Multifidelity regression of sparse plasma transport data available in disparate physical regimes

Lucas J. Stanek [1,*] Shaunak D. Bopardikar [2,†] and Michael S. Murillo [1,‡]

[1]*Department of Computational Mathematics, Science and Engineering, Michigan State University, Michigan 48824, USA*

[2]*Department of Electrical and Computer Engineering, Michigan State University, Michigan 48824, USA*

(Received 17 June 2021; revised 9 September 2021; accepted 22 November 2021; published 22 December 2021)

Physical data are typically generated by experiments and computations in limited parameter regimes. When datasets generated using such disparate methods are combined into one dataset, the resulting dataset is typically sparse, with dense "islands" in a potentially high-dimensional parameter space, and predictions must be interpolated among such islands. Using plasma transport data as our example, we propose a multifidelity Gaussian-process regression framework that incorporates physical data from multiple sources at multiple fidelities. The impact of the proposed framework varies from little improvement over simpler approaches to qualitatively changing the prediction with consistently increased confidence in regions lacking high-fidelity data. By varying low- and high-fidelity data sources, we demonstrate an approach for determining when multifidelity Gaussian-process regression adds value over single-fidelity regression and therefore when its additional computational costs are merited. We also examine the case in which the outputs of the low- and high-fidelity models correspond to different physical quantities, one of which may be intrinsically computationally cheaper to compute. We conclude by analyzing strategies for sampling high-fidelity data for use in this framework, and we develop a simple sampling approach for reducing regression error across large gaps in data.

DOI: 10.1103/PhysRevE.104.065303

I. INTRODUCTION

The generation of high-fidelity (HF) data requires substantial resources that limit the volume of data that can be generated. Moreover, the size and scope of datasets are constrained by the experimental accessibility of physical regimes and by the applicability and efficiency of computational models. These limitations can be addressed by combining data from several sources to form a dataset that contains multiple, separated point clouds that can be "interpolated." For example, the equation of state can be measured in one regime with a laser-heated diamond anvil cell [1] and computed in another regime with accurate electronic-structure methods [2]. Or, one may combine experimental data obtained along a shock Hugoniot with computational data available only at very low temperatures. Computational models for equations of state, atomic properties, and charged-particle transport [3–6] can also be combined to create a larger dataset.

Combining data sources in this way creates two challenges. First, predictions will be based on data in potentially very different physical regimes. Second, while it is natural to consider adding lower-fidelity (LF) data, which can be generated cheaply, to datasets to cover a parameter space more uniformly, it is not clear how to exploit such LF data in making predictions.

Machine-learning (ML) methods offer promising alternative frameworks for interpolating physical data [7–9]. ML treats the interpolation problem as regression in a high-dimensional space using nontraditional techniques such as neural networks. Gaussian-process regression (GPR) [10] is a nonparametric ML technique that interpolates data in multiple dimensions; importantly, GPR provides an uncertainty estimate that can be used to suggest where new data points should be acquired.

Here, we will explore GPR as an approach for interpolating physical data. In particular, we will examine the situation in which there are islands of HF data in parameter space, possibly from different sources, and we will fill the space between these islands with easier-to-compute LF data. Such an approach utilizes multifidelity (MF) extensions [11] of GPR. Here, we use GPR to refer to the methodology described in Ref. [10], and MF-GPR to refer to its MF extensions [11–14].

The generality of MF-GPR methods enables their use in many disciplines and applications [15–22]. The original MF-GPR framework has been improved to reduce the risk of overfitting during the training procedure [14], to include nonlinear relationships between LF and HF models [12,13], and to address concerns that arise with diverse data structures and dataset selection [23,24].

This manuscript is organized as follows. As described in the following section, we will illustrate our ML ideas using the example of ionic transport coefficients. The methods we used to generate our dataset of ionic transport coefficients are discussed in Sec. II A. In Sec. II B, we compare

*staneklu@msu.edu
†shaunak@msu.edu
‡murillom@msu.edu

single-fidelity regression methods and highlight the benefit of GPR over simple cubic spline regression. We then transition to MF regression and discuss the formulation of MF-GPR, as introduced by Kennedy and O'Hagan [11], in Sec. II C. Using toy examples, we show when MF-GPR adds value over single-fidelity GPR; we also show where improvements to this formulation are needed. We conclude Sec. II D by reporting a table of computation times and regression errors for MF-GPR and single-fidelity GPR to assess the cost-benefit tradeoff for these methods.

Section III illustrates an approach for choosing an LF model that is the most appropriate for an MF-GPR setting and examines how this choice impacts the resulting MF-GPR fit. A natural choice for LF and HF models are those with the same output quantity (e.g., both predict the viscosity of a system). However, the outputs of both models need not be the same quantities. We explore the use of models in MF-GPR that have different output quantities, as well as different levels of computational complexity.

In Sec. IV A, we compare regression errors resulting from single-fidelity GPR and MF-GPR analyses of sparse, disparate plasma transport-coefficient datasets. We find that while MF-GPR may result in modestly smaller errors compared to single-fidelity GPR, the uncertainty of the MF-GPR prediction is consistently much smaller. Finally, in Sec. IV B, we compare three approaches for sampling HF data to reduce the MF-GPR regression error. For a fixed number of HF data points, a simple approach we explored outperforms sampling from a uniform grid. We offer conclusions and discuss potential areas for future work in Sec. V.

II. DATASET AND REGRESSION METHODS

In this section, we discuss our dataset and review the ML approaches we will employ in Secs. III and IV. We begin in Sec. II A by describing plasma transport-coefficient data and the fidelities of several commonly used models based on the approximations they employ. In Secs. II B and II C, we describe the GPR methodology, including standard, single-fidelity GPR and its MF generalization. Our goals in Secs. II B and II C are to answer the following questions: what value does GPR add compared to simpler regression methods? And, how does including data from multiple levels of fidelity impact a prediction?

A. Ionic transport-coefficient dataset

For our study, we chose to explore MF-GPR in the context of plasma ionic transport coefficients because plasmas span many orders of magnitude in density, temperature and nuclear charge. Plasmas can include many species, which makes it difficult to use a single (computational or experimental) method to make accurate predictions. Computational methods that are typically used can be divided into LF and HF methods by examining the underlying assumptions of the models. Moreover, we can usually identify a limited parameter regime in which each model is HF. These delineations occur because the theoretical models that underpin the computational methods are known to have high accuracy only in certain limits (e.g., asymptotically at high temperature); methods that are not asymptotically accurate in a parameter regime are designated as LF there. The limiting regimes typically depend on multiple dimensionless parameters (e.g., the Coulomb coupling parameter and the degeneracy parameter) that rely on some combination of nuclear charge, density, and temperature of the system. We will use just the temperature of the system to specify the limiting regimes since models developed at extremes of temperature tend to have very different assumptions.

TABLE I. HF and LF models for the self-diffusion and viscosity coefficients in each temperature regime. Each LF model is used across the entire temperature range.

Coeff.	T (eV)	HF	LF
D	$T < O(10^1)$	DFT-MD [25–28]	HMP [30]
	$O(10^1) < T < O(10^3)$	–	HMP [30]
	$T > O(10^3)$	SMT [29]	HMP [30]
η	$T < O(10^1)$	DFT-MD [25–28]	YGBI [32]
	$O(10^1) < T < O(10^3)$	–	YGBI [32]
	$T > O(10^3)$	YVM [31]	YGBI [32]

Data in the low-temperature regime, loosely defined here as $T < O(10^1)$ eV, and in a high-temperature regime, defined here as $T > O(10^3)$ eV, will be generated using appropriate LF and HF models.

For the self-diffusion transport coefficient D, we will use the following HF models to generate data. At low temperatures, the HF data are obtained from density functional theory molecular dynamics (DFT-MD) simulations [25–28], which accurately calculate the electronic structure on-the-fly. At high temperatures, the Stanton-Murillo transport (SMT) model [29], which uses numerically computed cross-sections and an effective interaction potential, is employed. The LF model used across the entire temperature range is given by Hansen, McDonald, and Pollock (HMP) for a one-component plasma (OCP) [30].

Similarly, for viscosity η, we use one HF model at low temperatures and a different HF model at high temperatures. Once again, the HF data at low temperatures are obtained from DFT-MD simulations. We employ the Yukawa viscosity model (YVM) [31], which is based on a quasiuniversal form fit to MD data, as our HF model at high temperatures. Our LF model is derived from a correspondence between an OCP system and a Yukawa system. The correspondence is obtained from the Gibbs-Bogolyubov inequality [32]; this model will be referred to as the YGBI model.

The HF and LF models for the self-diffusion and viscosity coefficients in each temperature range are summarized in Table I. These models are used in analyses presented in Sec. IV A.

B. Single-fidelity regression

To provide a baseline to which results of MF-GPR can be compared in later sections, we first consider approaches that require only one level of data fidelity, i.e., single-fidelity approaches. We consider cubic-spline regression and GPR. Cubic-spline regression is a parametric regression method that aims to determine the optimal *parameters* that define a cubic-spline fit to data. In contrast, GPR is a nonparametric

regression approach that determines the optimal *function* that is fit to data. We begin with a brief overview of GPR that will provide a framework for understanding its MF generalization.

We introduce GPR with a discussion of prior and posterior distributions. Before observing the data, we have some prior beliefs about functions that are suitable. These functions are drawn from a *prior distribution*: a distribution of random functions that are consistent with our prior beliefs about the data. An example of a prior distribution is one in which the distribution of functions have zero mean at each input point and vary smoothly over the entire input space. For plasma transport data, we could impose constraints on our prior distribution of functions to enforce nonnegativity and that the functions reflect the known behaviors of different transport coefficients (e.g., increasing with temperature). After constructing a prior distribution, a *posterior distribution* is created by using available data to constrain the random functions by ensuring that they pass through the observed data points. As we will see, the mean and the covariance matrix of a posterior distribution are the prediction and uncertainty estimates of GPR.

Defining the prior and posterior distributions for GPR requires a kernel function that defines a measure of similarity among the input variables of a dataset. The kernel function determines the representation of the functions from the prior and posterior distributions (e.g., smoothness, periodicity, etc.). A common choice of kernel function, that we will use here, is the squared-exponential kernel

$$k(x_i, x_j; \sigma^2, \ell) = \sigma^2 \exp\left(-\frac{1}{2\ell^2}\|x_i - x_j\|^2\right), \tag{1}$$

where for d-dimensional data, we have m points $x_i \in \mathbb{R}^d$ and n points $x_j \in \mathbb{R}^d$. Evaluating the kernel $k(x_i, x_j; \sigma^2, \ell)$ gives the ijth entry of the kernel matrix (or covariance matrix) $K \in \mathbb{R}^{m\times n}$. The hyperparameters of (1) are the variance σ^2 and the length scale ℓ; they will be compactly denoted as the set $\theta \in \{\sigma^2, \ell\}$. These hyperparameters reveal the strength and extent of correlations in the data. As we will see, the values of the hyperparameters are particularly useful for quantifying the quality of MF-GPR methods.

A single-fidelity GPR problem is posed as follows: *given a set of n training points in d dimensions, represented by the columns of a matrix* $X_{SF} \in \mathbb{R}^{d\times n}$ *and the corresponding (scalar) output values* $y \in \mathbb{R}^n$ *of the unknown function at each training point, predict the value of the unknown function at a set of m test points* $X_* \in \mathbb{R}^{d\times m}$. As shown in Ref. [10], the posterior distribution of the unknown function using GPR at the new set of data points X_* is a multivariate Gaussian with mean $\boldsymbol{\mu}_*$ and covariance Σ_* given by

$$\boldsymbol{\mu}_*(X_*) = K(X_*, X_{\rm SF}; \theta) K(X_{\rm SF}, X_{\rm SF}; \theta)^{-1} y, \tag{2}$$

$$\Sigma_*(X_*) = K(X_*, X_*; \theta) - K(X_*, X_{\rm SF}; \theta) K(X_{\rm SF}, X_{\rm SF}; \theta)^{-1} K(X_{\rm SF}, X_*; \theta), \tag{3}$$

where the hyperparameters θ of the kernel function are determined by optimizing the log-likelihood function $\mathcal{L}(y, X_{\rm SF}, \theta)$, as discussed in Ref. [10]. The function $\mathcal{L}$ measures the likelihood that the observations are given by the values y at training locations $X_{\rm SF}$ for a given value of θ.

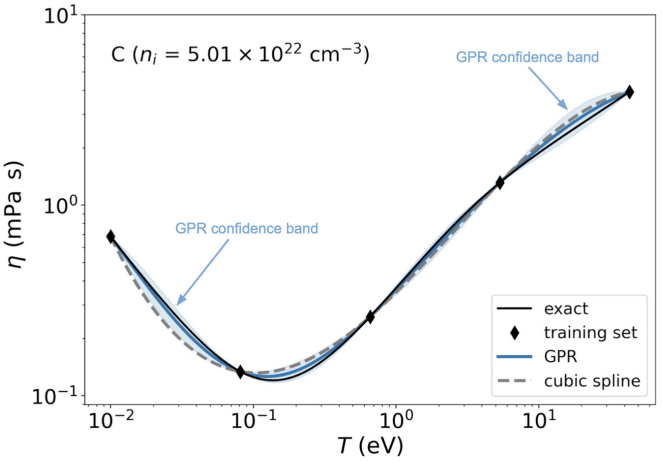

FIG. 1. Comparison of GPR and cubic-spline regression for a single-fidelity viscosity dataset using the YVM for the element C at $n_i = 5.01 \times 10^{22}$ cm^{-3}. The training points (black diamonds) were fit using both GPR (blue line) and a cubic spline (gray dashed line). The shaded bands show a 95% confidence interval around the GPR fit. Locations of future HF training points are suggested by the confidence band.

We now examine a simple example of GPR and compare with a cubic-spline interpolation. Viscosity data were generated using the YVM for the element C at $n_i = 5.01 \times 10^{22}$ cm^{-3}, and fits to these data using GPR [33] and cubic-spline regression [34] are shown in Fig. 1. The GPR fit, denoted as "GPR," corresponds to $\boldsymbol{\mu}_*(X_*)$ from Eq. (2); the shaded bands around $\boldsymbol{\mu}_*(X_*)$ correspond to a 95% confidence interval and are computed from Eq. (3). The fit generated using cubic-spline regression on the same dataset is denoted as "cubic spline." For all GPR fits, the data were first scaled to unit variance and zero mean. The hyperparameter optimization routine was carried out using the limited-memory quasi-Newton algorithm [35] with 15 random restarts, and a measurement noise with a variance of 10^{-6} was added to ensure that the kernel matrix $K(X_{\rm SF}, X_{\rm SF}; \theta)$ for computing the posterior distribution would be guaranteed to be positive-definite (and therefore, invertible) during fitting. Both the cubic-spline regression and GPR methods produced accurate fits, as shown by comparison to the underlying true solution, which is denoted with a black line in Fig. 1 and labeled "exact." A key difference between cubic splines and GPR is that the GPR method provides a confidence interval (shaded bands) around the GPR prediction—suggesting where additional data are needed to improve the prediction.

C. Multifidelity Gaussian-process regression

We now turn to the case where there are two sources of data, one LF and one HF. The outputs of the LF model are denoted as $y_{\rm LF} \in \mathbb{R}^{N_{\rm LF}}$ and are evaluated at $X_{\rm LF} \in \mathbb{R}^{d\times N_{\rm LF}}$. Similarly, the outputs from the HF model are denoted as $y_{\rm HF} \in \mathbb{R}^{N_{\rm HF}}$ and are evaluated at $X_{\rm HF} \in \mathbb{R}^{d\times N_{\rm HF}}$. Here, the numbers of LF and HF data points are denoted as $N_{\rm LF}$ and $N_{\rm HF}$, respectively.

To understand how the LF data can be used in HF predictions with an MF method, consider this simple procedure

with three steps. First, in step (a), we combine the LF and HF data into a single dataset with greater coverage than the HF data alone offer. In step (b), we use LF data to influence HF predictions by quantifying correlations between the LF and HF datasets with a correlation hyperparameter, ρ. Finally, step (c) of the procedure imposes a constraint that a prediction at a HF data point ignores the LF data.

Each part of the above procedure is addressed by the original MF-GPR formulation proposed by Kennedy and O'Hagan [11], which begins by assuming that there is a linear mapping between fidelities that is described by the autoregressive model

$$f_{\mathrm{HF}}(x) = \rho f_{\mathrm{LF}}(x) + \delta_{\mathrm{HF}}(x). \tag{4}$$

The function $\delta_{\mathrm{HF}}(x)$ is to be viewed as the error or *bias* between the HF data and a scaled value of the LF data, where the correlation hyperparameter ρ is the scaling term. Notice that if the LF and HF data are uncorrelated, i.e., $\rho = 0$, then $\delta_{\mathrm{HF}} = f_{\mathrm{HF}}$. The key idea in this approach is to use the LF and HF data to learn the parameters governing the unknown functions f_{LF} and δ_{HF} and the hyperparameter ρ to be able to predict the value of f_{HF} at a test point x. The functions f_{LF} and δ_{HF} are typically assumed to be realizations of independent Gaussian processes with zero mean and a kernel matrix K. This means that on a test set X_*, $f_{\mathrm{LF}}(X_*)$, and $\delta_{\mathrm{HF}}(X_*)$ are independent Gaussian random variables that are normally distributed as per

$$f_{\mathrm{LF}}(X_*) \sim \mathcal{N}[\mathbf{0}, K(X_*, X_*; \theta_{\mathrm{LF}})], \tag{5}$$

$$\delta_{\mathrm{HF}}(X_*) \sim \mathcal{N}[\mathbf{0}, K(X_*, X_*; \theta_{\mathrm{HF}})], \tag{6}$$

where θ_{LF} and θ_{HF} denote the hyperparameters for the LF and HF models, respectively. The notation $\mathcal{N}(\mathbf{0}, \Sigma)$ denotes a multivariate Gaussian random variable with mean $\mathbf{0}$ and covariance Σ. Because f_{LF} and δ_{HF} are independent, it follows that [36]

$$f_{\mathrm{HF}}(X_*) \sim \mathcal{N}[\mathbf{0}, \rho^2 K(X_*, X_*; \theta_{\mathrm{LF}}) + K(X_*, X_*; \theta_{\mathrm{HF}})]. \tag{7}$$

For brevity, we denote

$$K_{11}(X, X') \equiv K(X, X'; \theta_{\mathrm{LF}}), \tag{8}$$

$$K_{12}(X, X') \equiv \rho K(X, X'; \theta_{\mathrm{LF}}), \tag{9}$$

$$K_{21}(X, X') \equiv K_{12}(X, X'), \tag{10}$$

$$K_{22}(X, X') \equiv \rho^2 K(X, X'; \theta_{\mathrm{LF}}) + K(X, X'; \theta_{\mathrm{HF}}). \tag{11}$$

Equations (5)–(7) can be jointly written as [11,12]

$$\begin{bmatrix} f_{\mathrm{LF}}(X_{\mathrm{LF}}) \\ f_{\mathrm{HF}}(X_{\mathrm{HF}}) \\ f_{\mathrm{HF}}(X_*) \end{bmatrix} \sim \mathcal{N}\left\{ \begin{bmatrix} \mathbf{0} \\ \mathbf{0} \\ \mathbf{0} \end{bmatrix}, \begin{bmatrix} K_{11}(X_{\mathrm{LF}}, X_{\mathrm{LF}}) & K_{12}(X_{\mathrm{LF}}, X_{\mathrm{HF}}) & K_{12}(X_{\mathrm{LF}}, X_*) \\ K_{21}(X_{\mathrm{HF}}, X_{\mathrm{LF}}) & K_{22}(X_{\mathrm{HF}}, X_{\mathrm{HF}}) & K_{22}(X_{\mathrm{HF}}, X_*) \\ K_{21}(X_*, X_{\mathrm{LF}}) & K_{22}(X_*, X_{\mathrm{HF}}) & K_{22}(X_*, X_*) \end{bmatrix} \right\}. \tag{12}$$

The form of Eq. (12) reveals how the LF and HF data are combined (i.e., through K_{12} and K_{21}), completing step (a). Note that when the hyperparameter ρ, which couples the LF and HF models, is equal to zero, Eq. (12) reduces to two decoupled Gaussian processes. This means that when the LF and HF models are uncorrelated, the LF data will not influence the HF regression, resulting in one single-fidelity GPR at each fidelity level. Following the procedure for determining optimal hyperparameters θ for a kernel function, the hyperparameter ρ is also determined by optimizing a log-likelihood function, as discussed in Sec. 2.4 of Ref. [11]. As a result of this optimization procedure, if ρ turns out to have a large value, then there is substantial correlation between the LF and HF models. Otherwise, the LF and HF models are uncorrelated. Thus, the correlation hyperparameter ρ determined from the MF dataset directly quantifies the influence of the LF data on the HF fit, completing part (b) of the procedure mentioned above.

We have shown how data from LF and HF models can be combined into a single MF dataset and how the degree of influence of LF data on fits to HF data can be quantified using the correlation hyperparameter ρ. However, we still need to show how to produce a fit to HF data using Eq. (12), while also completing step (c) of the procedure mentioned above. By conditioning the joint Gaussian prior distribution Eq. (12), the predictive mean and the covariance matrix are obtained from the Gaussian posterior distribution

$$f_{*,\mathrm{HF}}|X_*, X_{\mathrm{LF}}, X_{\mathrm{HF}}, \mathbf{y} \sim \mathcal{N}[K_*\mathbf{K}^{-1}\mathbf{y}, K_{22}(X_*, X_*) - K_*\mathbf{K}^{-1}K_*^T], \tag{13}$$

where $f_{*,\mathrm{HF}}$ denotes the posterior distribution of the HF data, and

$$\mathbf{y} \equiv \begin{bmatrix} y_{\mathrm{LF}} \\ y_{\mathrm{HF}} \end{bmatrix}, \tag{14}$$

$$K_* \equiv [K_{21}(X_*, X_{\mathrm{LF}}) \quad K_{22}(X_*, X_{\mathrm{HF}})], \tag{15}$$

$$\mathbf{K} \equiv \begin{bmatrix} K_{11}(X_{\mathrm{LF}}, X_{\mathrm{LF}}) & K_{12}(X_{\mathrm{LF}}, X_{\mathrm{HF}}) \\ K_{21}(X_{\mathrm{HF}}, X_{\mathrm{LF}}) & K_{22}(X_{\mathrm{HF}}, X_{\mathrm{HF}}) \end{bmatrix}. \tag{16}$$

We note that the hyperparameters θ_{LF} and θ_{HF} of the kernels and ρ are all determined *simultaneously* by optimizing the log-likelihood function, as discussed in Refs. [11–13]. From Eq. (13), the MF predictive mean and covariance for the HF data are

$$\boldsymbol{\mu}_{*,\mathrm{HF}}(X_*) = K_*\mathbf{K}^{-1}\mathbf{y}, \tag{17}$$

$$\Sigma_{*,\mathrm{HF}}(X_*) = K_{22}(X_*, X_*) - K_*\mathbf{K}^{-1}K_*^T. \tag{18}$$

Note that when $X_* = X_{\mathrm{HF}}$, we have $\boldsymbol{\mu}_{*,\mathrm{HF}}(X_{\mathrm{HF}}) = y_{\mathrm{HF}}$ [37], which guarantees that the regression will pass through the HF data. This satisfies the constraint imposed in step (c) of the procedure and is due to the independence assumption of f_{LF} and δ_{HF}, as discussed in Ref. [38].

To highlight how the MF-GPR approach given by Eq. (4), which we denote as "linear MF-GPR," may add value over single-fidelity GPR, we consider the pedagogical case where

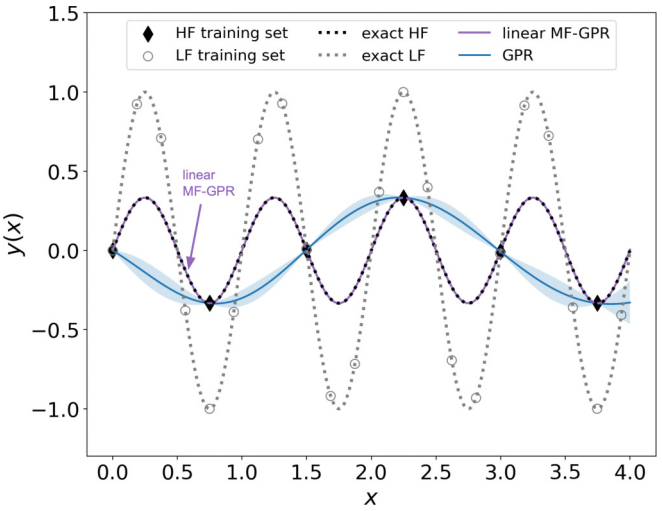

FIG. 2. Comparison of linear MF-GPR and single-fidelity GPR for a linear mapping between fidelities. The shaded bands represent a 95% confidence interval around a fit. The single-fidelity GPR result is shown as a blue line; single-fidelity GPR is used to fit only the HF data and does not recover the exact HF solution. The linear MF-GPR result is shown in purple; linear MF-GPR accurately predicts the exact HF solution by using the LF data in addition to the HF data, and this result overlaps the exact HF solution. The confidence interval for the linear MF-GPR fit is approximately the width of the thickness of the purple line.

the LF and HF models have the form

$$y_{\mathrm{LF}}(x) = \sin(2\pi x), \tag{19}$$

$$y_{\mathrm{HF}}(x) = \tfrac{1}{3}\sin(2\pi x), \tag{20}$$

for $x \in [0, 4]$. Note that the LF and HF models are linearly related by the factor of $1/3$ in Eq. (20). Predictions from single-fidelity GPR and linear MF-GPR are shown in Fig. 2, with $N_{\mathrm{HF}} = 6$ and $N_{\mathrm{LF}} = 22$. For all MF-GPR and GPR fits, the data were first scaled to unit variance and mean zero. The hyperparameter optimization routine was carried out using the limited-memory quasi-Newton algorithm for 15 random restarts, and a measurement noise with a variance of 10^{-6} was added to each kernel matrix to ensure a positive-definite matrix during fitting (see Ref. [39] for more information on the numerical implementation used here). In Fig. 2, the linear MF-GPR fit, denoted by a purple solid line, corresponds to $\boldsymbol{\mu}_{*,\mathrm{HF}}(X_*)$ from Eq. (17); the confidence bands around $\boldsymbol{\mu}_{*,\mathrm{HF}}(X_*)$ were computed from Eq. (18) and are approximately the width of the thickness of the purple line. The GPR fit, denoted by a blue solid line, corresponds to $\boldsymbol{\mu}_*(X_*)$ from Eq. (2), and the shaded confidence bands around $\boldsymbol{\mu}_*(X_*)$ are computed from Eq. (3). We see that inclusion of the LF data leads to a more accurate prediction, as linear MF-GPR recovers the exact HF solution. The GPR result, which is fit to only the HF data, is unable to recover the HF true solution. In addition, the 95% uncertainty band reported in Fig. 2 around the fit is much narrower with linear MF-GPR than with GPR, and the agreement of the linear MF-GPR fit with the HF true solution persists even beyond the last HF data point. It is important to note that all regression methods based on GPR will generate a fit that will regress to the mean of the data when the distance between a new test point and an HF training point is greater than the length-scale of the kernel(s).

This particular example can also be viewed through an information-theoretic lens. Observe that the LF and HF models have the same period of 1 s and therefore, according to the Nyquist-Shannon sampling theorem [40,41], the sampling period must be less than 0.5 s to reconstruct the HF model with sufficient accuracy. Note that the HF data by themselves do not satisfy the Nyquist-Shannon sampling rate. Thus, a GPR fit to the given HF data will be unable to recover the exact HF solution. If the LF model *is* sampled sufficiently to satisfy the Nyquist-Shannon sampling theorem, then it allows the linear MF-GPR model to recover the exact HF solution. If the LF model is not sampled sufficiently or if the LF model has a different frequency than the HF model, then the LF model is uncorrelated with the HF and, therefore, does not add any new information to the MF-GPR.

Lastly, we note that the LF model introduces bias in the resulting MF regression, and the MF-GPR fit is dependent on the choice of LF model; we compare various choices of LF models and their impact on MF-GPR in Sec. III.

Figure 2 illustrates how the autoregressive model Eq. (4) results in more accurate fits to HF data when the LF and HF models are related linearly. However, in many cases, the LF and HF models may be related nonlinearly, and schemes beyond the original MF-GPR approach [11] are needed. In recent years, there have been many improvements to the original MF-GPR approach that explore more efficient numerical schemes [42], transform input data to more accurately predict discontinuities in HF data [12], have the ability to learn a nonlinear mapping between LF and HF models [13], and more accurately propagate uncertainty between fidelity levels [14]. The approach proposed in Ref. [13] goes beyond the linear autoregressive scheme Eq. (4) by allowing for a spatially dependent nonlinear mapping between fidelities; we denote this mapping as $z(\cdot)$.

Following Ref. [13], the modified autoregressive equation that includes this mapping is

$$f_{\mathrm{HF}}(x) = z[x, f_{\mathrm{LF}}(x)] + \delta_{\mathrm{HF}}(x), \tag{21}$$

where $z(\cdot)$ is sampled from of a Gaussian process. Note that $z[x, f_{\mathrm{LF}}(x)]$ is now a Gaussian process of a Gaussian process and is referred to as a "deep GP" [43,44]. While the form of Eq. (21) has been shown to provide improvements over simpler models [43], computing the mean and covariance of the posterior distribution corresponding to Eq. (21) is often computationally intractable [44]. To address this intractability, the Gaussian-process prior $f_{\mathrm{LF}}(x)$ is often replaced with the corresponding posterior distribution $f_{*,\mathrm{LF}}(x)$ [42], resulting in a recursive multifidelity model (i.e., performing GPR at each fidelity level separately and then propagating the results to each successive level of fidelity).

Replacing $f_{\mathrm{LF}}(x)$ with $f_{*,\mathrm{LF}}(x)$ in Eq. (21) and using the independence assumption of $z[x, f_{\mathrm{LF}}(x)]$ and $\delta_{\mathrm{HF}}(x)$ results in a compact recursive multifidelity formulation [13]

$$f_{\mathrm{HF}}(x) = g[x, f_{*,\mathrm{LF}}(x)], \tag{22}$$

where the prior distribution g includes dependencies of both x and $f_{*,\mathrm{LF}}(x)$. It is shown in Ref. [13] that this recursive

multifidelity model Eq. (22) can be modeled by using a kernel of the form

$$k_g(x_i, x_j) = k_\rho(x_i, x_j; \theta_\rho) k_f[f_{*,\mathrm{LF}}(x_i), f_{*,\mathrm{LF}}(x_j); \theta_f] + k_\delta(x_i, x_j; \theta_\delta). \tag{23}$$

In contrast with the linear autoregressive model Eq. (4), the kernel k_ρ is now a spatially-dependent scaling factor responsible for measuring the correlations between the LF and HF models, k_f measures the correlations of the *outputs* of the GPR performed on the LF data, and k_δ accounts for the bias between the LF and HF data; in this work, each term in Eq. (23) is represented by a kernel of the form in Eq. (1).

The set of hyperparameters (variance and length scale) for each kernel is denoted by θ_ρ, θ_f, and θ_δ, respectively. Importantly, unlike the linear autoregressive formulation Eq. (4) where all hyperparameters at all fidelity levels are trained simultaneously, the hyperparameters at each fidelity level using the recursive formulation Eq. (22) are trained separately. This aspect greatly reduces computation costs associated with hyperparameter estimation. When the correlations between the LF and HF data are small, the product $k_\rho k_f$ will be close to zero, and the MF-GPR fit approximately recovers the GPR fit to the HF data. Recall that this was also the case for the correlation hyperparameter ρ in Eq. (4). The product $k_\rho k_f$ in Eq. (23) is plotted in Sec. III to reveal the effectiveness of different choices of LF models.

Next, we turn to the three steps for making an MF prediction using Eq. (22) with kernel Eq. (23). These are discussed in detail in Ref. [13]; for completeness, we summarize them here. Step 1 involves performing GPR on the lowest-fidelity data. This includes optimizing the kernel hyperparameters using the LF data. Step 2 takes as input the trained GPR model from Step 1, together with the HF data, to construct the posterior distribution according to the kernel in Eq. (23) (see Eq. (2.14) of Ref. [13]). The last step, Step 3, calculates the predictive mean and covariance by sampling the posterior distribution using numerical integration techniques (e.g., Monte Carlo [13,14]). Numerical integration is necessary because unlike the prior distributions of single-fidelity GPR and linear MF-GPR, the prior distribution in Eq. (22) may not be Gaussian. As a result, we will be unable to express its posterior distribution as a Gaussian. More details of the MF-GPR approach used in this work and its numerical implementation can be found in Refs. [13,14,39].

Recall that the LF and HF models given by Eqs. (19) and (20) are linearly related, with the quantity in Eq. (20) equal to the quantity in Eq. (19) multiplied by a coefficient of $1/3$. To highlight the limitations of the linear MF-GPR approach given by Eq. (4), we now consider LF and HF models of the form

$$y_{\mathrm{LF}}(x) = \sin(8\pi x), \tag{24}$$

$$y_{\mathrm{HF}}(x) = x \sin(8\pi x), \tag{25}$$

for $x \in [0, 1]$. Note that the coefficient by which Eq. (24) is multiplied to get Eq. (25) has been changed from $1/3$ to x. As a result of this mapping, we expect that predictions made using Eq. (17) will be of poor quality. This expectation is verified in Fig. 3, which shows a comparison between predictions from Eq. (4) and Eq. (22), with $N_{\mathrm{HF}} = 8$ and $N_{\mathrm{LF}} = 30$. We find that MF-GPR not only exhibits excellent agreement with the exact HF solution but also has far smaller confidence bands than those obtained with the linear MF-GPR model. Figure 3 illustrates the ability of MF-GPR to produce accurate results with limited HF data by incorporating additional data from an LF model that is not linearly related to the HF model.

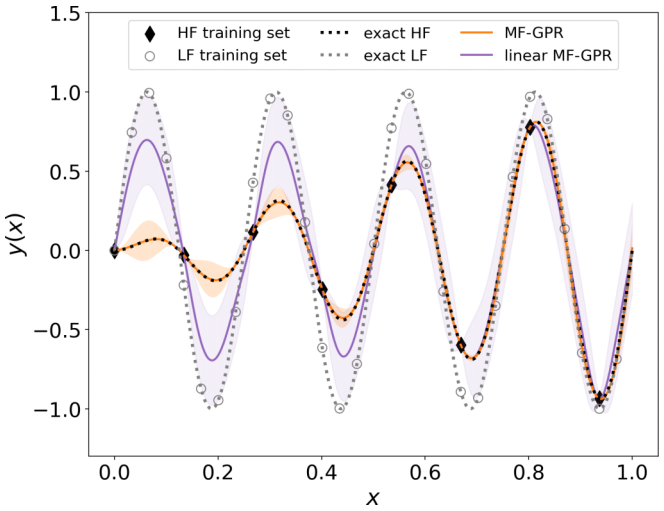

FIG. 3. Comparison of two MF-GPR approaches. One MF-GPR approach assumes a linear relationship between the fidelity levels [see Ref. [11] and Eq. (4)] and is denoted as "linear MF-GPR." The other approach assumes a nonlinear mapping between fidelity levels [see Ref. [13] and Eq. (22)] and is denoted as "MF-GPR"; the shaded bands represent a 95% confidence interval around the fit. The MF-GPR approach that assumes a nonlinear mapping between fidelities (orange solid line) is able to recover the underlying exact HF solution, in contrast to the MF-GPR approach that assumes a linear mapping between fidelity levels (purple solid line).

It would be undesirable to restrict MF-GPR approaches to plasma transport-coefficient data to linear relationships alone, as such data are known or derived to be accurate in certain physical regimes that need not be related linearly. The plasma transport coefficients we are considering illustrate this point; they are obtained using a variety of methods (recall Sec. II A) that have no simple, prescribed relationship to each other. Thus, we will use the nonlinear formulation of MF-GPR Eq. (22) throughout the reminder of this work, referring to it simply as "MF-GPR."

D. Error calculations and computation cost

We have shown the benefit of MF regression over single-fidelity techniques by considering toy examples. However, the computational cost of MF-GPR over single-fidelity GPR can not be disregarded. Thus, we would like to determine the cost-benefit tradeoff for using MF-GPR over single-fidelity GPR. To begin, we define an error metric to measure the regression error between the HF test set and MF-GPR/GPR predictions. The metric we use is the root-mean-square error (RMSE)

$$\mathrm{RMSE} = \sqrt{\frac{1}{N_{\mathrm{test}}} \sum_{i=1}^{N_{\mathrm{test}}} \| y_{i,\mathrm{true}} - y_{i,\mathrm{pred}} \|^2}, \tag{26}$$

TABLE II. Average computation time t and regression errors for single-fidelity GPR and MF-GPR Eq. (22) fits using the LF and HF model Eqs. (24) and (25). Each entry is an average over ten fits, and the hyperparameters for each fit were trained using the limited-memory quasi-Newton algorithm with 15 random restarts. For the RMSE values, the numbers in brackets denote the power of ten that the value in front of the brackets is multiplied by (e.g., $3.2[-1] = 0.32$). The column labeled t^*_{GPR} shows the computation time for single-fidelity GPR normalized by the computation time when $N_{\text{HF}} = 8$. The computational cost of single-fidelity GPR increases by a factor of two when the number of HF training points increases by roughly ten. We note that when $N_{\text{LF}} = 50$ and $N_{\text{HF}} = 13$, MF-GPR is six times more expensive than single-fidelity GPR but reduces the regression error by more than two orders of magnitude.

N_{LF}	N_{HF}	t^*_{GPR}	$t_{\text{MF-GPR}}/t_{\text{GPR}}$	RMSE_{GPR}	$\text{RMSE}_{\text{MF-GPR}}$
30	8	1	6 ± 1	3.2[−1]	8.7[−3]
34	9	1.0 ± 0.1	6 ± 1	3.5[−1]	1.5[−2]
38	10	1.0 ± 0.1	5 ± 1	3.8[−1]	9.4[−4]
43	11	1.3 ± 0.7	6 ± 3	3.5[−1]	1.1[−3]
50	13	1.1 ± 0.1	6 ± 1	3.6[−1]	2.0[−3]
60	15	1.1 ± 0.2	8 ± 2	1.1[−1]	1.9[−4]
75	19	1.4 ± 0.2	7 ± 2	9.0[−3]	1.5[−4]
100	25	1.2 ± 0.2	8 ± 2	5.4[−3]	9.9[−5]
150	38	1.7 ± 0.3	7 ± 2	3.1[−4]	7.7[−5]
300	75	2.1 ± 0.3	10 ± 2	2.1[−4]	6.5[−5]

where i denotes the location of a test point, N_{test} is the total number of test points, $y_{i,\text{true}}$ is the true solution at location i, and $y_{i,\text{pred}}$ is the value of the fit (MF-GPR or GPR) at location i. Table II compares the computational costs, which includes the costs of both hyperparameter training and predictions, and regression errors for the GPR and MF-GPR methods using the LF and HF model Eqs. (24) and (25). We find that while MF-GPR is roughly six to ten times more expensive than single-fidelity GPR, the MF-GPR method results in regression errors that are often a couple orders of magnitude lower than those obtained with single-fidelity GPR.

III. MULTIFIDELITY REGRESSION OF PLASMA TRANSPORT-COEFFICIENT DATA

In Secs. II B and II C, we have demonstrated the effectiveness and limitations of single-fidelity GPR and different MF-GPR approaches using toy examples. Additionally, in Sec. II D, we assessed the cost-benefit tradeoff between GPR and MF-GPR approaches. We illustrated the fact that relative to single-fidelity GPR, MF-GPR increases computation cost but decreases prediction error. While these toy examples were useful for building intuition and providing a baseline for computation-cost and error estimates, we will now consider real data generated for ionic plasma transport coefficients; we will begin by analyzing what role the choice of LF model plays in MF-GPR. The LF and HF models will be chosen from those listed in Table I.

We first consider two choices for the LF model for predicting the viscosity for the element C at $n_i = 5.01 \times 10^{22}$ cm^{-3}, as shown in Fig. 4. MF-GPR fits produced using the LF SMT model are shown in Fig. 4(a), and fits produced using the LF YGBI model are shown in Fig. 4(b). The HF training data were computed from the YVM. The inserts in Fig. 4 show the kernel matrix corresponding to $k_\rho k_f$ in Eq. (23).

In Fig. 4(a), with the SMT model used as the LF model, we see that the only nonzero values of the kernel matrix occupy the diagonal and quickly decay to zero a short distance from the diagonal, corresponding to a small length scale for the kernel. Because the entries of the kernel matrix have nearly zero magnitudes, the MF-GPR fit is nearly equivalent to the fit obtained by performing GPR on the HF data alone; this equivalence explains the overlap of the fits produced by GPR and MF-GPR.

In Fig. 4(b), with the YGBI model used as the LF model, two findings are of note. The first is that there are regions where the MF-GPR and GPR fits do not overlap; this is most clearly seen around $T = 0.2$ eV. Second, the entries of the kernel matrix are nonzero away from the diagonal, implying substantial correlations between the LF and HF data. However, the values are nearly constant throughout the matrix, differing from each other by at most by 1%. Thus, in contrast with the MF-GPR fit shown in Fig. 4(a), the MF-GPR fit shown in Fig. 4(b) includes information from the LF data and suggests correctly that the LF and HF data differ by an approximately constant shift.

A comparison of the sizes of the confidence bands for the MF-GPR results in Figs. 4(a) and 4(b) show that the MF-GPR fit in Fig. 4(b) is superior to that in Fig. 4(a). The choice of the YGBI model as the LF model for MF-GPR in Fig. 4(b) results in a superior fit because the YGBI model provides additional information that is used to improve the fit. This additional information can be seen in the kernel matrix computed from $k_\rho k_f$; an LF model for which kernel entries off the diagonal are nonzero improved the MF-GPR fit over the GPR fit more than an LF model for which the kernel entries are close to zero. Thus, we have found that the kernel matrix computed from $k_\rho k_f$ is a natural indicator of when an LF model is insufficient for MF-GPR and that a different, or more precise, LF model is needed to impact the MF-GPR fit. When kernel matrix entries decay rapidly to zero off the diagonal, it would be best to consider alternative LF models.

In Fig. 4, we considered LF and HF models that both predict the same quantity. ML models have been developed in which the LF and HF models do not predict the same quantity; for example, the prediction of rainfall using an elevation model has been examined [45,46]. As discussed in Refs. [45,46], a large amount of elevation data are available,

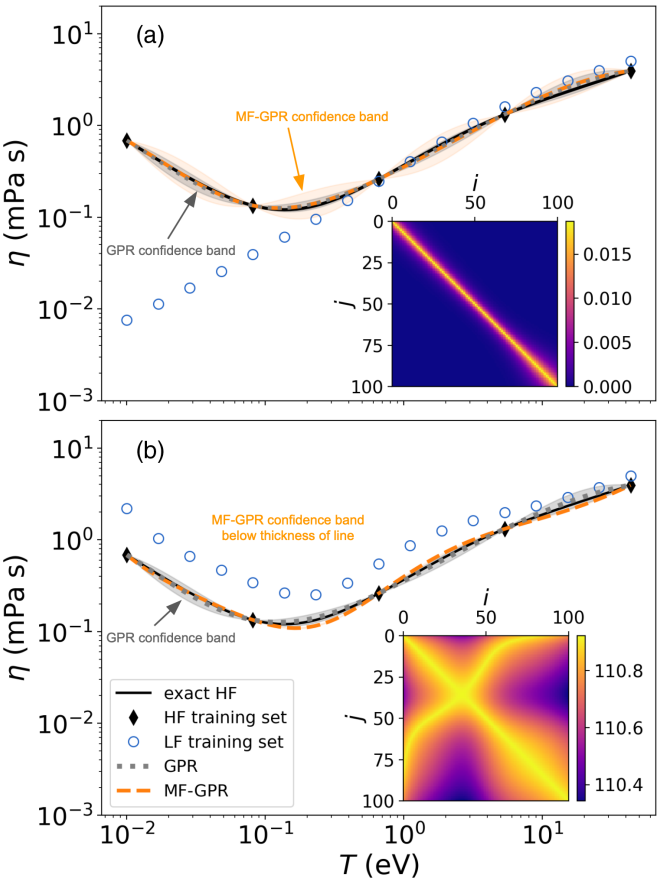

FIG. 4. MF-GPR prediction of the viscosity of the element C at $n_i = 5.01 \times 10^{22}$ cm^{-3} versus temperature. In both panels (a) and (b), GPR was performed using the HF training data computed from the YVM, and the GPR results are compared with those of an MF-GPR model constructed using data from both an HF model and an LF model. (a) The LF model is given by the SMT model. (b) The LF model is given by the YGBI model. The inserts in panels (a) and (b) display the kernel matrix from $k_\rho k_f$ with optimized hyperparameters. In panel (a), little correlation is found between the HF and LF models, as the only nonzero entries of the kernel matrix are on, or close to, the diagonal; in panel (b), however, the correlation is substantial, as demonstrated by the extent of the nonzero values off the diagonal of the kernel matrix, as shown in the insert.

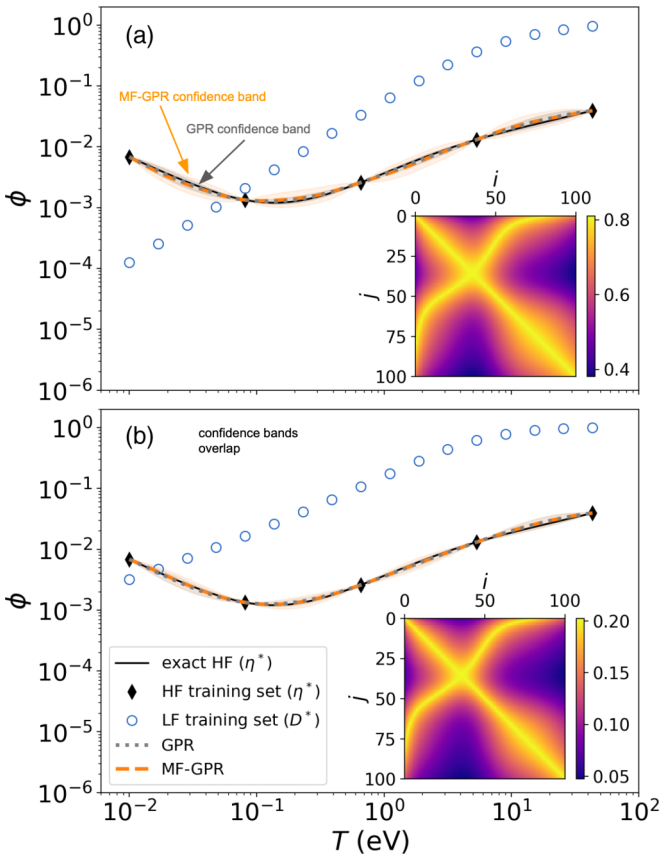

FIG. 5. Using self diffusion as the LF model to predict viscosity. The reduced transport coefficients $\phi \in \{D^*, \eta^*\}$ are shown for the element C at $n_i = 5.01 \times 10^{22}$ cm^{-3} versus temperature. In both panels (a) and (b), GPR was performed using the HF training data computed from the YVM, and the GPR results are compared with those of an MF-GPR model constructed using data from both an HF model and an LF model. (a) The LF model is the reduced self-diffusion coefficient D^* from the HMP model. (b) The LF model is D^* computed from the SMT model. The inserts display the kernel matrix from $k_\rho k_f$ with optimized hyperparameters.

but only a minimal amount of rainfall data are available; together, these data have been used to construct MF rainfall models. Similarly, a large amount of self-diffusion coefficient data and a minimal amount of viscosity data are available, and MF models of plasma transport coefficients could be constructed using both data sources. Thus, we also consider LF and HF models that do not predict the same quantity. In particular, we assess the validity of using the self-diffusion coefficient (LF model) as a predictor for the viscosity (HF model).

MF-GPR fits for viscosity of the element C at $n_i = 5.01 \times 10^{22}$ cm^{-3} using self-diffusion data for the LF model and viscosity data for the HF model are shown in Fig. 5; two different LF models for predicting the self-diffusion coefficient are considered. In both Figs. 5(a) and 5(b), HF data were calculated from the YVM model. In Fig. 5(a), the LF data were computed using the HMP model, and in Fig. 5(b), the LF data were computed from the SMT model. The transport coefficients have been reduced such that $D^* = D/\omega_p a_i^2$ and $\eta^* = \eta/m_i n_i \omega_p a_i^2$. Here, $\omega_p = (4\pi n_i Z^2 e^2/m_i)^{1/2}$ is the ion plasma frequency, and $a_i = (4\pi n_i/3)^{-1/3}$ is the ion-sphere radius, where n_i is the ion number density, Z is the mean ionization state, e is the elementary charge, and m_i is the ion mass. The inserts once again show the kernel matrix $k_\rho k_f$.

Figure 5 demonstrates that using self-diffusion coefficient data as our LF model and viscosity as our HF model substantially improves the MF-GPR model of the viscosity compared to using viscosity data for both models. What we mean by this is that the LF data used in both panels of Fig. 5 are more strongly correlated with the HF data than the LF data used in Fig. 4(a) are. A comparison between the kernel matrices shown in the insert of Fig. 4(a) and in the insert of both panels of Fig. 5 demonstrates this point; in contrast to Fig. 4(a), the kernel matrices shown in both panels of Fig. 5 have a nonzero value away from the diagonal. This means that the LF data used in both panels of Fig. 5 have a larger contribution to the

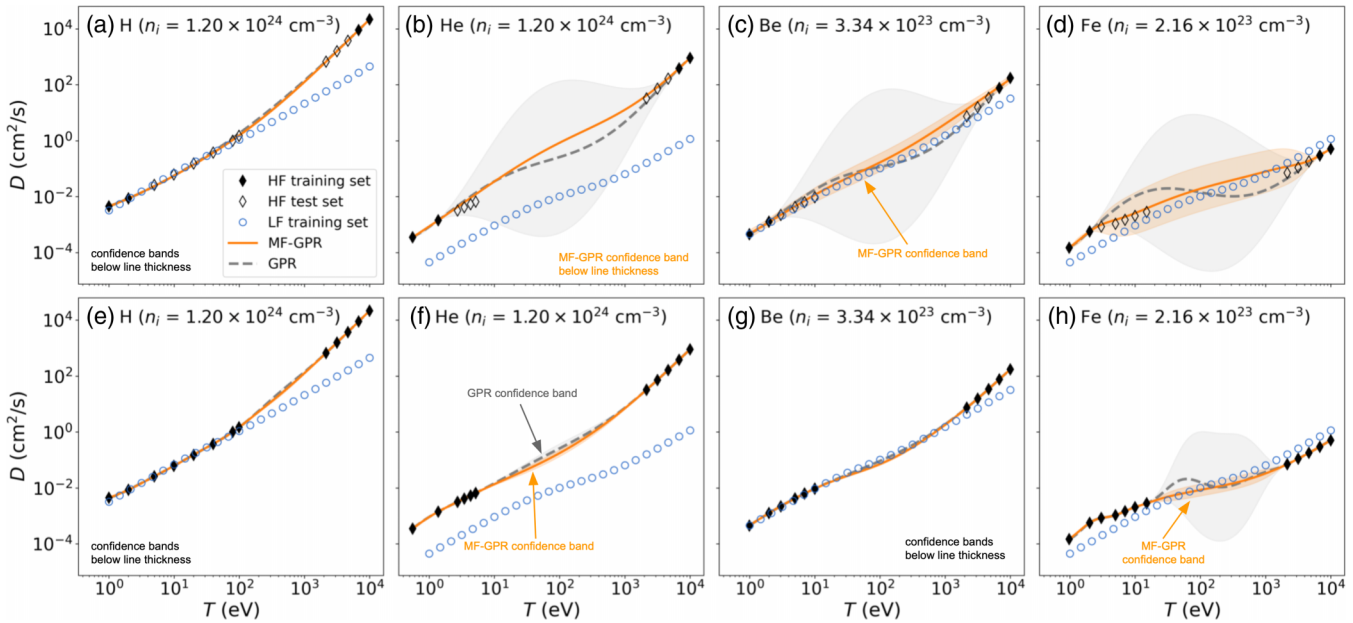

FIG. 6. MF-GPR and GPR fits, with 95% confidence intervals (shaded bands), of the self-diffusion coefficient versus temperature for multiple elements. The models used to generate this data are given in Table I. Panels (a)–(d) show MF-GPR and GPR fits obtained using a portion (filled diamonds) of the HF data (all diamonds). Panels (e)–(h) compare MF-GPR and GPR fits obtained using all of the available data; GPR is fit to only HF data, whereas MF-GPR uses both the LF and HF data. In general, the MF-GPR fit is less prone to spurious oscillations than the GPR fit, and the size of the uncertainty band is much smaller with MF-GPR than with GPR.

MF-GPR model than the LF data used in Fig. 4(a) does. Also note that the entries of the kernel matrix in both panels of Fig. 5 are not a constant value, in contrast with the entries in the kernel matrix shown in the insert in Fig. 4(b). Therefore, the LF and HF data used in both panels of Fig. 5 are not related by a shift but rather by a nonlinear relationship. Comparisons of the kernel matrix $k_\rho k_f$ provide valuable insight into the effectiveness of an LF model in an MF-GPR framework by quantifying the spatial extent of correlations and type of relationship between the low- and high-fidelity models e.g., linear or nonlinear. In particular, these comparisons revealed the effectiveness of using self-diffusion LF data to predict viscosity HF data. As self-diffusion data are more readily available and are cheaper to compute than viscosity data, Fig. 5 illustrates how MF-GPR provides improved estimates of viscosity at low computational cost where it has not been measured.

In addition to selecting a sufficient LF model for MF-GPR, it is imperative to include data in the LF and HF datasets that capture essential special features of a physical system. For example, it is possible that neither the LF data nor the HF data include information about features such as sudden changes (i.e., a jump discontinuity). For plasma transport coefficient data, sudden changes in quantities such as the electrical conductivity may result from a phase transition. In the absence of such data, MF-GPR is incapable of predicting a discontinuity. If this behavior is known in advance, then the LF and HF models should be sampled accordingly to ensure that the MF-GPR framework has sufficient training data near the discontinuity; then, an MF-GPR approach capable of handling a discontinuity, such as that described in Ref. [12], can be used.

IV. REGRESSION OF SPARSE DISPARATE DATA

In this section, we will use MF-GPR to predict transport coefficients when HF data are available in disparate physical regimes. We will consider a transport-coefficient dataset, which has "gap" regions, i.e., temperature ranges in which no HF data are available, as shown in Table I.

This section is organized as follows. First, we use MF-GPR to fit gapped transport-coefficient data as a function of temperature. Then, we consider a higher-dimensional feature space of ion number density and temperature. We conclude by varying the approach used to sample the HF dataset. We find that using a low-discrepancy sequence [47] to select data-sampling locations yields smaller regression errors than does sampling data on a uniform grid.

A. Self-diffusion and viscosity predictions versus temperature

We apply MF-GPR to gapped transport-coefficient data for the elements H, He, Be, and Fe. We first consider an HF training set that consists of only four data points—two points at both high and low temperatures—and thus features a large gap between the patches of HF data. Then, this gap is reduced in size by including all HF points in the training set.

This approach is illustrated in Fig. 6, which shows MF-GPR and GPR fits for the self-diffusion coefficient. In the top row, we note that multiple inflection points in the GPR predictions for He, Be, and Fe can be seen, while the MF-GPR fits are monotonically increasing. For Fe, a large oscillatory pattern is seen in the GPR fit. These oscillations are not physical and are likely due to the hyperparameters responsible for specifying the length scale of the kernel Eq. (1). With MF-GPR, oscillations do not appear, as the three terms in

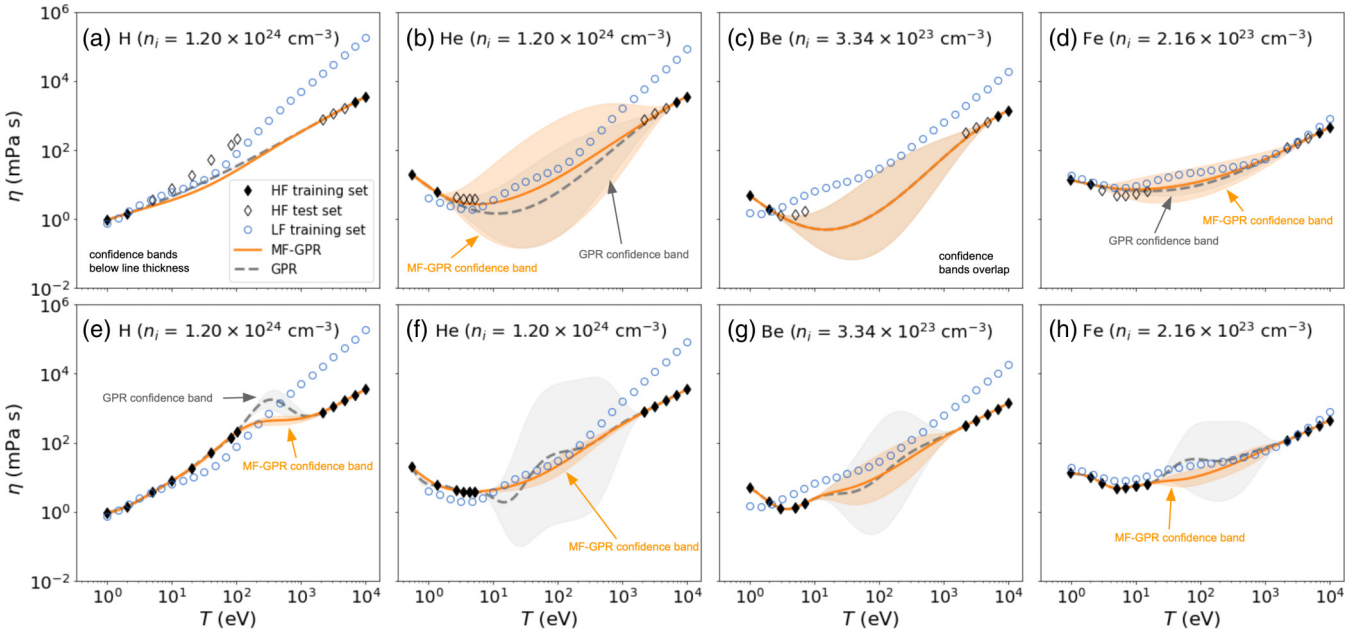

FIG. 7. MF-GPR and GPR fits, with 95% confidence intervals (shaded bands), of the viscosity coefficient versus temperature for multiple elements. The models used to generate this data are given in Table I. Panels (a)–(d) show MF-GPR/GPR fits obtained using a portion (filled diamonds) of the HF data (all diamonds). Panels (e)–(h) show MF-GPR fits obtained using all of the data. In general, the MF-GPR fit is less prone to oscillations than is the GPR fit which uses only HF data.

Eq. (23) do not restrict the form of the fit to a single length scale. Similar patterns are observed for the viscosity in Fig. 7.

B. Viscosity predictions versus temperature and number density

Because only a small amount of HF data were used in the work described in Sec. IV A, a well-defined error metric could not be reported. Therefore, we constructed an HF dataset in the n_i-T plane containing 900 points sampled on a grid. The data were generated using the YVM for H and Fe, and these data will act as a test set for the results described in this section. The dataset spans a temperature range of $T = 10^1$–10^4 eV and an ion number density of $n_i = 10^{18}$–10^{26} cm^{-3}.

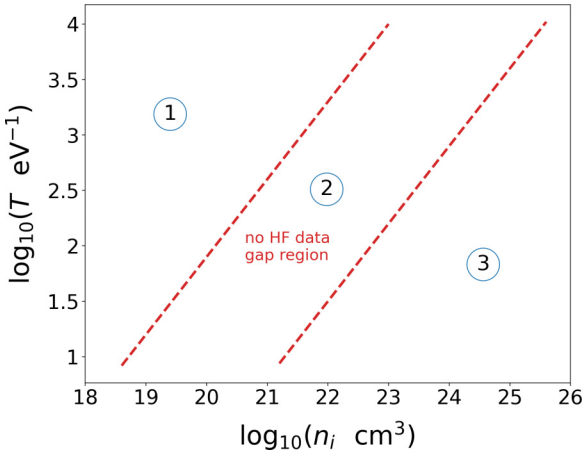

FIG. 8. The regions of the temperature and number-density space where HF and LF models were used to generate viscosity data for the MF training dataset. The red dashed lines indicate the divisions between the regions. In the regions labeled as "1" and "3," both HF and LF data are available. In the region labeled as "2," only LF data are available. The models used for each region are listed in Table III.

Next, we constructed an MF training dataset. When this MF training dataset is used together with the HF test set described above, we will be able to compute regression errors for GPR and MF-GPR using Eq. (26), now in n_i-T space. With the view of mimicking the scenario of datasets containing "gaps," as discussed in Sec. IV A, an MF training dataset was constructed to contain a region lacking HF data. We chose the YGBI model as our LF model and assumed that this LF model can be evaluated everywhere in the domain.

Figure 8 illustrates the concept of physical regimes that include an area or "gap" in which no HF data exist. The figure shows three regions, labeled as "1," "2," and "3." In regions 1 and 3, both the HF (YVM) and LF (YGBI) models can be evaluated. The area between the red dashed lines, denoted with a 2 and labeled as "no HF data," shows the region where no HF data are available. We refer to this region as the "gap

TABLE III. HF and LF models for the viscosity used in the temperature/number-density regions shown in Fig. 8. The same LF model is employed across all regions.

Region	HF	LF
1	YVM [31]	YGBI [32]
2	–	YGBI [32]
3	YVM [31]	YGBI [32]

TABLE IV. Sampling approaches used to sample the MF training dataset. The LF data were always evaluated on a grid; sampling methods for the HF data varied.

HF sampling	LF sampling	Description
Grid	Grid	HF and LF data were sampled on an evenly spaced grid in n_i and T.
Halton-23 [48]	Grid	HF data were sampled using a Halton-23 sequence. LF data were sampled on an evenly spaced grid in n_i and T.
Hybrid	Grid	The HF dataset includes the four extreme corners of the domain and data sampled using a Halton-23 sequence. The LF data were sampled on an evenly spaced grid in n_i and T.

region." A summary of the choices of LF and HF models for all of the regions shown in the figure are given in Table III.

Having defined the models used to generate the test and training datasets, we will describe, below in Sec. IV B 1, the three HF sampling approaches we used to create the MF training dataset.

1. Sampling methods for HF data

We used three approaches to sample the HF gapped dataset initially: an evenly spaced grid, a low-discrepancy sequence, namely, a Halton-23 sequence [48], and a hybrid method that used both approaches. For the LF data, we restricted the sampling approach to an evenly spaced grid. The details of each sampling approach are discussed below and summarized in Table IV.

To place data on an evenly spaced grid, we first specify the total number of HF data points (e.g., $N_{\mathrm{HF}} = 100$). Then, the grid spacing is computed by

$$\Delta x = \frac{x_u - x_\ell}{\sqrt{N_{\mathrm{HF}}}}, \tag{27}$$

where $x \in \{n_i, T\}$, and the subscripts "u" and "ℓ" denote the upper and lower bounds of x, respectively. Using Eq. (27) to determine the spacing between HF points is straightforward; however, we note that to refine the grid spacing by a factor of two, four times as many HF data points are needed. As a result, the evenly spaced grid approach becomes increasingly

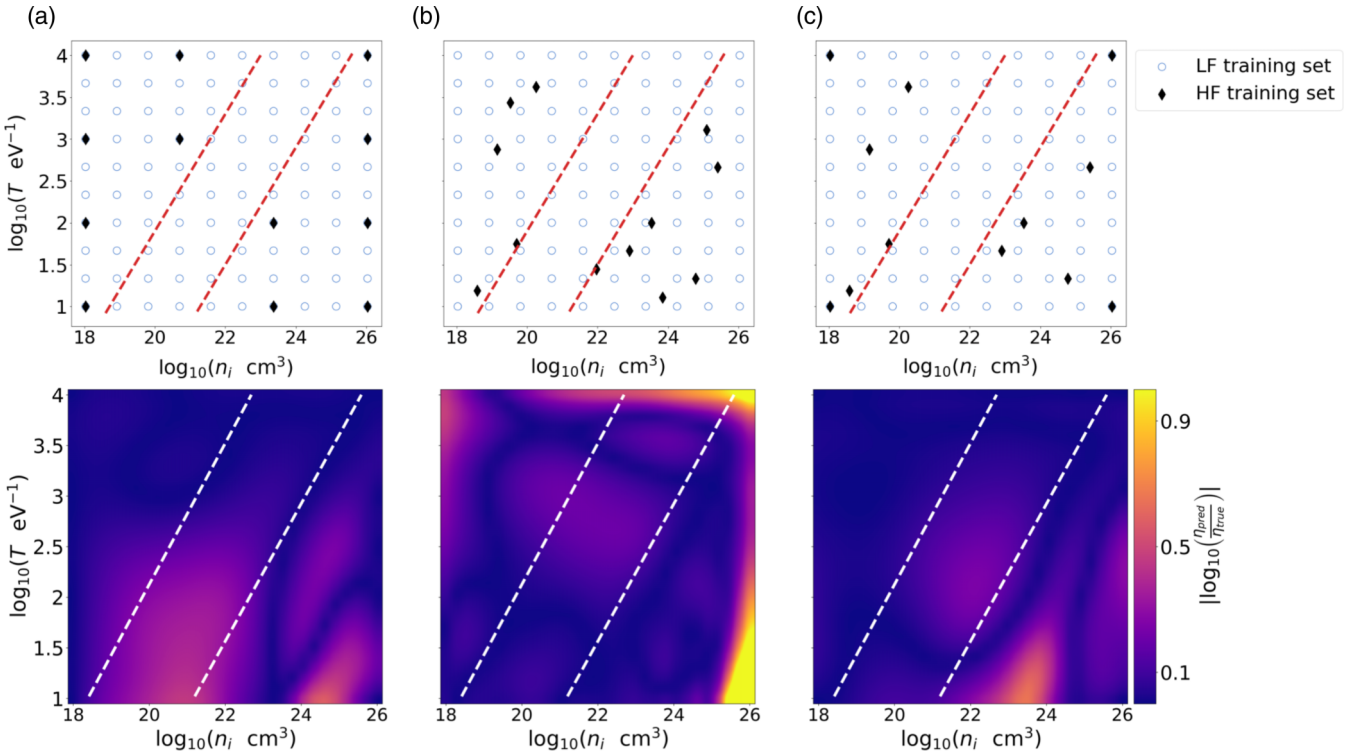

FIG. 9. MF-GPR prediction of the viscosity of the element Fe using $N_{\mathrm{LF}} = 100$ and $N_{\mathrm{HF}} = 12$. The HF data were sampled (a) on a uniformly spaced grid, (b) using a Halton-23 sequence, and (c) using our hybrid method. Top row: The locations of the HF training data (filled black diamonds) and LF training data (open blue circles) used to construct the MF training dataset are shown. The red dashed lines denote the boundaries between the regions shown in Fig. 8. Bottom row: The absolute differences between the predicted viscosities η_{pred} and the true viscosities η_{true} are shown. The hybrid sampling approach improves the prediction in the gap region between the dashed white lines. Note that the failure of the Halton-23 sampling approach to include the boundaries of the HF data in the training set results in large errors at the boundaries.

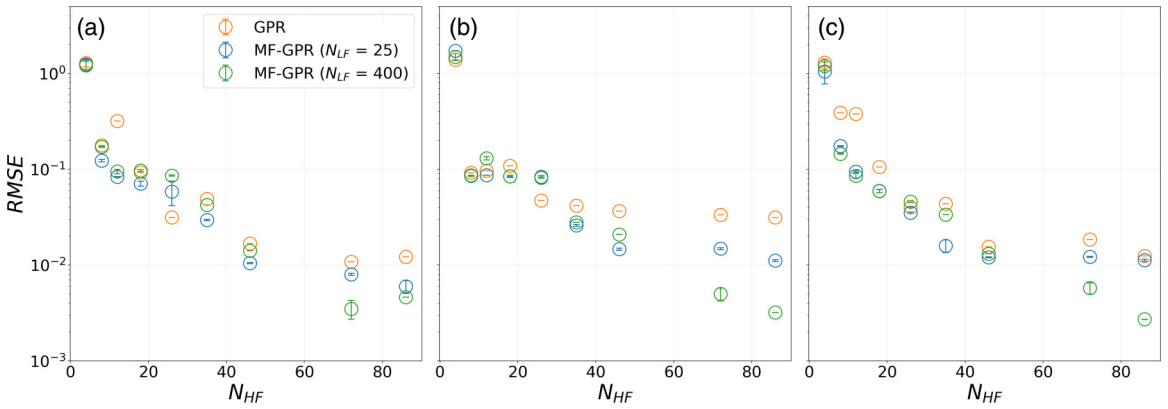

FIG. 10. The RMSE of $\log_{10}(\eta)$ for GPR and MF-GPR fits for the element H using different HF sampling methods. We sampled N_{HF} and N_{LF} points from the gapped dataset shown in Fig. 8. The models used to generate the data are specified in Table III. Each RMSE value was determined from an average of ten fits, and standard deviations for the values are shown as error bars. The HF data were sampled (a) on an evenly spaced grid, (b) using a Halton-23 sequence, and (c) using our hybrid method. We note that in most cases, MF-GPR outperformed GPR.

computationally expensive as the dimension of the input space increases.

Instead of restricting the locations of HF data to points on an evenly spaced grid, their locations may be determined randomly. However, two HF points chosen in this way could be extremely close together, and in such a circumstance, calculations would be repeated at roughly the same location in parameter space. By enforcing a constraint on the minimum distance between two points, calculating HF data at close locations can be avoided. An alternative to enforcing a constraint on the distance between data sampling locations is to use a low-discrepancy sequence to determine sampling locations; this is the second of our sampling methods. Low-discrepancy sequences consist of "quasirandom" numbers that are generated deterministically, and points constructed using these numbers as coordinates cover a domain more quickly and evenly than do points constructed with random numbers as coordinates. Here, we use a Halton-23 low-discrepancy sequence [48]. In the name "Halton-23," "23" denotes the bases 2 (for dimension n_i) and 3 (for dimension T); the bases 2 and 3 were chosen as they are mutually prime, which results in a uniform, limiting density of the points in the sequence [48].

While use of only a low-discrepancy sequence to determine HF sampling locations reduces the chance of performing repeated calculations, the edges of the domain may not be included in an HF dataset constructed in this way. If a specific domain is desired, then it is necessary to augment the low-discrepancy sequence locations with data along the domain boundary. To ensure coverage in a fixed domain, we used a hybrid sampling method. In this hybrid method, the four extreme corners of the domain of the HF dataset are sampled first. Then, the remainder of the allocated HF data points are sampled using a low-discrepancy sequence.

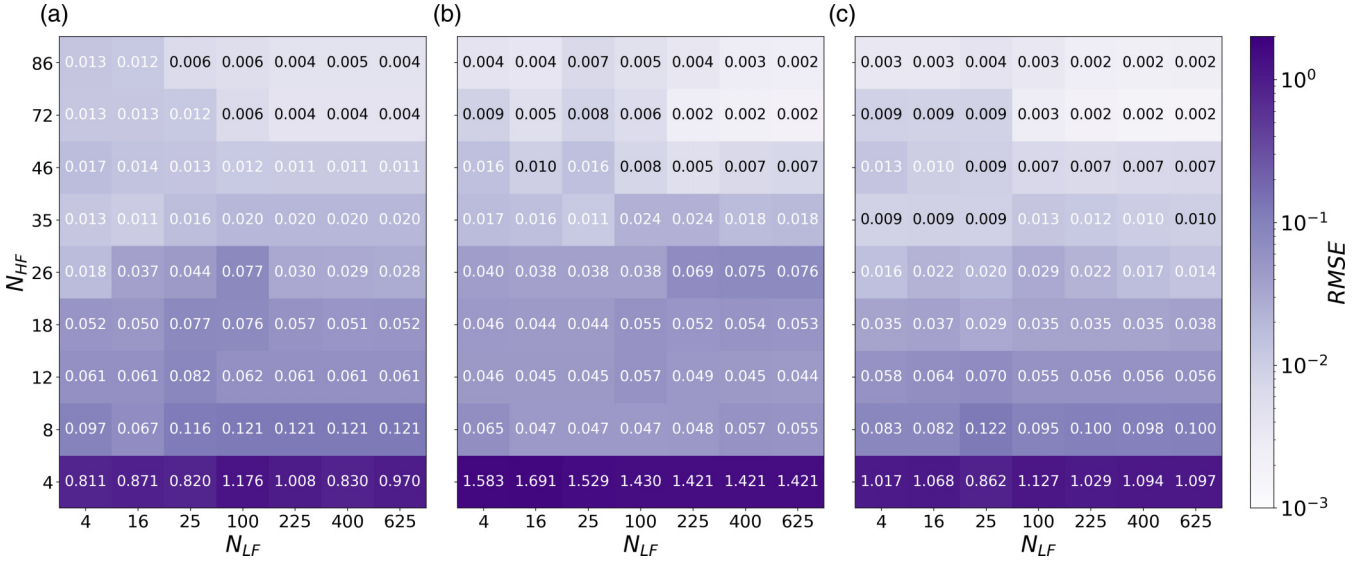

FIG. 11. The RMSE of $\log_{10}(\eta)$ for the element H using MF-GPR with different MF training sets constructed using various HF sampling approaches. (a) The HF data were sampled on a grid. (b) The HF data were sampled using a Halton-23 sequence. (c) The HF data were sampled using our hybrid approach.

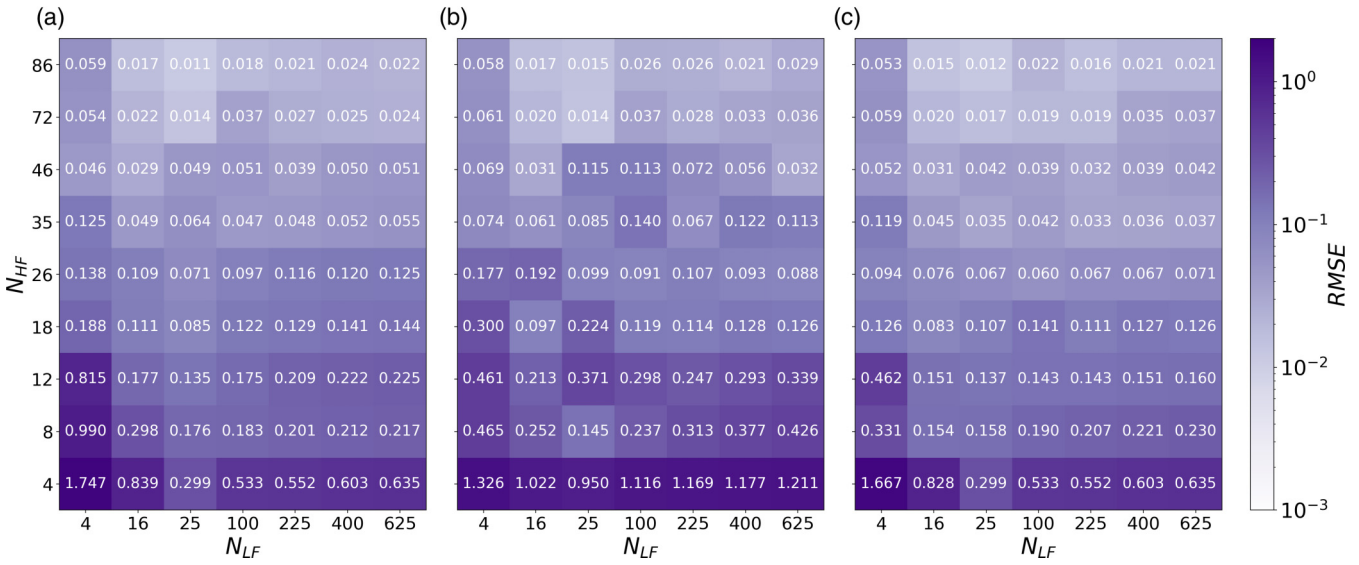

FIG. 12. The RMSE of $\log_{10}(\eta)$ for the element Fe using MF-GPR with different MF training sets constructed using various HF sampling approaches. (a) The HF data were sampled on a grid. (b) The HF data were sampled using a Halton-23 sequence. (c) The HF data were sampled using our hybrid approach.

The three sampling approaches we used are summarized in Table IV. In Fig. 9, we compare the MF-GPR prediction of the viscosity using each of these sampling methods, for $N_{\mathrm{LF}} = 100$ and $N_{\mathrm{HF}} = 12$. In the top row, we show the locations in the n_i-T plane where the HF data, indicated by filled black diamonds, and the LF data, indicated by open blue circles, were sampled using each method. The bottom row shows heat maps of the absolute error between the prediction and the true solution, for each sampling method; differences between the sampling methods are apparent. In particular, the regression error in the gap region is substantially smaller with the hybrid method than with the grid method.

2. *Regression error*

Fits produced using GPR and MF-GPR are shown in Fig. 10, with the HF sampling approach varied as described in Table IV. Each point in the figure shows an average of 10 fits, with error bars indicating one standard deviation from the average. For MF-GPR, the LF data were sampled from a grid, and the cases $N_{\mathrm{LF}} = 25$ and 400 are shown. The GPR fits were carried out using only the HF data from the MF dataset. We see that *the RMSE decreases as N_{HF} increases* for all methods, and that *the MF-GPR fit yields smaller RMSE values than does GPR*. In almost all cases, MF-GPR performs at least as well as GPR.

We next computed the RMSE of fits to HF viscosity data for different combinations of N_{HF} and N_{LF} for H and Fe. The results for H are displayed in Fig. 11, and for Fe, they are displayed in Fig. 12; each column in the figure corresponds to a different HF sampling method. We note that the RMSE values for $N_{\mathrm{HF}} = N_{\mathrm{LF}} = 4$ should be the same in columns (a) and (c), because the hybrid method first samples the four corners from the grid and then adds points sampled using the Halton-23 sequence. The average value of the RMSE for $N_{\mathrm{HF}} = N_{\mathrm{LF}} = 4$ in column (a) is within one standard deviation of the average value of the RMSE for the same case in column (c), and vice versa. Therefore, we do not consider these differences to be statistically significant. As shown in Figs. 11 and 12, fits generated using the hybrid sampling approach result in smaller RMSE values overall than do those generated using a simple grid approach. It is also worth noting that the pure Halton-23 method often produced higher RMSE values than did the grid method; this is because the boundaries of the domain were not sufficiently sampled in the HF training set. As a result, the MF-GPR fit tends to the mean of the HF data, and the largest errors are incurred near the boundaries, as shown in Fig. 9.

V. CONCLUSIONS AND OUTLOOK

We have investigated the use of MF-GPR to interpolate plasma transport data over a wide parameter space in which HF data are available in localized patches. We have examined the improvements in both the predicted mean and the predicted uncertainty that MF-GPR provides over GPR. We have seen that in most cases, MF-GPR results in a lower uncertainty than does single-fidelity GPR, sometimes by an order of magnitude. Examining the hyperparameters governing the structure of the $k_\rho k_f$ kernel reveals the improvement in the mean and uncertainty, or lack thereof, given by the LF data.

As a "black-box" regression method, MF-GPR provides increased reliability over single-fidelity methods, as trends from LF models are used during regression where HF data are sparse; the use of such LF trends enables MF-GPR to reduce the occurrence of nonphysical oscillations or inflection points that occur with single-fidelity GPR. In addition, confidence bands generated by MF-GPR and GPR suggest where additional HF data are needed once a fit has been produced; simpler regression methods do not offer this benefit.

From an experimental-design perspective, HF data are often sampled on a grid that is refined uniformly when finer resolution is needed [49]. We found that when performing MF-GPR, sampling HF data on a uniformly spaced grid can

bias length-scale hyperparameters and results in larger regression errors. Therefore, we developed a simple hybrid approach for initially sampling HF data that combines sampling both on a grid and using a low-discrepancy sequence, resulting in smaller regression errors.

The results here can be expanded upon in multiple ways. For example, the MF-GPR framework could be extended to include physically motivated constraints, such as enforcing nonnegativity [50]. Additionally, we restricted the work here to the self-diffusion and viscosity transport coefficients, but other transport coefficients, such as the thermal conductivity, the resistivity, and the interdiffusion coefficient in plasma mixtures, can also be investigated. The sampling methods described here can also be improved upon greatly and optimized for higher-dimensional feature spaces to avoid the curse of dimensionality. However, our approaches offer a starting point that highlights the importance of avoiding regressing beyond the bounds of available data in a GPR/MF-GPR setting.

Through the confidence intervals it provides, the GPR approach suggests where it would be most useful to generate additional data; the confidence of a fit would be improved most by obtaining additional HF data points in regions with the greatest uncertainties. Comparing GPR and MF-GPR results show the utility of generating LF data in parallel with HF datasets. In addition, it could be possible to improve the ML approach itself by developing customized kernels for this application [51–53].

ACKNOWLEDGMENTS

The authors thank Lisa Murillo for her careful language editing of the manuscript. L.J.S. and M.S.M. acknowledge support from the U.S. Air Force Office of Scientific Research Grant No. FA9550-17-1-0394.

[1] A. Dewaele, M. Mezouar, N. Guignot, and P. Loubeyre, High Melting Points of Tantalum in a Laser-Heated Diamond Anvil Cell, Phys. Rev. Lett. **104**, 255701 (2010).

[2] Y. Li, D. J. Siegel, J. B. Adams, and X.-Y. Liu, Embedded-atom-method tantalum potential developed by the force-matching method, Phys. Rev. B **67**, 125101 (2003).

[3] P. Grabowski, S. Hansen, M. Murillo, L. Stanton, F. Graziani, A. Zylstra, S. Baalrud, P. Arnault, A. Baczewski, L. Benedict, C. Blancard, O. Čertík, J. Clérouin, L. Collins, S. Copeland, A. Correa, J. Dai, J. Daligault, M. Desjarlais, M. Dharma-wardana *et al.*, Review of the first charged-particle transport coefficient comparison workshop, High Energy Density Phys. **37**, 100905 (2020).

[4] J. Gaffney, S. Hu, P. Arnault, A. Becker, L. Benedict, T. Boehly, P. Celliers, D. Ceperley, O. Čertík, J. Clérouin, G. Collins, L. Collins, J.-F. Danel, N. Desbiens, M. Dharma-wardana, Y. Ding, A. Fernandez-Pañella, M. Gregor, P. Grabowski, S. Hamel *et al.*, A review of equation-of-state models for inertial confinement fusion materials, High Energy Density Phys. **28**, 7 (2018).

[5] M. S. Murillo, J. Weisheit, S. B. Hansen, and M. W. C. Dharma-wardana, Partial ionization in dense plasmas: Comparisons among average-atom density functional models, Phys. Rev. E **87**, 063113 (2013).

[6] L. J. Stanek, R. C. Clay, M. W. C. Dharma-wardana, M. A. Wood, K. R. C. Beckwith, and M. S. Murillo, Efficacy of the radial pair potential approximation for molecular dynamics simulations of dense plasmas, Phys. Plasmas **28**, 032706 (2021).

[7] E. Bélisle, Z. Huang, S. Le Digabel, and A. E. Gheribi, Evaluation of machine learning interpolation techniques for prediction of physical properties, Comput. Mater. Sci. **98**, 170 (2015).

[8] M. A. Shandiz and R. Gauvin, Application of machine learning methods for the prediction of crystal system of cathode materials in lithium-ion batteries, Comput. Mater. Sci. **117**, 270 (2016).

[9] J. Schmidt, M. R. Marques, S. Botti, and M. A. Marques, Recent advances and applications of machine learning in solid-state materials science, npj Comput. Mater. **5**, 1 (2019).

[10] C. E. Rasmussen, Gaussian processes in machine learning, in *Summer School on Machine Learning* (Springer, Berlin, 2003), pp. 63–71.

[11] M. C. Kennedy and A. O'Hagan, Predicting the output from a complex computer code when fast approximations are available, Biometrika **87**, 1 (2000).

[12] M. Raissi and G. Karniadakis, Deep multifidelity Gaussian processes, arXiv:1604.07484 (2016).

[13] P. Perdikaris, M. Raissi, A. Damianou, N. D. Lawrence, and G. E. Karniadakis, Nonlinear information fusion algorithms for data-efficient multifidelity modelling, Proc. R. Soc. A **473**, 20160751 (2017).

[14] K. Cutajar, M. Pullin, A. Damianou, N. Lawrence, and J. González, Deep Gaussian Processes for Multi-fidelity Modeling (2019), arXiv:1903.07320.

[15] M. G. Fernández-Godino, C. Park, N.-H. Kim, and R. T. Haftka, Review of multifidelity models, AIAA **57**, 1786 (2019).

[16] A. I. J. Forrester, A. Sóbester, and A. J. Keane, Multi-fidelity optimization via surrogate modelling, Proc.: Math., Phys. Eng. Sci. **463**, 3251 (2007).

[17] A. Patra, R. Batra, A. Chandrasekaran, C. Kim, T. D. Huan, and R. Ramprasad, A multifidelity information-fusion approach to machine learn and predict polymer bandgap, Comput. Mater. Sci. **172**, 109286 (2020).

[18] J. Kou and W. Zhang, Multi-fidelity modeling framework for nonlinear unsteady aerodynamics of airfoils, Appl. Math. Model. **76**, 832 (2019).

[19] X. Meng and G. E. Karniadakis, A composite neural network that learns from multifidelity data: Application to function approximation and inverse PDE problems, J. Comput. Phys. **401**, 109020 (2020).

[20] N. Seryo, T. Sato, J. J. Molina, and T. Taniguchi, Learning the constitutive relation of polymeric flows with memory, Phys. Rev. Research **2**, 033107 (2020).

[21] T. Lee, I. Bilionis, and A. B. Tepole, Propagation of uncertainty in the mechanical and biological response of growing tissues using multifidelity Gaussian process regression, Comput. Methods Appl. Mech. Eng. **359**, 112724 (2020).

[22] S. Sarkar, S. Mondal, M. Joly, M. E. Lynch, S. D. Bopardikar, R. Acharya, and P. Perdikaris, Multifidelity and multiscale bayesian framework for high-dimensional engineering design and calibration, J. Mech. Design **141**, 121001-1 (2019).

[23] H. Liu, Y.-S. Ong, J. Cai, and Y. Wang, Cope with diverse data structures in multifidelity modeling: A Gaussian process method, Eng. Appl. Artific. Intell. **67**, 211 (2018).

[24] Z. Guo, L. Song, C. Park, J. Li, and R. T. Haftka, Analysis of dataset selection for multifidelity surrogates for a turbine problem, Struct. Multidisc. Optimiz. **57**, 2127 (2018).

[25] T. Sjostrom and J. Daligault, Ionic and electronic transport properties in dense plasmas by orbital-free density functional theory, Phys. Rev. E **92**, 063304 (2015).

[26] Z.-G. Li, Y. Cheng, Q.-F. Chen, and X.-R. Chen, Equation of state and transport properties of warm dense helium via quantum molecular dynamics simulations, Phys. Plasmas **23**, 052701 (2016).

[27] C. Wang, Y. Long, M.-F. Tian, X.-T. He, and P. Zhang, Equations of state and transport properties of warm dense beryllium: A quantum molecular dynamics study, Phys. Rev. E **87**, 043105 (2013).

[28] C. Wang, Z.-B. Wang, Q.-F. Chen, and P. Zhang, Quantum molecular dynamics study of warm dense iron, Phys. Rev. E **89**, 023101 (2014).

[29] L. G. Stanton and M. S. Murillo, Ionic transport in high-energy-density matter, Phys. Rev. E **93**, 043203 (2016).

[30] J. P. Hansen, I. R. McDonald, and E. L. Pollock, Statistical mechanics of dense ionized matter. iii. dynamical properties of the classical one-component plasma, Phys. Rev. A **11**, 1025 (1975).

[31] M. S. Murillo, Viscosity estimates of liquid metals and warm dense matter using the Yukawa reference system, High Energy Density Phys. **4**, 49 (2008).

[32] M. S. Murillo, Viscosity estimates for strongly coupled Yukawa systems, Phys. Rev. E **62**, 4115 (2000).

[33] GPy, GPy: A Gaussian process framework in python, http://github.com/SheffieldML/GPy (2012).

[34] 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, and E. Duchesnay, Scikit-learn: Machine learning in Python, J. Mach. Learn. Res. **12**, 2825 (2011).

[35] R. H. Byrd, P. Lu, J. Nocedal, and C. Zhu, A limited memory algorithm for bound constrained optimization, SIAM J. Sci. Comput. **16**, 1190 (1995).

[36] Recall that for two independent normally distributed random variables $A \sim \mathcal{N}(\boldsymbol{\mu_A}, \sigma_A^2)$ and $B \sim \mathcal{N}(\boldsymbol{\mu_B}, \sigma_B^2)$, then $C = A + B \sim \mathcal{N}(\mu_A + \mu_A, \sigma_A^2 + \sigma_B^2)$. Also, for some constant α, a random variable $D \sim \alpha\mathcal{N}(\boldsymbol{\mu}_D, \sigma_D^2) = \mathcal{N}(\boldsymbol{\mu}_D, \alpha^2\sigma_D^2)$.

[37] This can be most easily seen by using a single test point $x_* = x_{\mathrm{HF}}$ and single training points x_{LF} and x_{HF}.

[38] A. O'Hagan, A Markov property for covariance structures, Stat. Res. Rep. **98**, 1 (1998).

[39] A. Paleyes, M. Pullin, M. Mahsereci, C. McCollum, N. D. Lawrence, and J. González, Emulation of physical processes with Emukit, in Proceedings of the 2nd Workshop on Machine Learning and the Physical Sciences (NeurIPS), arXiv:2110.13293 [cs.LG].

[40] H. Nyquist, Certain topics in telegraph transmission theory, Trans. Am. Inst. Electr. Eng. **47**, 617 (1928).

[41] C. E. Shannon, Communication in the presence of noise, Proc. IRE **37**, 10 (1949).

[42] L. Le Gratiet and J. Garnier, Recursive co-kriging model for design of computer experiments with multiple levels of fidelity, Int. J. Uncert. Quant. **4**, 365 (2014).

[43] A. Damianou and N. D. Lawrence, Deep Gaussian processes, in *Proceedings of the 16th International Conference on Artificial Intelligence and Statistics (AISTATS)* (PMLR, Scottsdale, AZ, USA., 2013), pp. 207–215.

[44] A. Damianou, Deep Gaussian processes and variational propagation of uncertainty, Ph.D. thesis, University of Sheffield (2015).

[45] J. A. Hevesi, J. D. Istok, and A. L. Flint, Precipitation estimation in mountainous terrain using multivariate geostatistics. Part I: Structural analysis, J. Appl. Meteorol. **31**, 661 (1992).

[46] J. A. Hevesi, A. L. Flint, and J. D. Istok, Precipitation estimation in mountainous terrain using multivariate geostatistics. Part II: Isohyetal maps, J. Appl. Meteorol. **31**, 677 (1992).

[47] A sequence of points is said to be *low-discrepancy* if the proportion of points in the sequence falling into an arbitrary set is (on average) near-proportional to the measure of that set.

[48] J. H. Halton, Algorithm 247: Radical-inverse quasirandom point sequence, Commun. ACM **7**, 701 (1964).

[49] C. Caruso and F. Quarta, Interpolation methods comparison, Comput. Math. Appl. **35**, 109 (1998).

[50] A. Pensoneaulta, X. Yangb, and X. Zhua, Nonnegativity-enforced Gaussian process regression, Theo.App. Mech. Lett. **10**, 182 (2020).

[51] F. D. Swesty, Thermodynamically consistent interpolation for equation of state tables, J. Comput. Phys. **127**, 118 (1996).

[52] F. X. Timmes and F. D. Swesty, The accuracy, consistency, and speed of an electron-positron equation of state based on table interpolation of the Helmholtz free energy, Astrophys. J. Suppl. Ser. **126**, 501 (2000).

[53] A. M. Michalak, A Gibbs sampler for inequality-constrained geostatistical interpolation and inverse modeling, Water Resour. Res. **44** (2008).