_{max}$	$v_{min}$	$v_{max}$	$\sigma$	$t$
<i>Sim1</i>	PM	26	26	150	400	0.80	2.5	54	30

Same notation as explained in Table 1; *PM* denotes the Powder Model





**Fig. 5** Dataset showing simulated melt pool depth using the Powder Model

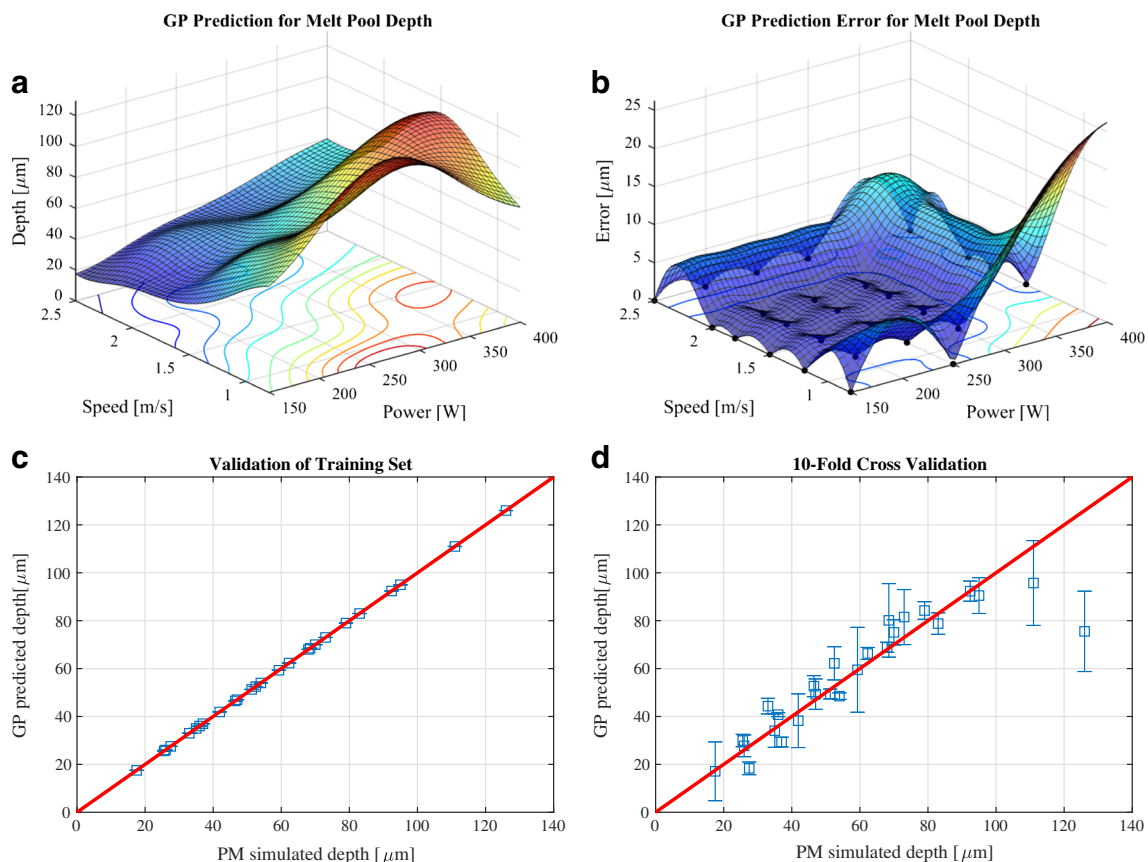
on the ideal red line,  $MAPE_{train} \approx 0 \mu m$  and no error bars are shown. This means that the model is predicting the exact same depths as the training dataset with no uncertainty in

predictions. As explained in Section 2.4, this is not a coincidence but rather a property of GP models, which represents a good verification of our developed framework.

The 10-fold cross-validation results also show good performance of the model,  $MAPE_{10CV} = 6.91 \mu m$ , and thus this surrogate model can be subsequently employed to make quick predictions, within a degree of confidence, for process optimization purposes. This is an important step forward since it would take less than 1 second to make predictions using the GP model as opposed to up to several days using the high-fidelity Powder Model even using high performance computing facilities.

### 5 L-PBF process planning

In this section, we will use the tools developed in previous sections for L-PBF process planning. More specifically, we will use two different criteria from the L-PBF literature to identify windows of processing parameters within which keyhole laser melting or thermal conduction modes occur. This is important since it has been suggested in the



**Fig. 6** Visualization of results for the GP model trained with simulation dataset *Sim1*. **a** Mean value of the predictions. **b** Standard deviation of the predictions; black dots represent the locations of the

training data points showing very low values of the standard deviation. **c** Training set validation results. **d** Tenfold cross-validation results

literature that keyhole mode conditions can potentially lead to the presence of voids and porosity in L-PBF parts [42–44]. Consequently, starting with identification of processing windows for 316L stainless steel, we would like to advocate the concept that every material should have a standardized datasheet that distinguishes the preferred conduction mode processing from the undesirable keyhole mode based on L-PBF parameters, such as a laser power–scanning speed phase diagram.

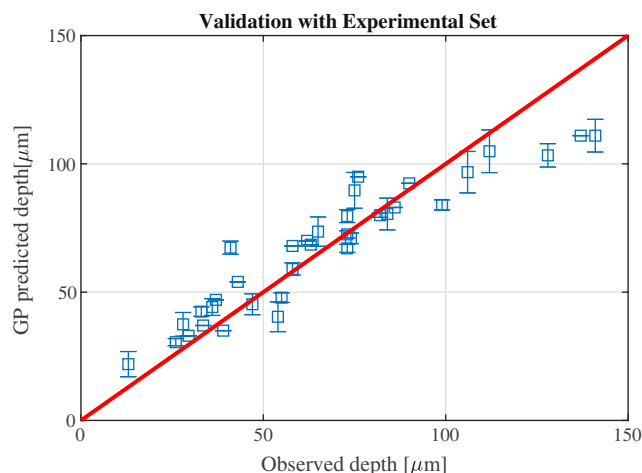
In this section, we make use of the statistical model that we presented previously, the experimental observations and high-fidelity computationally-expensive simulations to demonstrate consistency and validity of the UQ framework.

### 5.1 Validating the surrogate model of the high-fidelity simulation model with experimental observations

The statistical surrogate model from Section 4 (which was trained with L-PBF high-fidelity simulations and successfully cross-validated) is now validated against experimental observations using the pre-processed dataset described in Section 3. It is important to distinguish between the cross-validation conducted in Section 4, and the experimental validation being conducted in this section. Cross-validation refers to ensuring that the surrogate model predictions represent an accurate approximation of the predictions made using the high-fidelity simulation model. This does not involve any comparison with experimental observations. The validation in this section, on the other hand, refers to ensuring that the surrogate model predictions are in agreement with experimental observations. This step is important since the surrogate model will be used in lieu of the high-fidelity simulation model henceforth.

To proceed with this assessment, predictions from the surrogate model were calculated at the same L-PBF processing parameters from the pre-processed experimental dataset in Section 3. These predictions are then compared with the actual experimental measurements. The results are presented in Fig. 7 where the abscissa is the experimental measurement and the ordinate is the GP surrogate model prediction at the corresponding combination of L-PBF processing parameters. It should be noted that predictions were only calculated for experimental datapoints that lie within the domain of dataset *Sim1*, in order to avoid extrapolation error as a misleading indicator of inadequate performance.

It is clear from Fig. 7 that most points are close to the ideal red line, indicating that the surrogate model is in good agreement with the experiments. The computed mean absolute predictive error  $MAPE = 9.35 \mu\text{m}$ , confirming that the surrogate model trained with computationally-expensive simulations is able to adequately capture the L-PBF process and generalize over different settings. The error magnitude is acceptable since it is comparable to those of experimental



**Fig. 7** Comparison between predictions of GP surrogate model trained with high-fidelity simulations and experimental observations

measurement techniques. It is also important to point out that these validation results are achieved after using only a limited number of high-fidelity simulations to train the statistical surrogate model.

### 5.2 Processing windows

The final step in the methodology is to determine distinct windows of processing parameters within which keyhole laser melting or thermal conduction modes occur. Based on the works by King et al. [9, 15, 17] and King et al. [10], the depth of the molten pool is controlled by conduction of heat into the underlying solid material and vapor recoil pressure that creates a small depression. However, under high energy deposition, the mechanism of melting changes from conduction to keyhole laser melting. In this mode, the depth of the molten pool is controlled by evaporation of the metal and higher vapor recoil pressures. Keyhole mode laser melting results in melt pool depths that can be much deeper than observed in conduction mode. To build parts with good quality, it is desired to avoid deep keyhole mode during L-PBF fabrication.

King et al. [15] presented two different criteria to identify keyhole mode in L-PBF:

- When melt pool depth is equal or larger than half of the melt pool width, that is:  $d \geq \frac{w}{2}$ , or  $\frac{2d}{w} \geq 1$ .
- When normalized enthalpy is  $\frac{\Delta H}{h_s} \geq 30$ .

In the first bullet, the quantity  $\frac{2d}{w}$  can be thought of as a stochastic process due to the random nature of both melt pool depth and width, and therefore can be modeled through a GP statistical framework as in previous sections. Consequently, we can use prediction results from a statistical surrogate model based on  $\frac{2d}{w}$  to identify combinations of

laser power and speed (processing windows) within which either conduction or keyhole mode occur.

The second bullet follows from a physics-based analysis and derives the following equation,

$$\frac{\Delta H}{h_s} = \frac{AP}{\rho h_{sm} \sqrt{\pi D v} \left(\frac{\sigma}{4}\right)^3}, \tag{11}$$

where  $A$  is absorptivity of the laser into the material,  $\rho$  is density,  $h_{sm}$  is enthalpy at melting per unit of mass,  $D$  is thermal diffusivity, and processing parameters:  $P$  is laser power,  $v$  is scanning speed, and  $\sigma$  is laser beam size (beam width at 4 times its standard deviation).

Both criteria are taken into consideration in order to identify the processing windows. For the first case, we used the same experimental dataset introduced in Section 3, given that the single tracks experiments had been already characterized for melt pool width, in addition to depth. Therefore, the exact same process was followed, with the only difference that instead of training the statistical model with melt pool depth, it was fitted with the quantity  $\frac{2d}{w}$ . Figure 8a displays contour lines corresponding to different values of  $\frac{2d}{w}$  that can be readily used to identify processing with  $\frac{2d}{w} \geq 1$ .

For the second criterion, Eq. 11 was calculated over the same domain used in the first case, with the following values of constants for 316L stainless steel [15]:  $A = 0.4$ ,  $\rho = 7.98 \text{ kg/m}^3$ ,  $h_{sm} = 1.2 \times 10^6 \text{ J/kg}$  and  $D = 5.38 \times 10^{-6} \text{ m}^2/\text{s}$ . Additionally, the laser beam size was held constant at  $\sigma = 52 \mu\text{m}$  in order to keep consistency with the single-track experiments used. Figure 8b shows the corresponding contour map for normalized enthalpy at combinations of laser power and scanning speed and identifies those with  $\frac{\Delta H}{h_s} \geq 30$ .

Both panes from Fig. 8 aim to provide windows of L-PBF parameter settings that will produce melt pools within

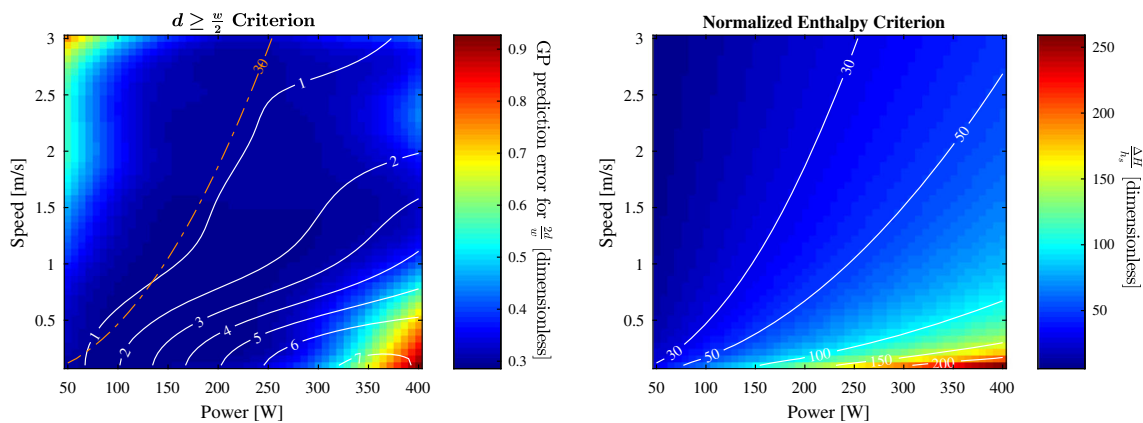
a desired thermal mode. It is clear that conduction-mode regions in both maps ( $\frac{2d}{w} < 1$  and  $\frac{\Delta H}{h_s} < 30$ ) are very similar for laser powers up to 225 W, which can be seen in Fig. 8a by comparing the dashed line with the contour line at level 1.

The normalized enthalpy criterion is derived from a physics-based perspective and it is a reduced order model that does not capture all the complicated physics and phenomena that occur during the L-PBF process. This is a strong reason for the disparity with the results from the  $\frac{2d}{w}$  criterion, whose nature is completely experimental. However, normalized enthalpy is useful as an indicator of acceptable consistency and strong performance of the whole GP statistical surrogate modeling framework.

Therefore, we finalize by defining results from Fig. 8a as the mentioned datasheet for L-PBF processing of 316L stainless steel for combinations of laser power and scanning speed, at laser beam size  $\sigma \approx 52 \mu\text{m}$  and layer thickness  $t = 50 \mu\text{m}$ .

Finally, we emphasize the following aspect regarding generality of the proposed framework: the GP model for determining processing windows was trained using single-track experiments, where the melt pool and, in turn, the thermal profile have attained a steady state. Therefore, the current model cannot be generalized to every material or scan pattern.

During an AM process, the conditions of melting can change due to thermal history effects, such as when doing adjacent tracks, overhangs, thin walls, etc. In the case when material properties are such that the heat is dissipated quickly (materials with high thermal conductivity) and the residual heat has no significant effect, single track data can help in narrowing the window of optimal process parameters (corresponding to the contour line at value 1 in Fig. 8a). In the case when the material properties or the scan pattern



**Fig. 8** L-PBF Process windows specifying conduction and keyhole mode behavior. **a** Using  $d \geq \frac{w}{2}$  criterion. The color scale on the right bar represents standard deviation of the GP predictions while the figures on the white contour lines display the mean value of the GP

prediction. The dashed line is superimposed on this plot to benchmark the result against the normalized enthalpy criterion. **b** Using normalized enthalpy  $\frac{\Delta H}{h_s} \geq 30$  criterion

is such that the thermal profile has not reached a steady state or is exhibiting residual thermal history effect, the model would need to be re-trained using data relevant to the new conditions.

In this study, we have dealt with the former case and formally validated our results. For the latter case, it would be an interesting and meaningful avenue to explore in the future, and one definitely within the capability of the proposed Gaussian process framework.

## 6 Conclusion

Process planning in L-PBF additive manufacturing for determining optimal process parameter settings has been the focus of many recent research efforts. Most of the existing approaches rely either on conducting large numbers of experiments to identify optimal parameter windows or on utilizing high-fidelity simulation models to assist in process planning. Both approaches are either cost- or time-intensive, and sometimes both. Statistical surrogate modeling can be an effective approach to alleviate some of these burdens by offering a computationally efficient alternative to expensive experiments or computationally intensive high-fidelity simulation models. We develop Gaussian process-based surrogate modeling to achieve this task, with limited available experimental data and high-fidelity simulations in L-PBF process.

The melt pool depth for a series of 96 single-track 316L stainless steel deposits at different processing parameters was first experimentally characterized. The data was used to create a GP-based response surface that can be used to predict melt pool depths at new unobserved parameter settings.

Next, a computationally efficient GP surrogate model was developed for a high-fidelity L-PBF simulation model using a training dataset of 26 simulations at different processing parameter settings. The surrogate model was cross-validated, showing good performance demonstrated by a low mean absolute predictive error  $MAPE \approx 6 \mu m$ .

Finally, the developed and validated surrogate modeling frameworks were used for L-PBF additive manufacturing process planning through identifying processing windows within which the melt pool at the laser-powder interface exhibits the desirable thermal conduction mode as opposed to the keyhole mode, based on two different criteria from literature. Both criteria resulted in similar processing windows which confirms the validity of our framework.

**Acknowledgments** This work was performed under the auspices of the U.S. Department of Energy by Lawrence Livermore National Laboratory under contract DE-AC52-07NA27344, and was partially supported by an Early Stage Innovations grant from NASA's Space

Technology Research Grants Program, Grant No. NNX15AD71G. This work was also partially funded through a Laboratory Directed Research and Development (LDRD) grant, Grant No. 15-ERD-037. LLNL Release No. LLNL-JRNL-726754.

## Compliance with Ethical Standards

**Conflict of interests** The authors declare that they have no conflicts of interest.

## References

1. Tapia G, Elwany A (2014) A review on process monitoring and control in metal-based additive manufacturing. *J Manuf Sci Eng* 136(6):060801
2. Gu DD, Meiners W, Wissenbach K, Poprawe R (2012) Laser additive manufacturing of metallic components: materials, processes and mechanisms. *Int Mater Rev* 57(3):133–164
3. American Society of Testing Materials (2012) ASTM F2792 - 12a: standard terminology for additive manufacturing technologies. Standard, ASTM. [Online]. Available from <http://www.astm.org/Standards/F2792.htm>
4. Wohlers TT, Wohlers Associates, Campbell RI, Caffrey T (2016) Wohlers Report 2016: 3D Printing and Additive Manufacturing State of the Industry: Annual Worldwide Progress Report. Wohlers Associates, USA. ISBN 9780991333226
5. The Minerals Metals & Materials Society (TMS) (2015) Modeling across scales: a roadmapping study for connecting materials models and simulations across length and time scales. TMS, Warrendale. ISBN 9780692376065. [www.tms.org/multiscalestudy](http://www.tms.org/multiscalestudy)
6. Frazier WE (2010) Direct digital manufacturing of metallic components: vision and roadmap. Direct digital manufacturing of metallic components: affordable, durable, and structurally efficient airframes, at Solomons Island. Austin, pp 9–11
7. National Institute of Standards and Technology (NIST) Measurement science roadmap for metal-based additive manufacturing, 2013 Online. Available from [https://www.nist.gov/sites/default/files/documents/el/isd/NISTAdd\\_Mfg\\_Report\\_FINAL-2.pdf](https://www.nist.gov/sites/default/files/documents/el/isd/NISTAdd_Mfg_Report_FINAL-2.pdf). Accessed 10 Jun 2015
8. Bourell DL, Leu MC, Rosen DW (2009) Roadmap for additive manufacturing: identifying the future of freeform processing. The University of Texas at Austin, Austin
9. King WE, Anderson AT, Ferencz RM, Hodge NE, Kamath C, Khairallah SA, Rubenchik AM (2015a) Laser powder bed fusion additive manufacturing of metals; physics, computational, and materials challenges. *Appl Phys Rev* 2(4):041304
10. Khairallah SA, Anderson AT, Rubenchik A, King WE (2016) Laser powderbed fusion additive manufacturing: physics of complex melt flow and formation mechanisms of pores, spatter, and denudation zones. *Acta Mater* 108:36–45
11. Megahed M, Mindt H-W, N'Dri N, Duan H, Desmaison O (2016) Metal additive-manufacturing process and residual stress modeling. *Integr Mater Manuf Innov* 5(1):1–33
12. Panwisawas C, Qiu C, Anderson MJ, Sovani Y, Turner RP, Attallah MM, Brooks JW, Basoalto HC (2017) Mesoscale modelling of selective laser melting: thermal fluid dynamics and microstructural evolution. *Comput Mater Sci* 126:479–490
13. Markl M, Körner C (2016) Multi-scale modeling of powder-bed-based additive manufacturing. *Annu Rev Mater Res* 46:1–34
14. Gürtler F-J, Karg M, Leitz K-H, Schmidt M (2013) Simulation of laser beam melting of steel powders using the three-dimensional volume of fluid method. *Phys Procedia* 41:881–886

15. King WE, Barth HD, Castillo VM, Gallegos GF, Gibbs JW, Hahn DE, Kamath C, Rubenchik AM (2014) Observation of keyhole-mode laser melting in laser powder-bed fusion additive manufacturing. *J Mater Process Technol* 214(12):2915–2925
16. Kleijnen JPC (1975) A comment on blanning's metamodel for sensitivity analysis: the regression metamodel in simulation. *Interfaces* 5(3):21–23
17. King WE, Anderson AT, Ferencz RM, Hodge NE, Kamath C, Khairallah SA (2015b) Overview of modelling and simulation of metal powder bed fusion process at Lawrence Livermore National Laboratory. *Mater Sci Technol* 31(8):957–968
18. Dai Donghua, Dongdong Gu (2014) Thermal behavior and densification mechanism during selective laser melting of copper matrix composites: simulation and experiments. *Mater Des* 55:482–491
19. Rasmussen CE, Williams CKI (2006) Gaussian processes for machine learning. MIT Press, Cambridge
20. Tapia G, Elwany AH, Sang H (2016) Prediction of porosity in metal-based additive manufacturing using spatial Gaussian process models. *Addit Manuf* 12:282–290
21. Tapia G, Johnson L, Franco B, Karayagiz K, Ma J, Arroyave R, Karaman I, Elwany A (2017) Bayesian calibration and uncertainty quantification for a physics-based precipitation model of nickel-titanium shape-memory alloys. *J Manuf Sci Eng* 139(7):071002
22. Friedman J, Hastie T, Tibshirani R (2001) The elements of statistical learning, vol 1. Springer Series in Statistics, New York
23. Mao R, Zhu H, Zhang L, Chen A (2006) A new method to assist small data set neural network learning. In: Sixth international conference on intelligent systems design and applications, ISDA06, 2006, vol 1. IEEE, New York, pp 17–22
24. O'Hagan A (2013) Polynomial chaos: a tutorial and critique from a statistician's perspective. *SIAM/ASA J Uncert Quantif* 20:1–20
25. Liu Pu, Lusk MT (2002) Parametric links among monte carlo, phase-field, and sharp-interface models of interfacial motion. *Phys Rev E* 66(6):061603
26. Büche D, Schraudolph NN, Koumoutsakos P (2005) Accelerating evolutionary algorithms with Gaussian process fitness function models. *IEEE Trans Appl Rev Syst Man Cybern Part C* 35(2):183–194
27. Christen A, Sansó B (2008) Advances in the design of Gaussian processes as surrogate models for computer experiments. Technical report, Tech. Report 5, University of California, Santa Cruz CA
28. O'Hagan A (2006) Bayesian analysis of computer code outputs: a tutorial. *Reliab Eng Syst Saf* 91(10):1290–1300
29. Conti S, Gosling JP, Oakley JE, O'Hagan A (2009) Gaussian process emulation of dynamic computer codes. *Biometrika* 96(3):663–676
30. Bastos LS, O'Hagan A (2009) Diagnostics for Gaussian process emulators. *Technometrics* 51(4):425–438
31. Gelfand AE, Diggle P, Guttorp P, Fuentes M (2010) Handbook of spatial statistics. CRC Press, Boca Raton
32. Cressie NAC (1993) Statistics for spatial data. Wiley, New York
33. Higdon D, Gattiker J, Williams B, Rightley M (2008) Computer model calibration using high-dimensional output. *J Amer Stat Assoc* 103(482):570–583
34. Higdon D, Kennedy M, Cavendish JC, Cafo JA, Ryne RD (2004) Combining field data and computer simulations for calibration and prediction. *SIAM J Sci Comput* 26(2):448–466
35. Kamath C (2016) Data mining and statistical inference in selective laser melting. *Int J Adv Manuf Technol* 1–19
36. Tapia G, Elwany AH (2015) Prediction of porosity in SLM parts using a MARS statistical model and bayesian inference. In: Proceedings of the solid freeform fabrication symposium. Austin, pp 1205–1219
37. Stein ML (2012) Interpolation of spatial data: some theory for Kriging, Springer Science & Business Media, New York
38. Gong H, Gu H, Zeng K, Dilip JJS, Pal D, Stucker B, Christiansen D, Beuth J, Lewandowski JJ (2014) Melt pool characterization for selective laser melting of Ti-6Al-4V pre-alloyed powder. In: Proceedings of the solid freeform fabrication symposium. Austin, pp 256–267
39. Thijs L, Verhaeghe F, Craeghs T, Van Humbeeck J, Kruth J-P (2010) A study of the microstructural evolution during selective laser melting of Ti-6Al-4V. *Acta Mater* 58(9):3303–3312
40. Yadroitsev I, Gusarov A, Yadroitsava I, Smurov I (2010) Single track formation in selective laser melting of metal powders. *J Mater Process Technol* 210(12):1624–1631
41. Kruth J-P, Froyen L, Van Vaerenbergh J, Mercelis P, Rombouts M, Lauwers B (2004) Selective laser melting of iron-based powder. *J Mater Process Technol* 149(1):616–622
42. Aboulkhair NT, Everitt NM, Ashcroft I, Tuck C (2014) Reducing porosity in AlSi10Mg parts processed by selective laser melting. *Addit Manuf* 1:77–86
43. Thijs L, Kempen K, Kruth J-P, Van Humbeeck J (2013) Fine-structured aluminium products with controllable texture by selective laser melting of pre-alloyed AlSi10Mg powder. *Acta Mater* 61(5):1809–1819
44. Qiu C, Panwisawas C, Ward M, Basoalto HC, Brooks JW, Attallah MM (2015) On the role of melt flow into the surface structure and porosity development during selective laser melting. *Acta Mater* 96:72–79