



Full length article

Data-augmented modeling for yield strength of refractory high entropy alloys: A Bayesian approach

Brent Vela^{a,*}, Danial Khatamsaz^a, Cafer Acemi^a, Ibrahim Karaman^a, Raymundo Arróyave^{a,b}^a Materials Science and Engineering Department, Texas A&M University, College Station, TX, 77840, USA^b J. Mike Walker '66 Department of Mechanical Engineering, Texas A&M University, College Station, TX, 77840, USA

ARTICLE INFO

Keywords:

Refractory high entropy alloys
 Temperature-dependent yield strength
 Alloy design
 Bayesian optimization
 Information fusion

ABSTRACT

Refractory high entropy alloys (RHEAs) have gained significant attention in recent years as potential replacements for Ni-based superalloys in gas turbine applications. Improving their properties, such as their high-temperature yield strength, is crucial to their success. Unfortunately, exploring this vast chemical space using exclusively experimental approaches is impractical due to the considerable cost of the synthesis, processing, and testing of candidate alloys, particularly at operation-relevant temperatures. On the other hand, the lack of reasonably accurate predictive property models, especially for high-temperature properties, makes traditional Integrated Computational Materials Engineering (ICME) methods inadequate. In this paper, we address this challenge by combining machine-learning models, easy-to-implement physics-based models, and inexpensive proxy experimental data to develop robust and fast-acting models using the concept of Bayesian updating. The framework combines data from one of the most comprehensive databases on RHEAs (Borg et al., 2020) with one of the most widely used physics-based strength models for BCC-based RHEAs (Maresca and Curtin, 2020) into a compact predictive model that is significantly more accurate than the state-of-the-art. This model is cross-validated, tested for physics-informed extrapolation, and rigorously benchmarked against standard Gaussian process regressors (GPRs) in a toy Bayesian optimization problem. Such a model can be used as a tool within ICME frameworks to screen for RHEAs with superior high-temperature properties. The code associated with this work is available at: <https://codeocean.com/capsule/7849853/tree/v2>.

1. Introduction

Refractory high entropy alloys (RHEAs) have gained significant attention over the past 10 years [1] due to their ability to retain their strength at temperatures much higher than the operating limit of most other structural alloys [2,3]. Very recently, initiatives such as ARPA-E's *ULTIMATE* program [4] have further encouraged research on refractory regions of the HEA space. Specifically, RHEAs could potentially replace state-of-the-art Ni-based superalloys in gas turbine engines, increasing their operating temperature and enhancing their efficiency. However, to supersede Ni-based superalloys, improvements must be made in the properties of these alloys. These properties include ductility, oxidation resistance, and, of particular interest to this work, *high temperature yield strength* (HT YS), while maintaining a density comparable to Ni-based superalloys.

Early research on RHEAs focused on equimolar compositions. Senkov et al. [5] pioneered research into RHEAs with the discovery that the equimolar RHEAs NbMoTaW and VNbMoTaW both possess competitive HT YS (405 MPa and 477 MPa at 1600 °C, respectively).

Although some equimolar RHEAs have been deemed to have attractive properties such as high melting points, high-temperature mechanical strength, and wear resistance, further improvements in their properties can be realized by deviating from equimolarity. For example, in our previous work [6], considering every permutation of the above alloys that Senkov et al. studied, we designed three high-strength, lightweight, non-equimolar RHEAs. Similarly, based on the results for the equimolar VNbZrTi alloys published in [7–9], Chen et al. [10] tuned the compositions within this chemistry space to improve strength and ductility.

Breaking away from conventional equimolar RHEAs and exploring the vast combinatoric chemistry space to find optimally strong alloys is a difficult task [11], especially considering the difficulty and the lack of high-temperature (>1000 °C) tensile testing results for RHEAs in the literature [3]. The RHEA design space is so vast and the cost of HT YS tests are so high, due to the requirement of testing under high vacuum levels, as a result of oxygen sensitivity of refractory alloys, large sample requirements of existing high-temperature tensile testing

* Corresponding author.

E-mail address: brentvela@tamu.edu (B. Vela).<https://doi.org/10.1016/j.actamat.2023.119351>

Received 5 August 2023; Received in revised form 15 September 2023; Accepted 15 September 2023

Available online 19 September 2023

1359-6454/© 2023 Acta Materialia Inc. Published by Elsevier Ltd. All rights reserved.

List of Symbols and Abbreviations

Symbol/Abrev.	Description
RHEA	Refractory High Entropy Alloy
HT	High temperature
RT	Room temperature
ICME	Integrated computational materials engineering
YS	Yield strength
HV	Vickers hardness
SSS	Solid Solution Strengthening
HTP	High-throughput
DFT	Density functional theory
GPR	Gaussian Process Regressor
HGPR	Gaussian Process Regressor
BO	Bayesian Optimization
δ	Lattice strain
G	Shear modulus
R	Metallic radius
E	Young's modulus
T_m	Solidus Temperature
VEC	Valence Electron Concentration
B	Bulk modulus
G	Shear modulus
ν	Poisson's Ratio
χ_{Allen}	Allen Electronegativity
$S_{conf.}$	Configurational Entropy
ρ	Density
C_{11} or C_{12} or C_{44}	Elastic constants

systems, and difficult-to-process nature of refractory alloys to obtain wrought microstructures, that conventional ICME methods for alloy design become impractical.

Conventional ICME methods for predicting YS can be broadly placed into the following categories, arranged in order of expense: (1) data-hungry empirical/machine-learning models based on statistical analysis [12] (2) computationally expensive atomistic and/or macroscale simulations to predict the yield strength [13–15], (3) and fast-acting analytical models that capture the microstructure and/or dislocation behavior of the alloys [16–18].

In terms of phenomenological models belonging to the first category, a simple empirical model often employed in the high-entropy alloy (HEA) domain is the Tabor relationship [19–21]. This model provides an estimate of the yield strength (YS). According to the Tabor relationship, the YS of a metal is approximately one third of its Vickers hardness (HV), defined as $9.81 \times HV/3 [MPa] \approx YS [MPa]$. Several authors have corroborated this relationship within body-centered cubic (BCC) refractory high-entropy alloys (RHEAs), thus it is often used as a cost-effective experimental surrogate in swift experimental endeavors. As a case in point, Ferreirós et al. [19], in the absence of YS measurements, relied on room temperature (RT) HV experiments as an upper limit for YS and the Maresca–Curtin model [22] as the lower limit. They deduced that the yield strength of the VNbCrMo alloy would not exceed 2.1 GPa at room temperature or fall below 0.85 GPa at 1000 °C. Kustas et al. [23] validated the utility of the Tabor relationship against the publicly accessible HEAs dataset from Gorsse et al. [24]. When uniaxial YS measurements (compression and tension) with HV measurements, they substantiated the validity of the Tabor relationship for HEAs, thus enabling them to avoid YS measurements and to use HV as an economical proxy for YS instead.

Moving on to the second category, we encounter models such as the one by Rao et al. [16], which serves as an extension of the Suzuki

solid solution strengthening (SSS) model to RHEAs [17]. Their atomistic simulations revealed that the characteristic preservation of YS at HT in RHEAs could be traced back to a significant energy barrier that hampers kink migration. This barrier originates from local chemistry variations along screw dislocation lines and from the dipole dragging stress connected with screw dislocation movement. However, the very need for these atomistic simulations restricts this model's applicability in the high-throughput (HTP) prediction design of RHEAs.

Third, we examine models that are analytical and provide information about SSS in RHEAs [18]. A notable example was proposed by Senkov et al. [25], who introduced a simple SSS model based on the disparities in atomic size and shear modulus among the constituent elements of the RHEA. Although this model is compatible with HTP and provides a reasonable initial approximation for the YS of RHEAs, it does not account for thermally activated mechanisms [25].

In light of the strengths and weaknesses of each of these models, our objective in this study is to amalgamate the precision and reliability of experimental data (akin to the first category) with the HTP attributes of analytical models (as in the third category).

In order to better verify the Tabor relationship for RHEAs, in this work we have taken the publicly available dataset of HEAs compiled by Borg et al. [26] (which itself is an expansion of the dataset of Gorsse et al. [24]) and compared the HV measurements to the YS measurements. Fig. 1b demonstrates the utility of the Tabor relationship in BCC RHEAs. RT HV measurements are successfully related to RT YS measurements via the Tabor relationship. The Pearson correlation coefficient between YS and HV is $\rho = 0.74$, indicating that there is a meaningful correlation between the two properties. Despite the simplicity and accuracy of the Tabor relationship, it is not appropriate for the HTP design of RHEAs because it is a data-hungry empirical model, i.e. many alloys must be synthesized, their homogeneity must be assured (if not, a homogenization heat treatment should be performed in a high vacuum without forming and volatilizing certain refractory oxides), and a hardness experiment must be performed in order to estimate the yield strength of an alloy.

Perhaps the most important model for HTP screening of RHEAs is the Maresca–Curtin model [22]. In their work, Maresca et al. postulate that, in RHEAs with high lattice strains (δ) and/or extremely high shear moduli (G), the dominant high temperature strengthening mechanism at high temperature is SSS caused by the sluggish movement of *edge dislocations* through a complex energy landscape within the RHEA matrix. This is in contrast to the typical theory of strengthening in BCC refractory alloys, which assumes screw dislocations are responsible for solid solution strengthening. This screw-controlled theory is appropriate when RHEAs have little lattice strain (δ) and/or moderate to low shear moduli (G). This theory is captured in models such as those proposed by Rao [16] and Maresca [27].

Despite this, Baruffi et al. [28] justify the use of the Maresca–Curtin model for alloy design by: (1) demonstrating that it acts as a lower bound for yield strength in RHEAs regardless of whether the deformation mechanism is screw or edge dislocation mediated; (2) and emphasizing that edge-controlled theory only relies on Vegard's law to predict critical quantities whereas the screw-controlled theory requires DFT-derived quantities such as solute misfit volumes and elastic constants. For these two reasons, the Maresca–Curtin model is widely used in the RHEA community to guide the design and development of RHEAs [20,29,30].

As the Maresca–Curtin model has been shown to be safe for design and is widely used by the HEA community, it is important to benchmark it against a chemically diverse data set. The largest publicly available data set of RHEAs to date is that compiled by Borg et al. [26]. This data set captures various mechanical properties associated with a chemically diverse set of HEAs. There are 468 single-phase BCC RHEAs present in the database. Of these single-phase BCC RHEAs, 365 have YS data available and 100 have HV data available. Approximately 63% of these BCC RHEAs with YS data have a lattice strain (δ) greater than

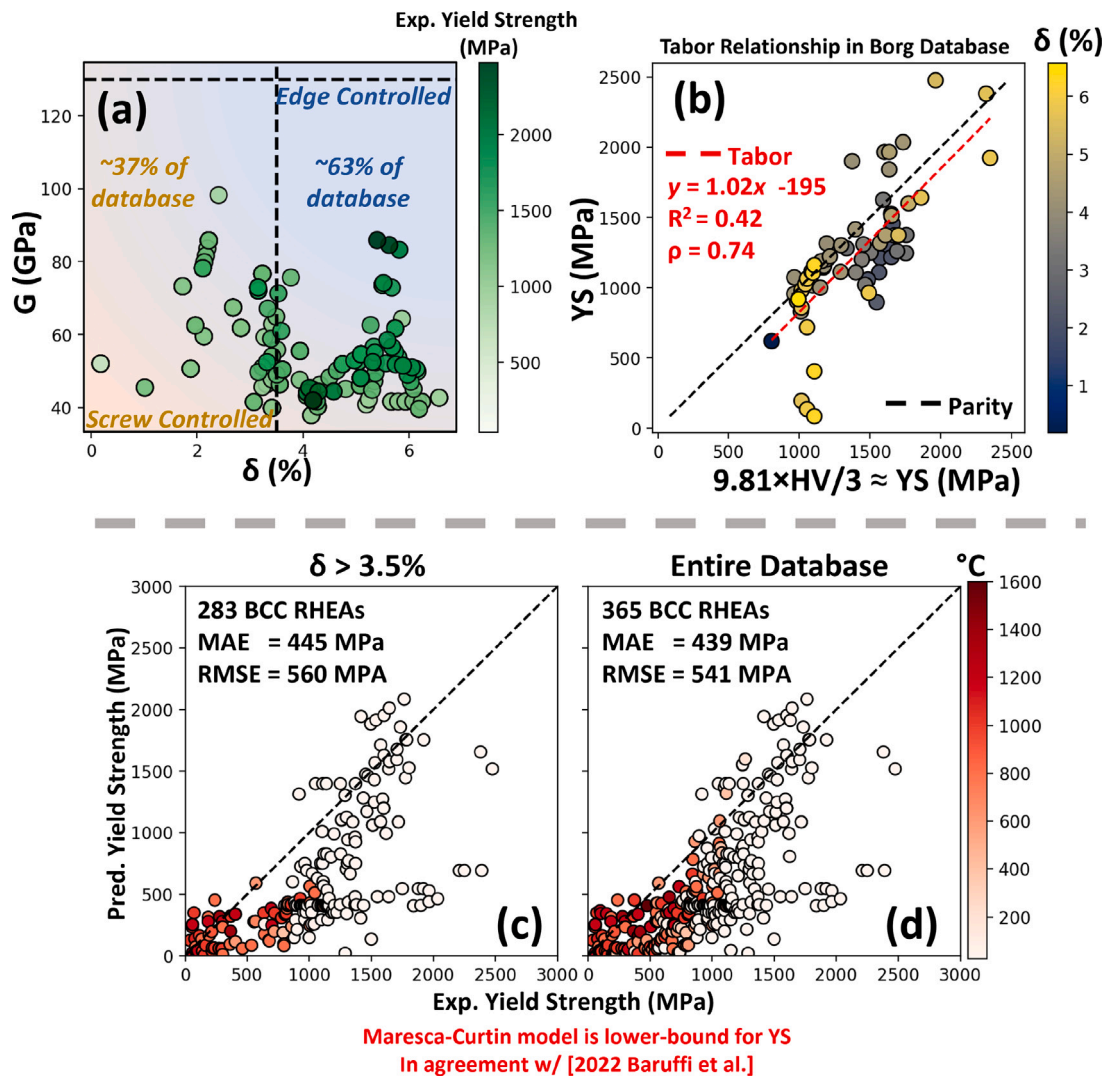


Fig. 1. (a) RHEAs with high shear moduli and/or high lattice mismatch are more likely to have yield mechanisms dictated by the glide of edge dislocations. RHEAs with standard shear moduli and/or standard lattice mismatch are more likely to have yield mechanisms dictated by screw dislocations. Approximately 63% of the Borg database is appropriate for the Maresca–Curtin model. (b) Tabor relationship as observed in the Borg dataset [26] (c) Parity plot for yield strength predictions from the Maresca–Curtin model when benchmarked against RHEAs in the Borg dataset that likely exhibit the same deformation mechanism as captured by the model. (d) Parity plot for yield strength predictions from the Maresca–Curtin model when benchmarked against all single-phase RHEAs in the Borg dataset without consideration for the likely deformation mechanism.

3.5%, indicating the YS of these alloys is edge-controlled and the Maresca–Curtin model is applicable [28], as shown in Fig. 1a.

Figs. 1c and d demonstrate that regardless of strengthening mechanisms (edge or screw-controlled) the Maresca–Curtin model underpredicts the YS of RHEAs in general. This is indeed in line with assertions from Baruffi et al. [28] that the Maresca–Curtin model acts as a lower bound for YS. When compared against the entire database, the mean absolute error (MAE) and the root-mean-square error (RMSE) of the Maresca–Curtin model are 469 MPa and 546 MPa, respectively. However, one cannot dismiss the Maresca–Curtin model as wholly inaccurate, as it evidently captures useful information about the composition and temperature dependence of the yield strength of RHEAs.

The Maresca–Curtin model is a HTP model that is safe to use in design, however, when benchmarked against a database of 365 RHEAs, its accuracy is low. The Tabor model is a simple and accurate empirical relationship that enables Vickers hardness to act as a proxy for yield strength, although, as mentioned above, it is not appropriate for HTP exploration of RHEA design spaces due to its reliance on synthesis, processing, and experiments. Given the Maresca–Curtin model, 100 experimental observations of hardness, and 365 observations of

yield strength, this paper seeks to address the following question: *How does one use all information at hand to better model the temperature-dependent yield strength of arbitrary RHEAs in a HTP manner?* We propose equipping Hierarchical Gaussian Process Regressors (HGPRs) [21] with informative priors in order to create a model that fuses information from multiple types of experimental data and analytical models.

In the context of HGPRs, Karumuri et al. [21] proposed a nested-GPR framework where a Gaussian Process regressor (GPR) is employed to derive feature values for a subsequent GPR. The authors exploited the Tabor relationship, using the Vickers hardness (HV) values – which were inferred from a GP – as a noisy feature in a subsequent GP for yield strength (YS). In doing so, they effectively utilized HV as a cost-effective surrogate experiment for YS experiments, circumventing the need for direct HV measurements on a specific alloy to estimate its YS. Rather, the optimal estimate of HV, based on the trained GP regressor, replaces actual experimental values. Nonetheless, this approach solely harnesses two types of experimental measurements (HV and YS) without incorporating insights from high-throughput (HTP) analytical models.

Regarding the use of HTP analytical models as informative priors during GP regression, to better predict the electric double layer

capacitance of carbon electrodes, Pan et al. [31] modified the mean function of the GPR, imbuing the model with physics learned from a semi-empirical model. The authors found that the physics-informed GPR performed better during cross-validation and was less likely to predict nonphysical electric-double layer capacitance values during extrapolation.

Liu et al. [32] recently implemented a novel approach to predict yield strength in fcc HEAs. They utilized experimental data to correct the Varvenne–Curtin model [33]—a model similar to the Maresca–Curtin model, specifically tailored for FCC HEAs. Specifically, the authors used Bayesian inference to update their belief about the temperature dependence of HEA elastic constants by calibrating model parameters in the Varshni model [34] for the temperature dependence of elastic constants. A quantitative relationship between the HEA composition and the elastic constants is predicted by an ensemble of support vector regression models. These elastic constants were then fed to the mechanistic Varvenne–Curtin model. Using Bayesian-inference to calibrate the parameters of Varvenne–Curtin was shown to improve the accuracy of the model. However, this framework is limited by the inflexibility of the mechanistic model. If the strengthening mechanism were to change in certain regions of the alloy space, the model would be unable to adapt even when provided experimental yield strength data. In the case of RHEAs, this modeling method would be inappropriate in design spaces where it is uncertain whether strengthening is attributed to mechanisms involving screw or edge dislocations. In this regard, Bayesian inference via Gaussian process regression equipped with physics-informed priors is a promising alternative to Bayesian inference via calibration of mechanistic models.

In this work, we propose combining data-hungry machine-learning models with fast-acting analytical models (recall the 3 aforementioned approaches to modeling yield strength) to develop a data-efficient model capable of rapid correction when new data are introduced. The proposed modeling approach is two-fold: (1) Using Hierarchical GP regression as proposed by [21] Karumari et al. to predict hardness values for arbitrary chemistries such that hardness can be used as an informative feature during GP regression of yield strength; (2) Using the Maresca–Curtin–Tabor model and the Maresca–Curtin model as physics-informed priors during Hierarchical GP regression of hardness and yield strength, respectively. In this way, we successfully fuse useful information from HTP analytical models, cheap hardness experiments, and ground-truth yield strength experiments.

We explore the utility of the proposed method with following three analyses:

1. We successfully demonstrate that equipping HGPR models with physics-informed priors increases accuracy and extrapolating ability during GPR regression. Specifically, a data-sparse 2-Fold cross-validation demonstrates that, despite data-sparse conditions, HGPR with informative priors have better extrapolating abilities along compositional and temperature axes.
2. Similarly, we demonstrate the improved extrapolating abilities of physics-informed HGPRs by hiding high-temperature YS data from standard GPRs and HGPRs with priors and demonstrate that the HGPRs with priors perform physics-informed extrapolation whereas the standard GPR will provide non-physical results i.e. will predict that yield strength does *not* monotonically decrease with temperature.
3. Furthermore, we demonstrate the utility of physics-informed HGPRs in Bayesian optimization (BO) schemes by performing a design exercise within the Nb–Ta–W alloy system. Specifically, we show that the BO schemes equipped with information from proxy HV experiments and informative priors converge on the maximum yield strength faster than BO schemes that utilize standard GPRs.

2. Methods

Gaussian process regression

A Gaussian Process Regression (GPR) is a powerful statistical tool in probabilistic modeling of objective functions in various fields due to cheap computational costs and flexibility in training, updating, and evaluation [35]. Given N observations, $\{\mathbf{X}_N, \mathbf{y}_N\}$, where $\mathbf{X}_N = (\mathbf{x}_1, \dots, \mathbf{x}_N)$ and $\mathbf{y}_N = (f(\mathbf{x}_1), \dots, f(\mathbf{x}_N))$, a GPR probabilistic modeling at an unobserved location $\mathbf{x}$ is determined by a normal distribution as:

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

where

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

with k as a real-valued kernel function, $K(\mathbf{X}_N, \mathbf{X}_N)$ as a $N \times N$ matrix with m, n entry as $k(\mathbf{x}_m, \mathbf{x}_n)$, and $K(\mathbf{X}_N, \mathbf{x})$ is a $N \times 1$ vector with m th entry as $k(\mathbf{x}_m, \mathbf{x})$. The term σ_n^2 accounts for stochasticity in observations or observation errors.

The squared exponential function is a popular choice as the kernel:

$$k(\mathbf{x}, \mathbf{x}') = \sigma_s^2 \exp \left(- \sum_{h=1}^d \frac{(\mathbf{x}_h - \mathbf{x}'_h)^2}{2l_h^2} \right) \quad (3)$$

where d is the dimensionality of the feature space, σ_s^2 is the signal variance, and l_h , where $h = 1, 2, \dots, d$, is the characteristic length-scale that determines the correlation strength between observations within dimension h . Due to the widespread use [36–38] and simplicity of the SE kernel, we will investigate the effect of proxy experiments and informative prior means using GPRs equipped with this very standard kernel.

Typically, a GPR mean function is initially set to zero, disregarding the opportunity to incorporate valuable prior knowledge into the model. In some cases, the mean function is shifted so a GPR defaults to the average of training data as the best prediction if no observation is available. One approach to incorporate prior knowledge in the presence of a prior model in constructing a GPR is to augment the prior model as the GPR mean function, which we call a GPR with *informative prior*. Then, it is trained with experimental observations to override the prior model predictions, but it defaults to the prior model predictions at regions with no observations. Mathematically, a GPR with an informative prior is built over the discrepancy between the prior model and the experimental observations. Note that if $f \sim \text{GP}(\mu, k)$, then $f' = f - \mu$ is a zero-mean GP. Thus, a zero-mean GP is constructed to represent the discrepancy between experimental observations and prior, f' , and prior, μ , is added to f' to obtain posterior on f . This approach guarantees that a GPR override prior if observations exist, else, exploits prior for probabilistic modeling.

In this work, we employ GPR with informative prior to creating a machine-learning model that combines information from the Maresca–Curtin model as the prior, μ , with limited experimental HV data and temperature-dependent YS data for RHEAs. Let the ground-truth (a function that captures the true experimental composition-temperature dependence of RHEAs) be denoted as $f_{\text{exp}}(\mathbf{x})$, where $\mathbf{x}$ is a vector that defines the RHEA and temperature of the tensile test. Let the GPR model built over the discrepancy between experimental data and prior, $f_{\text{exp}}(\mathbf{x}) - \mu(\mathbf{x})$, be denoted as f_{dis} . Then the GPR built with a combination of informative prior and experimental data is represented as $f = f_{\text{dis}} + \mu$.

The effect of these informative priors can be seen in the simplified toy problem depicted in Fig. 2. Here the ground-truth function is a sigmoid deactivation function, depicted in black. To a first approximation, this sigmoid deactivation function can be approximated by the pink line with the negative slope. When this pink line is used as the prior mean function during GPR, the model is better able to extrapolate than the uninformed model. Specifically, in regions of the domain where there are no observations of the ground-truth function, both models

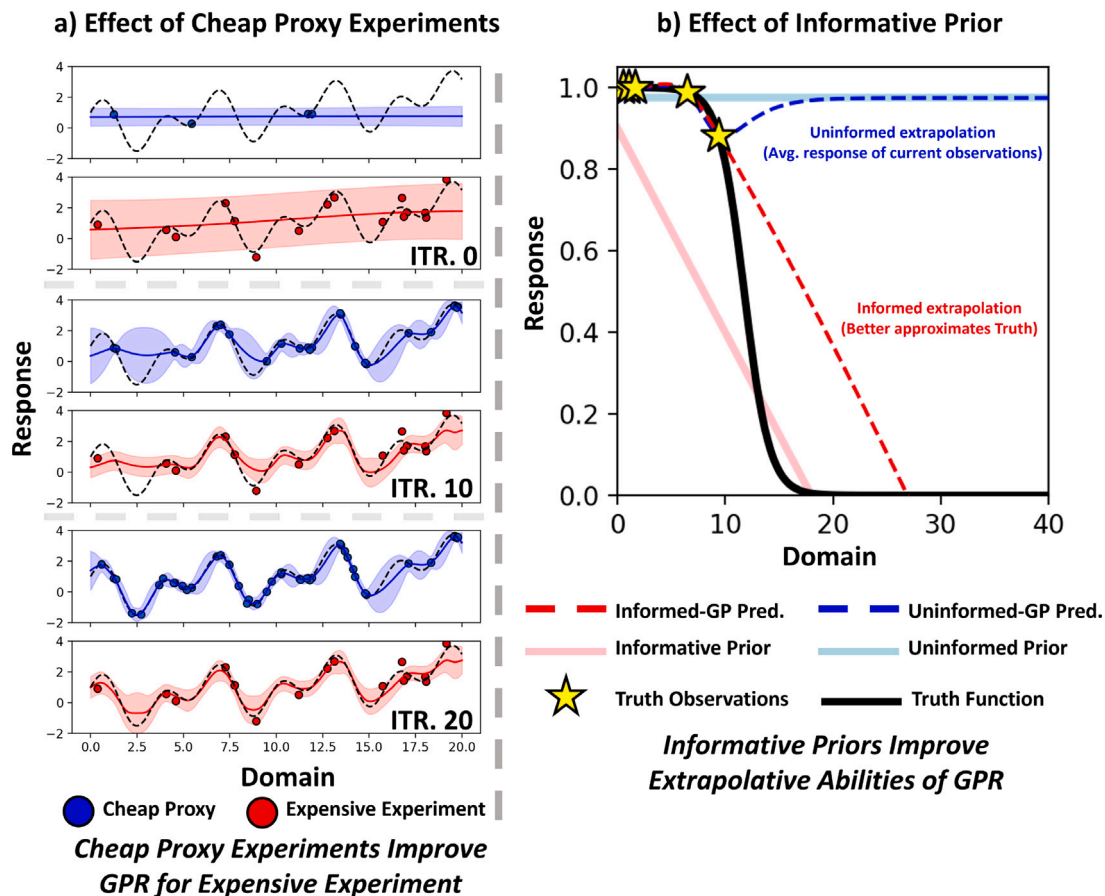


Fig. 2. Toy problems demonstrating the two proposed methods in 1D. (a) Demonstration of Hierarchical GP regression and the effect of cheap proxy experiments. The Gaussian process for the cheap proxy experiment (depicted as the blue line) improves as more proxy experiments are performed (depicted as the blue circles). This model for the proxy experiment is used to estimate the values of the proxy experiment at locations in the domain where there are observations of the expensive experiment (depicted as red circles). Readers are encouraged to view the animated version of this figure in our repository: <https://codeocean.com/capsule/7849853/tree/v2>. (b) Demonstration of informative and uninformative priors. The light-blue and pink lines are the uninformative and informative priors, respectively. When predicting far from observations of the ground truth, both models will default to their prior mean function. When using an informative prior, defaulting to the mean function will yield better extrapolating abilities than in the case of the uninformed prior.

will default to their prior mean function. In the case of the uninformed GP, the model will default to the average value of all ground-truth observations (average of the yellow stars). However, this prediction deviates significantly from the ground-truth model. In the case of the informed GPR, the model will default to the informative prior. In this case, all observations of the truth function shown to the model are underpredicted by the prior (pink line). The model learns that the prior systematically underpredicts the ground truth and thus defaults to a version of the prior mean that is shifted up. While this model also deviates from the ground-truth, the informed GPR is better able to approximate the ground-truth function even when data is sparse.

Furthermore, Karumuri et al. [21] showed that GPRs can be used in series in a technique known as Hierarchical GP regression in order to leverage cheap proxy experiments as informative features when performing regression for more expensive experiments. Specifically, a GPR can be built to estimate values for the cheap proxy experiment across a domain. Once the proxy experiment has been estimated across the domain, it can be used as an informative feature when performing GPR on a more expensive target. In their work, Karumuri et al. quantify the uncertainty of this nested model by deriving semi-analytical approximations of the marginal likelihood and the posterior predictive distribution. In this work, we only impute hardness values and use them as an informative feature in a yield strength model and do not account for uncertainty in the hardness inputted to the yield strength model, as described by the workflow shown in Section 2.4.

Consider the synthetic problem depicted in Fig. 2a that demonstrates this procedure. There are 15 observations of the expensive ground-truth function (red points), i.e. the function to be modeled. No more queries to the ground truth can be made, however, there is an unlimited budget of queries that can be made on the cheap proxy experiment (blue points). The proxy and the ground truth share a linear correlation with Gaussian noise. When there are limited queries of the proxy experiment, the GPR for the truth (red line) is not able to approximate the ground-truth function well, as shown in Fig. 2a in iteration 0. As more queries are made to the cheap proxy, the GPR for the proxy (blue line) improves and this, in turn, improves the fidelity of the GPR for the truth. By the 20th iteration of proxy experiments, the GPR for the truth has improved significantly. The code for this toy problem, implementation of these methods, and animated versions of 2a can be found in the following Code Ocean repository: <https://codeocean.com/capsule/7849853/tree/v2>.

The two aforementioned toy problems demonstrate the proposed methods in a simplified manner. In this work, we seek to apply these two methods to the Borg dataset and a toy optimization problem in order to improve both GPR and BO of YS. Specifically, (1) a GPR will be built for HV and will be used to predict the HV of alloys that have YS experiments available. These HV predictions will be used as a feature in the GPR model for YS. (2) The GPR for both the HV and YS models will be improved by leveraging informative priors that will help the GPRs during extrapolation. The prior for the yield strength model is the Maresca–Curtin model. The prior for hardness

Table 1
Explanation of elemental properties used during featurization of RHEAs.

R	Metallic radius
E	Young's modulus
T_m	Solidus
VEC	Valence Electron Concentration
B	Bulk modulus
G	Shear modulus
ν	Poisson's Ratio
χ_{Allen}	Allen Electronegativity
$S_{conf.}$	Configurational Entropy
ρ	Density
C_{11} or C_{12} or C_{44}	Elastic constants

GP is the Maresca–Curtin–Tabor model, i.e. yield strength predictions from the Maresca–Curtin are converted to hardness values via the Tabor relationship.

2.1. Implementation of Maresca–Curtin model and Maresca–Curtin–Tabor model

The prior belief of the temperature-dependent yield strength of RHEAs in this work is based on the analytical model presented by Maresca and Curtin in Ref. [22]. The model captures the behavior of an edge dislocation within a randomly distributed solute field found in a BCC HEA. The authors conducted a statistical analysis of the energy required for thermally activated edge glide and derived the following equations as a result:

$$\tau_{y0} = 0.040\alpha^{-1/3}\bar{\mu}\left(\frac{1+\bar{\nu}}{1-\bar{\nu}}\right)^{4/3}\left[\frac{\sum_n c_n \Delta V_n^2}{b^6}\right]^{2/3} \quad (4)$$

$$\Delta E_b = 2.00\alpha^{1/3}b^{-3}\bar{\mu}\left(\frac{1+\bar{\nu}}{1-\bar{\nu}}\right)^{2/3}\left[\frac{\sum_n c_n \Delta V_n^2}{b^6}\right]^{1/3} \quad (5)$$

$$\tau_y(T, \dot{\epsilon}) = \tau_{y0}\left[-\frac{1}{0.55}\left(\frac{k_B T}{\Delta E_b}\ln\frac{\dot{\epsilon}_0}{\dot{\epsilon}}\right)^{0.91}\right] \quad (6)$$

$$\sigma_y(T, \dot{\epsilon}) = M\tau_y \quad (7)$$

In these equations, the variables are defined as follows: the line tension parameter, α , which has been fixed at 1/12 for edge dislocations [22]; $\bar{\mu}$ represents the average shear modulus of the alloy, while $\bar{\nu}$ is its average Poisson ratio. The Burger vector, b , is related to the BCC edge dislocation in the random alloy, and the misfit volume of the n th solute, ΔV_n , can be approximated using the rule-of-mixtures [22]. Another important variable is the zero-temperature yield stress, τ_{y0} , and the energy barrier for thermally-activated flow, ΔE_b . The reference strain rate, $\dot{\epsilon}_0$, is generally set to 10^4 s^{-1} , while the applied strain rate, $\dot{\epsilon}$, is typically 10^3 s^{-1} and has been set to this value in the current work. The equation also includes the Taylor factor for edge glide in a random BCC polycrystal, M , as well as the Boltzmann constant, k_B , and the yield strength, $\sigma_y(T, \dot{\epsilon})$, which is calculated at a finite temperature, T , and strain rate, $\dot{\epsilon}$. In order to estimate HV, the Tabor relationship can be applied to the results from the Maresca–Curtin model. Specifically, the HV can be estimated by $HV = YS \times (3/9.81)$

2.2. Feature selection

Following the feature selection methodology by Khan et al. [39], we first query a suite of features that are believed to be correlated to solid solution strengthening. The suite of features relevant to SSS was taken from the work of Wen et al. [40]. A truncated list of these features can be seen in Fig. 3a. These feature capture information directly relevant to SSS. For example, features such as the atomic radius R_m capture information about size factors. Features such as bulk modulus (B) and shear modulus (G) capture bond/modulus information. Features such

as VEC and χ capture electronic behavior. While there may exist more relevant features, these features have been used in similar works where the yield strength of HEAs is modeled with machine-learning [40–42].

Next, the mutual information [43] between each feature and the target (HV and temperature-dependent YS) are evaluated. Mutual information describes the amount of information gained about one variable from knowing the value of another variable [43]. In the context of feature selection, features with high mutual information have a strong predictive ability for the target variable, whereas features with low mutual information have limited predictive ability for the target variable. The relevant features are then rank-ordered by their mutual information with the target, and the top n are selected, where n is any integer value. In this work $n = 8$, however, the optimal n will vary from case to case.

To eliminate overtly colinear features, any features with a Pearson's correlation coefficient above $\rho > .85$ is discarded and replaced with the feature with the next highest mutual information. This threshold of 0.85 is arbitrary and can be modified depending on model performance. Finally, features related to processing and test conditions are included regardless of their mutual information score. This information is important to include, as processing and test conditions can have a significant impact on the observed outcomes. Furthermore, if predictions are to be used to make decisions concerning which alloy merits scale-up synthesis and processing for tensile testing, this information is relevant to those predictions.

It should be noted that the features selected in this work may not be the most optimal to model solid solution strengthening in RHEAs, however, the goal of this work is to probe the effect of equipping HGPRs with physics-informed prior mean functions in order to fuse information from HV experiments, YS experiments, and analytical models.

2.3. Calculation of selected features

The optimal features for the HV and YS models fall into three categories: compositionally-weighted averages, compositionally-weighted average property differences (mismatches), and maximum compositionally-weighted average property differences (min–max mismatches). Averages are estimated by the rule of mixtures, as shown in Eq. (8) where P_i represents the property of interest for the i -th element and x_i represents the mole fraction of the i -th element in the RHEA. The dummy variable P would be interchanged with R, B, G, ν , χ , C_{11} , C_{12} , C_{44} , and VEC. Elemental values for any property P are then passed to Eqs. (8)–(10). This set of elemental properties P are described in Table 1. Following Eq. (8), averages are denoted by a bar superscript. Given the averages, the mismatches can be calculated according to Eq. (9). The mismatches capture the deviation of a property with respect to the ROM average of the property of the elemental constituents. Mismatches are denoted by the δ prefix. Likewise, the min–max-mismatch can be calculated according to Eq. (10). Following Eq. (10), the min–max-mismatch captures the largest property mismatch that can occur within the RHEA. Furthermore, additional features were considered such as the configurational entropy ($S_{conf.}$), the Zener Ratio ($2\overline{C_{44}}/(\overline{C_{44}} - \overline{C_{44}})$), and ductility indicators such as the Pugh Ratio ($\overline{B}/\overline{G}$) and Cauchy Pressure ($\overline{C_{12}} - \overline{C_{44}}$). While other features can be considered, the objective of this work is not to necessarily determine the best set of features possible to model SSS in RHEA, but rather to demonstrate the advantages of equipping HGPRs with physics-informed prior mean functions in order to fuse information from HV experiments, YS experiments, and analytical models.

$$\overline{P} = \sum_{i=1}^N P_i x_i \quad (8)$$

$$\delta P = \sqrt{\sum_{i=1}^N x_i \left(1 - \frac{P_i}{\overline{P}}\right)^2} \quad (9)$$

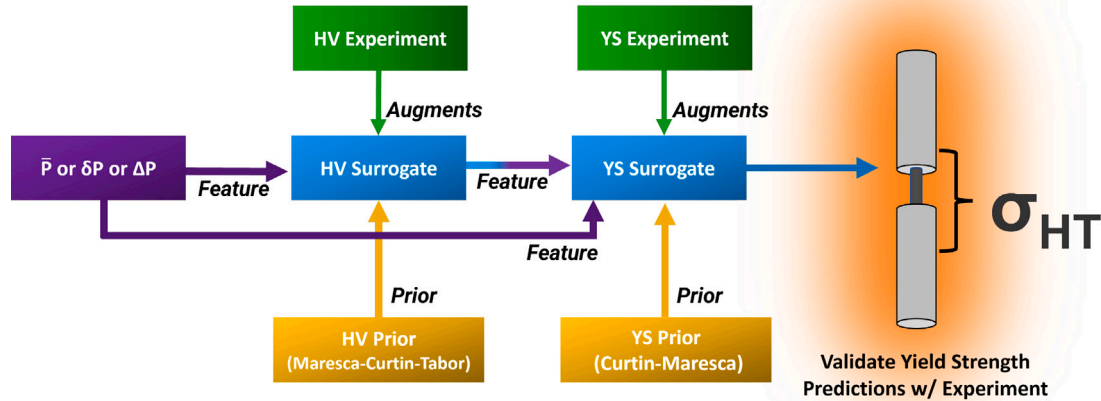
$$\Delta P = \max \left[x_i \left(1 - \frac{P_i}{\overline{P}}\right)^2 \right] - \min \left[x_i \left(1 - \frac{P_i}{\overline{P}}\right)^2 \right] \quad (10)$$

a) Feature Selection Protocol

Hardness Feature Selection		Yield Strength Feature Selection	
Feature	Mutual Info.	Feature	Mutual Info.
δR	0.86	HV	0.74
$\bar{R}$	0.73	δB	0.40
$\bar{E}$	0.56	$\bar{B}$	0.37
$\bar{T}_m$	0.54	δR	0.32
$\overline{VEC}$	0.53	$\bar{\rho}$	0.25
δB	0.53	ΔB	0.22
ΔG	0.49	R	0.21
$\delta \chi_{\text{Allen}}$	0.49	Test Temperature	0.19 *
ΔR	0.45	ΔR	0.16
$\bar{B}$	0.43	$\bar{T}_m$	0.16
S_{conf}	0.41	$\overline{C_{12}-C_{44}}$	0.14
$\bar{\rho}$	0.40	$\bar{G}$	0.14
ΔB	0.40	$\Delta \nu$	0.14
$\delta \nu$	0.26	S_{conf}	0.11
$2\overline{C_{44}} \div (\overline{C_{11}} - \overline{C_{12}})$	0.24	$2\overline{C_{44}} \div (\overline{C_{11}} - \overline{C_{12}})$	0.07 *
$\overline{C_{12}-C_{44}}$	0.24	Processing method	0.07 *
* Processing Method	0.00	$\delta \chi_{\text{Allen}}$	0.07
		Tensile/Compressive	0.06 *
		$\bar{B}/\bar{G}$	0.02

* Features related to processing and test conditions

The **information source** augments the **prior model**. The output of the **posterior model** (surrogate) can then be used as a **physics-based feature** for subsequent models.



b) Machine-Learning Workflow

Fig. 3. (a) Feature selection methodology for both GPs in the hierarchical GP regression framework. (b) Features relevant to SSS solution strengthening are calculated and used as inputs to a GP regressor for HV. This HV GP regressor is equipped with the Maresca-Curtin-Tabor model as its prior mean function. The HV GP regressor is then used to impute HV values for arbitrary compositions for which there is YS data available. This circumvents the need to perform a HV experiment on an alloy that has YS data available but no HV data available. In this way, HV can be used as an informative feature for YS GP regressors. These imputed HV values and SSS-relevant features are then used as inputs to the YS GP regressor. The prior mean function of the YS GP regressor is the Maresca-Curtin model. The data posterior model can be considered a version of the Maresca-Curtin model augmented by experimental data. These YS predictions are then validated against experiments using 3-Fold cross-validation.

2.4. Machine-learning workflow

With these features calculated, they are used in the following machine learning framework. In the first step of the framework, features related to solid solution strengthening (SSS) are identified, scaled using a standard scaler [44], and utilized as inputs to a GP regressor for hardness (HV). This HV GP regressor incorporates the Maresca-Curtin-Tabor model as its prior mean function. By imputing HV values for

compositions that have yield strength (YS) data but lack HV data, the need for additional HV experiments is bypassed. This allows HV to serve as an informative feature for the YS GP regressors. The imputed HV values, along with the SSS-relevant features, are subsequently employed as inputs to the YS GP regressor, which employs the Maresca-Curtin model as its prior mean function. The resulting data posterior model can be considered an augmented version of the Maresca-Curtin

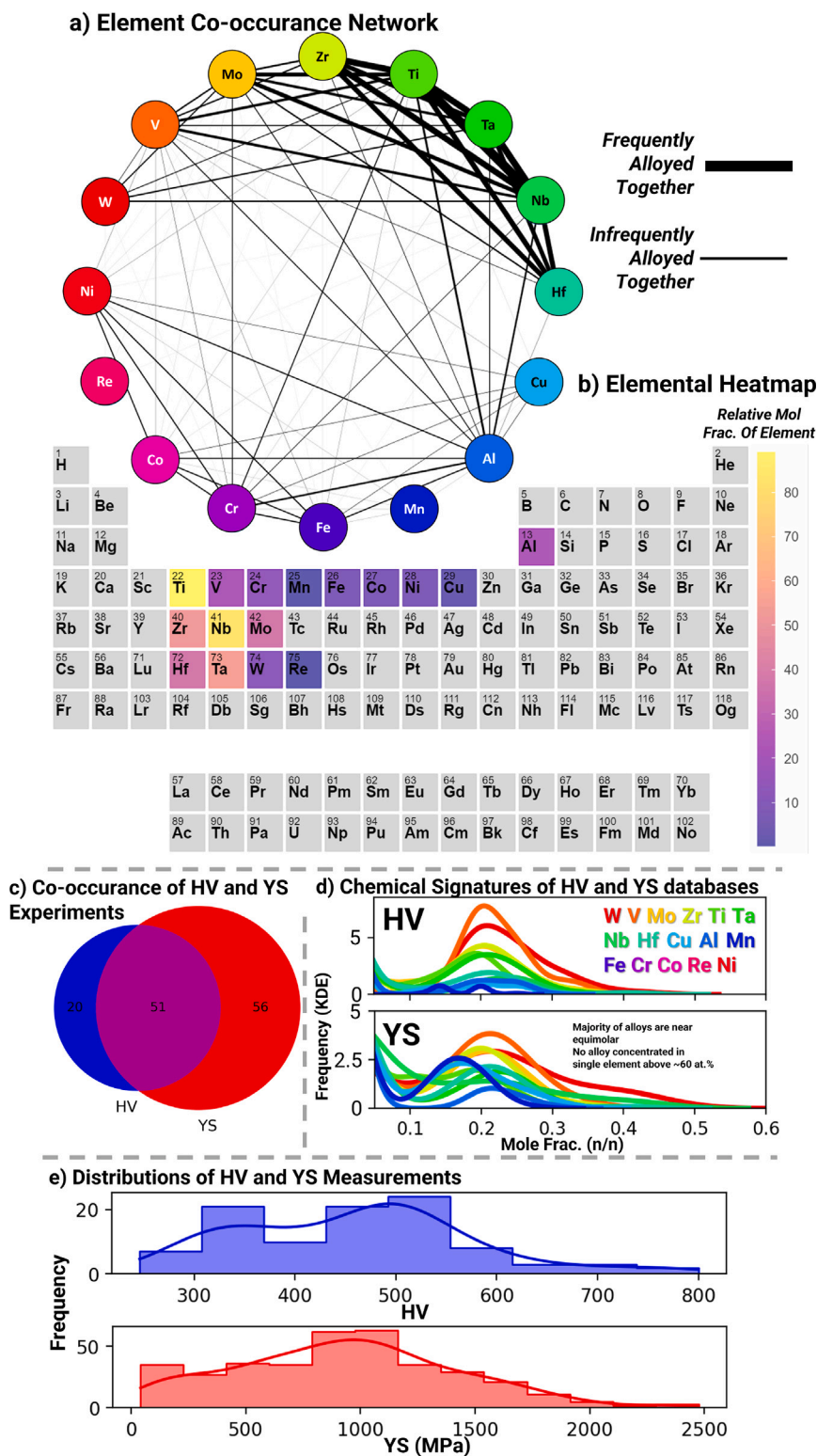


Fig. 4. Description of datasets used in this work. (a) Element co-occurrence network demonstrating the frequency in which certain elements are alloyed with each other. It is evident from this graph that the refractory elements (Mo,Zr,Ta,Nb,Hf) are most commonly alloyed together. (b) Elemental prevalence plotted as a heatmap on the periodic table. (c) Venn diagram demonstrating the occurrence and co-occurrence of HV and YS experiments in the database. There are 107 compositions with yield strength experiments and 364 temperature-dependent yield strength experiments. Regarding hardness, there are 71 compositions with HV experiments and 99 hardness experiments in total. There are 51 compositions that have data available for HV and YS. (d) Chemical signatures of the HV and YS datasets. The majority of elements appear in alloys at concentrations near 20 at.%. There are no alloys concentrated in a single element beyond 60 at.%. This indicates that the majority of the database are truly high entropy alloys. (e) The distribution of HV and YS measurements.

model, incorporating experimental data. The accuracy of the YS predictions is then assessed through 2-Fold cross-validation, comparing the predicted values against experimental measurements. The code associated with this workflow can be found at the following Code Ocean repository <https://codeocean.com/capsule/7849853/tree/v2>.

2.5. Model evaluation and error metrics

In order to benchmark the predictive ability of the models developed in this work, four error metrics are used: mean absolute error (MAE), root mean squared absolute error (RMSE), Spearman rank coefficient (R_s), and the Kendall rank coefficient (τ). The MAE captures the average absolute error between predicted and actual values, serving as a metric to assess the average prediction error of a model. RMSE, on the other hand, captures the square root of the average of squared errors between predicted and actual values. The RMSE is more sensitive to larger discrepancies than the MAE.

While MAE and RMSE are important error metrics that attempt to quantify the discrepancy a user should expect from a model, these metrics can be misleading when used alone [45]. For example, these metrics do not provide any indication that a model is consistent with physics-dictated trends. There may be cases, for example, when MAE and RMSE are extremely low, but the rank ordering is poor, and that is the result of a nonphysical (oftentimes but not necessarily overfit) model that happens to minimize MAE and RSME (as in the case of Ref. [46]). In the context of alloy design, the optimal solution is sought. Finding the optimal solution requires that the model is able to correctly rank order candidate designs. For this reason, in addition to the MAE and RMSE, we use the two common ranking ordering metrics (R_s and τ) to evaluate the models in this work.

R_s is commonly used to evaluate the extent to which the association between two variables can be characterized by a monotonic function [47]. In this work, it is used to probe the physicality of yield strength predictions in addition to each model's ability to find optimal solutions. R_s captures the monotonicity of the relation between two variables i.e. how well the relationship between two variables could be represented using a monotonic function. $R_s = 1$ or -1 indicates perfect rank ordering and perfect monotonicity. $R_s = 0$, on the other hand, indicates no rank ordering ability and no monotonicity. Mathematically, R_s can be calculated according to Eq. (13), where d_i is the difference between the two ranks of each observation, and n is the total number of observations.

Similarly, τ captures the probabilities of observing concordant and non-discordant pairs given a dataset [48]. Considering that we are modeling an arbitrary property y for two alloys labeled alloy i and alloy j . The ground truth for this property in each of these alloys is y_i and y_j while our predictions of y in these alloys is x_i and x_j . If $y_i > y_j$ and $x_i > x_j$, then the pairs (x_i, x_j) and (y_i, y_j) are concordant. If this is not the case, the pair is discordant. This is expressed in Eq. (14).

A successful model will have an acceptable MAE and RSME and competitive values for R_s and τ . With this suite of error metrics, the magnitude of the error is considered, but the rank ordering ability of the model is also taken into account. In accordance with best practices [49], in order to mitigate any bias associated with how the database is split during cross-validation or any bias associated with the random initialization of the GPRs, a distribution of these error metrics are reported as opposed to a single error metric per model. In this way, the model performance can be evaluated and compared in a more generalized sense.

$$\text{MAE} = \sum_{i=1}^n |x_i - y_i| \quad (11)$$

$$\text{RMSE} = \sqrt{\frac{\sum_{i=1}^n (x_i - y_i)^2}{n}} \quad (12)$$

$$\rho = 1 - \frac{6 \sum_{i=1}^n d_i^2}{n(n^2 - 1)} \quad (13)$$

$$\tau = \frac{2}{n(n-1)} \sum_{i < j} \text{sng}(x_i - x_j) \text{sng}(y_i - y_j) \quad (14)$$

2.6. Description of dataset

The dataset used to train the HGPRs in this work is a curated version of the Borg dataset. Specifically, the dataset has been filtered to only consider single-phase BCC RHEAs. This dataset is provided in the following Code Ocean repository associated with this work: <https://codeocean.com/capsule/7849853/tree/v2>. This filtered dataset is described as follows. Fig. 4a presents an element co-occurrence network, illustrating the frequency of alloying specific elements together. It is clearly observed from this graph that the refractory elements, namely Mo, Zr, Ta, Nb, and Hf, are frequently combined in alloys. Fig. 4b demonstrates the element prevalence in the dataset. Specifically, Ti, Nb, and Ta are the most represented elements in the dataset. In general, refractory alloys are more represented in the dataset than 3d transition metals. In Fig. 4c, a Venn diagram displays the occurrence and co-occurrence of HV and YS experiments on unique chemistries within the database. The dataset contains 107 unique compositions with yield strength experiments, 71 compositions with HV experiments, and a total of 51 compositions that have data available for both HV and YS measurements. Fig. 4d showcases the chemical signatures of the HV and YS datasets, revealing that a significant number of elements appear in alloys with concentrations around 20 at.%, while no alloys exhibit a single element concentration exceeding 60 at.%. This observation indicates that the majority of the database consists of high entropy alloys as opposed to chemically complex alloys. Finally, Fig. 4e presents the distribution of HV and YS measurements, providing insights into the spread of values for these properties. Such visualizations are in line with the best practices for machine learning in materials science proposed by Wang and Sparks et al. [49].

Regarding the limitations of this dataset, it should be noted that interstitial oxygen content of the RHEAs captured in the Borg dataset is rarely reported. Accounting for interstitial oxygen content is important as RHEAs suffer from extreme oxygen sensitivity [3]. Interstitial content has been shown to increase the strength and decrease ductility of RHEAs [50]. Furthermore, while processing information is reported categorically (as-cast, annealed, wrought, powder-metallurgy, etc.), information regarding grain size and detailed descriptions of the resultant microstructures are not reported. This is important as larger grain sizes [51,52] dendritic microstructures have been shown to affect the strength of RHEAs. These factors potentially account for the observed discrepancy between experimental YS values and YS values predicted using the Maresca–Curtin model.

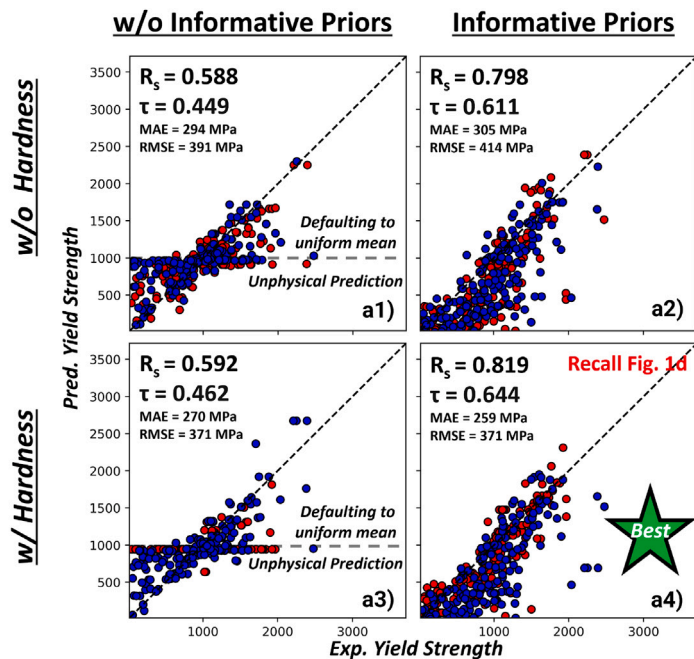
3. Results

3.1. Temperature-series aware 2-fold cross validation

In order to benchmark the efficacy of (1) using hardness estimates from GPRs as an informative feature in yield strength models and (2) using physics-informed priors during GPR, four models were cross-validated. The four models are comprised of every combination of the following conditions: uses hardness as an informative feature or not, and uses informative priors during GPR or not. The resultant models are shown in Fig. 5. Specifically, the double-control model in Fig. 5a1 is a standard Gaussian process equipped with a SE kernel. The model shown in Fig. 5a3 is a standard Gaussian process equipped with a SE but uses hardness as a feature as predicted by a GPR trained on hardness data captured in the Borg dataset. The model shown in Fig. 5a2 is a GPR equipped a SE kernel and uses the Maresca–Curtin model as the prior mean function. The model shown in Fig. 5a4 is the double-test model i.e. a GPR that uses the Maresca–Curtin model as its informative prior and is also informed by hardness predictions that themselves come from a GPR that uses the Maresca–Curtin–Tabor model as an informative prior for hardness.

Regarding cross-validation of these four models, the yield strength data-set contains 39 temperature series, meaning 39 unique alloy

a) Parity Plots for Single Instance of 2-Fold Cross-Validation



b) Rank-Ordering Error Metric Distributions for 200 Instances of 2-Fold Cross-Validation

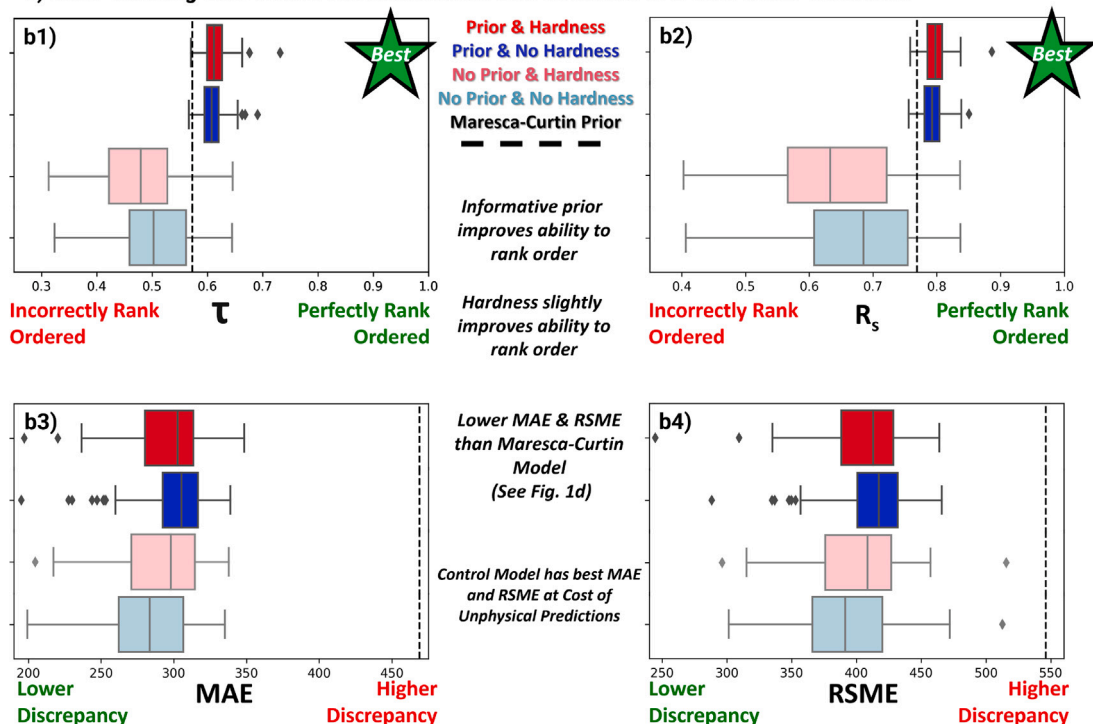


Fig. 5. 3-Fold Cross validation results for various yield strength models tested in this work. (a1) 3-Fold cross-validation results for the double-control model i.e. the standard Gaussian process regression for yield strength. (a2) 3-Fold cross-validation results for a Gaussian process that uses the Maresca-Curtin model as a prior during GP regression. (a3) 3-Fold cross-validation results for a Gaussian process that uses GP-imputed hardness as an informative feature for yield strength predictions. (a4) 3-Fold cross-validation results for the double test model i.e. a yield strength GPR equipped with the Maresca-Curtin model as its prior and uses hardness as predicted by a GPR equipped with the Maresca-Curtin-Tabor model as its prior. (b) Distributions of the 4 error metrics for the 4 models benchmarked in this work.

chemistries have been tested at various temperatures. During k-Fold cross-validation, these temperature series may be split across training and testing sets. When the model has seen some temperature-dependent data for an alloy, it will interpolate between temperature-dependent points, leading to improved accuracy when predicting the test set. While interpolation is a cornerstone in machine learning, it is important

to understand the limitations of the proposed model when it is expected to predict the yield strength of alloys for which no temperature-dependent data is available. To that end, we performed a modified k-Folds cross-validation where temperature series are not allowed to be split across training and testing sets. Unique compositions were selected with k-fold cross-validation, then the temperature-dependent

Table 2

Error metrics of the 4 benchmarked models.

Model	R_s	τ	MAE	RSME
Prior & Hardness	0.798 ± 0.017	0.614 ± 0.021	296 ± 28	408 ± 36
Prior & No Hardness	0.793 ± 0.016	0.609 ± 0.019	302 ± 21	415 ± 26
No Prior & Hardness	0.635 ± 0.105	0.476 ± 0.074	292 ± 28	400 ± 34
No Prior & No Hardness	0.673 ± 0.098	0.503 ± 0.069	283 ± 27	392 ± 34

data was added back to the testing and training sets. This prevention of data leakage is in line with the best practices for machine learning in materials science proposed by Wang and Sparks et al. [49].

In order to rigorously benchmark the models under data-sparse conditions, 2-Fold cross-validation was performed, i.e. 1/2 of the database is allocated as training data, and the other 1/2 is allocated as test data. Such data-sparse conditions are common when exploring the RHEA property space, thus it is important to benchmark models under such conditions. Furthermore, this ensures that the models are benchmarked against a sufficiently large test set (compared to the training set). As mentioned in Section 2.5, in order to mitigate bias associated with how the database is split during cross-validation or bias associated with stochasticity of the GPRs, 200 different instances of 2-Fold cross-validation were performed. Each of the 200 instances of 2-Fold cross-validation used a database that was partitioned in a different way. This way we ensure that the effects of random initialization of the model are mitigated, and we are able to determine how each model performs in general.

Fig. 5 demonstrates cross-validation results for the four models. As mentioned in Section 2.5, the MAE and RMSE are important error metrics to consider, however, when considered alone, these metrics can be misleading. One must consider the physicality of predictions as well as MAE and RMSE. For example, in the parity plots shown in Fig. 5a1 and a3, the models without informative priors provide a uniform prediction of yield strength for many alloys. This uniform prediction is the mean of all yield strength observations within the training set, represented by the horizontal dashed line in Figs. 5a1 and a3. This happens because the model has not been trained in that region of the input space, and thus the model has defaulted to its prior mean function. Similar visualization of this can be seen in Figs. 2b and 6. In the case of alloy design and optimization, the physicality and rank-ordering ability of models is more important than simply minimizing MAE and RMSE. A successful model will have an acceptable MAE and RMSE and have optimum rank-ordering ability as quantified by R_s and τ .

With this in mind, 5b demonstrates that alloys with informative priors (dark red and dark blue) are better able to rank-order alloys according to yield strength than GPRs without informative priors (light blue and light red). The average error metrics for the models that use a prior and hardness, a prior and no hardness, no prior and hardness, and no prior nor hardness are reported in Table 2. The distribution of MAE and RSME for all 4 of these models are similar however the models with informative priors outperform the models without informative priors with regards to the rank ordering coefficients R_s and τ .

3.2. Physics-informed extrapolation

Conventional GPRs often violate physical constraints as these models are not physics-informed [31] and instead solely rely on available data. In a situation where data is not available, GPRs predictions will return to the prior mean function. Specifying an informative prior mean function during GPR can help ensure physical predictions during extrapolation, as shown in Fig. 6 where the yield strength of alloys in the Borg dataset are plotted against their respective testing temperatures. The yellow stars represent yield strength measurements on the MoNbTaVW Senkov alloy. The circular points represent other alloys in the data set. The color axis represents the cosine similarity

between these alloys and the Senkov alloy i.e. alloys that are similar to the MoNbTaVW alloy are depicted as blue whereas alloys that are dissimilar to the MoNbTaVW alloy are depicted in green. In order to benchmark the extrapolating ability of informed and uninformed GPRs, yield strength observations above 600 °C, 1000 °C, and 1400 °C were hidden from a conventional GPR and GPR using the Maresca–Curtin model as its prior. When high-temperature data is hidden from any GPR, the model will default to the prior mean function for regions of the feature space where it has not been trained. What is important is whether or not the model will default to a reasonable prior or not.

In the case of Fig. 6a, the prior mean function is the average of all yield strength observations shown to the model, depicted as the orange/blue/green horizontal lines. As the model extrapolates further from the data, the model's prediction coincides with the uninformative prior. This results in the model predicting that yield strength does not monotonically decrease with temperature, which is a nonphysical prediction. Furthermore, it is clear that the extrapolation is highly dependent on the training data as the HT predictions from the 3 models (orange, green, and blue) do not coincide. In the case of Fig. 6b, in the absence of data, the model will default to the Maresca–Curtin prior, which predicts that YS monotonically decreases with temperature. Such a benchmark is relevant as high-temperature yield strength data is exceedingly difficult to obtain experimentally. *These results indicate that GPRs with physics-informed priors are safer to use during extrapolation than uniform priors.*

3.3. Significance to Bayesian optimization

This framework holds significance in the context of recent initiatives focused on Integrated Computational Materials Engineering (ICME)-enabled closed-loop design platforms [53–55] and autonomous materials discovery [56–58]. Bayesian optimization (BO) is a commonly used optimization scheme in the context of autonomous science and self-driven laboratories [59]. BO relies on building surrogate models to approximate the underlying material property and guide the search for compositions with optimal properties. The most common choice of surrogate model for BO is the GP regressor. However, the accuracy of these GP regressors depends highly on the availability of high-quality experimental data. In many cases, high-quality experimental data may be limited or expensive to obtain, while low-quality data from cheaper proxy experiments or HTP analytical models may be readily available but less reliable.

For instance, in the case of optimizing yield strength in RHEAs, often times there are budget constraints on the number of alloys that can be synthesized and the number of high-temperature tensile tests that can be performed. This framework can be applied so that (1) HTP analytical models can be used as priors during BO and (2) cost-effective hardness experiments can be conducted in order to investigate promising compositions without committing to high-temperature yield strength experiments.

Consider the toy optimization problem present in Fig. 7. Yield strength is to be optimized within the Nb–Ta–W alloy space. For this toy problem, the Ground-ground-truth yield strength model will be the Maresca–Curtin model, i.e. the goal of the BO scheme is to find where the Maresca–Curtin model is maximum in the chemistry space. BO schemes with two models are compared: a standard GPR and a GPR equipped with an informative prior and aided by HV proxy experiments. The Ground-ground-truth for the Proxy HV experiment is the Maresca–Curtin–Tabor model and the prior for the proxy GPR is ROM HV. The acquisition function for both schemes is the Expected Improvement [60]. In the case of the HGPR, the proxy next-best HV experiment are selected according to the expected improvement of the yield strength values. This ensures that the optimizer allocates hardness experiments to search for alloys with the highest yield strength and not the highest hardness, despite those two properties being correlated.

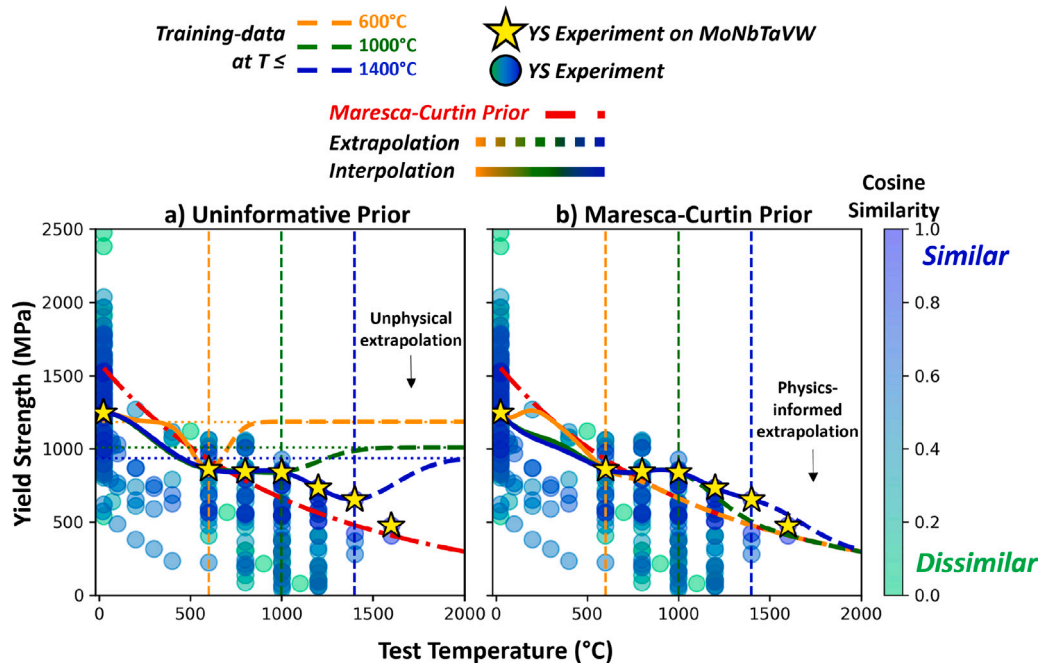


Fig. 6. The effect of physics-informed priors during extrapolation. (a) The models that use the mean of all training data as the prior predict that yield strength increases with temperature in regions where it has not seen training data. Such nonphysical predictions are unsafe for alloy design. (b) The models that use the Maresca-Curtin model as the prior predict that yield strength monotonically decreases with temperature in regions where it has not seen training data. Incorporating this physics into the model allows for safer prediction during extrapolation.

This code associated with this toy problem can be found in the following Code Ocean repository: <https://codeocean.com/capsule/7849853/tree/v2>.

In the initial iteration (0th iteration), when employing the GPR with a zero mean, the predicted yield strength is evenly distributed across the entire design space. There is no prior indication about which region of the design space likely contains high-strength candidates. However, when utilizing the yield strength obtained from the ROM as a prior, it is evident that the W-Ta edge of the alloy space likely contains the maximum. By iteration 4, the standard GPR is still in the process of learning how the objective varies within the design space, while the HGPR with an informative prior has already found the maximum value via focusing its queries of the ground-truth and Proxy models in the W-rich region of the design space, as indicated by the prior. The BO scheme with the standard GPR converges after 10 iterations. The BO scheme with the standard GPR spent iterations in the Nb-W edge of the space, whereas the informed GPR spent all but one iteration exploring the W-Ta binary system.

Additionally, in Fig. 7b, a plot of optimization progress is presented, showcasing the maximum value discovered by each scheme against the iteration number. It is evident that the Bayesian optimization scheme equipped with the HGPR and informative priors converges to the true maximum value more rapidly, on average, compared to the scheme employing the standard GPR. Furthermore, the optimization progress is less sensitive to initial conditions, as demonstrated by the smaller standard deviation associated with the green line in Fig. 7b.

For instance, in a scenario where there are budget constraints on experimental campaigns aimed at optimizing refractory high entropy alloy (RHEA) compositions for high yield strength at elevated temperatures, cost-effective hardness experiments can be conducted within the design space. The results of these hardness experiments can then be utilized to train a hardness Gaussian Process (HV-GP), which can serve as an informative feature in a yield strength Gaussian Process (YS-GP). In this way, hardness can be used as a cheap proxy experiment for yield strength as proposed by Karumuri et al. [21].

4. Summary and conclusion

In order to replace Ni-based superalloys as the current structural material in jet engine turbine blades, RHEAs must be designed to meet a stringent set of property constraints relevant to the extreme environments in the hot zones of the engine. A property of particular interest to this work is high-temperature yield strength of RHEA. However, the vast number of potential RHEA compositions and the high cost associated with high-temperature tensile testing render conventional ICME methods for modeling yield strength impractical. These conventional ICME methods for predicting yield strength can be broadly categorized into three groups, arranged in increasing order of expense: (1) Data-hungry empirical and machine-learning models, (2) computationally expensive multiscale modeling, (3) and fast-acting yet frequently inaccurate analytical models. In this work, we combine the best aspects of categories 1 and 3 in order to create a model that is capable of exploiting multiple sources of experimental information yet is HTP.

In this work, we improved upon previous works by equipping hierarchical GP regressors [21] with informative priors. In this way, we combine data-hungry machine-learning models with fast-acting analytical models to develop a data-efficient model capable of iterative correction as new data is introduced. This methodology was benchmarked against the Borg dataset, which is the largest publicly available dataset of RHEA mechanical properties to-date.

A double-control GP model (not using imputed hardness as a feature, nor using informative priors), two single-test models (either using hardness as a feature or informative priors) and a double-test GP model (using imputed hardness as a feature and imperative priors) were compared. During a data-sparse 2-fold cross-validation, the double-control model performed the best *on average*. Specifically, the double-test model had the highest rank-ordering coefficient distributions (R_s and τ) indicating more physically relevant predictions and a better ability to identify optimally strong alloys. Additionally, the double-test had competitive MAE and RMSE distributions, indicating accurate predictions from the model in addition to the high Spearman Ranking coefficient, indicating correct rank ordering of alloys by strength.

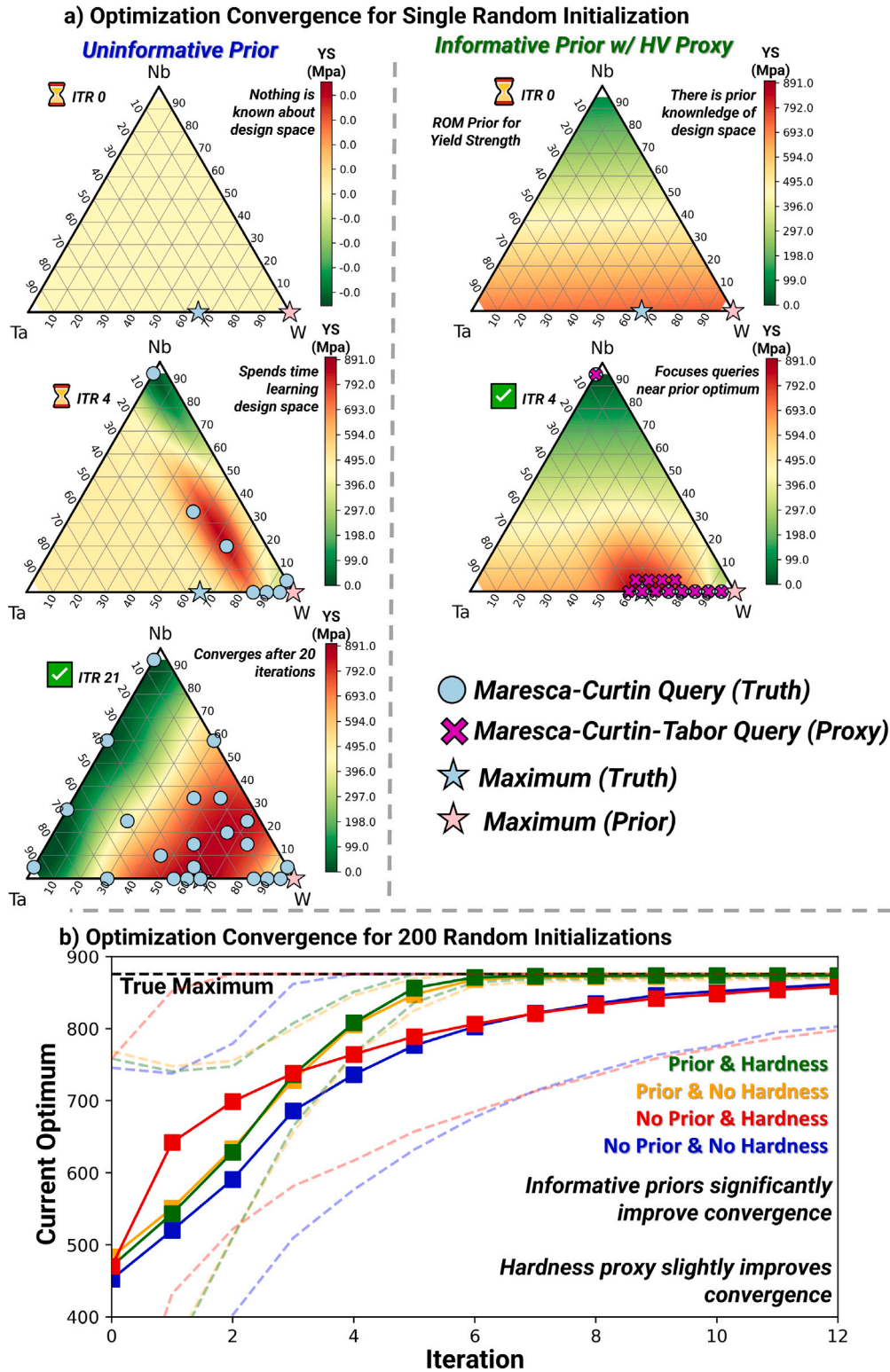


Fig. 7. (a) A simplified optimization problem using the Nb-Ta-W alloy space as an example. In the initial iteration (0th iteration), when the GPR with a zero mean is employed, the predicted yield strength is uniformly distributed throughout the design space. However, when the yield strength obtained from the rule-of-mixtures (ROM) is used as a prior, the 0th iteration clearly indicates the most promising point for further investigation, which is located near the W corner of the alloy space. In iteration 4, the standard GPR is still learning how the objective varies in the design space, whereas the HGPR equipped with an informative prior has only queried the ground-truth and Proxy in the W-rich region of the design space. Readers are encouraged to view the animated version of this figure at our repository: <https://codeocean.com/capsule/7849853/tree/v2>. (b) Optimization progress plot demonstrating the maximum value discovered by each scheme vs iteration. The BO scheme equipped with an HGPR with informative priors converges on the truth maximum quicker on average than the BO scheme with the standard GPR.

Next, the extrapolating abilities of the double-control and double-test models were visualized by withholding high-temperature yield strength data from the model. The results showed that HGPRs with priors exhibit physics-informed extrapolation, whereas the standard GPR produces non-physical outcomes i.e. predicting that yield strength does not monotonically decrease with temperature.

Finally, we showcased the effectiveness of this framework in Bayesian optimization (BO) approaches by conducting a design exercise using the Nb–Ta–W alloy system. The results illustrate that BO schemes that incorporate information from proxy hardness experiments and analytical models converge to the maximum yield strength more rapidly compared to BO schemes that rely on standard GPRs.

To summarize, the use of informative priors in Gaussian Process (GP) regression improves both the accuracy and physicality of yield strength predictions, especially during extrapolation. By incorporating informative priors, we effectively combine analytical models with experimental data, utilizing the available experimental information and relying on the analytical model when data is scarce. Furthermore, when informative priors are applied in Hierarchical GP regression frameworks, the benefits become even more significant, leading to improved performance and more reliable extrapolations.

Code availability

The data and code associated with this work can be accessed at the following Code Ocean Repository: <https://codeocean.com/capsule/7849853/tree/v2>.

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.

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.

Acknowledgments

The authors would like to thank Prof. Bill Curtin for his insights and recommendations while discussing this work. The authors acknowledge the support from the U.S. Department of Energy (DOE) ARPA-E ULTIMATE Program through Project DE-AR0001427, the NASA ESI Program through Grant No. 80NSSC21K0223, and US Army Research Laboratory Combat Capabilities Development Command (DEVCOM-ARL) under Contract No. W911NF2220106 (HTMDEC). DK acknowledges the support of National Science Foundation (NSF) through Grant No. CDSE-2001333. RA acknowledges the support from NSF Grant No. NSF-DMREF-2119103. BV acknowledges the support of NSF through Grant No. 1746932 (GRFP) and 1545403 (NRT-D³EM). The authors would like to acknowledge the GRACE supercomputing facility of Texas A&M University for providing the computing resources to carry out the research reported in this article.

References

- [1] M. Srikanth, A.R. Annamalai, A. Muthuchamy, C.-P. Jen, A review of the latest developments in the field of refractory high-entropy alloys, *Crystals* 11 (6) (2021) <http://dx.doi.org/10.3390/cryst11060612>, URL <https://www.mdpi.com/2073-4352/11/6/612>.
- [2] O. Senkov, S. Gorsse, D. Miracle, High temperature strength of refractory complex concentrated alloys, *Acta Mater.* 175 (2019) 394–405, <http://dx.doi.org/10.1016/j.actamat.2019.06.032>, URL <https://www.sciencedirect.com/science/article/pii/S1359645419304033>.
- [3] K.C. Atli, I. Karaman, A short review on the ultra-high temperature mechanical properties of refractory high entropy alloys, *Front. Met. Alloys* 2 (2023) <http://dx.doi.org/10.3389/fmal.2023.1135826>, URL <https://www.frontiersin.org/articles/10.3389/fmal.2023.1135826>.
- [4] Ultrahigh temperature impervious materials advancing turbine efficiency (ULTIMATE), 2020, URL <https://arpa-e.energy.gov/technologies/programs/ultimate>.
- [5] O. Senkov, G. Wilks, J. Scott, D. Miracle, Mechanical properties of Nb₂₅Mo₂₅Ta₂₅W₂₅ and V₂₀Nb₂₀Mo₂₀Ta₂₀W₂₀ refractory high entropy alloys, *Intermetallics* 19 (5) (2011) 698–706, <http://dx.doi.org/10.1016/j.intermet.2011.01.004>, URL <https://www.sciencedirect.com/science/article/pii/S0966979511000185>.
- [6] B. Vela, C. Acemi, P. Singh, T. Kirk, W. Trehern, E. Norris, D.D. Johnson, I. Karaman, R. Arróyave, High-throughput exploration of the WMoVTaNbAl refractory multi-principal-element alloys under multiple-property constraints, *Acta Mater.* 248 (2023) 118784.
- [7] T. Butler, K. Chaput, J. Dietrich, O. Senkov, High temperature oxidation behaviors of equimolar NbTiZrV and NbTiZrCr refractory complex concentrated alloys (RCCAs), *J. Alloys Compd.* 729 (2017) 1004–1019.
- [8] O. Senkov, S. Rao, K. Chaput, C. Woodward, Compositional effect on microstructure and properties of nbtizr-based complex concentrated alloys, *Acta Mater.* 151 (2018) 201–215.
- [9] N. Stepanov, N.Y. Yurchenko, D. Shaysultanov, G. Salishchev, M. Tikhonovsky, Effect of Al on structure and mechanical properties of AlxNbTiVZr (x = 0, 0.5, 1, 1.5) high entropy alloys, *Mater. Sci. Technol.* 31 (10) (2015) 1184–1193.
- [10] Y. Chen, Z. Xu, M. Wang, Y. Li, C. Wu, Y. Yang, A single-phase V_{0.5}Nb_{0.5}ZrTi refractory high-entropy alloy with outstanding tensile properties, *Mater. Sci. Eng. A* 792 (2020) 139774, <http://dx.doi.org/10.1016/j.msea.2020.139774>, URL <https://www.sciencedirect.com/science/article/pii/S0921509320308480>.
- [11] Y. Ye, Q. Wang, J. Lu, C. Liu, Y. Yang, High-entropy alloy: challenges and prospects, *Mater. Today* 19 (6) (2016) 349–362, <http://dx.doi.org/10.1016/j.matod.2015.11.026>, URL <https://www.sciencedirect.com/science/article/pii/S1369702115004010>.
- [12] U. Bhandari, M.R. Rafi, C. Zhang, S. Yang, Yield strength prediction of high-entropy alloys using machine learning, *Mater. Today Commun.* 26 (2021) 101871, <http://dx.doi.org/10.1016/j.mtcomm.2020.101871>, URL <https://www.sciencedirect.com/science/article/pii/S2352492820328828>.
- [13] B. Yin, W.A. Curtin, First-principles-based prediction of yield strength in the RhIrPdPtNiCu high-entropy alloy, *npj Comput. Mater.* 5 (1) (2019) 14.
- [14] Z. Liang, Y. Wu, Y. Miao, W. Pan, Y. Zhang, Composition design and tensile properties of additive manufactured low density Hf-Nb-Ta-Ti-Zr high entropy alloys based on atomic simulations, *Materials* 16 (11) (2023) 4039.
- [15] S. Mishra, S. Maiti, B. Rai, Computational property predictions of Ta-Nb-Hf-Zr high-entropy alloys, *Sci. Rep.* 11 (1) (2021) 4815.
- [16] S. Rao, C. Woodward, B. Akdim, O. Senkov, D. Miracle, Theory of solid solution strengthening of BCC chemically complex alloys, *Acta Mater.* 209 (2021) 116758, <http://dx.doi.org/10.1016/j.actamat.2021.116758>, URL <https://www.sciencedirect.com/science/article/pii/S1359645421001385>.
- [17] H. Suzuki, Solid solution hardening in body-centred cubic alloys, *Dislocations Solids* 4 (1980) 191–217.
- [18] I. Toda-Caraballo, P.E.R.-D. del Castillo, Modelling solid solution hardening in high entropy alloys, *Acta Mater.* 85 (2015) 14–23, <http://dx.doi.org/10.1016/j.actamat.2014.11.014>, URL <https://www.sciencedirect.com/science/article/pii/S1359645414008581>.
- [19] P. Ferreirós, S. von Tiedemann, N. Parkes, D. Gurah, D. King, P. Norman, M. Gilbert, A. Knowles, Vnbcmo refractory high-entropy alloy for nuclear applications, *Int. J. Refract. Met. Hard Mater.* 113 (2023) 106200, <http://dx.doi.org/10.1016/j.ijrmhm.2023.106200>, URL <https://www.sciencedirect.com/science/article/pii/S0263436823001002>.
- [20] L. Raman, A. Anupam, G. Karthick, C.C. Berndt, A.S.M. Ang, S.V.S. Narayana Murthy, D. Fabijanic, B.S. Murty, R.S. Kottada, Strengthening mechanisms in Cr-MoNbTiW refractory high entropy alloy, *Mater. Sci. Eng. A* 819 (2021) 141503, <http://dx.doi.org/10.1016/j.msea.2021.141503>, URL <https://www.sciencedirect.com/science/article/pii/S0921509321007723>.
- [21] 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>.

- [22] F. Maresca, W.A. Curtin, Mechanistic origin of high strength in refractory BCC high entropy alloys up to 1900K, *Acta Mater.* 182 (2020) 235–249, <http://dx.doi.org/10.1016/j.actamat.2019.10.015>, URL <https://www.sciencedirect.com/science/article/pii/S1359645419306755>.
- [23] A.B. Kustas, M.R. Jones, F.W. DelRio, P. Lu, J. Pegues, P. Singh, A. Smirnov, J. Tiarks, E.D. Hintsala, D.D. Stauffer, J.K. Román-Kustas, M. Abere, E.M. White, D.D. Johnson, I.E. Anderson, N. Argibay, Extreme hardness at high temperature with a lightweight additively manufactured multi-principal element superalloy, *Appl. Mater. Today* 29 (2022) 101669, <http://dx.doi.org/10.1016/j.apmt.2022.101669>, URL <https://www.sciencedirect.com/science/article/pii/S2352940722003031>.
- [24] S. Gorsse, M. Nguyen, O. Senkov, D. Miracle, Database on the mechanical properties of high entropy alloys and complex concentrated alloys, *Data Brief* 21 (2018) 2664–2678, <http://dx.doi.org/10.1016/j.dib.2018.11.111>, URL <https://www.sciencedirect.com/science/article/pii/S235234091831504X>.
- [25] O. Senkov, J. Scott, S. Senkova, D. Miracle, C. Woodward, Microstructure and room temperature properties of a high-entropy TaNbHfZrTi alloy, *J. Alloys Compd.* 509 (20) (2011) 6043–6048.
- [26] C.K. Borg, C. Frey, J. Moh, T.M. Pollock, S. Gorsse, D.B. Miracle, O.N. Senkov, B. Meredig, J.E. Saal, Expanded dataset of mechanical properties and observed phases of multi-principal element alloys, *Sci. Data* 7 (1) (2020) 430.
- [27] F. Maresca, W.A. Curtin, Theory of screw dislocation strengthening in random BCC alloys from dilute to “high-entropy” alloys, *Acta Mater.* 182 (2020) 144–162, <http://dx.doi.org/10.1016/j.actamat.2019.10.007>, URL <https://www.sciencedirect.com/science/article/pii/S1359645419306676>.
- [28] C. Baruffi, F. Maresca, W. Curtin, Screw vs. edge dislocation strengthening in body-centered-cubic high entropy alloys and implications for guided alloy design, *Mrs Commun.* (2022) 1–8.
- [29] C. Lee, Y. Chou, G. Kim, M.C. Gao, K. An, J. Brechtel, C. Zhang, W. Chen, J.D. Poplawsky, G. Song, et al., Lattice-distortion-enhanced yield strength in a refractory high-entropy alloy, *Adv. Mater.* 32 (49) (2020) 2004029.
- [30] R. Feng, B. Feng, M.C. Gao, C. Zhang, J.C. Neufeld, J.D. Poplawsky, Y. Ren, K. An, M. Widom, P.K. Liaw, Superior high-temperature strength in a supersaturated refractory high-entropy alloy, *Adv. Mater.* 33 (48) (2021) 2102401.
- [31] R. Pan, M. Gu, J. Wu, Physics-informed Gaussian process regression of in operando capacitance for carbon supercapacitors, *Energy Adv.* (2023).
- [32] S. Liu, K. Lee, P.V. Balachandran, Integrating machine learning with mechanistic models for predicting the yield strength of high entropy alloys, *J. Appl. Phys.* 132 (10) (2022) 105105.
- [33] C. Varvenne, A. Luque, W.A. Curtin, Theory of strengthening in fcc high entropy alloys, *Acta Mater.* 118 (2016) 164–176, <http://dx.doi.org/10.1016/j.actamat.2016.07.040>, URL <https://www.sciencedirect.com/science/article/pii/S1359645416305481>.
- [34] Y.P. Varshni, Temperature dependence of the elastic constants, *Phys. Rev. B* 2 (1970) 3952–3958, <http://dx.doi.org/10.1103/PhysRevB.2.3952>, URL <https://link.aps.org/doi/10.1103/PhysRevB.2.3952>.
- [35] C.E. Rasmussen, C.K.I. Williams, *Gaussian Processes for Machine Learning (Adaptive Computation and Machine Learning)*, The MIT Press, 2005.
- [36] S.V. Kalinin, M. Valletti, R.K. Vasudevan, M. Ziatdinov, Exploration of lattice Hamiltonians for functional and structural discovery via Gaussian process-based exploration–exploitation, *J. Appl. Phys.* 128 (16) (2020) 164304, <http://dx.doi.org/10.1063/5.0021762>, arXiv:https://pubs.aip.org/aip/jap/article-pdf/doi/10.1063/5.0021762/13992868/164304_1_online.pdf.
- [37] L. Chen, H. Tran, R. Batra, C. Kim, R. Ramprasad, Machine learning models for the lattice thermal conductivity prediction of inorganic materials, *Comput. Mater. Sci.* 170 (2019) 109155.
- [38] M. Gheytaianzadeh, A. Baghban, S. Habibzadeh, A. Esmaeili, O. Abida, A. Mohaddespour, M.T. Munir, Towards estimation of CO₂ adsorption on highly porous MOF-based adsorbents using gaussian process regression approach, *Sci. Rep.* 11 (1) (2021) 15710.
- [39] T.Z. Khan, T. Kirk, G. Vazquez, P. Singh, A. Smirnov, D.D. Johnson, K. Youssef, R. Arróyave, Towards stacking fault energy engineering in FCC high entropy alloys, *Acta Mater.* 224 (2022) 117472, <http://dx.doi.org/10.1016/j.actamat.2021.117472>, URL <https://www.sciencedirect.com/science/article/pii/S135964542100851X>.
- [40] C. Wen, C. Wang, Y. Zhang, S. Antonov, D. Xue, T. Lookman, Y. Su, Modeling solid solution strengthening in high entropy alloys using machine learning, *Acta Mater.* 212 (2021) 116917, <http://dx.doi.org/10.1016/j.actamat.2021.116917>, URL <https://www.sciencedirect.com/science/article/pii/S1359645421002974>.
- [41] S. Gao, Z. Gao, F. Zhao, Machine learning and visualization assisted solid solution strengthening phase prediction of high entropy alloys, *Mater. Today Commun.* 35 (2023) 105894, <http://dx.doi.org/10.1016/j.mtcomm.2023.105894>, URL <https://www.sciencedirect.com/science/article/pii/S2352492823005858>.
- [42] X. Huang, C. Jin, C. Zhang, H. Zhang, H. Fu, Machine learning assisted modelling and design of solid solution hardened high entropy alloys, *Mater. Des.* 211 (2021) 110177, <http://dx.doi.org/10.1016/j.matdes.2021.110177>, URL <https://www.sciencedirect.com/science/article/pii/S0264127521007322>.
- [43] J.R. Vergara, P.A. Estévez, A review of feature selection methods based on mutual information, *Neural Comput. Appl.* 24 (2014) 175–186.
- [44] 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.
- [45] T. Chai, R.R. Draxler, Root mean square error (RMSE) or mean absolute error (MAE), *Geosci. Model Dev. Discuss.* 7 (1) (2014) 1525–1534.
- [46] J.M. Davila Delgado, L. Oyedele, Deep learning with small datasets: using autoencoders to address limited datasets in construction management, *Appl. Soft Comput.* 112 (2021) 107836, <http://dx.doi.org/10.1016/j.asoc.2021.107836>, URL <https://www.sciencedirect.com/science/article/pii/S1568494621007584>.
- [47] J.H. Zar, Spearman rank correlation: overview, *Wiley StatsRef: Stat. Ref. Online* (2014).
- [48] M. Sanderson, I. Soboroff, Problems with Kendall's tau, in: *Proceedings of the 30th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval*, 2007, pp. 839–840.
- [49] A.Y.-T. Wang, R.J. Murdock, S.K. Kauwe, A.O. Olynyk, A. Gurlo, J. Brzgoch, K.A. Persson, T.D. Sparks, Machine learning for materials scientists: an introductory guide toward best practices, *Chem. Mater.* 32 (12) (2020) 4954–4965.
- [50] C. Jin, X. Li, J. Kang, H. Li, H. Wang, Effect of interstitial oxygen/nitrogen on mechanical and wear properties of TiZrHfNb refractory high-entropy alloy, *J. Alloys Compd.* 960 (2023) 170863, <http://dx.doi.org/10.1016/j.jallcom.2023.170863>, URL <https://www.sciencedirect.com/science/article/pii/S0925838823021667>.
- [51] C.-C. Juan, M.-H. Tsai, C.-W. Tsai, W.-L. Hsu, C.-M. Lin, S.-K. Chen, S.-J. Lin, J.-W. Yeh, Simultaneously increasing the strength and ductility of a refractory high-entropy alloy via grain refining, *Mater. Lett.* 184 (2016) 200–203, <http://dx.doi.org/10.1016/j.matlet.2016.08.060>, URL <https://www.sciencedirect.com/science/article/pii/S0167577X16313398>.
- [52] A. Kanchi, K.V. Rajulapati, B.S. Rao, D. Sivaprasam, R.C. Gundakaram, Influence of thermomechanical processing on microstructure and mechanical properties of MoNbTaW refractory high-entropy alloy, *J. Mater. Eng. Perform.* 31 (10) (2022) 7964–7972.
- [53] S. Thapliyal, M. Komarasamy, S. Shukla, L. Zhou, H. Hyer, S. Park, Y. Sohn, R.S. Mishra, An integrated computational materials engineering-anchored closed-loop method for design of aluminum alloys for additive manufacturing, *Materialia* 9 (2019) <http://dx.doi.org/10.1016/j.mtla.2019.100574>.
- [54] Q. Sun, G. Zhou, Z. Meng, M. Jain, X. Su, An integrated computational materials engineering framework to analyze the failure behaviors of carbon fiber reinforced polymer composites for lightweight vehicle applications, *Compos. Sci. Technol.* 202 (2021) 108560, <http://dx.doi.org/10.1016/j.compscitech.2020.108560>, URL <https://www.sciencedirect.com/science/article/pii/S0266353820323526>.
- [55] S. Thapliyal, M. Komarasamy, S. Shukla, L. Zhou, H. Hyer, S. Park, Y. Sohn, R.S. Mishra, An integrated computational materials engineering-anchored closed-loop method for design of aluminum alloys for additive manufacturing, *Materialia* 9 (2020) 100574, <http://dx.doi.org/10.1016/j.mtla.2019.100574>, URL <https://www.sciencedirect.com/science/article/pii/S2589152919303709>.
- [56] J.H. Montoya, K.T. Winther, R.A. Flores, T. Bligaard, J.S. Hummelshøj, M. Aykol, Autonomous intelligent agents for accelerated materials discovery, *Chem. Sci.* 11 (32) (2020) 8517–8532.
- [57] M.M. Noack, G.S. Doerk, R. Li, J.K. Streit, R.A. Vaia, K.G. Yager, M. Fukuto, Autonomous materials discovery driven by Gaussian process regression with inhomogeneous measurement noise and anisotropic kernels, *Sci. Rep.* 10 (1) (2020) 17663.
- [58] S.T. Bukkapatnam, Autonomous materials discovery and manufacturing (AMDM): A review and perspectives, *IIEE Trans.* 55 (1) (2023) 75–93, <http://dx.doi.org/10.1080/24725854.2022.2089785>, arXiv:[10.1080/24725854.2022.2089785](https://arxiv.org/abs/10.1080/24725854.2022.2089785).
- [59] F. Häse, L.M. Roch, C. Kreisbeck, A. Aspuru-Guzik, Phoenix: A Bayesian optimizer for chemistry, *ACS Cent. Sci.* 4 (9) (2018) 1134–1145, <http://dx.doi.org/10.1021/acscentsci.8b00307>, PMID: 30276246, arXiv:[10.1021/acscentsci.8b00307](https://arxiv.org/abs/10.1021/acscentsci.8b00307).
- [60] F. Archetti, A. Candelieri, *Bayesian Optimization and Data Science*, Springer, 2019, pp. 57–72.