

PAPER • OPEN ACCESS

## Guided patchwork kriging to develop highly transferable thermal conductivity prediction models

To cite this article: Rinkle Juneja and Abhishek K Singh 2020 *J. Phys. Mater.* **3** 024006

View the [article online](#) for updates and enhancements.

You may also like

- [On The Partition Dimension of a Lollipop Graph and a Generalized Jahangir Graph](#)  
Maylinda Purna Kartika Dewi and Tri Atmojo Kusmayadi
- [Isometric Path Partition Number of Butterfly Network and X-Trees](#)  
R Prabha and R Kalaiyarasi
- [Experimental Investigation and Modeling of the Performance of Pure and Mixed Surfactant Inhibitors: Partitioning and Distribution in Water-Oil Environments](#)  
Yakun Zhu and Michael L. Free

### ECS Toyota Young Investigator Fellowship



For young professionals and scholars pursuing research in batteries, fuel cells and hydrogen, and future sustainable technologies.

At least one \$50,000 fellowship is available annually.  
More than \$1.4 million awarded since 2015!



Application deadline: January 31, 2023

**Learn more. Apply today!**



## PAPER

## OPEN ACCESS

## RECEIVED

18 December 2019

## REVISED

1 February 2020

## ACCEPTED FOR PUBLICATION

21 February 2020

## PUBLISHED

26 March 2020

# Guided patchwork kriging to develop highly transferable thermal conductivity prediction models

Rinkle Juneja and Abhishek K Singh

Materials Research Centre, Indian Institute of Science, Bangalore 560012, India

E-mail: [abhishek@iisc.ac.in](mailto:abhishek@iisc.ac.in)**Keywords:** machine learning, thermal conductivity, krigingSupplementary material for this article is available [online](#)

Original content from this work may be used under the terms of the [Creative Commons Attribution 4.0 licence](#).

Any further distribution of this work must maintain attribution to the author(s) and the title of the work, journal citation and DOI.



## Abstract

The machine learning models developed on a dataset comprising particular class of materials show poor transferability across different classes. The problem can be partially solved by increasing the variability in the dataset at the cost of prediction accuracy. To develop a model on a highly variable database, we propose a localized regression based patchwork kriging approach for capturing most of the complex details in the data. In this approach, the data is partitioned into smaller regions with shared patches of few datapoints across the neighboring boundaries. Local regression functions are developed in each partition with a constrain to give similar performance at the boundary. Out of 17 different properties tried for partitioning the data, the decomposition with respect to target output  $\kappa_l$  gave local models with unprecedented accuracies. The partitioning with respect to  $\kappa_b$  however, requires its estimate for any unknown compound beforehand. To address this, we developed a global model for the entire database. The global model accurately predicts the order of magnitude of  $\kappa_l$  for the compounds in the dataset and hence, directs them towards a particular partition for more accurate prediction. We define this stepwise approach as guided patchwork kriging, which can be applied to develop highly accurate transferable prediction models.

## 1. Introduction

Fueled by the developments in algorithms and advancement in computational resources, data-driven machine learning (ML) methods are revolutionizing the materials science domain. Using these ML-based approaches, there is a rich profusion of prediction models for resource extensive materials properties [1–11]. Among these, lattice thermal conductivity  $\kappa_l$  is of great relevance in variety of applications such as thermoelectrics, barrier coatings, and electronic devices [12–16]. Significant efforts and notable contributions have been made by the scientific community for the efficient and accelerated prediction of  $\kappa_b$  using data-assisted methods. For example, high-throughput modeling was employed on a dataset of half-Heuslers from AFLOW repository and three low  $\kappa_l$  materials were proposed for experimental verification [9]. Automatic Gibbs Library has been implemented within the AFLOW and Materials Project frameworks for high-throughput screening of desired  $\kappa_l$  materials, calculated under quasiharmonic Slack model [17]. 221 low  $\kappa_l$  candidates were proposed using Bayesian optimization, starting from the rocksalt, zinblende, and wurtzite-type 101 compounds [10]. By combining finite-temperature phonon calculations with high-throughput screening,  $\kappa_l$  was estimated for oxide and fluoride perovskites. Moreover, physical insights such as the variation of  $\kappa_l$  slower than usual  $T^{-1}$  dependence for cubic perovskites were unraveled [18]. Semiempirical models for isotropic and anisotropic  $\kappa_l$  have been proposed for accelerated high-throughput screening [19–21]. Recently, a machine learning model was developed using four high-throughput assisted descriptors directly related to the physics of  $\kappa_b$ , which outperformed the Slack model [22]. Undoubtedly, these studies over the past few years present a significant advance in expediting the search for desired  $\kappa_l$  materials and provide several new physical insights. However, for the single target output property such as  $\kappa_b$ , the developed

prediction models are different. These models differ mainly with respect to the class of compounds and type of descriptors in the dataset.

The ML-based predictions depend solely on the type of data employed for training. Hence, the models focused on particular class of materials show poor transferability across different classes. Increasing the variability in the data, however, will result in high prediction errors compared to focused models due to its incapability to adapt every complex detail in the dataset. So far, there has been no approach discussed in literature, which addresses the generalizability of prediction models for a given target output property of interest. In this work, for the target output  $\kappa_b$ , we propose the use of a localized regression based patchwork kriging approach for a class independent dataset. The compilation of dataset is done such that it covers diverse chemical and structural space, having compounds with elements belonging to any group of the periodic table, covering all the 7 crystal systems, and having 4 orders of magnitude variation in the target output  $\kappa_b$ . As expected, the model developed on this data using elemental and structural descriptors gives high train/test root mean square error (RMSE) of 0.24/0.25 for the log-scaled  $\kappa_b$ . To improve the prediction accuracy, we employ the patchwork kriging, where the dataset is partitioned into smaller local subsets with respect to a property [23, 24]. The neighboring local regions are patched together to share some datapoints at the boundary and the local models on either side of the boundary are expected to produce the same response at the boundary points. Out of 17 different possible properties tested for creating partitions, the target property  $\kappa_b$  turns out to be the best in terms of providing unprecedented accuracies. Since the decomposition of data is with respect to target output, it requires an estimation of  $\kappa_b$  for the unknown compounds. For that, global model developed on the entire dataset was used, which predicts the correct order of magnitude of  $\kappa_b$  and thus can direct the datapoints to specific partitions for improved predictions by local models. The developed approach is therefore guided patchwork kriging, where the global model directs the datapoints to specific partitions for more accurate prediction of target output.

## 2. Methodology

The electronic structure calculations were performed using density functional theory [25, 26], as implemented in the VASP package [27, 28]. Generalized gradient approximation (GGA) of Perdew–Burke–Ernzerhof (PBE) was employed to account for electronic exchange and correlation. The projector augmented wave (PAW) potentials were used to represent the ion–electron interactions [29, 30]. The energy cutoff for the plane waves was taken to be 1.3 times the default value specified in the PAW pseudopotentials. The criteria for energy convergence was set to  $10^{-8}$  eV. The dynamical stability was assessed by phonon dispersion calculated using finite difference method, as implemented in the PHONOPY [31]. By solving the phonon Boltzmann transport equation, lattice thermal conductivity is calculated, as implemented in the PHONO3PY [32, 33].

For making the generalized model for capturing wide variation of  $\kappa_b$ , patchwork kriging method is employed [23, 24]. Patchwork kriging regression is a supplement to the family of well established kriging/Gaussian process regression (GPR) [34–36]. Instead of fitting a particular parametric function to the input data, the Gaussian processes (GP) provide a non-parametric approach for constructing a regressor function by using the input data based on Bayesian inferences. In GP, to each input vector  $\mathbf{x}$ , a random variable  $f(\mathbf{x})$  is assigned, which is assumed to follow a Gaussian distribution defined by a mean function  $m(\mathbf{x})$  and covariance function  $k(\mathbf{x}, \mathbf{x}')$ . For a finite number of variables  $\mathbf{X} = (\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_N)$ , the joint prior distribution is also Gaussian:

$$p(\mathbf{f}|\mathbf{X}) = \mathcal{N}(\mathbf{f}|\boldsymbol{\mu}, \mathbf{K}),$$

where  $\mathbf{f} = (f(\mathbf{x}_1), f(\mathbf{x}_2), \dots, f(\mathbf{x}_N))$ ,  $\boldsymbol{\mu} = (m(\mathbf{x}_1), m(\mathbf{x}_2), \dots, m(\mathbf{x}_N))$  and  $K_{ij} = k(\mathbf{x}_i, \mathbf{x}_j)$ . The mean function  $m(\mathbf{x})$  are generally taken to be 0. The covariance  $k$  are positive definite symmetric kernel functions, which decide the shape/smoothness of functions. The GP prior combined with the observations  $\mathbf{y}$  gives the GP posterior distribution  $p(\mathbf{f}|\mathbf{X}, \mathbf{y})$ . The posterior distribution is used to obtain posterior predictive distribution of  $\mathbf{f}$  at an arbitrary new input  $\mathbf{x}_*$ , denoted by  $\mathbf{f}_*$ . The joint distribution of observations  $\mathbf{y}$  and predictions  $\mathbf{f}_*$  is

$$\begin{pmatrix} \mathbf{y} \\ \mathbf{f}_* \end{pmatrix} \sim \mathcal{N}\left(\mathbf{0}, \begin{pmatrix} \mathbf{K}_y & \mathbf{K}_* \\ \mathbf{K}_*^T & \mathbf{K}_{**} \end{pmatrix}\right), \quad (1)$$

where  $\mathbf{K}_y = k(\mathbf{X}, \mathbf{X} + \sigma^2\mathbf{I})$ ,  $\mathbf{K}_* = k(\mathbf{X}, \mathbf{X}_*)$ , and  $\mathbf{K}_{**} = k(\mathbf{X}_*, \mathbf{X}_*)$ . Using the Gaussian conditioning rules, the predictive posterior distribution of  $\mathbf{f}_*$  is given by

$$\begin{aligned} p(\mathbf{f}_*|\mathbf{X}_*, \mathbf{X}, \mathbf{y}) &= \int p(\mathbf{f}_*|\mathbf{X}_*, \mathbf{f})p(\mathbf{f}|\mathbf{X}, \mathbf{y}) d\mathbf{f} \\ &= \mathcal{N}(\mathbf{f}_*|\boldsymbol{\mu}_*, \boldsymbol{\Sigma}_*). \end{aligned} \quad (2)$$

The posterior predictive distribution is also a Gaussian with mean  $\mu_*$  and  $\Sigma_*$  given by

$$\begin{aligned}\mu_* &= \mathbf{K}_*^T \mathbf{K}_y^{-1} \mathbf{y} \\ \Sigma_* &= \mathbf{K}_{**} - \mathbf{K}_*^T \mathbf{K}_y^{-1} \mathbf{K}_*\end{aligned}\quad (3)$$

In Patchwork kriging, the input domain of  $f(\mathbf{x})$  is partitioned into  $P$  local regions  $\{\Omega_p; p = 1, 2, \dots, P\}$  and inferences are made in each partition with additional continuity constraints.  $f_p(\mathbf{x})$  is defined as local approximation of  $f(\mathbf{x})$  in the  $p^{\text{th}}$  partition, which follows Gaussian prior distribution. The continuity conditions are imposed in such a way that two neighboring partitions share some datapoints, so that predictions of local models does not show discontinuity at the boundary.

Now to develop the prediction models, the input domain  $\mathbf{X}$ , also known as descriptors, and the covariance function  $k(\mathbf{x}_i, \mathbf{x}_j)$  needs to be selected. The performance of the model depends crucially on the choice of descriptors and the covariance function. The filtering of most relevant subset of features was carried out using the least absolute shrinkage and selection operator (LASSO) [34, 35]. It minimizes the sum of squared errors along with a constraint on the sum of the absolute values of model parameters (L1 norm). Assuming the linear relationship between the predicted variable vector  $\mathbf{y}$  and the input feature vectors  $\mathbf{X}$ , the LASSO estimate is defined by [34, 35]

$$\hat{\beta}(\lambda) = \arg \min_{\beta} (\|\mathbf{y} - \mathbf{X}\beta\|_2^2 + \lambda \|\beta\|_1), \quad (4)$$

where  $\lambda \geq 0$  is the shrinkage parameter. Depending upon the value of  $\lambda$ , some of the coefficients  $\beta$  may become zero, thereby providing the reduced set of relevant features. Out of different possibilities for the covariance functions, the automatic relevance determination (ARD) squared exponential covariance function gave best accuracy. For the given data points  $x_i$  and  $x_j$  and hyperparameter vector  $\theta$ , it is given by [36]

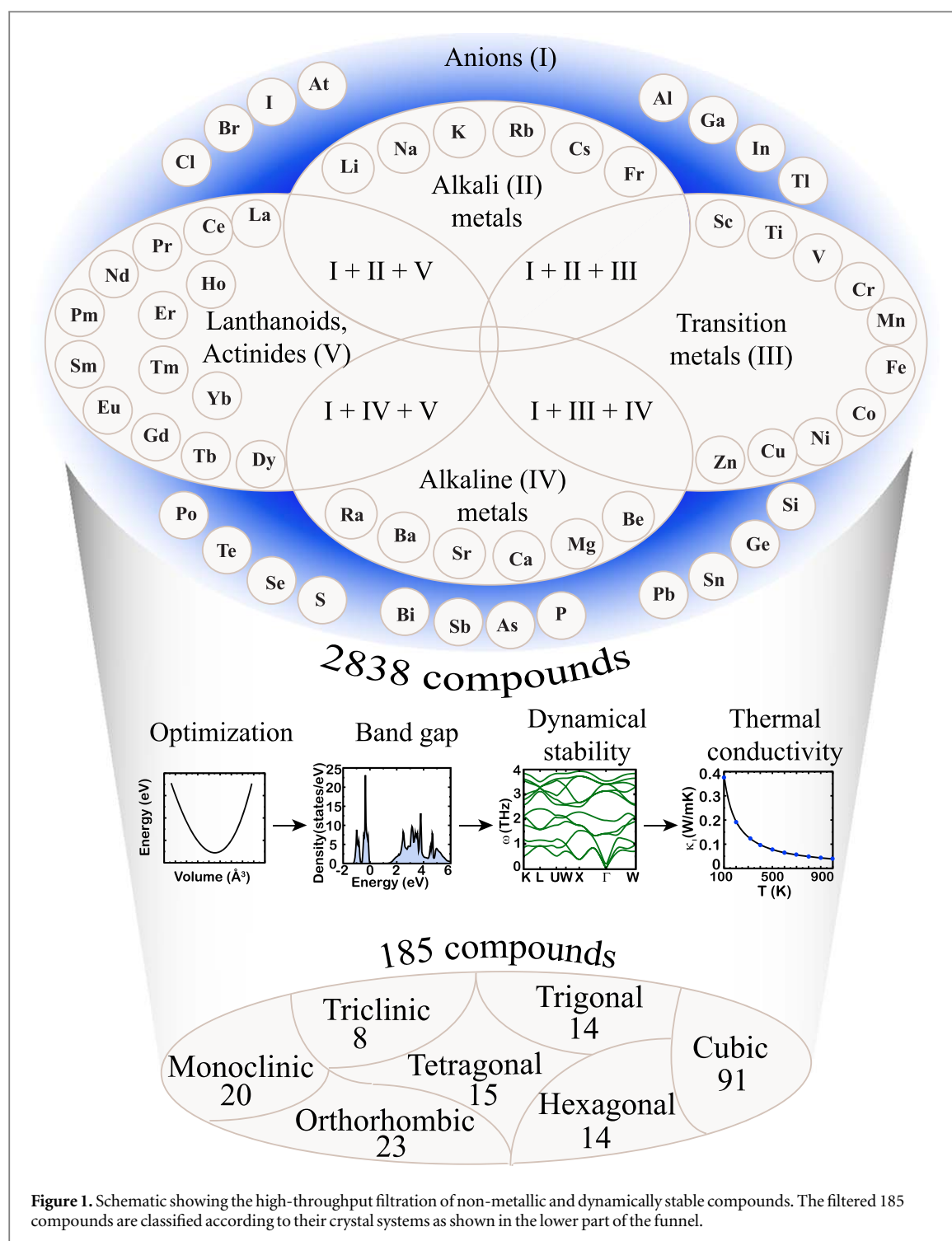
$$k(x_i, x_j | \theta) = \sigma_f^2 \exp\left(-\frac{1}{2} \sum_{m=1}^d \frac{(x_{im} - x_{jm})^2}{\sigma_m^2}\right), \quad (5)$$

where  $\sigma_f$ ,  $d$ , and  $\sigma_m$  are the standard deviation, number of predictors in the data, and length scale for each predictor  $m$ . The statistical parameters namely standard deviation and length scale are also known as hyperparameters  $\theta$  [36]. The ARD covariance function has the advantage of providing a separate length scale for each predictor, which improves the flexibility of the model. The parameter  $\sigma$  controls the numerical stability of the covariance function and length scale captures impact of each data point. The hyperparameters for the covariance functions were optimized using Broyden–Fletcher–Goldfarb–Shanno (BFGS) algorithm [37]. The performance of the developed model was assessed by statistical regression metric such as root mean square error (RMSE) and the coefficient of determination ( $R^2$ ), which determines the proportion of variability captured by input descriptors in the target output values.

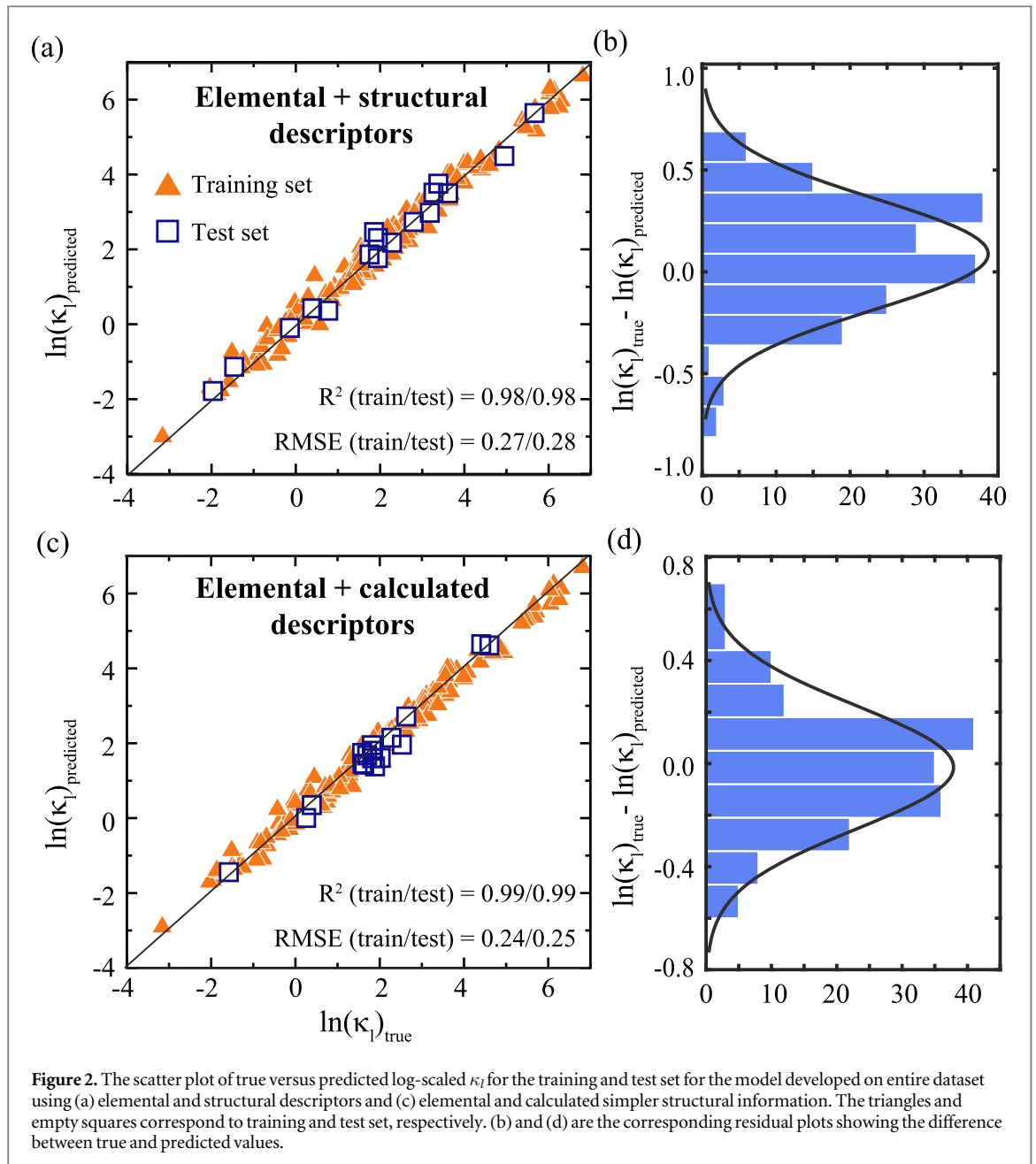
### 3. Results and discussion

With the objective to develop a generalized prediction model for  $\kappa_t$ , we utilized the original dataset of 2162 compounds [22] along with additional 676 compounds from the Materials Project [38, 39]. These additional compounds further diversify the spanned structural and chemical space. The dataset is independent of any particular class of materials and covers wide variability in various aspects. For instance, it consists of compounds with elements belonging to different groups of the periodic table (alkali, alkaline earth, transition metals, lanthanides, actinides, anions) and covering all 7 crystal systems with different possible space groups. After the high-throughput screening for fully optimized, non-metallic, and dynamically stable compounds, a total of 185 candidates are filtered. The screened compounds cover all the 7 crystal systems, as seen in figure 1. There are 20 monoclinic, 8 triclinic, 23 orthorhombic, 15 tetragonal, 14 trigonal, 14 hexagonal, and 91 cubic compounds in the screened dataset. Additionally, the response variable  $\kappa_t$  at room temperature for these screened compounds covers four orders of magnitude variation, thereby further enriching the diversity in the dataset. The phonon dispersion and  $\kappa_t$  for the dynamically stable additional compounds are shown in the supporting information figures S1–S2 and S3–S4, respectively available at [stacks.iop.org/JPMATER/3/024006/mmedia](https://stacks.iop.org/JPMATER/3/024006/mmedia).

To check the prediction accuracy for this dataset, we developed the ML model for the log-scaled  $\kappa_t$  at room temperature. For developing the ML prediction model, a certain set of descriptors  $\mathbf{X}$  for the constituent compounds are required, which are mapped to the target property of interest  $\mathbf{y}$ . These descriptors can be either directly related to the physics of target output [22] or they can be in the form of elemental and structural properties to represent the compounds. For  $\kappa_t$ , Seko *et al* [6] have proposed a combination of elemental and structural descriptors, which result in ML models with unprecedented accuracies for the log-scaled  $\kappa_t$ . The elemental descriptors include the mean and standard deviation of various physical properties of constituent



elements such as boiling point ( $B.P.$ ), melting point ( $M.P.$ ), specific heat ( $c_p^s$ ), molar specific heat ( $c_p$ ), molar volume ( $V_m$ ), heat of fusion ( $H_f$ ), heat of vaporization ( $H_v$ ), Pauling electronegativity ( $\chi_p$ ), first ionization energy ( $I.E._1$ ), group and period in the periodic table, elemental thermal conductivity ( $\kappa_{ele}$ ), atomic mass ( $M$ ), covalent radius ( $r_{cov}$ ), van der Waals radius ( $r_{vdw}$ ), and density ( $\rho$ ). The structural features include generalized radial distribution function (GRDF) and bond order parameter (BOP). The large variance among descriptors is nullified by standardizing them to have zero mean and unit variance. The relevant set of descriptors was selected via ten-fold cross-validated LASSO. Using this descriptor set, a GPR model is developed for predicting the log-scaled  $\kappa_l$ . Out of various covariance functions, ARD squared exponential covariance function turns out to be the best performing. The training of the model is carried out on 90% of the data and the remaining 10% is used for the testing of the developed model. The splitting of data into training and test set is done over 2000 random trials and the hyperparameters of the covariance function are tuned for each random trial. The performance of the developed models is estimated by evaluating RMSE and  $R^2$  for the training as well as the test sets. The best



optimized model has train/test RMSE and  $R^2$  of 0.27/0.28 and 0.98/0.98, respectively. Figure 2(a) presents the regression plot showing the variation of true versus predicted log-scaled  $\kappa_i$  for the developed model, where the triangles and empty squares correspond to the training set and the test set, respectively. The difference between true and predicted log-scaled  $\kappa_i$ , also known as residuals, for the whole data set is shown in figure 2(b), which is very high.

Since the structural descriptors such as GRDF and BOP are relatively complex to calculate, we searched for alternative simpler descriptors, which can provide the similar information about the structure without compromising the performance of the prediction model. For example, the GRDF is calculated by defining a pairwise function  $g_n(r_{ij})$  for distance  $r_{ij}$  [6] between atom  $i$  and neighboring atom  $j$  as

$$\text{GRDF}_n^i = \sum_j g_n(r_{ij}). \quad (6)$$

GRDF represents the variation in density as a function of distance from a reference atom, which could possibly be captured by quantifying bonding and anti-bonding strength. We calculated the average bond strength by integrating the crystal orbital Hamilton population (COHP) [40, 41]. The bond strength has been shown to be directly correlated with  $\kappa_i$  for zincblende, rock salt and CsCl structure classes [42], where increasing bond strength results in high  $\kappa_i$ . The another structural descriptor BOP is evaluated by calculating the average spherical harmonics  $Q_{lm}$  as [6]

$$Q_l = \left( \frac{4\pi}{2l+1} \sum_{m=-l}^l |Q_{lm}|^2 \right)^{1/2}. \quad (7)$$

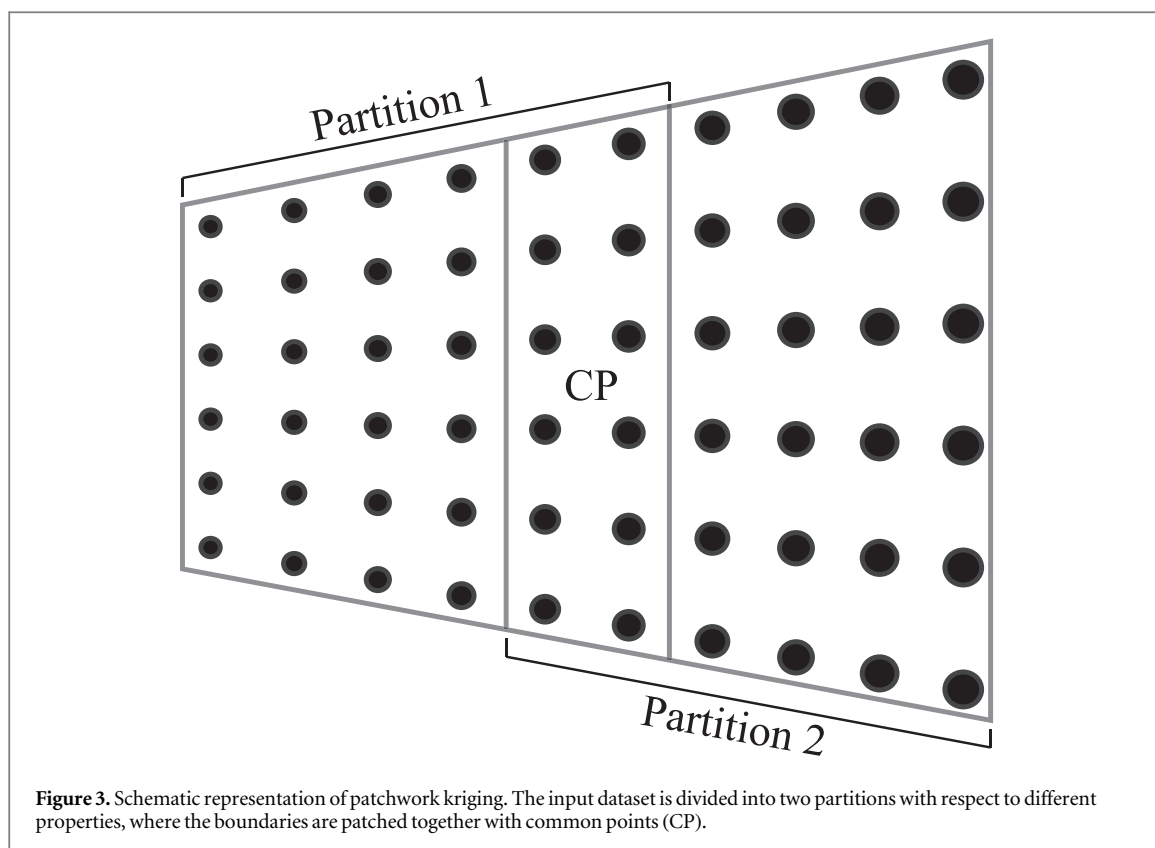
BOP represents the local environment around a reference atom due to its surrounding atoms, which could alternatively be described by average coordination number and average bond distances. While bond distances are inversely related to  $\kappa_l$  [42], coordination number has been shown as a potential descriptor for capturing anharmonicity [19]. Other simpler parameter volume, which has inverse correlation [22] with  $\kappa_l$  has also been included as descriptor. Hence, we replaced the complex structural descriptors GRDF and BOP with the new simpler descriptors average bond distances, average bond strength, average coordination number, and volume of cell, which have strong correlations with  $\kappa_l$ . The GPR model developed with these new descriptors gives the improved accuracy compared to previously developed model using GRDF and BOP, thereby implying that the new descriptors can provide better structural representation for compounds without any additional computational complexity. In this case, the best optimized model has train/test RMSE and  $R^2$  of 0.24/0.25 and 0.99/0.99, respectively. Figure 2(c) presents the regression plot showing the variation of true versus predicted log-scaled  $\kappa_l$  by the developed model and figure 2(d) shows the corresponding residuals plot.

As expected, the model developed on the entire dataset gives high prediction errors compared to state-of-the-art models reported for  $\kappa_l$  [6, 22]. To develop a generalized model with improved prediction accuracy, we propose the use of localized regression-based approach. The underlying notion is that in a dataset with wide variability, the local details of the input space may have significant effect on the prediction of target output. Several local approximations have been developed in the literature for capturing the large variability in the input space of big datasets. For example, sparse approximations such as sparse Gaussian process (SGP), where the full dataset is replaced by a smaller subset of relevant inputs [43–45]; local approximations such as mixture of experts [46–48], where input data is divided into smaller subspaces to build local models. The major advantage of the local approaches is to take into consideration every possible detail in the data. However, the disadvantage is discontinuity in the prediction by local models at the boundary between the local regions. There have been several ways to smooth out the discontinuity [47, 49–51]. One of such approaches with high numerical stability is patchwork kriging [23, 24], where the neighboring local regions are patched together to share some datapoints at the boundary and the local models on either side of the boundary are expected to produce the same response at the boundary points. Inspired by the success of patchwork kriging for the capturing the complexities in the big datasets [23, 24], we applied this algorithm to our dataset.

Efficient identification of the property with respect to which data is partitioned to build several local models is one of the key factors deciding the prediction accuracy. Here, we tried several properties for partitioning the data, which include elemental descriptors such as  $B.P.$ ,  $M.P.$ ,  $c_p^g$ ,  $c_p$ ,  $H_p$ ,  $H_v$ ,  $\chi_p$ ,  $I.E._1$ ,  $\kappa_{ele}$ ,  $M$ ,  $r_{cov}$ ,  $r_{vdw}$  and  $\rho$ ; and structural descriptors such as  $B.S.$ ,  $B.D.$ , and  $V_{cell}$ . Partitioning with respect to the target output  $\kappa_l$  was also considered. Since the dataset size in the present case is not very large, we created only two partitions corresponding to each of these properties. As the large number of partitions would result in very less datapoints in each partition, it may lead to overfitted local prediction models. The schematic representation of the patchwork kriging is illustrated in figure 3. The two partitions corresponding to all the 17 properties are created in the ascending order of the magnitude of the properties with some common points (CP) at the shared boundary. For example, in the case of partitions corresponding to  $\kappa_b$ , partition 1 and 2 contain compounds with log-scaled  $\kappa_l$  ranging from  $-3.16$  to  $1.96$  and  $1.18$  to  $6.81$ , respectively. At the shared boundary of these two partitions, there were 13 datapoints common with log-scaled  $\kappa_l$  ranging from  $1.18$  to  $1.91$ . The choice of the range for creating the common boundary and the corresponding number of common datapoints across partitions was decided in such a way to have nearly equal and sufficient data in each partition. For all the properties, the selected range and corresponding number of common datapoints are given in the supporting information table S1.

To develop the local prediction models in each partition corresponding to all the considered properties, a set of suitable descriptors has to be selected. We used ten-fold cross validated LASSO to select the relevant subset of elemental and structural descriptors for all the partitions. The descriptor set for both partitions corresponding to different properties is given in the supporting information table S2.

The descriptor set in the partitions created with respect to 17 properties is different. This may be attributed to the fact that each partition contains different local information. By using the descriptor sets for respective partitions, the prediction models are developed using GPR for the log-scaled  $\kappa_l$ . For all the cases, the ARD squared exponential covariance function turns out to be the best performing. The training of the model in each case is carried out on 90% of the data and the remaining 10% is used for the testing of the developed model. The splitting of data into training and testing set is done over 2000 random trials and the hyperparameters of the covariance function are tuned for each random trial. The performance of the developed models is estimated by RMSE and  $R^2$  for the training and the test sets, which is given for all the cases in table 1.



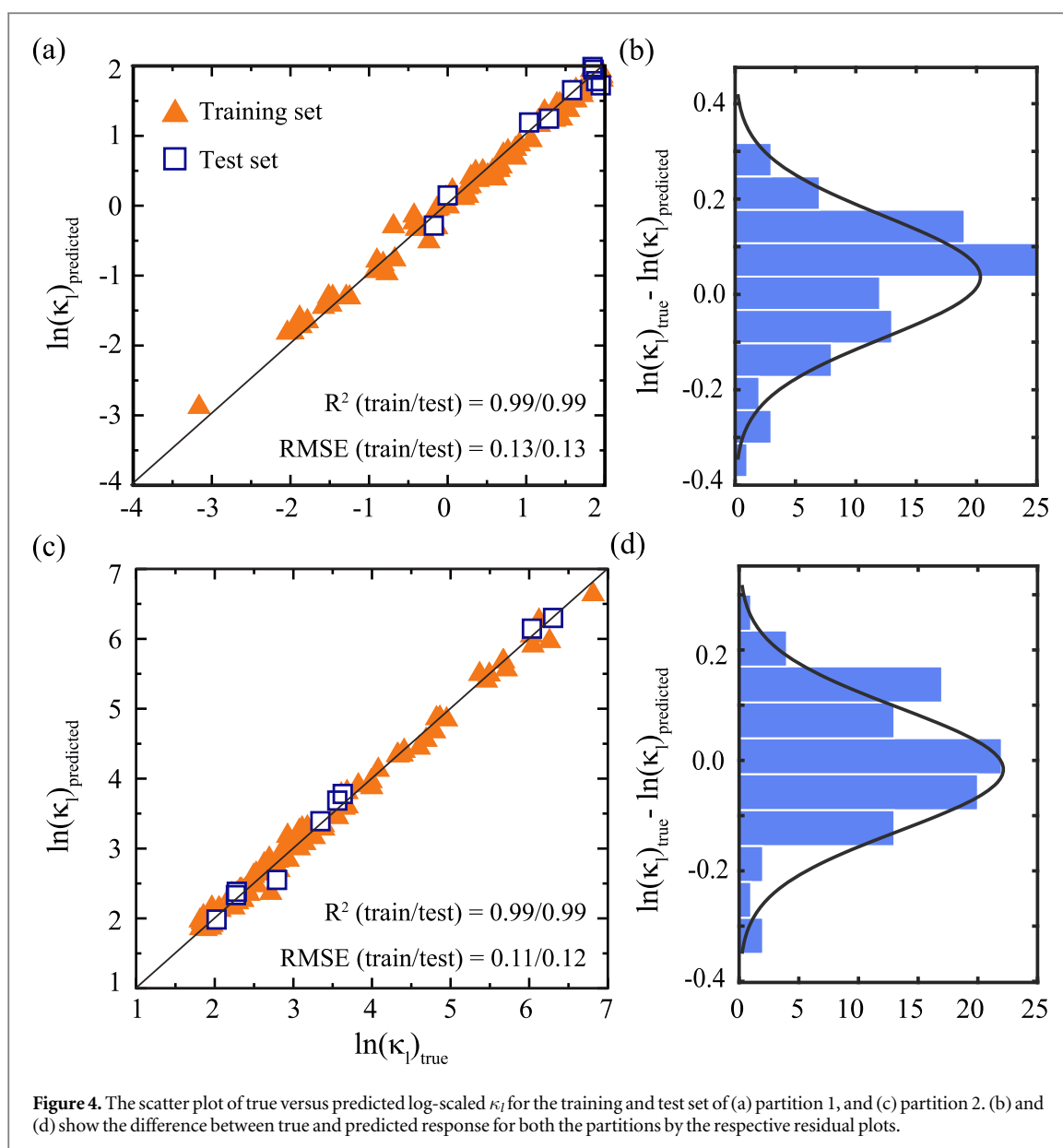
**Figure 3.** Schematic representation of patchwork kriging. The input dataset is divided into two partitions with respect to different properties, where the boundaries are patched together with common points (CP).

**Table 1.** Performance of GPR models developed on each partition corresponding to different properties.

| Properties   | Partition 1       |                    | Partition 2       |                    |
|--|-------------------|--------------------|-------------------|--------------------|
|  | RMSE (train/test) | $R^2$ (train/test) | RMSE (train/test) | $R^2$ (train/test) |
| Boiling point ( $B.P.$ )                               | 0.16/0.17         | 0.99/0.99          | 0.14/0.15         | 0.99/0.99          |
| Melting point ( $M.P.$ )                               | 0.20/0.21         | 0.99/0.99          | 0.14/0.15         | 0.99/0.99          |
| Specific heat ( $c_p^S$ )                              | 0.15/0.15         | 0.99/0.99          | 0.16/0.16         | 0.99/0.99          |
| Molar specific heat ( $c_p$ )                          | 0.20/0.21         | 0.99/0.99          | 0.21/0.21         | 0.98/0.98          |
| Heat of fusion ( $H_f$ )                               | 0.25/0.26         | 0.98/0.98          | 0.21/0.21         | 0.99/0.99          |
| Heat of vaporization ( $H_v$ )                         | 0.14/0.15         | 0.99/0.99          | 0.18/0.20         | 0.99/0.99          |
| Pauling electronegativity ( $\chi_p$ )                 | 0.12/0.13         | 0.99/0.99          | 0.20/0.22         | 0.99/0.99          |
| First ionization energy ( $I.E._1$ )                   | 0.16/0.17         | 0.99/0.99          | 0.19/0.20         | 0.99/0.99          |
| Elemental thermal conductivity ( $\kappa_{ele}$ )      | 0.27/0.28         | 0.98/0.98          | 0.12/0.13         | 0.99/0.99          |
| Atomic mass ( $M$ )                                    | 0.14/0.15         | 0.99/0.99          | 0.23/0.24         | 0.98/0.98          |
| Covalent radius ( $r_{cov}$ )                          | 0.16/0.17         | 0.99/0.99          | 0.11/0.13         | 0.99/0.99          |
| van der Waals radius ( $r_{vdw}$ )                     | 0.27/0.28         | 0.98/0.98          | 0.12/0.13         | 0.99/0.99          |
| Density ( $\rho$ )                                     | 0.16/0.17         | 0.99/0.99          | 0.26/0.27         | 0.98/0.98          |
| Bond strength ( $B.S.$ )                               | 0.26/0.27         | 0.98/0.98          | 0.20/0.20         | 0.99/0.99          |
| Bond distance ( $B.D.$ )                               | 0.19/0.19         | 0.99/0.99          | 0.15/0.16         | 0.99/0.99          |
| Volume of cell ( $V_{cell}$ )                          | 0.21/0.22         | 0.99/0.99          | 0.16/0.16         | 0.99/0.99          |
| Log-scaled lattice thermal conductivity ( $\kappa_l$ ) | 0.13/0.13         | 0.99/0.99          | 0.11/0.12         | 0.99/0.99          |

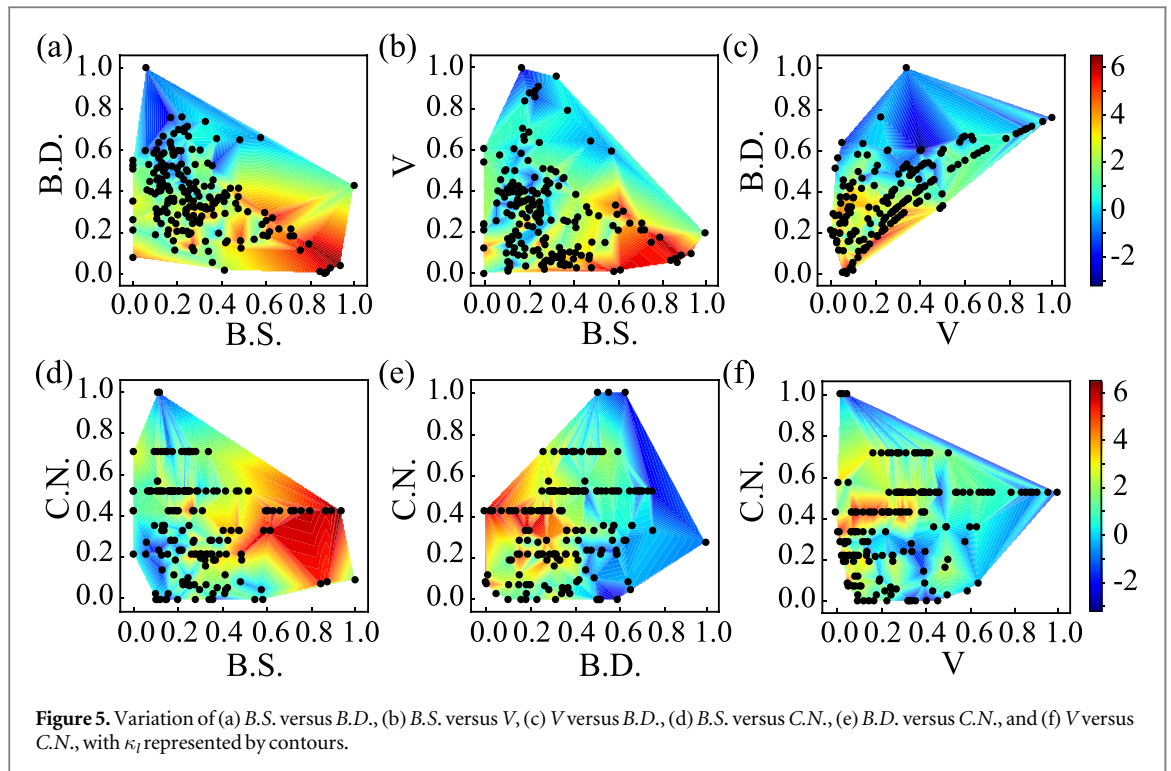
The notable differences observed in the prediction accuracies corresponding to different partitions with respect to various properties indicate the effect of local structure of the data on the performance of the model. For most of the cases, the model performs well only in one of the partitions. However, the partitioning corresponding to  $B.P.$ ,  $c_p^S$ ,  $r_{cov}$ , and  $\kappa_l$  gives local models with high accuracies in both the partitions. Among these, decomposition with respect to target output  $\kappa_l$  leads to local models with the highest accuracy. The partition 1 in this case has train/test RMSE and  $R^2$  of 0.13/0.13 and 0.99/0.99, while partition 2 has train/test RMSE and  $R^2$  of 0.11/0.12 and 0.99/0.99. Figure 4(a) presents the regression plot showing the variation of true versus predicted log-scaled  $\kappa_l$  for the developed model, where the triangles and empty squares correspond to the training set and test set, respectively. The high value of  $R^2$  clearly indicates the good predictive capability of our model covering wide variability. The difference between true and predicted log-scaled  $\kappa_l$  for the whole data set is





shown in figure 4(b). Figure 4(c) presents the regression plot showing the variation of true versus predicted log-scaled  $\kappa_l$  by the developed model for partition 2 and figure 4(d) shows the corresponding residual plot. The prediction accuracies achieved for the highly variable data using this approach are comparable to models developed on particular class and type of materials. For example, using the mean and standard deviation of elemental descriptors and BOP, the best model developed on 110 compounds for log-scaled  $\kappa_l$  gives RMSE of 0.096 [6]. In another model developed on 93 compounds, the difference between experimental and predicted values of  $\kappa_l$  has been reported to lie within a factor of 1.5–2 [52]. Using the descriptors related directly to the physics of  $\kappa_l$ , a prediction model developed on 120 compounds gives RMSE of 0.21 [22]. A prediction model developed on 100 experimentally available  $\kappa_l$  compounds, a train/test RMSE and  $R^2$  of 0.17/0.21 and 0.93/0.93, respectively, has been reported [53].

Since the highly accurate models are obtained by partitioning the data with respect to target output  $\kappa_l$ , employing this patchwork kriging approach for unknown compounds needs an estimation of its  $\kappa_l$  so as to decide the particular partition for its accurate prediction. For that, we used the model developed on the entire dataset for classification of compounds to particular partitions (figure 2(c)). This model is named as global model. The global model predicts correct order of magnitude of  $\kappa_l$  for all the compounds in the dataset. Therefore, the model can be used to direct/guide unknown data to particular partition for accurate prediction of  $\kappa_l$ . The global model is tested on 10 independent data points, which were not present in initial training set. The global model accurately guides each datapoint to the correct partition, as presented in the supporting information table S4. Hence, we propose a two-step guided patchwork kriging approach, where the global



**Table 2.** Steps for guided patchwork kriging.

**Input:**

1. Data  $D$  with descriptors  $X$  and location of target output  $y$

**Output1:**

2.  $\mu_*$  and  $\Sigma_*$  from the global model

**Re-prediction:**

3. Decomposition of data into partitions

**Output2:**

4. **for** each partition  $p$  **do**

5.  $\mu_*^p$  and  $\Sigma_*^p$

6. **end for**

model will guide the datapoints to a particular partition for subsequent accurate prediction. The pseudo-code for the guided patchwork kriging is summarized in table 2.

To make sure that the developed models are not overfitted, we checked the learning curves by plotting the prediction accuracy as a function of increasing training data size. The learning curves for the best global model, partition 1, and partition 2 model are shown in the supporting information figures S5(a), (b), and (c), respectively. The convergence of train and test RMSE around 90% of training data size ensures no overfitting. Furthermore, we checked the average performance of all the three best models, as shown in the Supporting Information figure S6. The average is taken for the developed model over 2000 different train-test splits. Figures S6(a), (b), and (c) show the true versus predicted  $\kappa_l$  for the average global, partition 1, and partition 2 model, respectively. The average test RMSE of 2000 models for global, partition 1, and partition 2 model is 0.81, 0.56, and 0.41, respectively. The average performance of local models are again better than the global model. The corresponding learning curves are shown in figures S6(d)–(f).

Analysis of the machine learning outcomes gives very useful physical insights and also a correct description of the underlying physics related to  $\kappa_l$ . For all the partitions corresponding to 17 different properties, average bond distances came out to be one of the most relevant descriptors. Additionally, the average bond strength, average coordination number and volume of cell were also found to be relevant descriptors in most of the partitions. To get the intuitive insights, we plotted these three descriptors against each other as shown in figure 5, where the magnitude of  $\kappa_l$  is represented by the color bar. The smaller *B.S.* and larger *B.D.* give low  $\kappa_l$  and vice versa, as shown in figure 5(a). This is because large bond distances in a structure imply that the atoms in the system are not tightly packed, hence the bonding strength between the atoms is weak. The group velocity of

phonons can be approximated as  $v_g \sim \sqrt{\frac{k}{M}}$  [54]. The weaker bond strength will lead to weaker force constant and hence small phonon group velocities, which will result into low  $\kappa_l$ . The model reveals this physical picture accurately. Similar trend has been reported for zincblende, rock salt and CsCl structure based materials [42]. Furthermore, figures 5(b) and (c) show the inverse correlation of  $V_{cell}$  with  $\kappa_l$ . Weak *B.S.* or large *B.D.* and large  $V_{cell}$  gives low  $\kappa_l$ . The inverse dependence of  $V$  with  $\kappa_l$  has also been reported for a large class of materials [22]. Figures 5(d)–(f) show, respectively, the behavior of *B.S.*, *B.D.*, and  $V_{cell}$  as a function of *C.N.*. High *C.N.* and weaker *B.S.* result in low  $\kappa_l$  and vice versa, as shown in figure 5(d). The opposite trend is observed for *C.N.* and *B.D.* (Figure 5(e)), as *B.D.* and *B.S.* are inversely related to each other. The observed behavior of  $\kappa_l$  unravels that *C.N.* can shed some light on anharmonicity of a system. Anharmonicity in a system signifies deviation from perfect harmonic behavior. The systems where the atoms are weakly bonded usually have more anharmonicity compared to strongly bonded closely packed systems. Hence, for the soft anharmonic lattices, *B.S.* will be weak and *B.D.* will be large, leading to high *C.N.* Hence, *C.N.* could be a good descriptor for anharmonicity. Figure 5(f) shows the regions of high  $\kappa_l$  for small  $V_{cell}$  and low *C.N.* This is because smaller volume corresponds to tightly packed stiff lattice, which will result in large phonon group velocities and hence high  $\kappa_l$ . The effect of coordination number on  $\kappa_l$  has also been illustrated for group I–V–VI<sub>2</sub> compounds, where octahedral-coordinated compounds has lower  $\kappa_l$  by a factor of 4 compared to tetrahedrally bonded compounds [55]. These descriptors, thereby, provide a simpler route for screening low or high  $\kappa_l$  compounds. Hence, the developed ML model captures the relationship between bonding chemistry of structures and  $\kappa_l$ .

## 4. Conclusion

In summary, we proposed patchwork kriging approach for developing the prediction model for a dataset with huge variability, where several local models are developed over the partitioned domains with shared patches of datapoints across neighboring boundaries. The partitioning of the data with respect to 17 properties is carried out. The best local models were obtained, when the data was partitioned based on magnitude of  $\kappa_l$ . The involvement of the target output for creating the partitions requires its estimate beforehand. We used a global model developed on the entire dataset, which directs the data points to a specific partitions for accurate prediction of  $\kappa_l$ . When applied to 10 unseen datapoints, the global model accurately guides them to correct partitions. We call this step-wise property prediction approach as guided patchwork kriging. Additionally, the analysis of machine learning models and descriptors correctly captures the expected correlations between chemistry of bonding characteristics and  $\kappa_l$ . Our approach can be extended to develop highly transferable models for prediction of any physical or chemical properties of class-independent, highly variable, comprehensive datasets.

## Acknowledgments

RJ thanks DST for INSPIRE fellowship (IF150848). We thank the Materials Research Centre, Thematic Unit of Excellence, and Supercomputer Education and Research Centre, Indian Institute of Science, for providing computing facilities. RJ and AKS thank the support from Institute of Excellence (IoE) MHRD grant.

## Notes

The authors declare no competing financial interests.

## ORCID iDs

Abhishek K Singh  <https://orcid.org/0000-0002-7631-6744>

## References

- [1] Toyoura K, Hirano D, Seko A, Shiga M, Kuwabara A, Karasuyama M, Shitara K and Takeuchi I 2016 *Phys. Rev. B* **93** 054112
- [2] Xue D, Balachandran P V, Hogden J, Theiler J, Xue D and Lookman T 2016 *Nat. Commun.* **7** 11241
- [3] Deml A M, O'Hayre R, Wolverton C and Stevanović V 2016 *Phys. Rev. B* **93** 085142
- [4] Faber F A, Lindmaa A, Von Lilienfeld O A and Armiento R 2016 *Phys. Rev. Lett.* **117** 135502
- [5] Kim C, Pilia G and Ramprasad R 2016 *Chem. Mater.* **28** 1304–11
- [6] Seko A, Hayashi H, Nakayama K, Takahashi A and Tanaka I 2017 *Phys. Rev. B* **95** 144110
- [7] Pilia G, Mannodi-Kanakkithodi A, Uberuaga B, Ramprasad R, Gubernatis J and Lookman T 2016 *Sci. Rep.* **6** 19375
- [8] Rajan A C, Mishra A, Satsangi S, Vaish R, Mizuseki H, Lee K R and Singh A K 2018 *Chem. Mater.* **30** 4031–8

- [9] Carrete J, Li W, Mingo N, Wang S and Curtarolo S 2014 *Phys. Rev. X* **4** 011019
- [10] Seko A, Togo A, Hayashi H, Tsuda K, Chaput L and Tanaka I 2015 *Phys. Rev. Lett.* **115** 205901
- [11] Mishra A, Satsangi S, Rajan A C, Mizuseki H, Lee K R and Singh A K 2019 *J. Phys. Chem. Lett.* **10** 780–5
- [12] Snyder G J and Toberer E S 2008 *Nat. Mater.* **7** 105–14
- [13] Geža V, Jakovičs A, Gendelis S, Usijonoks I and Timofejevs J 2017 Thermal conductivity of disperse insulation materials and their mixtures *IOP Conf. Ser.: Mater. Sci. Eng.* **251** 012012
- [14] Wilcox W 1963 *IEEE Trans. Electron Devices* **10** 308–13
- [15] Arik M, Becker C A, Weaver S E and Petroski J 2004 Thermal management of leds: package to system *3rd Int. Conf. on Solid State Lighting* vol 5187 (International Society for Optics and Photonics) pp 64–76
- [16] Pop E 2010 *Nano Res.* **3** 147–69
- [17] Toher C, Plata J J, Levy O, De Jong M, Asta M, Nardelli M B and Curtarolo S 2014 *Phys. Rev. B* **90** 174107
- [18] van Roekeghem A, Carrete J, Osés C, Curtarolo S and Mingo N 2016 *Phys. Rev. X* **6** 041061
- [19] Miller S A *et al* 2017 *Chem. Mater.* **29** 2494–501
- [20] Yan J, Gorai P, Ortiz B, Miller S, Barnett S A, Mason T, Stevanović V and Toberer E S 2015 *Energy Environ. Sci.* **8** 983–94
- [21] McKinney R, Gorai P, Toberer E S and Stevanovic V 2019 *Chem. Mater.* **31** 2048–57
- [22] Juneja R, Yumnam G, Satsangi S and Singh A K 2019 *Chem. Mater.* **31** 5145–51
- [23] Park C and Huang J Z 2016 *J. Mach. Learn. Res.* **17** 1–29
- [24] Park C and Apley D 2018 *J. Mach. Learn. Res.* **19** 269–311
- [25] Kohn W and Sham L J 1965 *Phys. Rev.* **140** A1133
- [26] Sham L and Schlüter M 1983 *Phys. Rev. Lett.* **51** 1888
- [27] Kresse G and Furthmüller J 1996 *Phys. Rev. B* **54** 11169
- [28] Kresse G and Furthmüller J 1996 *Comput. Mater. Sci.* **6** 15–50
- [29] Blöchl P E 1994 *Phys. Rev. B* **50** 17953
- [30] Kresse G and Joubert D 1999 *Phys. Rev. B* **59** 1758
- [31] Togo A and Tanaka I 2015 *Scr. Mater.* **108** 1–5
- [32] Chaput L, Togo A, Tanaka I and Hug G 2013 *Phys. Rev. Lett.* **110** 265506
- [33] Togo A, Chaput L and Tanaka I 2015 *Phys. Rev. B* **91** 094306
- [34] Friedman J, Hastie T and Tibshirani R 2001 *The Elements of Statistical Learning* (New York: Springer Series in Statistics)
- [35] Murphy K P 2012 *Machine Learning: A Probabilistic Perspective* (Cambridge, MA: Massachusetts Institute of Technology Press) Cambridge, MA
- [36] Rasmussen C E 2004 Gaussian processes in machine learning *Advanced Lectures on Machine Learning* (Berlin: Springer) pp 63–71
- [37] Williams T 1982 *J. Oper. Res. Soc.* **33** 675–6
- [38] Jain A *et al* 2013 *APL Mater.* **1** 011002
- [39] Ong S P, Richards W D, Jain A, Hautier G, Kocher M, Cholia S, Gunter D, Chevrier V L, Persson K A and Ceder G 2013 *Comput. Mater. Sci.* **68** 314–9
- [40] Dronskowski R and Blöchl P E 1993 *J. Phys. Chem.* **97** 8617–24
- [41] Deringer V L, Tchougréeff A L and Dronskowski R 2011 *J. Phys. Chem. A* **115** 5461–6
- [42] Zeier W G, Zevalkink A, Gibbs Z M, Hautier G, Kanatzidis M G and Snyder G J 2016 *Angew. Chem.* **55** 6826–41
- [43] Csató L and Oppel M 2002 *Neural Comput.* **14** 641–68
- [44] Quiñonero-Candela J and Rasmussen C E 2005 *J. Mach. Learn. Res.* **6** 1939–59
- [45] Snelson E and Ghahramani Z 2006 Sparse Gaussian processes using pseudo-inputs *Advances in Neural Information Processing Systems* (San Francisco, CA: Morgan Kaufmann) pp 1257–64
- [46] Tresp V 2001 Mixtures of Gaussian processes *Advances in Neural Information Processing Systems* (San Francisco, CA: Morgan Kaufmann) pp 654–60
- [47] Rasmussen C E and Ghahramani Z 2002 Infinite mixtures of Gaussian process experts *Advances in Neural Information Processing Systems* Inc. San Francisco (San Francisco, CA: Morgan Kaufmann) pp 881–8
- [48] Snelson E and Ghahramani Z 2007 Local and global sparse Gaussian process approximations *Proceedings of the Eleventh International Conference on Artificial Intelligence and Statistics* PMLR San Juan, Puerto Rico pp 524–31
- [49] Gramacy R B and Lee H K H 2008 *J. Am. Stat. Assoc.* **103** 1119–30
- [50] Tresp V 2000 *Neural Comput.* **12** 2719–41
- [51] Das K and Srivastava A N 2010 Block-gp: scalable Gaussian process regression for multimodal data *2010 IEEE Int. Conf. on Data Mining (Piscataway, NJ)* (IEEE) pp 791–6
- [52] Zhang Y and Ling C 2018 *NPJ Comput. Mater.* **4** 25
- [53] Chen L, Tran H, Batra R, Kim C and Ramprasad R 2019 *Comput. Mater. Sci.* **170** 109155
- [54] Tritt T M 2005 *Thermal Conductivity: Theory, Properties, and Applications* (New York: Springer Science & Business Media)
- [55] Morelli D, Jovovic V and Heremans J 2008 *Phys. Rev. Lett.* **101** 035901