

Research paper

Data-augmented modeling in laser powder bed fusion: A Bayesian approach

Peter Morcos^{a,1}, Brent Vela^{a,*,1}, Cafer Acemi^a, Alaa Elwany^{a,b}, Ibrahim Karaman^a,
Raymundo Arróyave^{a,b,c}

^a Department of Materials Science and Engineering, Texas A&M University, College Station, TX, USA

^b Department of Industrial and Systems Engineering, Texas A&M University, College Station, TX, USA

^c J. Mike Walker '66 Department of Mechanical Engineering, Texas A&M University, College Station, TX, 77840, USA

ARTICLE INFO

Keywords:

Bayesian updating

Information fusion

High throughput

Laser powder bed fusion

Processing maps

ABSTRACT

The laser powder bed fusion (LPBF) technique has become increasingly prominent in metal additive manufacturing. However, tuning parameters for printing defect-free parts requires labor-intensive experimental work and computationally expensive simulations. Moreover, to calibrate LPBF models against the experimental data, typically MCMC methods or similar methods are used which is also time-consuming. These procedures are viable when calibrating LPBF models against data for individual chemistries but are not efficient for alloy design. A rapid method to calibrate LPBF models is needed to design for printable alloys. We address this challenge by integrating a low-fidelity analytical thermal model, a machine learning model, and proxy experimental data to create an accurate and rapidly-trained model that leverages the principles of Bayesian updating. As a case study in 'printability extrapolation', a dataset of 195 single-tracks on 16 unique chemistries was used to probe the method's ability to predict melt-pool dimensions on 'unseen' chemistries. As a case study in 'printability interpolation' the framework was deployed on two compositions that were studied rigorously in the literature for their printability, namely, the ultra-high strength martensitic steel alloy AF9628 and the nickel super alloy 718. The interpolative/extrapolative abilities of the proposed method were compared to a set of 4 control models under data sparse conditions.

1. Introduction

Among the different additive manufacturing (AM) technologies, laser-powder bed fusion (L-PBF) has gained interest for the manufacturing of metallic parts for various applications [1,2]. In L-PBF, parts are manufactured from alloy powder in a layer-by-layer manner. After depositing each layer using a recoater or a roller blade, a laser beam source selectively melts powder particles forming a thin layer of the 3D part. Upon depositing all layers, the 3D part is manufactured.

For industrial applications, attaining high-quality parts via laser-powder bed fusion (L-PBF) requires a meticulous selection of processing parameters to avoid defects and ensure defect-free outcomes. Common defects in L-PBF include keyholing, lack of fusion, and balling, all of which contribute to porosity, adversely affecting layer adhesion and the mechanical integrity of the as-printed product. To preemptively identify unsuitable processing parameters, single-track experiments are a valuable preliminary screening method. These initial tests are crucial for detecting potential issues in printing parameters before committing resources to produce larger test pieces, such as tensile coupons or cubes.

During L-PBF AM, melt pool dimensions are crucial for assessing print quality – most porosity-induced defects can be correlated with the characteristics of the melt pool – and are typically measured through optical or scanning electron microscopy (SEM). The width of the melt pool, often measured first due to simpler sample preparation, is considered the 'easily seen/measured' dimension of the melt pool. In contrast, measuring the depth of the melt pool requires cross-sectioning, which is more costly and complex, leading to its classification as an 'unseen' dimension. Consequently, one objective of this study is to simplify melt pool depth quantification. Furthermore, a more accurate assessment of the melt pool characteristics under AM could in turn be used to a more efficient determination of the regions in the processing space more susceptible to the formation of macroscopic solidification defects.

Various works have noted a strong linear correlation between melt pool widths and depths [3,4], however to-date only Ref. [3] has exploited this correlation to use melt pool width as a proxy measurement for depth. Both approaches use in-situ monitoring of the melt pool width to predict depth. Ref. [4] uses an analytical model to predict the melt pool aspect ratio to convert width measurement to depth.

* Corresponding author.

E-mail address: brentvela@tamu.edu (B. Vela).

¹ Authors contributed equally.

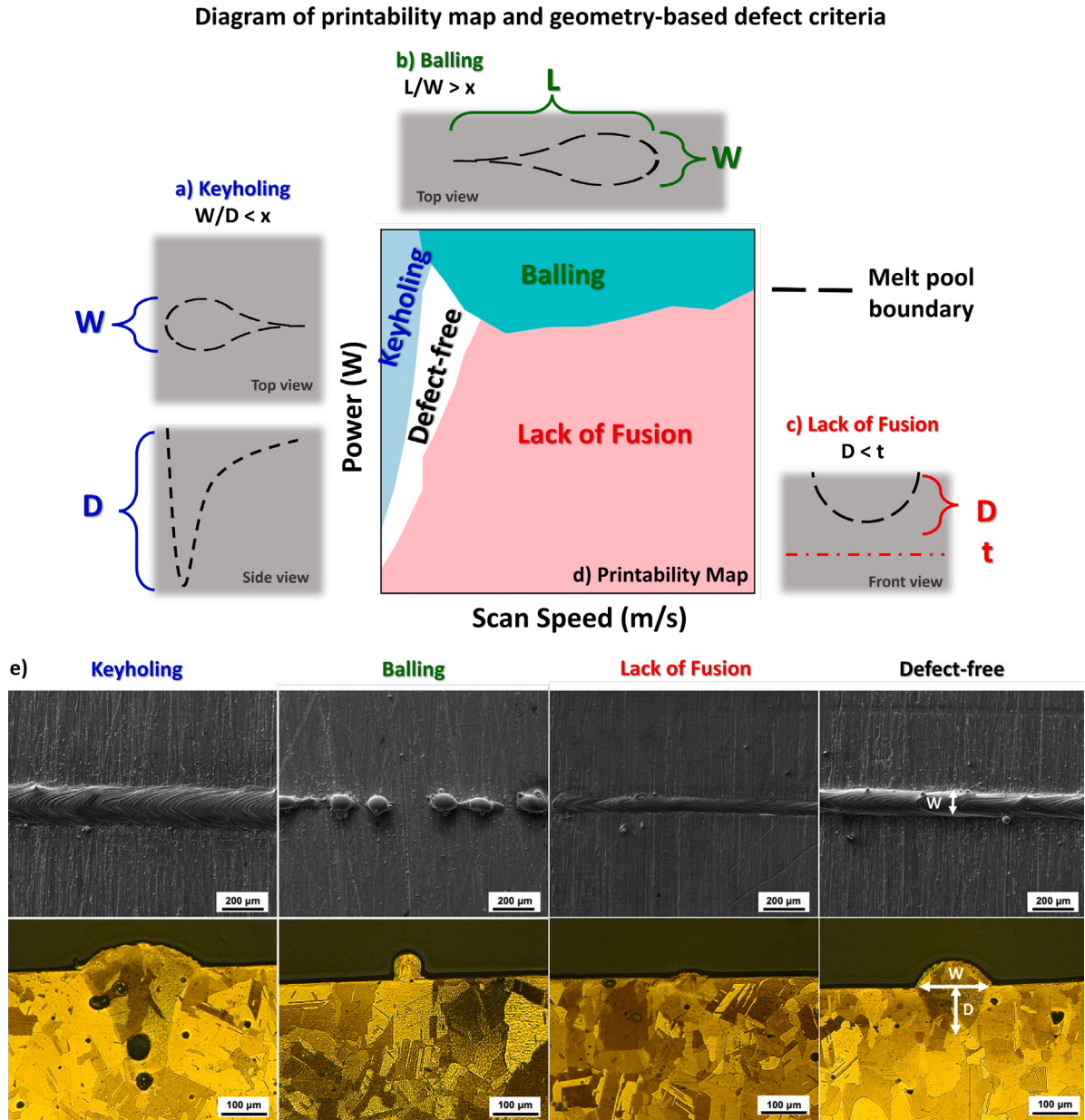


Fig. 1. Example of a printability map and the defect criteria used to generate printability maps. (a) Diagram of a melt pool that results in a keyhole. The width-to-depth aspect ratio (W/D) is small indicating a keyhole. The exact threshold for keyholing, x , is determined experimentally. (b) Diagram of a melt pool that results in balling. The length-to-width aspect ratio (L/W) is large, indicating an elongated melt pool. Elongated melt pools are susceptible to Rayleigh instability which results in balling. The exact threshold for keyholing, x , is determined experimentally. (c) Diagram of a melt pool that results in lack of fusion. The depth (D) is less than the specified layer thickness (t), indicating that there will be insufficient binding between layers during LPBF. (d) The complete printability map that shows porosity defects for a composition as a function of power and scan speed. (e) Experimental images of porosity defects. The experimental examples were obtained and reproduced from Ref. [5].

Ref. [3] relates in-situ melt pool signatures (including width) to depth via an SVM machine learning model. In this work, we take a Bayesian approach to this concept, relating width to depth via Hierarchical Gaussian Process regression (HGPR) instead. The benefits of this method are detailed in Section 2.1.

In terms of the usefulness of these measurements, once the melt pool dimensions are determined, they are employed to categorize a single-track as either defect-free or not based on various defect criteria, as shown in Fig. 1a–c. Typically, these defect criteria are based on the melt pool geometry, i.e. width, depth, and length. Experimental images of these porosity defects (modified from [5]) are shown in Fig. 1. After classifying multiple single-tracks printed with different processing parameters, a printability map can be generated. A printability map (or processing map) is a visual tool that delineates defect-containing and

defect-free regions in the processing parameter design space. Fig. 1d shows a typical representation of a printability map. As shown in Fig. 1d, the processing space is often defined by the most two significant process parameters; laser power and scanning speed. This makes it easier to visualize printability maps as 2D plots.

As demonstrated in previous studies, single-tracks serve as useful proxy experiments for more complex builds, such as cube and block builds, in terms of screening for porosity defects [6,7]. While we acknowledge that single-tracks do not capture the full thermal histories associated with multi-layer builds, they do provide a first approximation of printability and can effectively narrow down processing parameters that may lead to defect-free builds. This is particularly useful in the scenario of alloy design, where there is a need for high-throughput models that can efficiently screen large alloy spaces for

compositions that are resistant to porosity defects, including keyholing, balling, and lack of fusion. For more information between the relationship between single-track and multi-layer builds, we refer readers to Refs. [2,6].

Two approaches have been used in the literature for generating printability maps: data-driven approaches and *a priori* physics-based approaches.

Data-driven approaches rely on experimental measurements and observations to determine the boundaries of different defect regions via interpolation and/or machine learning models. Morcos et al. [8] reviewed the printability maps for W and W-based alloys and found that most printability maps are data-driven and do not make use of machine learning models for interpolation and extrapolation [9–11]. On the other hand, Wilkinson et al. [12] and Wentai et al. [13] used experimental data from Ti-6Al-4V single tracks to train an unsupervised convolutional neural network model (CNN) while Zhang et al. [14] treated printability maps as a classification problem and used support vector machine (SVM) model to delineate the processing space into 4 different regions based on experimental data. Likewise, in Ref. [15] the authors review commonly used machine learning models used in the context of AM. While many of these models perform well, these models have significant drawbacks when compared to Bayesian models.

Bayesian machine learning methods offer several distinct advantages: Bayesian methods have the unique capability to systematically improve prior models as new data becomes available, making them particularly suitable for integrating analytical models with experimental data. This is important for machine learning in data-sparse conditions where experimental data must be supplemented by prior knowledge. Alloy design for AM is one such data-sparse scenario as designers must consider a large pool of candidate compositions and processing conditions using modeling trained on very little experimental data [16,17]. Furthermore, of the machine learning strategies reported in Ref. [15], Bayesian methods are the strategies that natively quantify uncertainty. Uncertainty quantification is crucial for alloy design, risk management, and qualification & certification processes. In fact, uncertainty quantification is a prerequisite of any Bayesian optimization method. For these reasons we prefer to use Bayesian models. In Refs. [16,17] we further explain our perspective of why alloy design and alloy design for AM should be conducted in a Bayesian manner.

Bayesian methods have been used to integrate physics-based thermal models with experimental measurements. Several studies were published discussing the printability of different alloys [18–22]. In these works, the Eagar-Tsai (E-T) model [23] is used to predict the melt pool dimensions at different laser power and scan speed values that cover the entire processing space of interest. While the ET model is physics-based, it is not without limitations. The model only simulates quasi-steady state heat transfer from a Gaussian beam on a semi-infinite medium. The ET model is incapable of considering temperature-dependent thermal properties. Furthermore, it only considers conduction, but does not capture more complex physics such as evaporative effects, recoil pressure, nor Marangoni convection. For this reason, several works [19,24–27] have corrected the ET model via Bayesian updating.

Most relevant to the current work are Refs. [24,25]. In these works the authors use experimental single-track data on various alloys to perform Bayesian calibration on an altered version of the ET model, i.e. a version of the ET model that considers temperature dependent properties. The authors calibrate two difficult-to-measure parameters used by the ET model: the apparent absorptivity (η) and powder bed porosity (ϕ). After MCMC calibration, the posterior distributions for these two quantities are obtained. Ref. [25] takes this analysis further by generating probabilistic printability maps. The advantages of these methods are their explainability and interpretability. For instance, by calibrating the model parameters of the E-T model against single-track data, it was possible for the authors to gain insight into difficult-to-measure quantities such as the absorptivity (η) and its dependence on

LBPF processing parameters. The disadvantages of these methods lie in their computational expense. The MCMC method limits its application in Bayesian optimization where a fast-acting surrogate model is needed.

In this study, we propose an alternative Bayesian approach for constructing printability maps. The primary objective of this framework is to reduce the amount of experimental data needed to produce an accurate printability map. The approach is 2-fold: (1) We correct the analytical ET model via Gaussian process regression. Specifically, we consider our prior belief of melt pool dimensions across the chemistry-processing space to be the ET model. We then update this prior belief in light of experimental melt pool measurements. This data-fusion enables accurate printability maps to be constructed using less experimental data. (2) Melt pool depth measurements require more sample preparation (e.g. cross-sectioning) than width measurements (top-surface imaging). We reduce the number of melt pool depth measurements required by exploiting the correlation between width and depth via HG-PRs, enabling width measurements to be used as proxy measurements for depth.

Compared to the state-of-the-art Bayesian method in Ref. [25], our framework presents some trade-offs. While our approach generates a non-parametric correction to the ET model rather than directly learning the underlying physics, it is computationally efficient and does not require MCMC methods. In Ref. [25], the authors calibrated the ET model with experimental data on a per-alloy basis. To calibrate their enhanced ET model for a single composition, they searched the parameter space with a burn-in period of 500 runs and drew a total of 10,500 samples for each run, necessitating the execution of the ET model 10,500 times to achieve an explainable calibrated model for a single alloy. The computational cost of MCMC calibration is comparable in other studies for single-alloy calibration.

In our method, the ET model (or its surrogate) must be queried for every point in the database and for each point in the design space where predictions are desired. For generating the printability map for a single alloy as shown in Figs. 8 and 9, we used 10,000 queries, which is comparable to Ref. [25]. However, the resolution of the printability map is arbitrary and can be adjusted as needed, unlike in Ref. [25], where decreasing the number of ET queries adversely affects the model's accuracy. We direct readers to Fig. 12 for an example of such useful but low resolution printability map. In our case, reducing the number of ET queries will only decrease the resolution of the predictions. Additionally, since we are not calibrating parameters but are instead interested in establishing a reasonable prior belief about melt pool dimensions as a function of chemistry and process parameters, a fast-acting surrogate or emulator of the ET model can be constructed to overcome this bottleneck.

Furthermore, the work in Ref. [25] rigorously addresses the printability of four alloys on a per-alloy basis, predicting printability as a function of processing parameters. In contrast, we demonstrate our method in printability extrapolation beyond the training set, predicting printability as a function of both chemistry and process parameters.

Both methods produce probabilistic printability maps, which are essential for qualification and certification. However, while our approach may lack the interpretability of a fully parametric model, interpretability is not always necessary in a design context. Instead, fast-acting accurate models with quantified uncertainty are required in alloy design [16,17].

The proposed method is particularly well-suited for the case study of predicting porosity defects because we have explicit priors from the ET model for width and depth, making it ideal for GP regression with informative priors. Likewise, width is a readily observable proxy for depth, motivating the use of Hierarchical GPs. However, in order to demonstrate the versatility of our method in addressing a range of defects during AM we have benchmarked our method against modeling of solid state cracking during AM [8,28,29].

Solid state cracking is a common issue during the AM of refractory metals such as W [8] or Mo [28]. Such solid state micro-cracks

causes premature failure in printed artifacts [29] and therefore must be eliminated to achieve usable printed parts. Mullin et al. [30] showed that single-track experiments on solid substrates can serve as a first approximation to the cracking behavior of alloys during AM processing. The authors showed that single-track experiments are indeed useful for screening compositions for the purpose of alloy design for AM [30]. Based on this analysis, we apply our proposed framework to the issue of solid state cracking using a dataset of crack number densities on 195 single-tracks printed on 16 unique refractory compositions. Specifically, we exploit correlations between melt pool dimensions and the linear crack number density. We show how when a GPR for the width is equipped with a physics-based prior, the subsequent predictions from a GPR for linear crack density are improved.

We explore the utility of this proposed method in 4 case studies:

1. **Printability Assessment:** We show by fusing the ET model with experimental data and inferring depth from width, a printability map for single alloys can be created using far fewer melt pool measurements. This amounts to informed extrapolation and interpolation within the processing space, resulting in a more resource-efficient approach to printability assessment.
2. **Printability Inference for Design:** After deploying the method in the processing-space, we extend the analysis to the union of the chemistry and processing design spaces. We show that the proposed method is better able to extrapolate printability predictions for alloys for which no single-track data is available. The method proves valuable in predicting the intrinsic printability of alloys prior to experimental efforts, contributing to a more informed and streamlined *printable*-alloy design process. This case study is conducted on 16 solid refractory multi-principal element alloy (RMPEA) substrates with at least 10 single-tracks per alloy.
3. **Crack-Resistance Inference for Design:** To demonstrate the utility of the proposed framework in predicting other defect modes common to AM, we benchmark the Bayesian method presented in this work against a dataset of linear crack number densities. The crack densities are taken from the same dataset of single-tracks on 16 unique RMPEAs.
4. **Probabilistic Printability Maps:** We show the benefit of using Bayesian models by creating data-driven probabilistic printability maps. In this example, the probability that a certain processing condition results in LOF is shown. Probabilistic printability maps are essential for qualification and certification. Furthermore, probabilistic predictions are important in any closed-loop decision making process like Bayesian optimization.

2. Materials and methods

2.1. Gaussian process regression

To perform the Bayesian updating mentioned in the introduction, we rely on Gaussian Process Regressors (GPR). GPRs are a class of Bayesian non-parametric models. GPRs consider an initial distribution of possible functions that *could* fit the observed data i.e. a GPR is a prior over functions [31]. These functions are defined by a prior mean and a kernel. Typically the prior mean is assumed to be either 0 for all values in the design domain or it is set to be the average of all observations, Y_N . When no observations are available in a certain region in the input space, a GPR will default to its prior mean predictions [32], as shown in Fig. 2a. The kernel (also known as the covariance function) defines the shape of the functions that fit the data [31]. The shape of the functions used to fit the data is controlled by the kernel's hyperparameters. These hyperparameters are selected by maximizing log marginal likelihood. Essentially, likelihood functions are used to evaluate the goodness-of-fit of functions that *could* fit the data, then the hyperparameters of the kernel are tuned to best fit the data. In summary, given data, a

prior distribution of functions, and a likelihood function, a posterior distribution of functions is obtained via Bayes rule (recall that Bayes rule provides a posterior given data, a prior, and a likelihood). This posterior distribution of functions is the trained GPR model.

Mathematically, N observations are represented as $\{\mathbf{X}_N, \mathbf{y}_N\}$, where $\mathbf{X}_N = (\mathbf{x}_1, \dots, \mathbf{x}_N)$ and $\mathbf{y}_N = (f(\mathbf{x}_1), \dots, f(\mathbf{x}_N))$. To obtain a GP prediction f_{gp} for an unobserved design $\mathbf{x}$, the following equations are employed:

$$f_{GP}(\mathbf{x}) | \mathbf{X}_N, \mathbf{y}_N \sim \mathcal{N}(\mu(\mathbf{x}), \sigma_{GP}^2(\mathbf{x})) \quad (1)$$

where

$$\begin{aligned} \mu(\mathbf{x}) &= K(\mathbf{X}_N, \mathbf{x})^T [K(\mathbf{X}_N, \mathbf{X}_N) + \sigma_n^2 I]^{-1} \mathbf{y}_N \\ \sigma_{GP}^2(\mathbf{x}) &= k(\mathbf{x}, \mathbf{x}) - K(\mathbf{X}_N, \mathbf{x})^T [K(\mathbf{X}_N, \mathbf{X}_N) + \sigma_n^2 I]^{-1} K(\mathbf{X}_N, \mathbf{x}) \end{aligned} \quad (2)$$

In the equations above, k is a real-valued covariance function (also known as a kernel function). The covariance matrix evaluated at all points in the design space is represented by $K(\mathbf{X}_N, \mathbf{X}_N)$. This matrix is of shape $N \times N$. An individual entry at m, n is represented by $k(\mathbf{x}_m, \mathbf{x}_n)$. The covariance of a single point with the rest of the design space is represented by $K(\mathbf{X}_N, \mathbf{x})$ which is a vector of shape $N \times 1$. To account for noise in the observation the σ_n^2 is multiplied by the identity matrix I and added to $K(\mathbf{X}_N, \mathbf{X}_N)$. Details on GPRs can be found in Ref. [31,32].

In this study, we employed an additive kernel composed of three individual kernels: the automatic relevance determination squared exponential kernel (SE), the dot product kernel, and the white noise kernel as implemented in Scikit-Learn [33]. Each of these kernels plays a distinct role in contributing to the overall model. The SE kernel accounts for the smooth localized variations in the data. The dot product kernel captures linear non-stationary relationships between features and the target. The white noise kernel accounts for uncorrelated noise in the data. For a more detailed example of the benefit of additive kernels, the reader is referred to Ref. [34].

$$k(\mathbf{x}, \mathbf{x}') = \exp\left(-\frac{(\mathbf{x} - \mathbf{x}')^2}{2l^2}\right) + (\sigma_0^2 + \mathbf{x} \cdot \mathbf{x}') + \sigma_n^2 \delta(\mathbf{x}, \mathbf{x}') \quad (3)$$

In Eq. (3), l is the characteristic length-scale of the RBF kernel. This hyperparameter controls the strength of correlation between 2 observations $\mathbf{x}'$ and $\mathbf{x}'$ over a distance in the input domain. The hyperparameter σ_0 controls the homogeneity of the dot product kernel. The hyperparameter σ_n is the noise level of the white noise kernel, and $\delta(\cdot, \cdot)$ is the Kronecker delta. More details on these kernels can be found in Refs. [33,35].

As mentioned, in regions of the input space where there is no training data present, the output of a GPR will be the prior mean function i.e. if the GPR predicts in regions far from data, it will rely on its prior prediction. In many GPR implementations, this prior mean function is typically set to be constant at 0 or at the average of all target training data. However, this approach is naïve and does not make use of physics-based priors. In the context of additive manufacturing, we often have low-fidelity thermal models that provide reasonable predictions of resultant melt pools as a function of chemistry and processing parameters. These low-fidelity models can serve as a prior belief melt pool geometry across the chemistry-process design space. In this scenario, in the absence of data, we rely on an 'educated guess' from the low-fidelity thermal model. When an observation is made, the GPR is trained with this experimental data. Upon training, the prior mean function is updated and becomes the posterior mean function (i.e. Bayesian updating). However, as the GPR predicts farther from regions in the design space where observations have been made, the GPR will begin to default to the prior prediction, as shown in Fig. 2a.

Informative priors can be added to any standard zero-mean GPRs, such as those implemented in the Sci-Kit Learn Python library, via the following procedures: First, define a prior for the design space. Next, given a set of ground-truth data, train the standard GPR on the discrepancy between the ground-truth and the prior, $y_{true} - y_{prior}$. Once the GPR has been trained, predict the discrepancy for the entire design

space and add back the prior, $(y_{true} - y_{prior}) + y_{prior} = y_{true}$. This will ensure that in regions in the domain where data has been observed, predictions from the GPR will rely on the training data, whereas in regions where no observations have been made, the predictions from the model will revert to the prior mean function, i.e. the posterior mean function is informed by both observation and the prior mean. This procedure is shown algorithmically in Code Snippet 1.

```

1 import matplotlib.pyplot as plt
2 import numpy as np
3 from sklearn.gaussian_process import
  GaussianProcessRegressor
4 from sklearn.gaussian_process.kernels import
  RBF, DotProduct
5 import random
6
7 #Continuous Domain
8 x = np.linspace(0, 40, 150)
9 #Ground-Truth
10 y = np.sin(0.2 * x) + 0.1 * np.sin(x) + 0.5 *
  np.sin(0.05 * x)
11 #Prior
12 y_prior = .75 * np.sin(0.2 * x - .5)
13 #Sample 10 points for training
14 inds = np.random.randint(60, size=10)
15 #Instantiate GPR
16 gpr = GaussianProcessRegressor(kernel=RBF(
  length_scale_bounds=(1,100))+DotProduct(),
  normalize_y=True)
17 #Find difference between prior and truth
18 y_disc = y[inds].reshape(-1, 1) - y_prior[
  inds].reshape(-1,1)
19 #Train GPR on discrepancy term
20 gpr.fit(x[inds].reshape(-1, 1),y_disc)
21 #Predict discrepancy term for entire domain
22 y_disc_pred = gpr.predict(x.reshape(-1, 1))
23 #Add back the prior to get final prediction
24 y_pred = y_disc_pred + y_prior
25 #Plot
26 plt.plot(x,y,linewidth=3,c="black",zorder=3)
27 plt.plot(x,y_prior,color="pink",linewidth=3)
28 plt.plot(x,y_pred,linestyle="--",color="red",
  zorder=4)
29 plt.scatter(x[inds],y[inds],marker="*",color=
  "yellow",edgecolors="k",s=200,zorder=15)
30 plt.show()

```

Listing 1: Python script that equipped a Gaussian Process Regressor with an informative prior. The results of this script are shown in Fig. 2

The difference between GPRs with informative and naive priors is evident in the simplified example illustrated in Fig. 2a. In this scenario, the ground-truth function is a linear combination of sinusoids, depicted in black. Consider a scenario in which we have a model that approximates the ground truth, depicted in pink. The typical prior for a standard zero-mean GPR is shown in light blue. This light blue line corresponds to the average of all current observations (yellow stars). In areas of the input space where there is no training data (i.e. no queries of the ground-truth) both models return the outputs of their respective prior mean functions. This corresponds to the right-hand side of Fig. 2a.

For the GPR equipped with a naive prior mean, as the model predicts in regions of the input domain where there is no training data the model's output will revert from a data-informed prediction to a prediction that is informed by the naive prior mean function i.e. a constant. However, the prediction from the naive prior notably diverges from the actual ground-truth model, as shown in the right-hand side of Fig. 2a where the blue and light-blue lines begin to coincide. For the GPR equipped with a physics-informed mean function, in regions in the domain where there is no training data, the model output

will revert from a data-informed prediction to a prediction that is informed by both the informative prior mean function (pink line) and the prior mean's average error. Note in Fig. 2a that the pink and red lines do not coincide. This is because the physics-based prior systematically underpredicted the ground-truth, therefore the prior mean function will be corrected by a constant term that is equal to the average deviation from the ground-truth. In previous works, we showed that GPRs equipped with physics-informed priors perform better during extrapolation and aid in the initial stages of Bayesian optimization [32].

Concerning inexpensive proxy experiments, previous works [32,36] have shown that if a correlation exists between an easy-to-measure feature (proxy) and an expensive feature (ground-truth experiment), a GP regressor can be created to model the proxy and this model can featurize a limited dataset of ground-truth experiments. As the proxy GPR is improved the GPR for the ground-truth is also improved [32,36].

In this work, we exploit the correlation between single-track width and depth using Hierarchical GP regression. Width measurements can be made from the top view of a substrate, however, depth measurements require cross-sectioning, grinding, polishing, and etching. This makes measuring single-track depth more difficult than measuring width. Recall that width and depth are highly correlated, as shown in Fig. 2c. For this reason, in Fig. 2d, a GPR is built to estimate width for arbitrary linear energy densities (LEDs). Once width can be estimated for an arbitrary LED, this predicted width can be used as a data-informed feature in a GPR for depth. Fig. 2d compares a GPR for depth trained on LED and a Hierarchical GP trained on width, and subsequently mapped back to LED. It is evident that the Hierarchical GP informed by width performs better than the standard GP in predicting depth. This implementation is shown in Code Snippet 2.

```

1 def hgpr(df, feats_1, feats_2, target_1,
  target_2, kernel1, kernel2, prior1, prior2
  ):
2
3   # Extract features and prior values for
  Target 1
4   X1_train = df[feats_1]
5   y1_prior_train = np.array(df[prior1])
6   y1_train = np.array(df[target_1])
7   beta1_train = y1_train - y1_prior_train
8
9   # Normalize features using MinMaxScaler
10  sclr_X = preprocessing.MinMaxScaler()
11  X1_train = sclr_X.fit_transform(X1_train)
12
13  # Initialize GPR model for Target 1
14  gpr1 = GaussianProcessRegressor(kernel=
  kernel1,
15    n_restarts_optimizer=10, normalize_y=
  True,
16    optimizer="fmin_l_bfgs_b",
  random_state=30)
17
18  gpr1.fit(X1_train, beta1_train)
19
20  # Predict proxy values for Target 2 based
  on proxy model for Target 1
21  df["Proxy"] = gpr1.predict(sclr_X.
  transform(df[feats_1])) + df[prior1]
22
23  # Construct new feature set for Target 2
  (including proxy value)
24  X2_train = df[feats_2 + ["Proxy"]]
25  y2_prior_train = np.array(df[prior2])
26  y2_train = np.array(df[target_2])
27  beta2_train = y2_train - y2_prior_train
28
29  # Normalize features for Target 2
30  X2_train = sclr_X.fit_transform(X2_train)
31

```

```

32 # Initialize GPR model for Target 2
33 gpr2 = GaussianProcessRegressor(kernel=
kernel2,
34     n_restarts_optimizer=10, normalize_y=
True,
35     optimizer="fmin_l_bfgs_b",
random_state=30)
36
37 gpr2.fit(X2_train, beta2_train)
38
39 # Return trained models for Target 1 (
proxy) and Target 2 (proxy-informed)
40 return gpr1, gpr2

```

Listing 2: Python script for the Deep GP informed by width proxy

The synthetic problems in Fig. 2 demonstrate the individual benefit of informative priors (Fig. 2a) and proxy experiments (Fig. 2d). In this work, we use both informative priors *and* proxy experiments to create printability maps. Specifically, we conduct two categories of case studies. In case study 1 we demonstrate how informative priors and proxy experiments can lessen the amount of data required to create a high-fidelity printability map for a given chemistry (i.e. interpolate printability as a function of processing conditions). Case study 2 is an extension of case study 1 in that we demonstrate how informative priors and proxy experiments can be used to generate printability maps for compositions for which we do not have any experimental data (i.e. extrapolate printability as a function of chemistry and processing conditions).

In previous works [32] we demonstrated the effectiveness of equipping hierarchical GPRs with physics-based priors. Specifically, we supplemented predictions from a GPR train on yield strength with predictions from a GPR trained on hardness. The naive prior mean functions of these GPRs were replaced with physical models relevant to yield strength and hardness, respectively. We differentiate the current work from our previous work by noting that this work is novel as thermal modeling and printability assessment is distinct from modeling mechanical properties such as hardness and yield strength. This work seeks to supplement difficult-to-obtain depth measurements with easier-to-obtain width measurements. Specifically, we use predictions from a GPR trained on width as a data-driven feature for a GPR trained on depth. The prior mean function of the width GPR is set to be the width prediction from the ET model. The prior mean function of the depth GPR is set to be the depth prediction from the ET model. We show that both proxy width experiments and priors based on analytical thermal models improve GP-based thermal modeling capabilities.

2.2. Feature selection

In machine learning, feature selection typically involves using statistical methods to identify the most relevant and discriminative features within a dataset. The goal is to find a subset of features that maximizes the model's predictive power while minimizing redundancy and overfitting. When modeling melt pool dimensions, we solve the heat conduction equation. The relevant thermophysical properties associated with this equation, along with processing parameters, serve as features for our GPR model. We have prior physics-based knowledge that these features are most suitable for a heat transfer model. Specifically, all inputs to the Eagar–Tsai (ET) model are also used as inputs for our GPR. While additional features could enhance the model's accuracy, we opt for simplicity by only considering inputs to the ET model as inputs to the GPRs used in this work. In this way, we focus on the effect of applying Hierarchical GPRs and informative priors to modeling single-track melt pool width and length. The inputs to the ET model are shown in Table 1.

Table 1

The inputs to the ET model and GPR models along with the information source of each of the inputs.

Property	Information source
Liquidus temperature	Equilibrium module
Thermal conductivity at liquidus	Property module
Density	Property module
Heat capacity	Property module
Absorptivity	Property module & Drude theory
Laser power	Independent variable
Scan speed	Independent variable

2.3. Calculation of features

The thermophysical properties required as inputs for the ET model and the GPRs were estimated using Thermo-Calc's TC-Python API. There are various CALPHAD databases that are tailored for specific alloy systems however for consistency, in this work we use the most generalized CALPHAD database, Thermo-Calc's high entropy alloy database TCHEA6. Specifically, all thermophysical properties are queried using the Property Module equipped with TCHEA6. Table 1 summarizes all the properties queried from Thermo-Calc. The absorptivity is the only property that is not directly queried from Thermo-Calc. Instead, the Absorptivity is estimated with Drude's empirical model which relates the electrical resistivity to absorptivity, as shown in Eq. (4). The electrical resistivity is calculated using Thermo-Calc's property module.

$$A = 0.365 \sqrt{\frac{\rho_0}{\lambda}} \quad (4)$$

2.4. Model evaluation and error metrics

To evaluate the proposed methods, 2 case studies will be used: printability interpolation and printability extrapolation, as shown in Fig. 3. In the case of interpolation, single-track experiments on 2 individual compositions will be used to train the models developed in this work. Specifically, single-track experiments on AF9628 and nickel alloy 718 were taken from Refs. [2,18]. In the case of AF9628, 52 width and 52 depth measurements were available. In the case of nickel alloy 718, 65 width and 65 depth measurements were available.

In this approach, 50% of the single-track data is allocated as the training dataset, while the remaining 50% is allocated as the testing dataset i.e. 2-fold cross-validation. For context, the standard selection of k is the less rigorous 5-fold cross-validation, where 80 at.% of the data is used for training [37]. To control for any biases that may arise from the way data is split during cross-validation, 250 instances of 2-fold cross-validation were conducted and a distribution of error metrics are presented. This strategy ensures that the manner in which the dataset is split does not unduly influence the results. By presenting a distribution of error metrics and rank ordering metrics we can better quantify the performance of each model.

For printability extrapolation, we used a dataset of 195 single-track width measurements and 176 single-track depth measurements across 16 unique refractory alloys. Each composition contains approximately 12 single-tracks. This data is provided in the repository associated with this work. Details on the experimental procedures used to generate the single-track data are given in Section 2.5. In order to establish a realistic benchmarking scheme, we stipulate that the cross-validation must be done at the level of alloys and not at the level of single-tracks. Specifically, from the database of single-tracks on 16 alloys, the single-tracks printed on 10 alloys will be used as training data and the remaining single-tracks on the 6 alloys will be used as testing data. This ensures that we are extrapolating the printability of alloys for which we have no training data. This is a realistic benchmarking scheme for

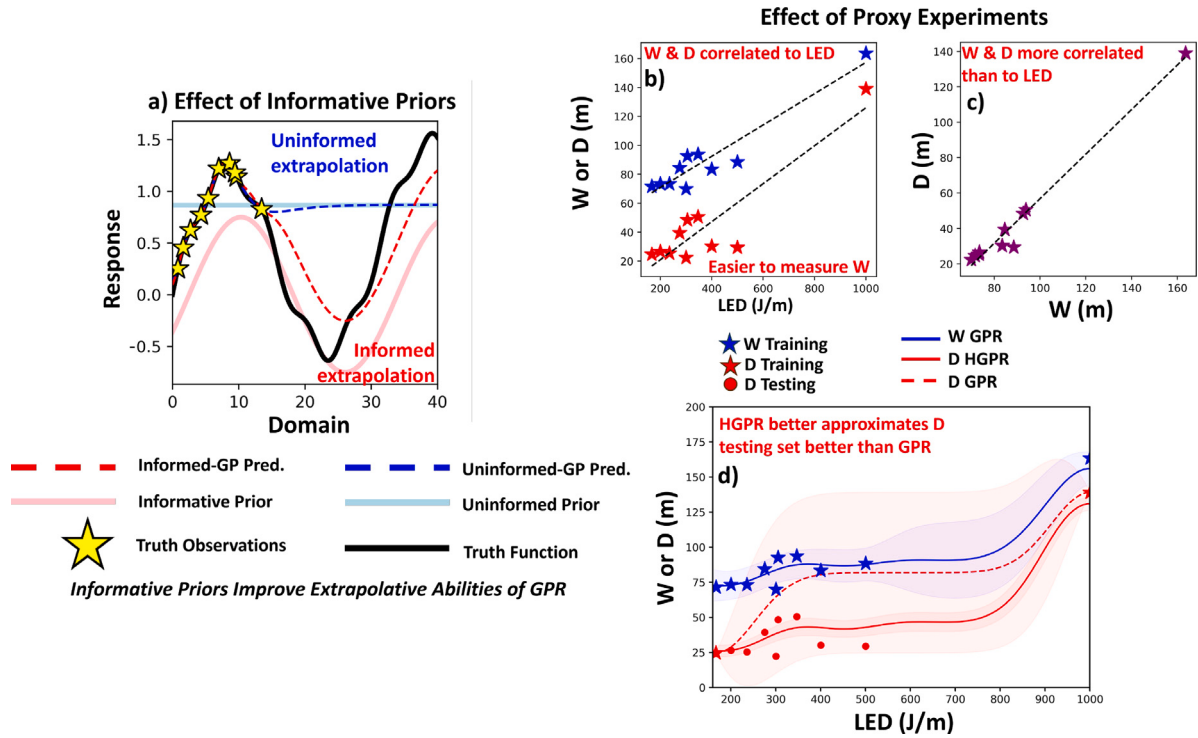


Fig. 2. (a) Illustration of physics-based and naive prior mean functions on the performance of GPRs. The light-blue and pink lines represent the naive and informative priors means, respectively. The black line represents the ground-truth. When making predictions in regions in the domain where there is no observed data, both models will revert to their prior mean function. However, equipping a GPR with an informative prior leads to improved extrapolation capabilities compared to using a naive prior. (b) The correlation between width and linear energy density (LED) of single-tracks on RMPEA substrates. It is evident that there is some degree of correlation between both width and depth and LED. (c) The correlation between width and depth of single-tracks on RMPEA substrates. It is evident width and depth are more correlated with each other than they are correlated with LED. (d) Comparison between standard GPR and Hierarchical GPR when predicting melt pool depth. The standard GPR (red dashed line) predicts an unrealistic trend between depth and LED. On the other hand, the Hierarchical GPR (informed by width predictions) better approximates the unseen training data (red circles). The code used to generate these plots is available at the following <https://colab.research.google.com/drive/122WiduyseKuztZupmGdXVOhqD1uk-Lqg?usp=sharing>. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

alloy design because in the design scenario one seeks to predict the printability of alloys *a priori* i.e. before synthesis and testing. Again, to control biases that may arise from the way the dataset is partitioned during cross-validation, 250 instances of this 6–10 test-train split were performed. Instead of a single error metric, a distribution of error metrics are reported, as shown in Figs. 11 and 12. In previous work, we have shown that such a benchmarking scheme provides a better, more generalized metric of a machine learning model's performance [32].

These benchmarking schemes are more rigorous and data-sparse than standard 5-fold or 10-fold cross validation. As for minimum data requirements, Bayesian models like GPRs technically do not have a strict lower bound for data quantity. With extremely sparse data, predictive accuracy may decrease, but the model compensates by increasing uncertainty estimates, ensuring the output remains informative. This characteristic ensures that even under sparse conditions, the model provides meaningful insights by conveying the confidence (or lack thereof) in its output. The model remains valuable by reflecting this uncertainty. Additionally, GPR-predicted uncertainty is reducible through systematic sampling of the design space, allowing for systematic improvements in model accuracy as more data becomes available [38]. This feature is crucial in high-stakes applications like alloy design, where understanding the model's confidence in its predictions is as valuable as the predictions themselves [17].

During cross-validation schemes for model benchmarking, two categories of metrics are used: error metrics and rank ordering metrics. In previous work [32], we showed the importance of considering both error metrics and rank ordering metrics during the benchmarking of ML models. Low values of error metrics (good) and poor rank ordering (bad) is an indication that the ML model is overfitted and yielding nonphysical predictions. Two models can have similar error metrics,

however the model with the higher rank ordering coefficient should be considered the more physically accurate model.

The two error metrics used in this work are mean absolute error (MAE), and the root mean squared error (RMSE). The MAE and RMSE are typical error metrics that capture the magnitude of the discrepancy between prediction and the ground-truth.

In addition to the two error metrics, two rank order metrics are used to evaluate the models in this work. The rank-ordering metrics used are the Spearman rank coefficient, which is represented by R_s , and Kendall's Tau Coefficient, which is represented by τ . R_s captures the ability of a model to sort a dataset based on the ordinal ranking of targets i.e. the ability of a model to correctly determine which design has the 1st highest property, 2nd highest property etc. In Eqn 7, d_i is the difference between the actual rank of a data-point, and n is the total number of data-points [39]. R_s is only defined from -1 to 1 . A value of 1 indicates a model's output has a perfectly monotonic relationship with the target values in a dataset [32]. A value of -1 indicates a model's output has a perfectly inverse monotonic relationship with the target values in a data [32]. A value of 0 indicates there is no monotonicity between the output of a model and the ordinal ranking of targets in a dataset [32].

Another ordinal sorting metric is τ . Kendall's τ is a statistic that measures the ordinal relationship between two variables. In the context of this work, consider a scenario where a model prediction x is used to approximate a ground-truth experiment y . Consider we have 2 alloys in the dataset, alloy i and alloy j . Let the prediction from the model be x_i and x_j for alloys i and j , respectively. Let the ground-truth property for alloys i and j be y_i and y_j , respectively. In the case of 2 data-points, if $y_i > y_j$ and $x_i > x_j$, then the model x has perfect rank ordering ability [32]. The formula for τ is given in Eq. (8).

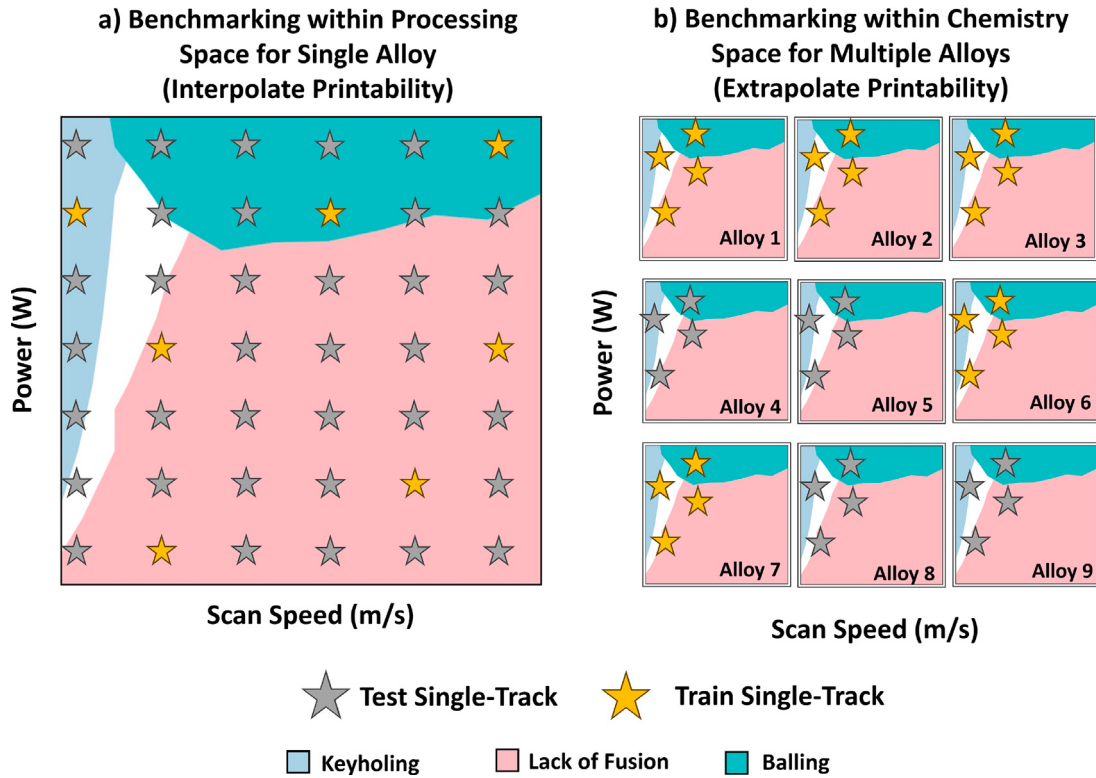


Fig. 3. Diagram of the cross-validation scheme used to benchmark the models proposed in this work. (a) The 'printability interpolation' cross-validation scheme. Databases of single-track experiments on individual powderized alloys will be considered. From the database, 50% of the single-tracks will be allocated as training data and 50% will be allocated as testing data i.e. 2-fold cross-validation. The models will be evaluated with error metrics in Eqns 5–8. This 2-fold cross-validation is repeated 250 times and the error metrics are reported as distributions. (b) The 'printability extrapolation' cross-validation scheme. A database of 195 single-track experiments on solid substrates of 16 RMPEAs is considered. From the database the single-tracks on 10 alloys will be used as training data and the remaining 6 alloys will be used as testing data.

It is not sufficient for a 'good' model to have a low MAE and RMSE. Instead, a model must have both low error metrics and high rank ordering coefficients. If a model has low error metrics but poor rank ordering coefficients, this could be indicative of over-fitting [32]. By considering both the error metrics and rank ordering metrics, the physicality of the model is evaluated [32]. Each model undergoes 250 instances of cross-validation and the distribution of MAE, RMSE, R_s , and τ value are reported. By reporting a distribution of metrics we are better able to control for biases that could arise from random initialization of GPRs or from the way data is split during cross-validation.

$$\text{MAE} = \sum_{i=1}^n |x_i - y_i| \quad (5)$$

$$\text{RMSE} = \sqrt{\frac{\sum_{i=1}^n (x_i - y_i)^2}{n}} \quad (6)$$

$$\rho = 1 - \frac{6 \sum_{i=1}^n d_i^2}{n(n^2 - 1)} \quad (7)$$

$$\tau = \frac{2}{n(n-1)} \sum_{i < j} \text{sng}(x_i - x_j) \text{sng}(y_i - y_j) \quad (8)$$

2.5. Experimental procedures

For the extrapolation case study, 195 single-tracks on 16 RMPEA solid-substrates (as opposed to powder) were used to benchmark the extrapolative methods proposed in this work. These 16 alloys were downselected throughout an alloy design campaign aimed at designing refractory alloys for both printability and high-temperature performance [40]. Between 12 to 13 single-tracks were printed on each of the 16 compositions. The compositions of these 16 RMPEAs are reported in Table 2.

The power and scan speed used to print these single-tracks was chosen based on the ET model. Specifically, the ET model was used

Table 2

Compositions of the 16 refractory multi-principal element alloys (RMPEAs) used in the extrapolation case study. Each composition was selected as part of an alloy design campaign aimed at optimizing printability and high-temperature performance [40].

Nb ₆₀ W ₄₀	Nb ₅ Ta ₅₀ V ₅ W ₄₀
Nb ₅₅ Ta ₅ W ₄₀	Nb ₂ Ta ₆₂ W ₃₆
Mo ₅ Nb ₅₅ Ta ₅ W ₃₅	Mo ₂ Ta ₅₈ W ₄₀
Mo ₂ Nb ₂ Ta ₆₄ W ₃₂	Mo ₂ Nb ₂ Ta ₅₂ W ₄₄
Mo ₂₀ Nb ₁₀ Ta ₂₅ W ₄₅	Mo ₁₅ Nb ₅ Ta ₃₀ W ₅₀
Mo ₁₄ Nb ₂₆ Ta ₂₆ W ₃₄	Mo ₁₀ Nb ₅ Ta ₃₀ W ₅₅
Mo ₁₀ Nb ₅₅ W ₃₅	Hf ₂ Nb ₄ Ta ₄₆ W ₄₈
Hf ₂ Nb ₁₂ Ta ₄₆ W ₄₀	Hf ₂ Mo ₂ Ta ₄₈ W ₄₈

as a first approximation to predict the printability maps of each of the alloys. The powers and scan speeds on the predicted defect decision-boundary in the printability map were selected by expert opinion. These solid-substrate single-track data are available in the Code Ocean repository associated with this work: <https://rb.gy/ed3l0n>

The selected RMPEA compositions were synthesized from high-purity elements (>99.9 wt%) using a Buehler AM200 vacuum arc melter (VAM) under an Ar atmosphere. The lowest melting point element, Vanadium (V), was added last to reduce the weight loss during synthesis due to evaporation. The arc melting chamber was evacuated and filled with Argon (Ar) gas three times after each material addition. Arc-melted coupons were flipped and remelted at least 10 times to ensure the compositional homogeneity of the alloys. To dissolve the dendritic microstructure observed in the as-cast condition, synthesized alloys were homogenized at 2000 °C for 28 h using a Centorr brand high-temperature furnace (LF Series, Model 22) under an Ar atmosphere. The heat treatment chamber was evacuated and filled with Ar gas three times before heat treatments. Heat-treated specimens were furnace-cooled down to room temperature. Then, the cross-sectional

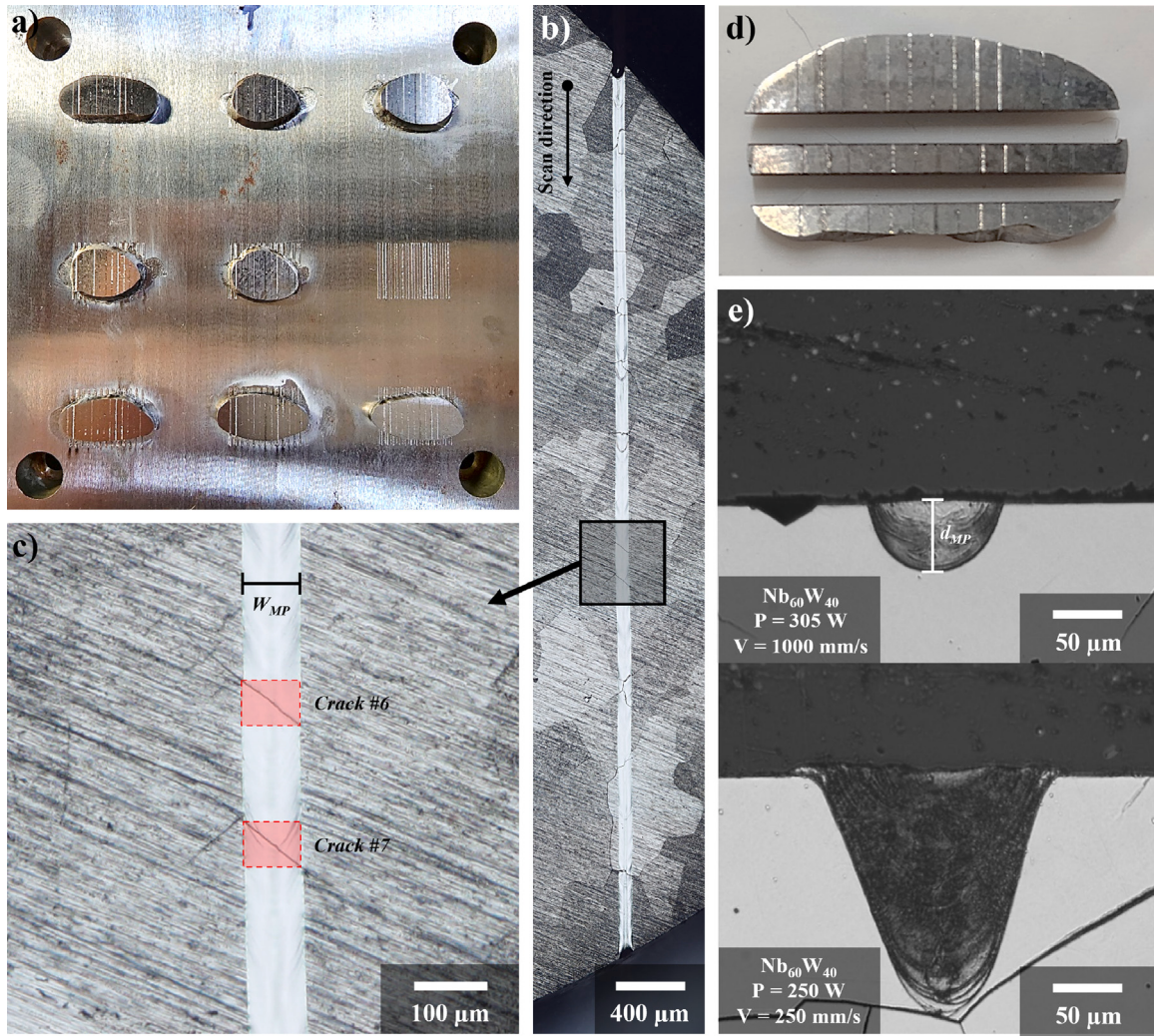


Fig. 4. (a) Prepared refractory MPEA substrates for single track experiments glued on the ground SS-316 plate, (b) top view image obtained after scanning laser using the parameters of 275 W power and 1650 mm/s scan speed on $\text{Nb}_{55}\text{Ta}_5\text{W}_{40}$ alloy, (c) higher magnification top view image taken from $\text{Nb}_{55}\text{Ta}_5\text{W}_{40}$ alloy showing the measurements for melt width and total number of cracks, (d) alloy after sectioning using wire-EDM to produce cross-sectional specimens to investigate melt depths, and (e) cross-sectional images taken from $\text{Nb}_{60}\text{W}_{40}$ alloy for melt depth measurements.

sample profiles were cut into 2-mm thick substrates using wire electrical discharge machining (wire-EDM). Specimen surfaces were polished using abrasive SiC polishing papers starting from 320 grit to produce a final polish with 1200 grit. After surface preparation, refractory MPEA substrates were glued on the ground SS-316 plate, as shown in Fig. 4a.

Single-track experiments were printed on the ground substrates without using powder using an EOS M290 LPBF system equipped with a fiber laser. The laser has a maximum power of 400 W, a wavelength of 1070 nm, a beam diameter of 80 μm , and a maximum speed of 7 m/s. The printing chamber was filled with Argon gas, and the oxygen level was set below 500 ppm. Eight substrates were mounted on a stainless steel square base plate with 10 cm side length. At least 12 different process parameters (laser power and scan speed) were used on each substrate depending on the substrate size. The process parameters were selected to be inside the printable region as well as close to the boundaries of each defect.

After printing the single tracks, the top views of the single tracks were imaged using Olympus DSX 500 optical microscope (OM) to measure melt width and linear crack number density, as shown in Fig. 4b and c. The average value of melt widths obtained from the 10 randomly selected locations was used as a representative value for each single track. The total number of cracks per single track and the overall single-track lengths were measured to obtain the linear crack

number densities. Then, specimens were cut using EDM in 2 different locations to investigate the cross-sections of the melt pools (Fig. 4d). The cross-sections of the single tracks were polished using abrasive SiC polishing papers starting from 320 grit to 1200 grit and placed into a vibratory polisher for 72 h in 0.04 μm colloidal silica suspension. Specimen surfaces were then etched using Murakami's reagent before the cross-sectional examination to observe the melt pool better. The average values of melt depths obtained from 2 different locations are recorded. Fig. 4e shows an example of melt depth measurements from alloy $\text{Nb}_{60}\text{W}_{40}$ using the cross-sections. Moreover, each single track was classified into keyhole, lack of fusion, balling, or defect-free single track.

3. Results and discussion

3.1. Single composition benchmarking

To assess the interpolative abilities of the proposed methods, we benchmarked 5 different models against single-track melt pool measurements for 2 distinct alloys: Alloy AF9628 and nickel alloy 718. These models are designated as M1 to M5 for clarity:

- **M1:** Melt pool width and depth calculated directly from the ET thermal model (i.e. analytical control model).

- **M2:** Two separate GPR models for predicting the melt pool width and depth (i.e. solely data-driven control model).
- **M3:** A GPR is used to predict width. This GPR is then used to impute width as an informative feature for a GPR trained on depth. Neither width nor depth GPRs are equipped with informative priors (i.e. model to probe the effect of using width as a proxy for depth).
- **M4:** Two separate GPR models for width and depth that use the predicted melt pool width/depth from the ET model as a prior. (i.e. model to probe the effect of using ET model as informative prior).
- **M5:** A GPR is used to predict width. This GPR is then used to impute width as an informative feature for a GPR trained on depth. Both GPRs use the predicted melt pool width or depth from the ET model as an informative prior (i.e. model to probe the effect of using informative priors and width as a proxy for depth).

Alloy AF9628 is an ultra-high strength martensitic steel alloy while nickel alloy 718 is a superalloy capable of operating at high-temperature. These alloys were selected as their printability has already been rigorously studied experimentally in Refs. [2,18]. Notably, these compositions exhibit significantly different thermophysical properties. Furthermore, the single-track experiments were conducted on different L-PBF machines for each alloy. Benchmarking against these 2 distinct datasets highlights the general applicability of our proposed method. As described in Section 2.4.

Figs. 5–8 show the distribution of the error metrics and rank-ordering metrics explained in Section 2.4 for alloy AF9628 and nickel alloy 718. For melt pool width, 3 models were employed corresponding to **M1** (ET calculations), **M2** (GPR model), and **M3** (GPR+prior). These distributions were generated from 250 instances of 2-fold cross-validation, meaning the model underwent training on 250 different sets of single tracks. Specifically, width measurements are randomly sampled from 50% of the measurements available in the dataset for each alloy. The remaining 50% of width data is used as the test set. Median values and interquartile ranges of the error metrics and rank ordering metrics are presented in Tables 3 and 4.

When benchmarked against single-track width experiments on AF9628, the proposed model (M3) has better MAE and RMSE values than the 2 control models, M1 and M2. Specifically, the median MAE of M1 is the highest (39.98 μm), followed by M2 (18.29 μm) and M3 (14.20 μm). Similarly, M3 has better RMSE values than the 2 control models, M1 and M2. The median RMSE of M1 is the highest (43.71 μm) followed by M2 (29.07 μm) and M3 (19.39 μm). The IQR of the MAE and RMSE for M2 are the largest (9.62 μm and 17.14 μm), indicating that the solely data-driven model is highly sensitive to the training data and exhibits significant variability. On the other hand, the IQR of the MAE and RMSE for M5 is relatively small compared to M2, indicating that the physics-informed M5 is more robust to variations in training data.

When considering Kendall's τ rank-ordering metric, it is evident that all models demonstrate acceptable performance. Notably, models M1 and M5 exhibit comparable median τ rank-ordering metrics (0.898 and 0.883), surpassing those of M2 (0.834). The median Spearman rank-ordering coefficient R_s for M5 is slightly higher (0.976) than R_s for M1 (0.973) whereas R_s for M2 is the lowest (0.950). These results indicate that the physics-informed models yield the most physical results.

Recall that low error metrics and high rank-ordering metrics indicate a model is both accurate and yielding physical predictions. M1 exhibited competitive rank-ordering metrics but poor error metrics, indicating that the ET model is physical, yet has systematic error. M2 had intermediate values for both rank ordering and error metrics. M5 exhibited the best error metrics and competitive rank ordering metrics, making it the best model. These results underscore the importance of physics-based models for LPBF with experimental data.

Fig. 6 illustrates the error distribution for melt pool depth using the 5 models. The MAE and RMSE for **M1** exhibited the highest median

errors (MAE = 35.58 μm and RMSE = 62.63 μm). This is likely because the model fails to capture the complex physics associated with the recoil pressure in the keyholing region [26]. The vanilla GPR model (M2) has slightly elevated median error metrics (MAE = 16.79 μm and RMSE = 24.62 μm) compared to the physics-informed models (M3 and M5). The model equipped with the informative prior (M3) has competitive error metrics (MAE = 13.87 μm and RMSE = 21.10 μm). The depth model that uses width proxy experiments (M4) has slightly elevated error metrics (MAE = 16.85 μm and RMSE = 23.72 μm) compared to the physics-informed models. Recall that M4 is also a solely data-driven model. Finally, the depth model that makes use of informative priors and proxy width experiments (M5) exhibits competitive metrics (MAE = 13.89 μm and RMSE = 20.92 μm).

Regarding rank-order metrics, the model informed by physics-based priors and proxy width experiments (M5) outperforms the other models ($\tau = 0.840$ and $R_s = 0.948$). The model informed by physics-based priors (M3) has comparable error metrics ($\tau = 0.834$ and $R_s = 0.944$). The solely data-driven models M2 and M4 has slightly decreased rank ordering metrics ($\tau = 0.814$ and $R_s = 0.934$ for M2 and $\tau = 0.809$ and $R_s = 0.934$ for M4). However in general the models have comparable rank-ordering ability.

These error metrics and rank-ordering metrics indicate that when interpolating depth within the processing space of AF9628, models that are equipped with informative priors consistently exhibit improved interpolative ability for depth compared to solely data-driven models. Furthermore, models M2–M5 outperformed M1 the analytical ET model in predicting depth.

This analysis is repeated on a dataset of single-track measurements of nickel alloy 718. Figs. 7 and 8 depict the error distribution and rank-ordering distribution for the 5 models trained on single-track measurements of nickel alloy 718. Despite the models being trained on a different dataset of single tracks printed using a different machine and laser setup, at varying process parameters and powder composition, the same trend in median error metrics and median rank-ordering metrics is observed. This suggests that our approach remains stable irrespective of chemistry. Again, these results indicate that models that are equipped with informative priors consistently exhibit improved interpolative ability compared to solely data-driven models.

While accurately predicting melt pool width and depth is important, ultimately the purpose of these predictions is to determine the printability of an alloy as shown in Fig. 1. To visualize the impact of different models on printability maps, 5 maps were generated using predicted melt pool width and depth from each of the 5 models. The resultant printability maps are shown in Figs. 9 and 10 for alloy AF9628 and nickel alloy 718, respectively. Each printability map utilized 10,000 sets of process parameters (laser power and scan speed) sampled using uniform grid sampling. Each map is trained on 50% of the single-tracks available in the dataset. The generated printability maps are compared to printability maps that use a computationally expensive MCMC-based calibration algorithm as reported in Refs. [2,18]. Essentially, given the same single-track data, we seek to show that our fast-acting GPR-based method can converge on calibrated maps similar to those reported in Refs. [2,18] in which the authors use a computationally expensive MCMC-based calibration method. Figs. 9 and 10 depict the overlap between the ground truth and printability maps generated using predicted melt pool dimensions from the different models. The balling region was excluded in both cases since it is typically determined using a support vector machine (SVM) model which is not affected by our calculations [5]. The black dashed lines delineate the ground truth boundaries, signifying that the closer the predicted regions are to the dashed lines, the more accurate the model. The red dashed lines are the defect-bounds predicted by the models M1–M5.

In Fig. 9a, the LOF predictions from the ET model are comparable to the predicted LOF after calibrating the ET model with the MCMC method reported by Seed et al. [21]. However, at high laser power and low scan speeds, the model fails to capture keyholing defects

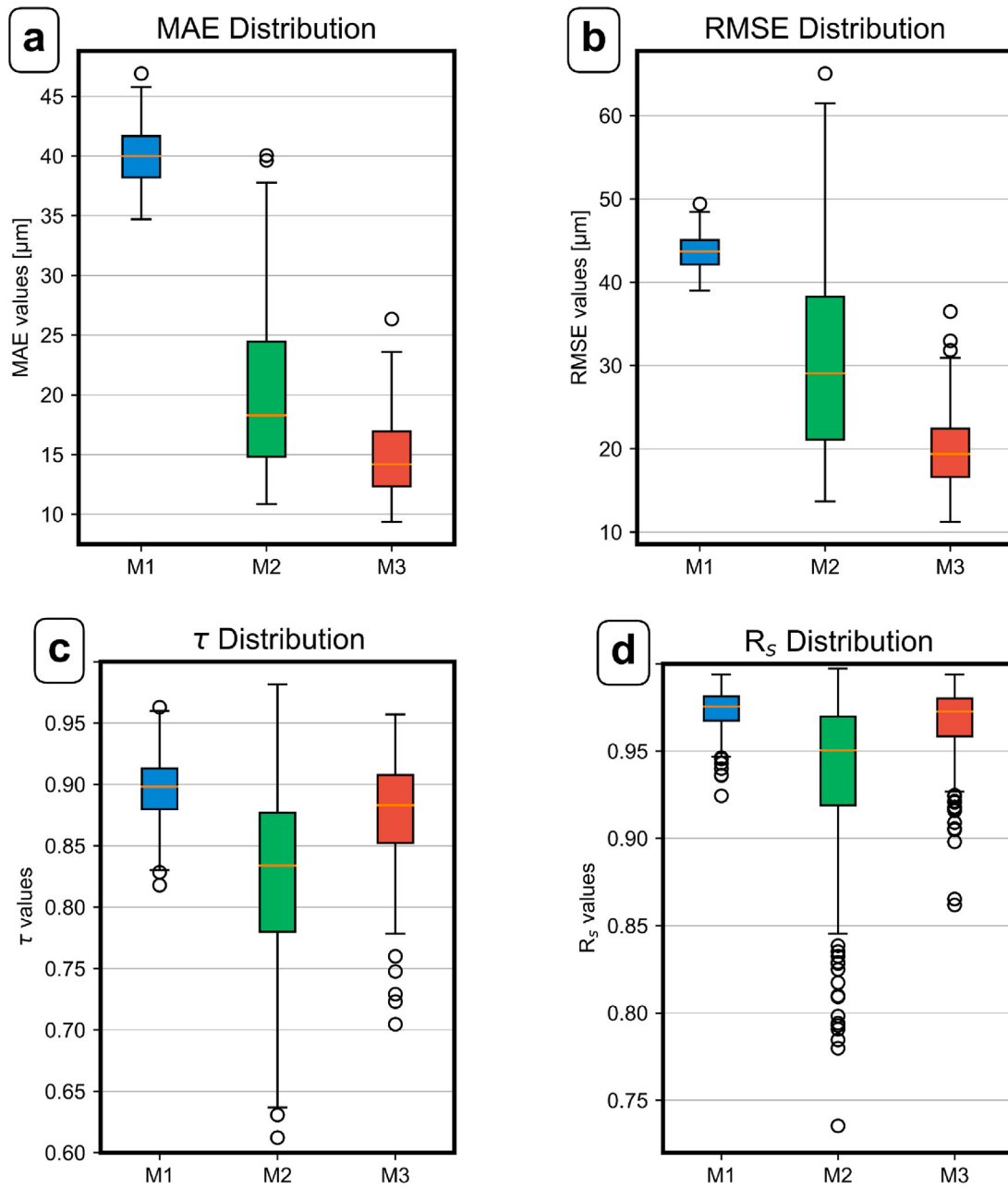


Fig. 5. Comparative analysis of the three models benchmarked in this study for alloy AF9628's melt pool width. In M1 (blue), the melt pool width as calculated from the ET model is used. In M2 (green), a GPR model is trained on 50% of the single track width data. In M3 (red), a GPR model is trained on 50% of the single track data and the prior mean function of the model is set to be the width prediction from the ET model. The distributions of (a) mean absolute error (MAE), (b) root mean square error (RMSE), (c) Kendall rank coefficient (τ), (d) Spearman rank coefficient (R_s), are shown for 250 instances. The proposed model (M3 — Red) has the lowest MAE and RMSE, and high values for τ and R_s . (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

Table 3

Median and interquartile range for the four error metrics calculated from training each model for 250 instances on different training and testing datasets. The four error metrics are for melt pool width of alloy AF9628 and nickel alloy 718. MAE: Mean absolute error, RMSE: Root mean square error, τ : Kendall rank coefficient, and R_s : Spearman rank coefficient.

	Model	MAE		RMSE		τ		R_s	
		Median	IQR	Median	IQR	Median	IQR	Median	IQR
AF9628	ET	39.98	3.46	43.71	2.90	0.898	0.03	0.976	0.01
	ET+GPR	18.29	9.62	29.07	17.14	0.834	0.10	0.950	0.05
	ET+Prior	14.20	4.58	19.39	5.84	0.883	0.06	0.973	0.02
Ni Alloy 718	ET	32.19	5.39	42.51	6.90	0.884	0.04	0.968	0.02
	ET+GPR	31.05	6.10	40.45	9.01	0.853	0.06	0.960	0.03
	ET+Prior	23.08	4.52	32.21	6.17	0.883	0.04	0.973	0.01

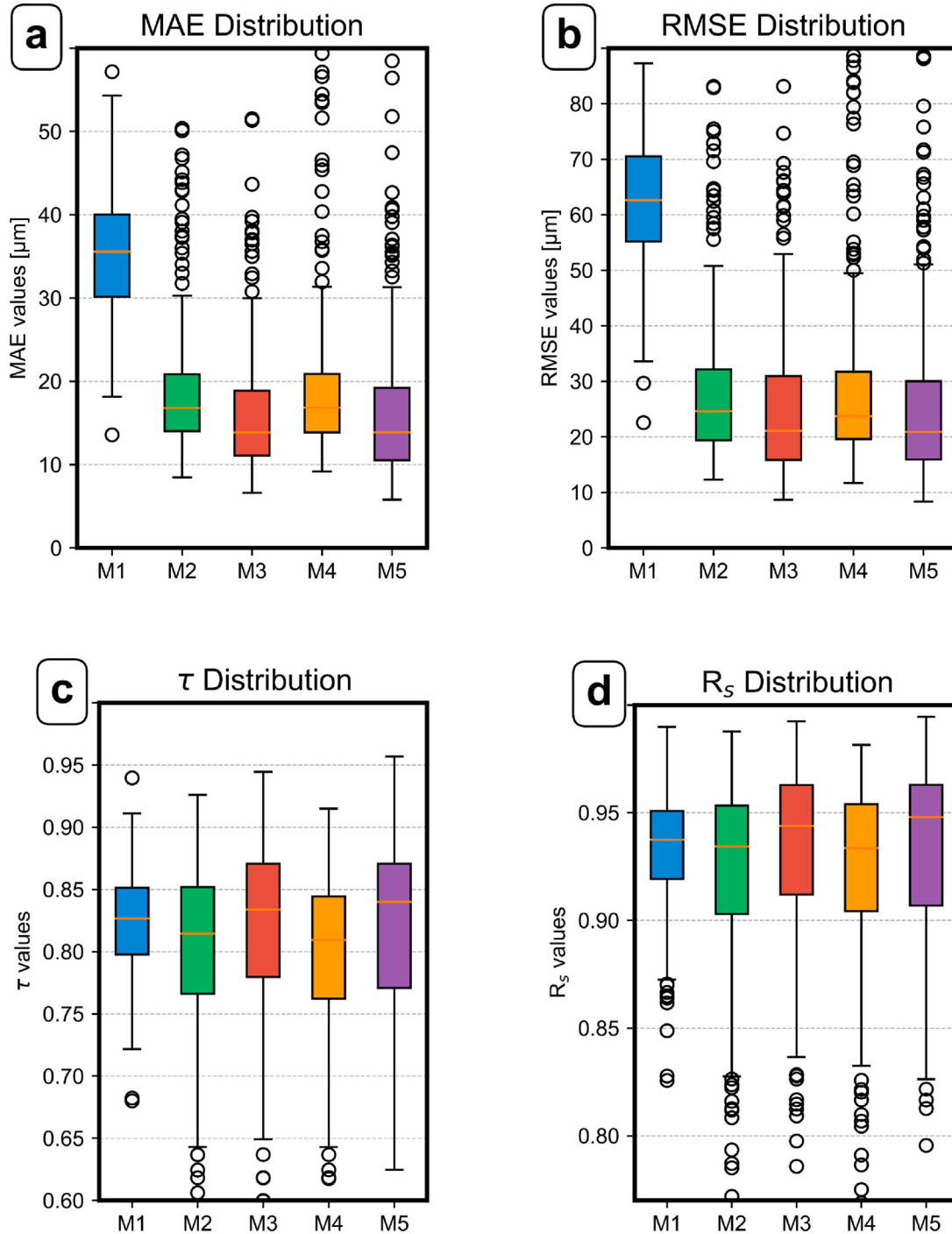


Fig. 6. Comparison of the interpolative ability of the 5 candidate models in this study using melt pool depth data from AF9628. In M1 (blue), melt pool depth is calculated from the ET model. In M2 (green), a GPR model is trained on 50% of the available single-track depth data. In M3 (red), proxy width predictions are used as a data-driven feature in a GPR model that is trained on 50% of the available single-track depth data. In M4 (orange), a GPR model is trained on 50% of the available single-track depth data and uses ET depth predictions as a prior mean function. M5 (purple) is a combination of M3 and M4 i.e. it uses data-informed width predictions (these width predictions are from a GPR that uses ET width as a prior) and leverages ET depth predictions as physics-based prior. The distributions of (a) mean absolute error (MAE), (b) root mean square error (RMSE), (c) Kendall rank coefficient (τ), (d) Spearman rank coefficient (R_s). The proposed model (M5 — purple) has the lowest MAE and RMSE, and the highest values for τ and R_s compared to the other four models. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

due to the lack of physics in the ET model. The ET model fails to predict that keyholing occurs in AF96 at all. Likewise, in Fig. 9c the HGPR model that makes use of width proxy experiments for depth also fails to predict a keyholing region in AF96. The vanilla GPR model in Fig. 9b generates a printability map that is in moderate agreement with

the MCMC method, however, we would expect the keyholing region to extend to higher velocities as power increases. Likewise, the GPR models equipped with informative priors for width and depth produce a printability map in 9d that is in moderate agreement with the MCMC-based map. In this case, the keyholing region is overpredicted at powers

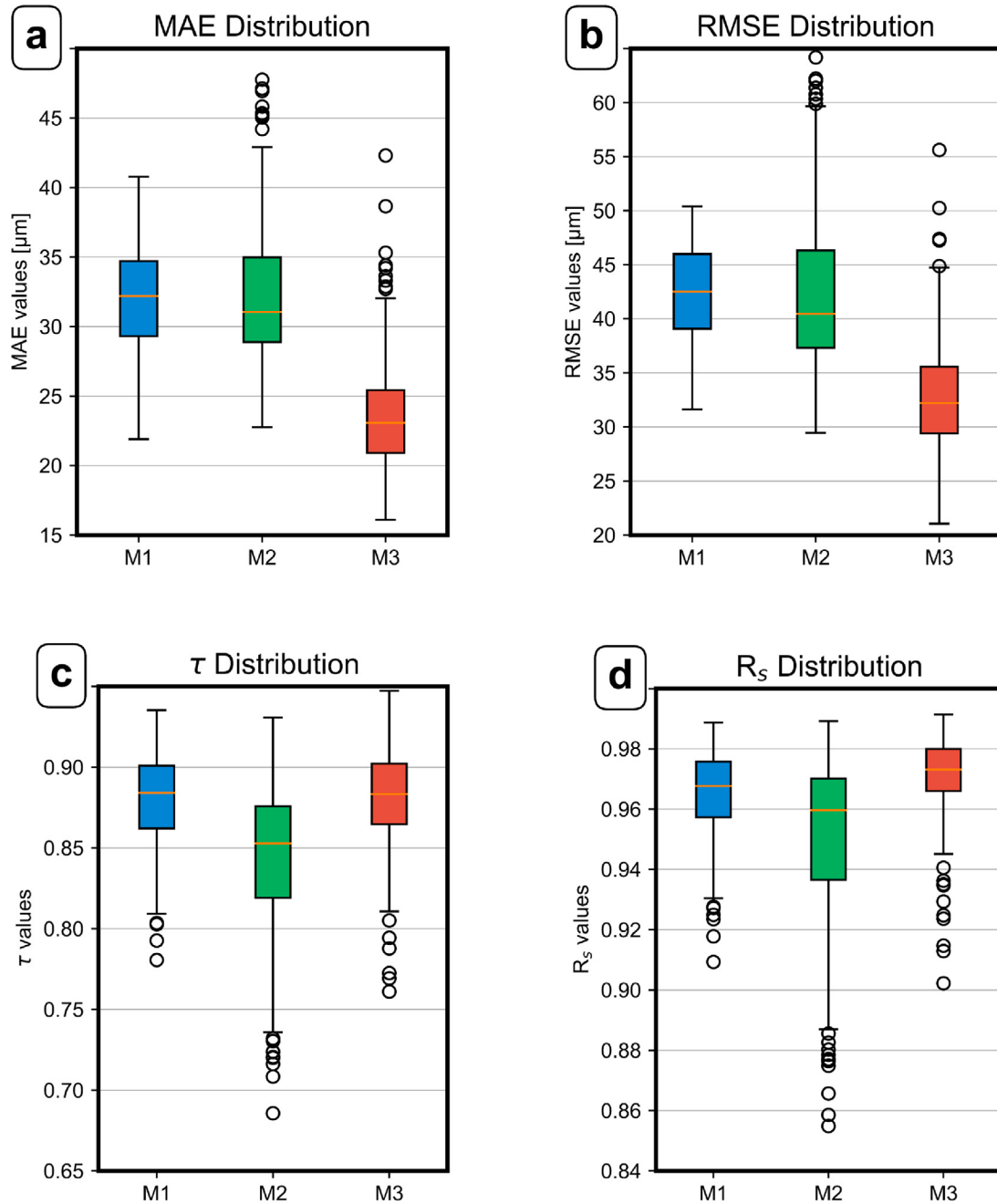


Fig. 7. Comparative analysis of the three models benchmarked in this study for nickel alloy 718's melt pool width. In M1 (blue), the error metrics associated with the width prediction from the ET model are shown. In M2 (green), the error metrics for a GPR model trained on 50% of the single track data are shown. In M3 (red), the error metrics associated with a GPR model trained on 50% of the single track data and equipped with ET-predicted width as a physics-based prior are shown. The distributions of (a) mean absolute error (MAE), (b) root mean square error (RMSE), (c) Kendall rank coefficient (τ), (d) Spearman rank coefficient (R_s), are shown for 250 instances of cross-validation. The proposed model (M3 — Red) has the lowest MAE and RMSE, and the highest τ and R_s , compared to the other two models. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

above 200 W. The model equipped with informative priors and proxy width experiments has the most accurate printability map in 9e. The model produces a map whose LOF and keyholing regions are in close agreement with the MCMC-based map.

In Fig. 10, the printability maps for nickel alloy 718 generated from the 5 models are compared to the printability map reported by Shoukr et al. [2]. No keyholing was captured in Fig. 10a, c, and d. This can be explained by the under-prediction of the melt pool depth when the ET model is solely used in Fig. 10a. Moreover, when the hierarchical GPR model was trained using experimental data without any prior, the model still failed to capture the keyhole defect region, as shown in Fig. 10c. In Fig. 10d, using the prior was more significant than the

experimental data, and this can be observed by the mismatch in the depth values in the lack of fusion region at high laser power; therefore, the keyhole region is still not captured. On the other hand, Fig. 10b and e show a keyhole region but Fig. 10e. However, when the experimental data was only used, the shape of the keyhole and lack of fusion regions differs based on the training set (in some cases, the keyhole defect is not captured in the printability map). Finally, when the hierarchical GPR model with a prior is used, both lack of fusion and keyhole regions agree with the true printability map.

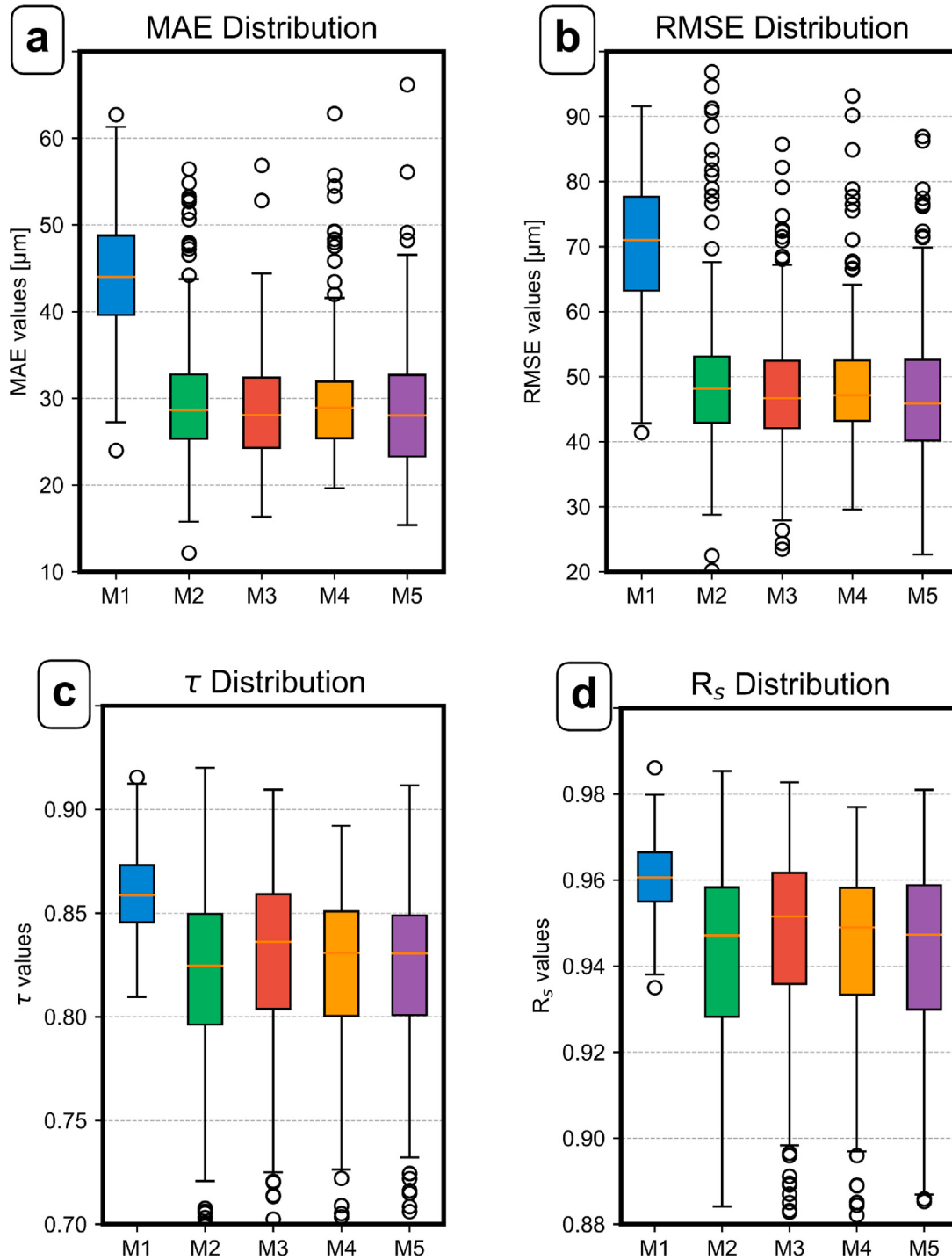


Fig. 8. Comparative analysis of the five models benchmarked in this study for the dataset of melt pool depths for nickel alloy 718. In M1 (blue), melt pool depth is calculated from the ET model. In M2 (green), a GPR model is trained on half of the available single-track depth data. In M3 (red), proxy width predictions are used as a data-driven feature for a depth-GPR model that is trained on 50% of the available single-track depth data. In M4 (orange), a GPR model is trained on 50% of the available single-track depth data *and* uses depth prediction from the ET model as a prior mean function. M5 is a combination of M2 and M4 i.e. it uses data-informed width prediction and leverages ET depth predictions as a physics-based prior. The distributions of (a) mean absolute error (MAE), (b) root mean square error (RMSE), (c) Kendall rank coefficient (τ), (d) Spearman rank coefficient (R_s). The proposed model (M5 — purple) has the lowest MAE and RMSE, and competitive values for τ and R_s compared to the other four models. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

3.2. Case study on refractory multi-principal element alloys system

The proposed framework was extended from interpolating the printability maps of single alloys for which we have limited data to extrapolating printability maps of multiple alloys for which we have no data at all. For this purpose melt pool width and depth from 195 single

track experiments printed on 16 solid RMPEA compositions were used to train and test the 5 models (M1–M5) as explained in Section 3.1. The 16 RMPEA compositions were downselected throughout an alloy design campaign targeted at designing printable refractory MPEAs. Details on their synthesis and single-track testing are provided in Section 2.5. The compositions of these alloys will be released at a later date, however,

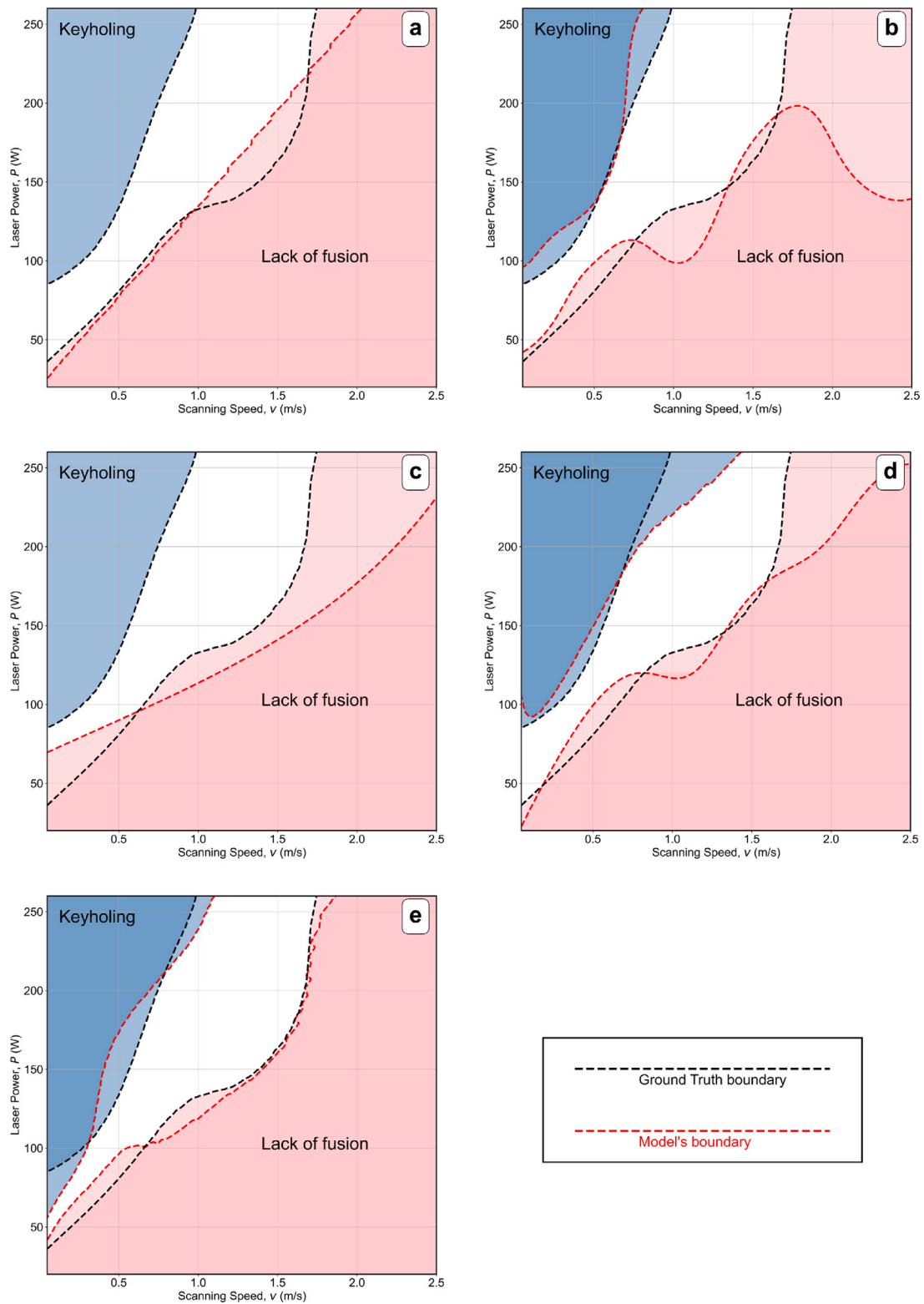


Fig. 9. Printability maps for AF96 generated using 10,000 data points of melt pool width and depth predicted using (a) M1, (b) M2, (c) M3, (d) M4, (e) M5. The black dashed lines show the boundary of the keyhole and lack of fusion regions from the printability map generated using the MCMC method reported in [18] denoted as the ground truth, while the red dashed lines show the boundary of the keyhole and lack of fusion regions generated from the different models used in the benchmarking. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

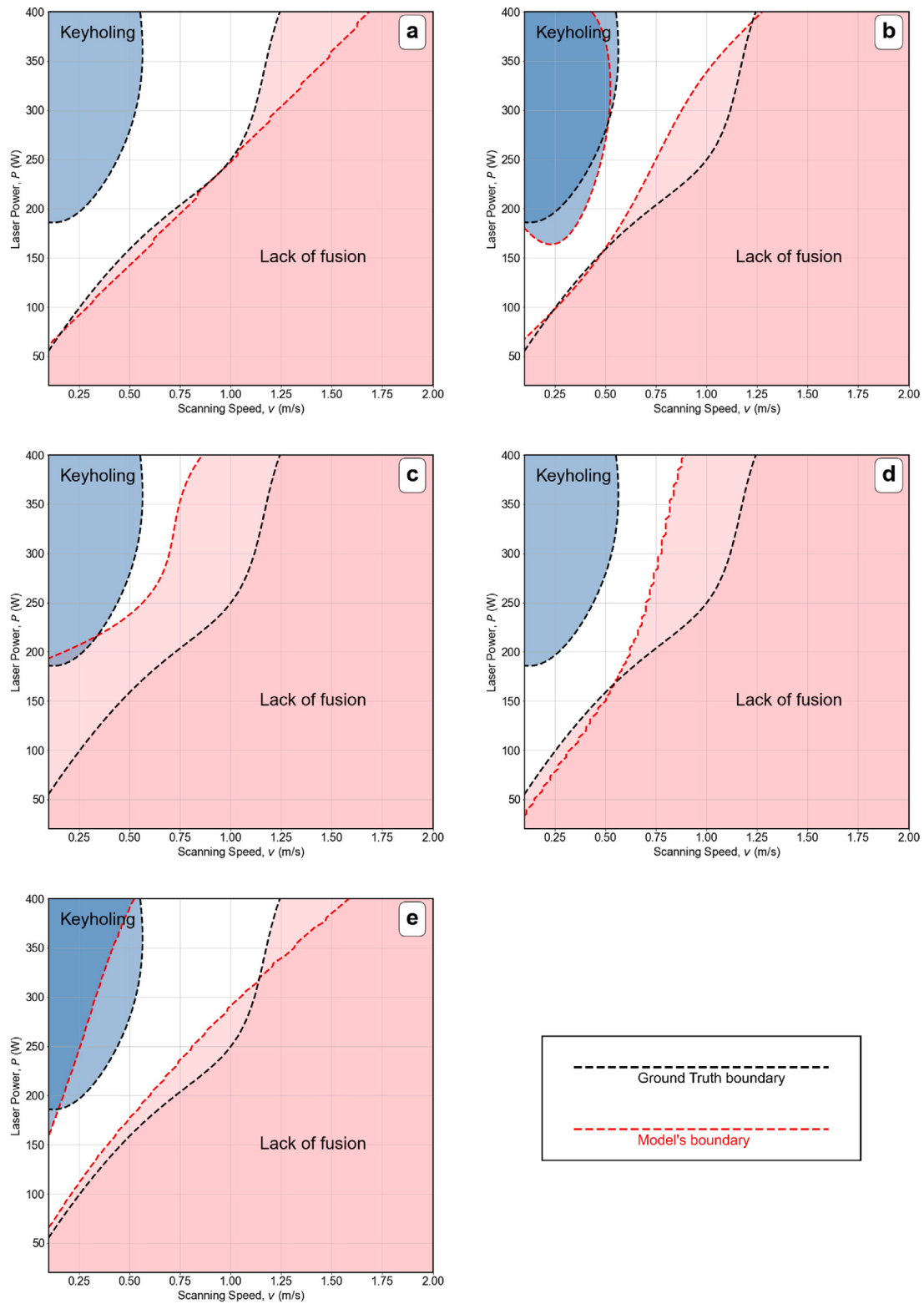


Fig. 10. Printability maps for nickel Alloy 718 generated using 10,000 data points of melt pool width and depth predicted using (a) Model 1, (b) Model 2, (c) Model 3, (d) Model 4, (e) Model 5. The black dashed lines show the boundary of the keyhole and lack of fusion regions from the printability map generated using the MCMC method reported in [2] denoted as the ground truth, while the red dashed lines show the boundary of the keyhole and lack of fusion regions generated from the different models used in the benchmarking. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

Table 4

Median and interquartile range for the four error metrics calculated from training each model for 250 instances on different training and testing datasets. The four error metrics are for melt pool depth of alloy AF9628 and nickel alloy 718. MAE: Mean absolute error, RMSE: Root mean square error, τ : Kendall rank coefficient, and R_s : Spearman rank coefficient.

Model		MAE		RMSE		τ		R_s	
		Median	IQR	Median	IQR	Median	IQR	Median	IQR
AF9628	ET	35.58	9.88	62.63	15.35	0.827	0.05	0.938	0.03
	GPR	16.79	6.84	24.62	12.79	0.814	0.09	0.934	0.05
	GPR+Prior	13.87	7.80	21.10	15.10	0.834	0.09	0.944	0.05
	GPR+Proxy	16.85	7.03	23.72	12.14	0.809	0.08	0.934	0.05
	GPR+Proxy+Prior	13.89	8.70	20.92	14.11	0.840	0.10	0.948	0.06
Ni Alloy 718	ET	44.02	9.17	71.01	14.43	0.859	0.03	0.961	0.01
	GPR	28.66	7.41	48.15	10.18	0.825	0.05	0.947	0.03
	GPR+Prior	28.07	8.12	46.70	10.38	0.836	0.06	0.952	0.03
	GPR+Proxy	28.90	6.54	47.15	9.27	0.831	0.05	0.949	0.02
	GPR+Proxy+Prior	28.02	9.40	45.86	12.45	0.831	0.05	0.947	0.03

Table 5

Median and interquartile range for the four error metrics calculated from training each model for 250 instances on different training and testing datasets. The four error metrics are for the melt pool width of RMPEAs. MAE: Mean absolute error, RMSE: Root mean square error, τ : Kendall rank coefficient, and R_s : Spearman rank coefficient.

Model	MAE μm		RMSE μm		τ		R_s	
	Median	IQR	Median	IQR	Median	IQR	Median	IQR
ET	57.95	2.53	60.82	2.39	0.64	0.04	0.82	0.02
ET+GPR	19.79	17.10	23.26	18.12	0.79	0.11	0.93	0.06
ET+Prior	9.61	2.49	12.09	2.70	0.82	0.03	0.95	0.02

the predicted thermo-physical properties of these alloys are provided in the code capsule associated with this work.

The cross-validation scheme is conducted at the alloy level as shown in Fig. 3. Specifically, from the 16 alloys, the single-tracks on 10 alloys are used as training data, and the single-tracks on the remaining 6 alloys are used as test data. This cross-validation scheme is repeated 250 times and a distribution of error metrics is reported.

Fig. 11a & b show the error metrics and rank ordering metrics for the 5 models when extrapolating melt pool width. The 2 GPR-based models, M2 and M3 have comparable MAE and RMSE metrics. However, both M2 and M3 significantly outperform the analytical control model M1 in predicting width. Regarding the rank-ordering metrics shown in Fig. 11c & d, the analytical control model M1 has the best metrics, followed by the GPR model equipped with the physics-based prior M3, and finally followed by the solely data-driven model M2. The results are summarized in Table 5. These results indicate that the proposed method has competitive error metrics and acceptable rank-ordering metrics.

Fig. 11a & b show the error metrics for the 5 models when extrapolating melt pool depth. Models M2-M5 outperform the analytical control model M1 concerning MAE and RMSE. The error metrics of M2-M5 are extremely similar. Despite this, in Fig. 11c & d, it is clear that the GPR that utilizes a physics-based prior (M3) has the best rank ordering coefficients out of models M2-M5. While M1 has the best rank ordering coefficient, recall that it also has the worst error metrics. The results are summarized in Table 6. These results indicate that M3 strikes an optimal balance between low error and physically realistic results.

3.3. Case study on linear crack density

The proposed framework was adapted to predict crack density in addition to melt pool dimensions. Crack number density data from 195 single-track experiments on the 16 solid RMPEA compositions in Table 2 were used for training and testing the models. Details concerning the measurement of the crack density are provided in Section 2.5.

In this case study, because we do not have an explicit prior for crack number density, we will demonstrate a hierarchical GP that uses

GP predictions for width to enhance GP predictions for crack number density. We will equip the GP for width with the ET model's prediction for width as an informative prior. M1 will be a vanilla GP trained only on crack number density using the features listed in Table 1. M2 will be a hierarchical GP. The first GP in this sequence will be a vanilla GPR trained on width. The predictions from the width GPR will be used as an input feature for a GPR trained on linear crack density (in addition to the features listed in Table 1). M3 is the double-test model, where a GP for width is equipped with the ET model as an informative prior for how width varies as a function of chemistry and processing parameters. The width prediction from this physics-informed GP is then passed as an input to a vanilla GP trained on crack number density. The code and data associated with this case study are provided in the repository associated with this work.

The cross-validation scheme mirrored the procedure outlined in Section 3.2, utilizing crack density data from 16 alloys. Of these data, the single-tracks on 10 alloys were used for training, while the single-tracks on the remaining 6 alloys served as the testing set. This benchmarking scheme represents the alloy design scenario where one wishes to predict the cracking tendency of an alloy that has yet to be synthesized. The cross-validation was repeated 250 times, and the resulting error metrics were analyzed and reported.

Fig. 13a and b present the error metrics for the three models when predicting crack number density, while Fig. 13c and d present the rank ordering metrics for the 3 models when predicting crack number density. In Fig. 13a, it is evident that M3 has the best MAE metrics in general. Interestingly, the MAE for M2 is the worst on average, indicating that hierarchical GP regression is only beneficial when the first GP layer in the sequence of GPs is physics-informed. The same conclusion can be made from Fig. 13b, indicating that the double-test model has the best error metrics, indicating our method outperforms the control models. The statistics associated with these MAE and RMSE distributions are reported in Table 7.

Fig. 13c and d show the rank ordering metrics associated with the three models. The distributions of the τ and R_s rank ordering metrics indicate that our double-test method (M3) outperforms the other models with median Kendall and Spearman rank coefficients of 0.5837 and 0.7668, respectively. This indicates that hierarchical GP regression with informative priors on the first GP layer in the hierarchy is a promising method for exploiting multiple sources of experimental and physics-based information during regression. The single-test model (i.e. M2, the vanilla hierarchical GP) performed the second best with regard to rank order. The control single vanilla GP (M1) performed the worst with regard to rank ordering. The statistics associated with these τ and R_s distributions are reported in Table 7.

To conclude this case study, we have showed that when width observations can supplement predictions for crack density, especially when the width predictions are enhanced with informative priors. This demonstrates that the proposed method is versatile and can be applied various defect modes prevalent in AM.

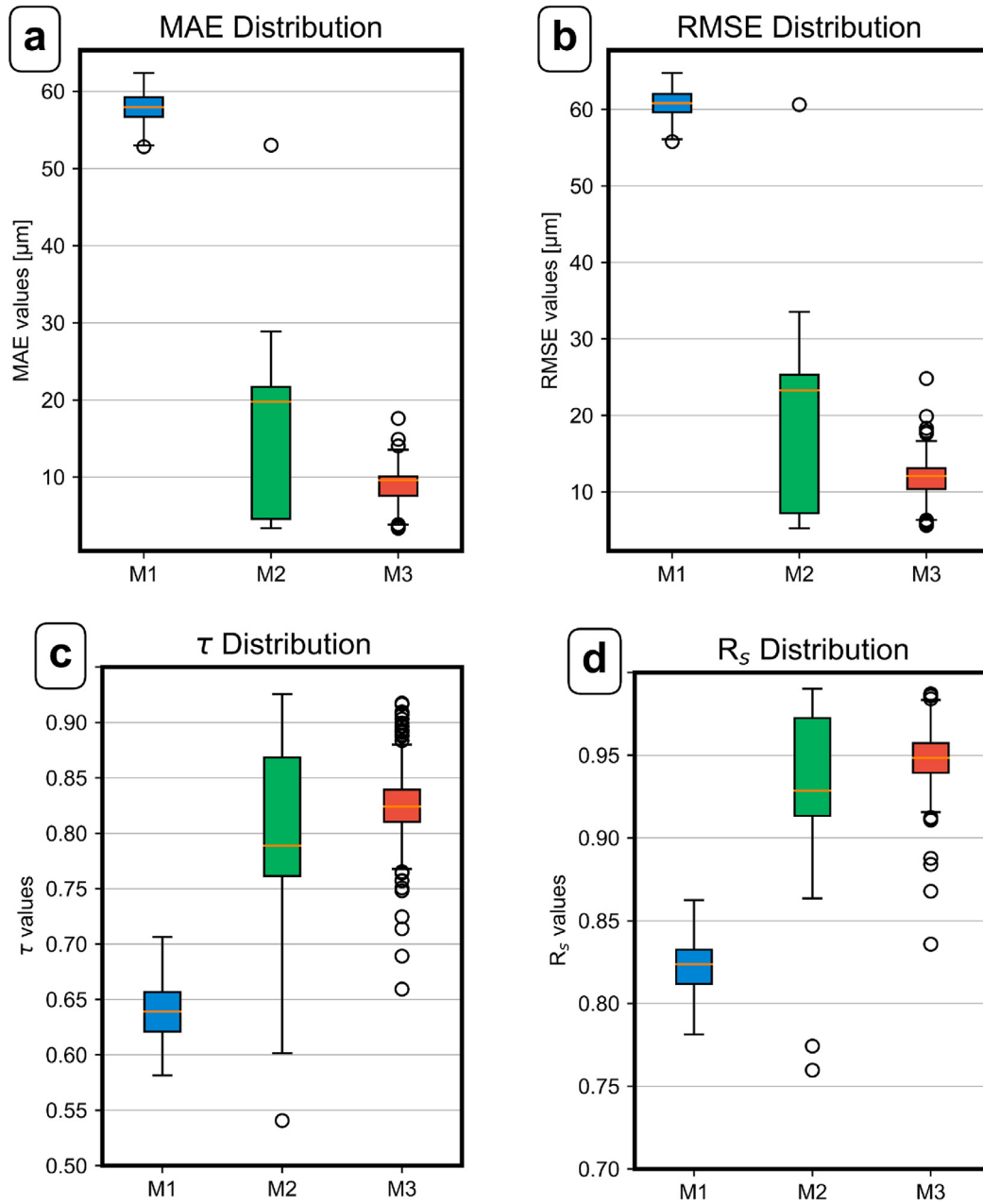


Fig. 11. Comparative analysis of the three models benchmarked in this study for melt pool width of RMPEAs. In M1 (blue), the melt pool width calculated from the ET model is used, in M2 (green), a GPR model is trained on 10 alloy compositions, while in M3 (red), a GPR model is trained on 10 alloy compositions and ET model is used as an informative prior. The distributions of (a) mean absolute error (MAE), (b) root mean square error (RMSE), (c) Kendall rank coefficient (τ), (d) Spearman rank coefficient (R_s), are shown for 250 instances. The proposed model (M3 — Red) has the lowest MAE and RMSE, and the highest τ and R_s . (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

Table 6

Median and interquartile range for the four error metrics calculated from training each model for 250 instances on different training and testing datasets. The four error metrics are for the melt pool depth of RMPEA. MAE: Mean absolute error, RMSE: Root mean square error, τ : Kendall rank coefficient, and R_s : Spearman rank coefficient.

Model	MAE (μm)		RMSE (μm)		τ		R_s	
	Median	IQR	Median	IQR	Median	IQR	Median	IQR
ET	25.58	1.28	38.29	2.19	0.57	0.05	0.75	0.04
GPR	21.69	15.89	27.66	13.09	0.73	0.09	0.90	0.06
GPR+Proxy	18.95	4.50	23.71	3.66	0.75	0.08	0.91	0.05
GPR+Prior	25.55	17.80	32.36	17.52	0.70	0.08	0.89	0.05
GPR+Proxy+Prior	16.33	4.47	23.04	5.00	0.74	0.06	0.90	0.04

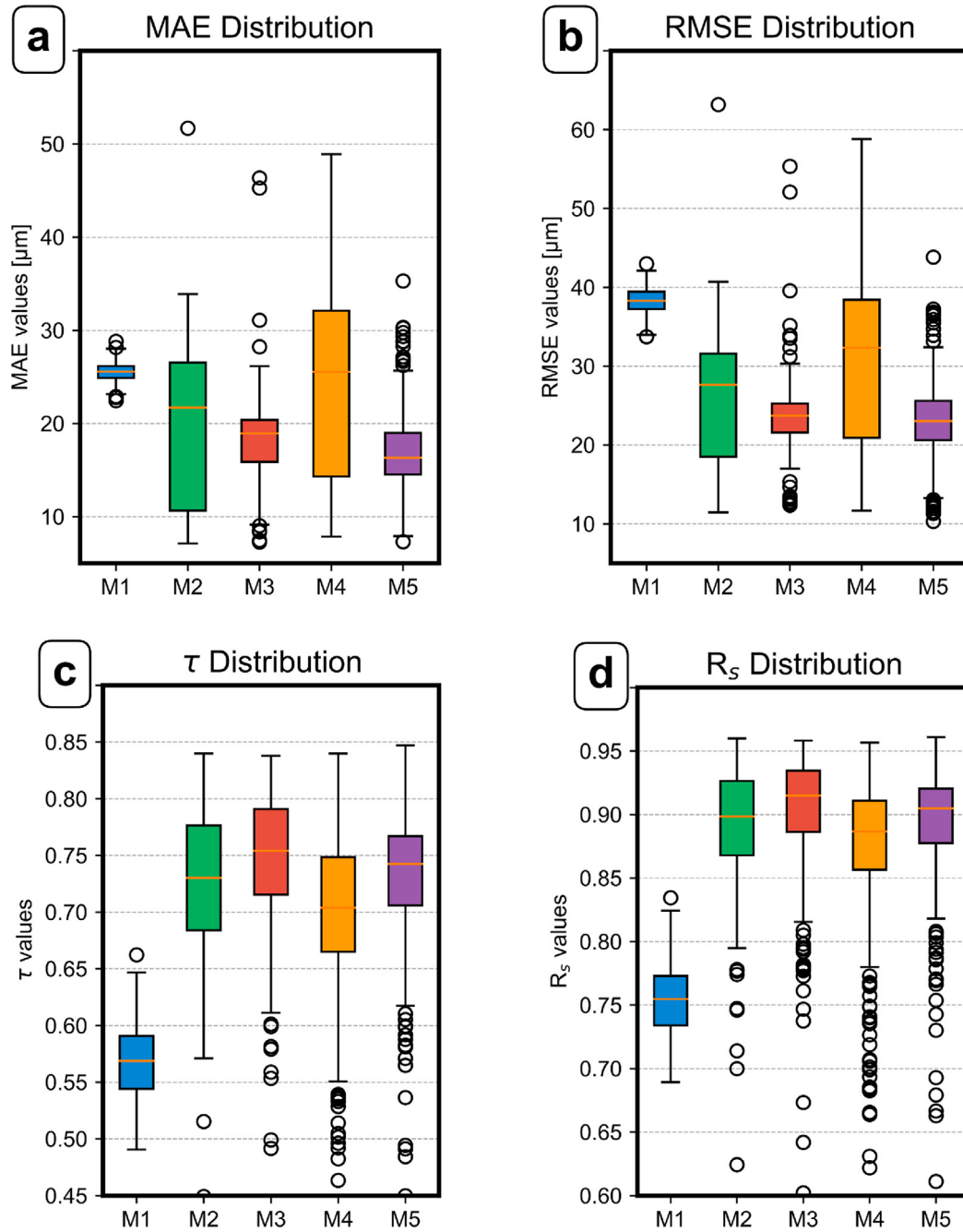


Fig. 12. Comparative analysis of the five models benchmarked in this study for the melt pool depth of RMPEAs. In M1 (blue), the melt pool width calculated from the ET model is used, in M2 (green), a GPR model is trained on 10 alloy compositions, in M3 (red), a GPR is used to impute width as an informative feature for a GPR trained on depth, in M4 (orange), GPR uses ET predictions as an informative prior, while M5 (Purple) combines ET predictions as an informative prior and proxy width. The distributions of (a) mean absolute error (MAE), (b) root mean square error (RMSE), (c) Kendall rank coefficient (τ), (d) Spearman rank coefficient (R_s). The proposed model (M5 — purple) has the lowest MAE and RMSE, and high values for τ and R_s when compared to the other four models. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

Table 7

Median and interquartile range for the four error metrics calculated from training each model for 250 instances on different training and testing datasets for crack number densities. The four error metrics are for the crack number density of RMPEAs. MAE: Mean absolute error, RMSE: Root mean square error, τ : Kendall rank coefficient, and R_s : Spearman rank coefficient.

Model	MAE (mm^{-1})		RMSE (mm^{-1})		τ		R_s	
	Median	IQR	Median	IQR	Median	IQR	Median	IQR
GPR	0.5651	0.0709	0.7169	0.0827	0.5669	0.0591	0.7502	0.0657
GPR+Proxy	0.5982	0.1331	0.7781	0.1859	0.5772	0.0624	0.7607	0.0618
GPR+(Proxy+Prior)	0.5342	0.0931	0.6531	0.1099	0.5837	0.0631	0.7668	0.0672

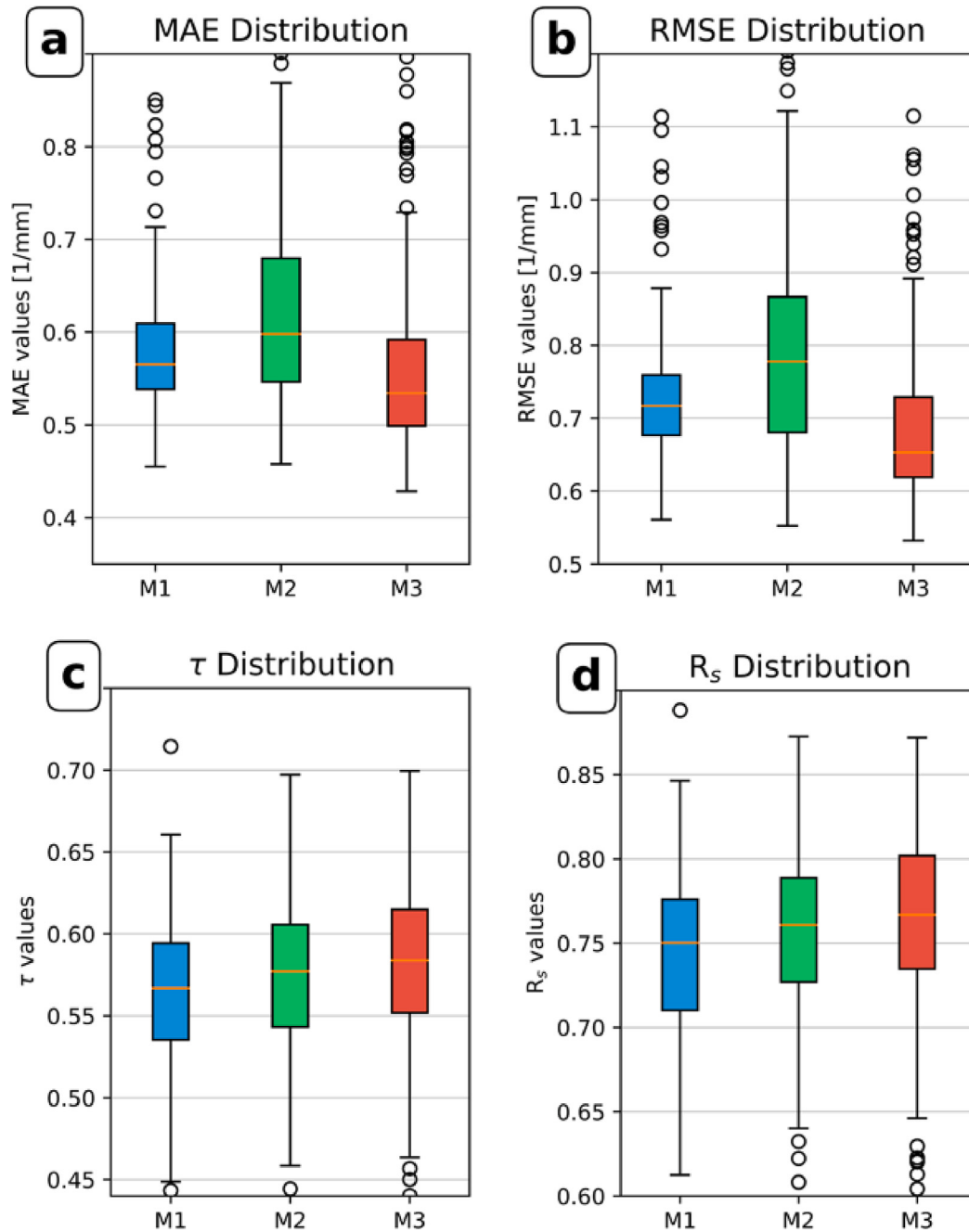


Fig. 13. Comparative analysis of the three models benchmarked in this study for the crack number density of RMPEAs. In M1 (blue), the crack number density calculated from the vanilla GP model is used. In M2 (green), a hierarchical GP model is used. This model consists of a vanilla GPR trained on width. These width predictions are used as an input feature for a GPR that is trained on linear crack number density (in addition to the features listed in Table 1). M3 (red) is the double-test model, where a GP for width is equipped with the ET model as an informative prior for how width varies as a function of chemistry and processing parameters. These improved width predictions are passed as input features to a vanilla GPR trained on linear crack density. The distributions of (a) mean absolute error (MAE), (b) root mean square error (RMSE), (c) Kendall rank coefficient (τ), (d) Spearman rank coefficient (R_s). The proposed model (M3 — red) has the lowest median MAE and RMSE, and high values for τ and R_s when compared to the other two models. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

3.4. Case study on probabilistic defect maps

In Section 3.1, we demonstrated that the proposed method more accurately recreates the ground truth printability maps reported in Ref. [2,18] compared to the control models. This is significant because our method is computationally more efficient, as it does not require MCMC calibration, unlike the model in Ref. [25]. However, it is important to note that the printability maps presented in Section 3.1 have deterministic decision boundaries and do not show the models'

uncertainty. A key reason we believe GPR models are highly suitable for alloy design is their ability to quantify uncertainty. We therefore deem it important to demonstrate how probabilistic defect maps can be made using these GPR-based models.

As discussed in the introduction, a major advantage of Bayesian methods is their ability to quantify uncertainty. In Gaussian Process Regression (GPR), predictions and their uncertainties are represented as normal distributions. Specifically, the model's prediction corresponds to the mean, while the uncertainty is expressed through the standard

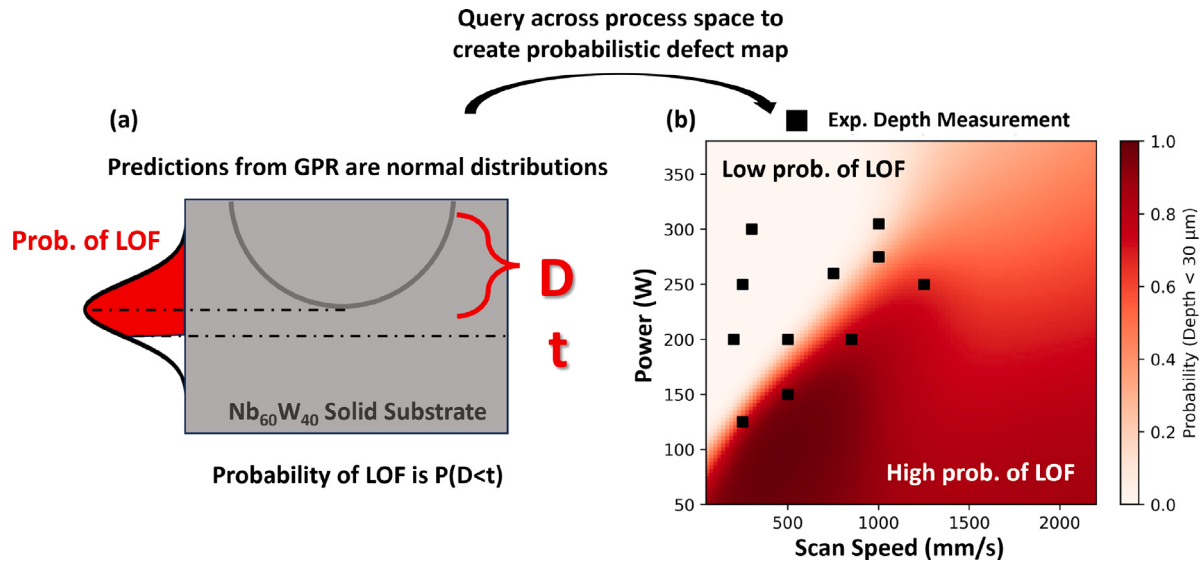


Fig. 14. (a) A schematic representation of GPR predictions, illustrating the mean melt-pool depth (solid line) and uncertainty (shaded area) represented by the standard deviation. The critical threshold for lack of fusion (LOF) defects is indicated by the layer thickness. The probability of a defect is derived from the area under the normal distribution that exceeds this threshold. (b) A probabilistic defect map showing the likelihood of LOF defects across varying power and velocity combinations. The probabilistic boundary between defect and non-defect regions softens, reflecting increased uncertainty, especially in regions with sparse data. The transition from deterministic to probabilistic boundaries demonstrates the ability of GPR models to quantify prediction uncertainty in alloy design applications.

deviation. For example, the left panel of Fig. 14 shows a schematic of GPR depth predictions, where the mean indicates the predicted depth, and the uncertainty is conveyed by the standard deviation's range.

In the scenario depicted in Fig. 14a, the mean depth prediction is below the layer thickness, indicating a lack of fusion. However, due to the uncertainty in the predictions, there remains a small probability that the depth will exceed the layer thickness, meaning this melt-pool might not be classified as a lack of fusion. This probability is determined by the area under the normal distribution that extends beyond the layer thickness threshold. Essentially, this is a typical probability problem, where the goal is to find the probability that a normally distributed random variable exceeds a given value. The solution involves calculating the cumulative probability from the standard normal distribution, which gives the probability of the variable being greater than the threshold.

To create a probabilistic defect map, we query the GPR model across a grid of power and velocity combinations. For each combination, the model provides a normal distribution, from which we calculate the probability that the melt pool depth exceeds the critical thickness threshold. Plotting these probabilities results in a probabilistic defect map, as shown in Fig. 14b.

In Fig. 14b, the boundary for lack of fusion (LOF) defects becomes probabilistic rather than deterministic, with a more diffuse transition. The decision boundary softens, particularly in the top-right corner where no single-track data points exist, and the probabilities tend toward a 50/50 split. In this way the uncertainty in the depth prediction and defect classification is quantified.

4. Summary and conclusion

Data-driven printability maps demand significant effort and resources, e.g. the procurement of powder, the printing of single tracks, and the precise measurement of melt pool dimensions. On the other hand, simulation-based printability maps require extensive computational resources to accurately capture the intricate physics inherent in the laser melting process, especially when employing high-fidelity models. Employing low-fidelity models to generate printability maps may sacrifice the accuracy of predictions, particularly when simulating phenomena such as keyholing. These limitations underscore the necessity of integrating both low-cost computational simulations and

high-fidelity experimental measurements to generate reliable printability maps. Despite existing methods in the literature [24,25], challenges persist. Specifically, a quickly-corrected data-driven model for printability is required for high-throughput alloy design.

In this study, we propose a novel approach that combines machine learning models with a low-fidelity thermal model, culminating in a computationally efficient yet accurate model capable of leveraging multiple data sources to generate printability maps in an HTP manner. Specifically, our approach has two aspects: (1) We propose using the ET model as an informative prior for width and depth during GP regression. (2) We propose exploiting the correlation between width and depth via Hierarchical GP regression to improve depth predictions using width measurements. We benchmarked these methods applied both separately and in tandem against two control models — the analytical ET model and a standard GP regressor. These models are described in detail below:

- **Model 1:** The melt pool width and depth are directly computed using the ET thermal model. This is the analytical control model.
- **Model 2:** Two distinct Gaussian Process Regression (GPR) models are employed for predicting melt pool width and depth. This approach is solely data-driven and does not incorporate physics-based priors.
- **Model 3:** A GPR predicts the width, which then serves as an informative feature for another GPR trained on depth. Neither the width nor depth GPRs integrate informative priors. Essentially, benchmarking this model probes the effect of using width as a proxy for depth.
- **Model 4:** Similar to M2, separate GPR models for width and depth are used. However, in this case, the predicted melt pool width/depth from the ET model serves as an informative prior.
- **Model 5:** Similar to M3, a GPR predicts width, informing another GPR trained on depth. However, similar to M4, both GPRs utilize the predicted melt pool width or depth from the ET model as informative priors. This model investigates the combined effect of informative priors and using width as a proxy for depth.

The models were evaluated using two error metrics and two rank-order metrics. By using both error and rank-ordering metrics we better evaluate the *physicality* of printability predictions and avoid over-fitting. All models underwent training under data-sparse conditions

using 2-fold cross-validation, repeated 250 times on distinct training and testing datasets to identify the optimal model.

Initially, our study focused on two distinct alloy compositions (AF9628 and Nickel Alloy 718), extensively studied in literature with rigorous calibration techniques applied to ET model parameters. Results revealed that the double-test model (Model 5) consistently outperformed other models on average, demonstrating the lowest errors and highest rank-order coefficient distributions, indicating accurate and reasonable melt pool predictions. Furthermore, validation was extended by generating printability maps for both alloys and comparing them to ground truth maps from the literature. Printability maps generated from melt pool predictions from Model 5 exhibited strong agreement with ground truth maps for both alloys.

Finally, we illustrated the efficacy of our framework in HTP alloy design by extrapolating printability predictions as a function of chemistry. Specifically, we use single-track experiments on 10 RMPEAs to train M2–M5. When then predict the melt pool dimension of single-tracks on 6 RMPEAs. This train-test splitting was performed 250 times better benchmark the model. In this way we probe the ability of these models to predict the printability of alloys for which no training data is available. Such a scenario is important for alloy design as printability predictions must be made *a priori* and can only be validated after synthesis and testing.

Finally, to demonstrate the versatility of the proposed framework in addressing common defects in AM we conducted a case study where we model the crack number density of 195 single-tracks printed on 16 unique refractory substrates. Specifically, we use the same benchmarking scheme as above, considering the single-tracks on 10 RMPEAs as training data and considering the single-tracks on the remaining 6 RMPEAs as testing data. We repeat the train-test process 250 times. Again, this benchmarking scheme represents the scenario in which we are conducting alloy design i.e. predicting the crack number density for single-tracks on compositions that have yet to be synthesized. Although we lack explicit priors for crack density, we exploit the correlation between melt pool width and crack density. By equipping the GPR for width with a physics-based prior, we enhance the predictions for linear crack density, demonstrating the method's broader applicability to a range of defects in AM. This benchmark is important because if linear crack densities can be predicted *a priori* for alloy-process pairs, then they can be minimized in the context of alloy design.

In conclusion, changing the prior mean function from an uninformative constant value to a physics-informed function during GP regression improves the accuracy and physical relevance of melt pool dimension predictions. These benefits are particularly evident during extrapolation. By updating the ET model prior with experimental melt pool dimensions, we effectively create a data-corrected version of the ET model. In this way we benefit from the fast-acting analytical nature of the ET model and from the fidelity of experimental melt pool data. When Hierarchical GP regression is used to supplement depth predictions with data-informed width predictions (and both of these models are equipped with informative priors), model performance increases more, indicating a synergy between these 2 methods. The lessons-learned from this work are summarized below:

- **Combining Analytical Models and Data:** By using the ET model as a physics-based prior response surface for melt pool width and depth and subsequently updating this response surface in light of experimental melt pool measurements, we effectively correct the ET model with data in a non-parametric manner. This approach is particularly effective when prior models are fast-acting (as is the case with the ET model) and experimental data is sparse.
- **Exploiting Correlation Between Width and Depth:** Melt pool width is easier to measure than melt pool depth. Hierarchical GP regression can exploit the correlation between width and depth. In this way, measurements of width (a 'seen' melt pool dimension) can inform predictions of depth (an 'unseen' melt pool

dimension). Recall in this context a Hierarchical GPR is a set of 2 GPRs: a GPR for width whose output is used as a feature in a GPR for depth. When these GPRs are equipped with the ET model as a prior, the performance of the Hierarchical GPR for depth is greatly improved.

- **Exploiting Correlation Between Width and Crack Number Density:** The melt pool width is an easily measurable quantity from single-tracks. The melt pool width and linear crack density are correlated. Hierarchical GP regression with informative priors showed improved accuracy in predicting crack number density over vanilla GP regressors. Being able to predict the crack number density of single-tracks prior to synthesizing the alloy is critical for alloy design as crack density can be used as a screening metric for printable alloys [30].

While this study focuses on printability maps derived from single-track experiments, we acknowledge that actual LPBF processes involve layer-by-layer accumulation. In multi-layer builds, complex thermal histories alter subsequent melt pools, making them quite different from single tracks. The method we propose, which uses Gaussian Process Regressions (GPRs) with informative priors, could theoretically be extended to more sophisticated models — such as those that account for thermal histories, like the Enhanced Analytical Solidification Model (EASM) [41]. However, this level of complexity lies beyond the scope of the current study and we recommend this as future work.

Porosity defects are not the only problem present during LPBF of alloys. There are a suite of problems that still must be addressed such as hot-cracking, thermal shock induced cracking, undesired microstructures, and distortions/warping caused by the high residual stresses. In this work we presented a model that address only one of these issues (porosity defects) however we posit that alloy design cannot be myopic but instead must consider all of these issues simultaneously. In previous work we posit that the area of the processing window that resistant to porosity defects should be used as a design metric. In this way it can be used with other printability metrics such as hot-cracking indicators and ductility indicators to design alloys that are resistant to porosity and cracking during LPBF. To better put the proposed tool in context of alloy design for printable alloys we conclude by referring the reader to Refs. [16,40] where we ideate and execute alloy design campaigns toward printable alloys.

CRedit authorship contribution statement

Peter Morcos: Writing – review & editing, Writing – original draft, Visualization, Software, Methodology, Investigation, Formal analysis, Data curation. **Brent Vela:** Writing – review & editing, Writing – original draft, Visualization, Validation, Methodology, Investigation, Formal analysis, Data curation, Conceptualization. **Cafer Acemi:** Writing – original draft, Investigation, Data curation. **Alaa Elwany:** Writing – review & editing, Supervision, Project administration, Funding acquisition. **Ibrahim Karaman:** Writing – review & editing, Supervision, Project administration, Funding acquisition. **Raymundo Arroyave:** Writing – review & editing, Supervision, Project administration, Funding acquisition.

Code availability

The code and data associated with this work can be found at: <https://rb.gy/ed3l0n>.

Declaration of Generative AI and AI-assisted technologies in the writing process

During the preparation of this work the author(s) used ChatGPT version 3.5 in order to ideate/brainstorm alternative sentence structures and stylistic choices in limited sections of the paper. Generative AI was used on a sentence-by-sentence basis and was not used to generate complex sentence structures. After using this tool/service, the author(s) reviewed and edited the content as needed and take(s) full responsibility for the content of the publication.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgments

The authors acknowledge the NASA-ESI Program under Grant No. 80NSSC21K0223 and National Science Foundation through Grant No. DMREF-2323611. The authors also wish to acknowledge the support from the U.S. Department of Energy (DOE) ARPA-E ULTIMATE Program through Project DE-AR0001427. BV acknowledges Grant No. NSF-DGE-1545403 (NSF-NRT: Data-Enabled Discovery and Design of Energy Materials, D³EM) and Grant No. 1746932. RA, AE and IK acknowledge support from NSF through Grant No. DMREF-2323611. High-throughput CALPHAD calculations were carried out in part at the Texas A&M High-Performance Research Computing (HPRC) Facility.

Data availability

Data is released in the code repository linked in the paper.

References

- ## Declaration of competing interest
- The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.
- ## Acknowledgments
- The authors acknowledge the NASA-ESI Program under Grant No. 80NSSC21K0223 and National Science Foundation through Grant No. DMREF-2323611. The authors also wish to acknowledge the support from the U.S. Department of Energy (DOE) ARPA-E ULTIMATE Program through Project DE-AR0001427. BV acknowledges Grant No. NSF-DGE-1545403 (NSF-NRT: Data-Enabled Discovery and Design of Energy Materials, D³EM) and Grant No. 1746932. RA, AE and IK acknowledge support from NSF through Grant No. DMREF-2323611. High-throughput CALPHAD calculations were carried out in part at the Texas A&M High-Performance Research Computing (HPRC) Facility.
- ## Data availability
- Data is released in the code repository linked in the paper.
- ## References
- [1] B. Soundararajan, D. Sofia, D. Barletta, M. Poletto, Review on modeling techniques for powder bed fusion processes based on physical principles, *Addit. Manuf.* 47 (2021) 102336, <http://dx.doi.org/10.1016/j.addma.2021.102336>, URL <https://www.sciencedirect.com/science/article/pii/S2214860421004942>.
 - [2] D. Shoukr, P. Morcos, T. Sundermann, T. Dobrowolski, C. Yates, J.R. Jain, R. Arróyave, I. Karaman, A. Elwany, Influence of layer thickness on the printability of nickel alloy 718: A systematic process optimization framework, *Addit. Manuf.* 73 (2023) 103646, <http://dx.doi.org/10.1016/j.addma.2023.103646>, URL <https://www.sciencedirect.com/science/article/pii/S2214860423002592>.
 - [3] C.K.P. Vallabh, S. Sridar, W. Xiong, X. Zhao, Predicting melt pool depth and grain length using multiple signatures from in-situ single camera two-wavelength imaging pyrometry for laser powder bed fusion, *J. Mater. Process. Technol.* 308 (2022) 117724, <http://dx.doi.org/10.1016/j.jmatprotec.2022.117724>, URL <https://www.sciencedirect.com/science/article/pii/S0924013622002357>.
 - [4] M. Naderi, J. Weaver, D. Deisenroth, N. Iyyer, R. McCauley, On the fidelity of the scaling laws for melt pool depth analysis during laser powder bed fusion, *Integr. Mater. Manuf. Innov.* 12 (1) (2023) 11–26.
 - [5] B. Zhang, R. Seede, L. Xue, K.C. Atli, C. Zhang, A. Whitt, I. Karaman, R. Arroyave, A. Elwany, An efficient framework for printability assessment in laser powder bed fusion metal additive manufacturing, *Addit. Manuf.* 46 (2021) 102018, <http://dx.doi.org/10.1016/j.addma.2021.102018>, URL <https://www.sciencedirect.com/science/article/pii/S2214860421001834>.
 - [6] J. Wang, R. Zhu, Y. Liu, L. Zhang, Understanding melt pool characteristics in laser powder bed fusion: An overview of single- and multi-track melt pools for process optimization, *Adv. Powder Mater.* 2 (4) (2023) 100137.
 - [7] L. Xue, K. Atli, S. Picak, C. Zhang, B. Zhang, A. Elwany, R. Arroyave, I. Karaman, Controlling martensitic transformation characteristics in defect-free NiTi shape memory alloys fabricated using laser powder bed fusion and a process optimization framework, *Acta Mater.* 215 (2021) 117017.
 - [8] P. Morcos, A. Elwany, I. Karaman, R. Arróyave, Additive manufacturing of pure tungsten and tungsten-based alloys, *J. Mater. Sci.* 57 (21) (2022) 9769–9806.
 - [9] X. Ren, H. Liu, F. Lu, L. Huang, X. Yi, Effects of processing parameters on the densification, microstructure and mechanical properties of pure tungsten fabricated by optimized selective laser melting: From single and multiple scan tracks to bulk parts, *International Journal of Refractory Metals and Hard Materials* 96 (2021) 105490, <http://dx.doi.org/10.1016/j.ijrmhm.2021.105490>, URL <https://www.sciencedirect.com/science/article/pii/S0263436821000226>.
 - [10] A. Iveković, M.L. Montero-Sistiaga, G. Vanmeensel, J.-P. Kruth, J. Vleugels, Effect of processing parameters on microstructure and properties of tungsten heavy alloys fabricated by SLM, *Int. J. Refract. Met. Hard Mater.* 82 (2019) 23–30, <http://dx.doi.org/10.1016/j.ijrmhm.2019.03.020>, URL <https://www.sciencedirect.com/science/article/pii/S0263436819300216>.
 - [11] D. Zhang, Q. Cai, J. Liu, R. Li, Research on process and microstructure formation of W-Ni-Fe alloy fabricated by selective laser melting, *J. Mater. Eng. Perform.* 20 (6) (2011) 1049–1054.
 - [12] T. Wilkinson, M. Casata, D. Barba, Automatic process mapping for Ti64 single tracks in laser powder bed fusion, in: S. Wagstaff, A. Anderson, A.S. Sabau (Eds.), *Materials Processing Fundamentals 2023*, Springer Nature Switzerland, Cham, 2023, pp. 199–209.
 - [13] W. Zhang, B. Abravanti, J. Hanson-Regalado, C. Koz, B. Duvvuri, K. Shimada, J. Beuth, L.B. Kara, Flaw detection in metal additive manufacturing using deep learned acoustic features, in: *Workshop of Conference on Neural Information Processing Systems*, 2020.
 - [14] C. Zhang, L. Xue, K.C. Atli, R. Arróyave, I. Karaman, A. Elwany, On the Fabrication of Defect-Free Nickel-Rich Nickel–Titanium Parts Using Laser Powder Bed Fusion, *J. Manuf. Sci. Eng.* 144 (9) (2022) 091011, <http://dx.doi.org/10.1115/1.4054935>, arXiv:https://asmedigitalcollection.asme.org/manufacturingscience/article-pdf/144/9/091011/6901568/manu_144_9_091011.pdf.
 - [15] P. Wang, Y. Yang, N.S. Moghaddam, Process modeling in laser powder bed fusion towards defect detection and quality control via machine learning: The state-of-the-art and research challenges, *J. Manuf. Process.* 73 (2022) 961–984, <http://dx.doi.org/10.1016/j.jmapro.2021.11.037>, URL <https://www.sciencedirect.com/science/article/pii/S15266112521008446>.
 - [16] S. Sheikh, B. Vela, V. Attari, X. Huang, P. Morcos, J. Hanagan, C. Acemi, I. Karaman, A. Elwany, R. Arroyave, Exploring chemistry and additive manufacturing design spaces: a perspective on computationally-guided design of printable alloys, *Mater. Res. Lett.* 12 (4) (2024) 235–263.
 - [17] R. Arróyave, D. Khatamsaz, B. Vela, R. Couperthwaite, A. Molkeri, P. Singh, D.D. Johnson, X. Qian, A. Srivastava, D. Allaire, A perspective on Bayesian methods applied to materials discovery and design, *MRS Commun.* 12 (6) (2022) 1037–1049.
 - [18] R. Seede, D. Shoukr, B. Zhang, A. Whitt, S. Gibbons, P. Flater, A. Elwany, R. Arroyave, I. Karaman, An ultra-high strength martensitic steel fabricated using selective laser melting additive manufacturing: Densification, microstructure, and mechanical properties, *Acta Mater.* 186 (2020) 199–214, <http://dx.doi.org/10.1016/j.actamat.2019.12.037>, URL <https://www.sciencedirect.com/science/article/pii/S1359645419308766>.
 - [19] L. Johnson, M. Mahmoudi, B. Zhang, R. Seede, X. Huang, J.T. Maier, H.J. Maier, I. Karaman, A. Elwany, R. Arróyave, Assessing printability maps in additive manufacturing of metal alloys, *Acta Mater.* 176 (2019) 199–210, <http://dx.doi.org/10.1016/j.actamat.2019.07.005>, URL <https://www.sciencedirect.com/science/article/pii/S1359645419304355>.
 - [20] L. Xue, K. Atli, S. Picak, C. Zhang, B. Zhang, A. Elwany, R. Arroyave, I. Karaman, Controlling martensitic transformation characteristics in defect-free NiTi shape memory alloys fabricated using laser powder bed fusion and a process optimization framework, *Acta Mater.* 215 (2021) 117017, <http://dx.doi.org/10.1016/j.actamat.2021.117017>, URL <https://www.sciencedirect.com/science/article/pii/S1359645421003979>.
 - [21] R. Seede, J. Ye, A. Whitt, W. Trehern, A. Elwany, R. Arroyave, I. Karaman, Effect of composition and phase diagram features on printability and microstructure in laser powder bed fusion: Development and comparison of processing maps across alloy systems, *Addit. Manuf.* 47 (2021) 102258, <http://dx.doi.org/10.1016/j.addma.2021.102258>, URL <https://www.sciencedirect.com/science/article/pii/S2214860421004188>.
 - [22] K. Atli, H. Boon, R. Seede, B. Zhang, A. Elwany, R. Arroyave, I. Karaman, Laser-based additive manufacturing of a binary Ni-wt, *J. Manuf. Process.* 62 (2021) 720–728, <http://dx.doi.org/10.1016/j.jmapro.2020.12.059>.
 - [23] T.W. Eagar, N. sien Tsai, Temperature fields produced by traveling distributed heat sources, 1983, URL

- [33] F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, J. Vanderplas, A. Passos, D. Cournapeau, M. Brucher, M. Perrot, E. Duchesnay, Scikit-learn: Machine learning in python, *J. Mach. Learn. Res.* 12 (2011) 2825–2830.
- [34] E. Schulz, M. Speekenbrink, A. Krause, A tutorial on Gaussian process regression: Modelling, exploring, and exploiting functions, *J. Math. Psych.* 85 (2018) 1–16, <http://dx.doi.org/10.1016/j.jmp.2018.03.001>, URL <https://www.sciencedirect.com/science/article/pii/S0022249617302158>.
- [35] C.E. Rasmussen, C.K. Williams, et al., *Gaussian Processes for Machine Learning*, vol. 1, Springer, 2006.
- [36] S. Karumuri, Z.D. McClure, A. Strachan, M. Titus, I. Bilonis, Hierarchical Bayesian approach to experimental data fusion: Application to strength prediction of high entropy alloys from hardness measurements, *Comput. Mater. Sci.* 217 (2023) 111851, <http://dx.doi.org/10.1016/j.commatsci.2022.111851>, URL <https://www.sciencedirect.com/science/article/pii/S0927025622005626>.
- [37] T. Fushiki, Estimation of prediction error by using K-fold cross-validation, *Stat. Comput.* 21 (2011) 137–146.
- [38] D. Khatamsaz, B. Vela, P. Singh, D.D. Johnson, D. Allaire, R. Arróyave, Bayesian optimization with active learning of design constraints using an entropy-based approach, *npj Comput. Mater. Sci.* 9 (1) (2023) 49.
- [39] J.H. Zar, Spearman rank correlation, *Encycl. Biostat.* 7 (2005).
- [40] C. Acemi, B. Vela, E. Norris, W. Trehern, K.C. Atli, C. Cleek, R. Arroyave, I. Karaman, Multi-objective, multi-constraint high-throughput design, synthesis, and characterization of tungsten-containing refractory multi-principal element alloys, *Acta Mater.* (2024) 120379.
- [41] J.C. Steuben, A.J. Birnbaum, J.G. Michopoulos, A.P. Iliopoulos, Enriched analytical solutions for additive manufacturing modeling and simulation, *Addit. Manuf.* 25 (2019) 437–447.