

Five Degree-of-Freedom Property Interpolation of Arbitrary Grain Boundaries via Voronoi Fundamental Zone Framework: Supplementary Information

Sterling G. Baird^{a,*}, Eric R. Homer^a, David T. Fullwood^a, Oliver K. Johnson^a

^a*Department of Mechanical Engineering, Brigham Young University, Provo, UT 84602, USA*

Contents

S0.1 Use of Interpolation Function	1
S1 Euclidean and Arc Length Distances	2
S2 Additional Interpolation Results	3
S2.1 Smaller Set Sizes of Input GBs	3
S3 Ensemble Interpolation Results	3
S3.1 Methods	3
S3.2 Results	4
S3.3 Possibility: Combining Ensemble with Gaussian Process Regression Mixture	5
S4 Barycentric Interpolation	5
S4.1 High-Aspect Ratios	5
S5 Kim Interpolation	6
S5.1 Details of Gaussian Process Regression Mixture	6
S5.2 Input Data Quality	8
S6 Olmsted Interpolation	10
Glossary	14

S0.1. Use of Interpolation Function

To facilitate easy application of the presented methods, a vectorized, parallelized, MATLAB implementation, `interp5DOF.m`, is made available in the Voronoi fundamental zone (VFZ) repository [1] with similar input/output structure to that of built-in MATLAB interpolation functions (e.g. `scatteredInterpolant()`, `griddatan()`). A typical function call is as follows: `ypred = interp5DOF(qm,nA,y,qm2,nA2,method)`. The argument `y` is a vector of known property values corresponding to the GBs defined by `(qm,nA)`, which respectively denote pairs of GB misorientation

*Corresponding author.

Email address: `ster.g.baird@gmail.com` (Sterling G. Baird)

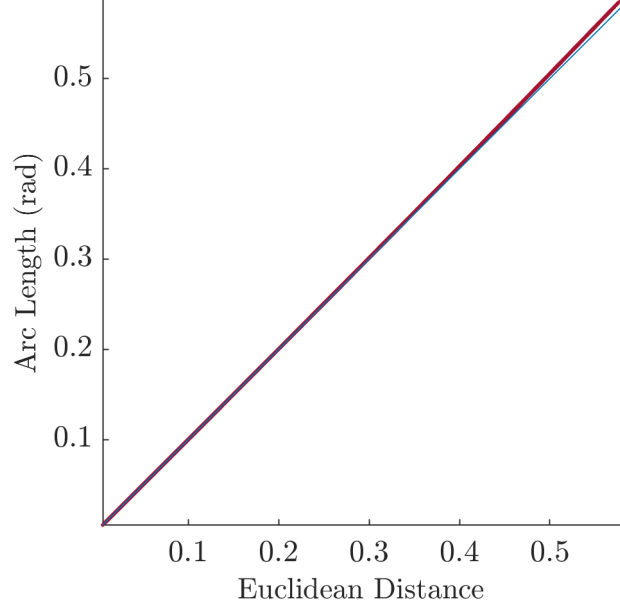


Figure S1: Parity plot of 8D Cartesian hyperspherical arc length vs. 8D Cartesian Euclidean distance for pairwise distances in a ($m\bar{3}m$) symmetrized set of 10 000 randomly sampled VFZ-GBOs. The max arc length is approximately 0.58 rad, indicating a max GBO distance of approximately 1.16 rad or 66.5° between any two points in the VFZ. The close correlation between arc length and Euclidean distance supports the validity of using Euclidean distance instead of arc length in the interpolation methods. This is *separate* from the correlation between VFZ-GBO Euclidean or arc length distances with the traditional GBO distance [3].

quaternions and boundary plane normals. The result, `ypred`, is a vector of predicted/interpolated property values corresponding to the prediction grain boundaries (GBs) defined by (`qm2,nA2`).

Internally, these are converted to grain boundary octonions (GBOs) and interpolation is performed using the selected `method`. For the validation function, these can be compared to the true grain boundary energies (GBEs) `ytrue`. The methods used in this work are '`pbary`', '`gpr`', '`idw`', and '`nn`', corresponding to planar barycentric, Gaussian process regression (GPR), inverse-distance weighting, and nearest neighbor (NN) interpolation, respectively. A placeholder template with instructions for implementing additional interpolation schemes is also provided in `interp5DOF.m`. See Francis et al. [2] and `five2oct.m` [1] treatments of conversions to GBO coordinates in the passive and active senses, respectively (Appendix A).

S1. Euclidean and Arc Length Distances

The close correlation between Euclidean and arc length distances in the Voronoi fundamental zone grain boundary octonion (VFZ-GBO) sense is shown in Figure S1 using pairwise distances of 10 000 VFZ-GBOs. This justifies our use of Euclidean distance as an approximation of hyperspherical arc length (and by extension, that a scaled Euclidean distance approximates a non-symmetrized GBO distance, see Eqs. (1)–(3) of the main paper). However, comparison with the original GBO metric [2] gives overestimation for some boundaries. This is an inherent feature of the VFZ framework that can be addressed via use of the ensemble methods described in Section 2.1.3 (see also Figures 3 and 4).

Additionally, the use of an isometry equivalence relationship in Morawiec [4] in a non-VFZ sense results in identical distance results within numerical tolerance (Figure S2).

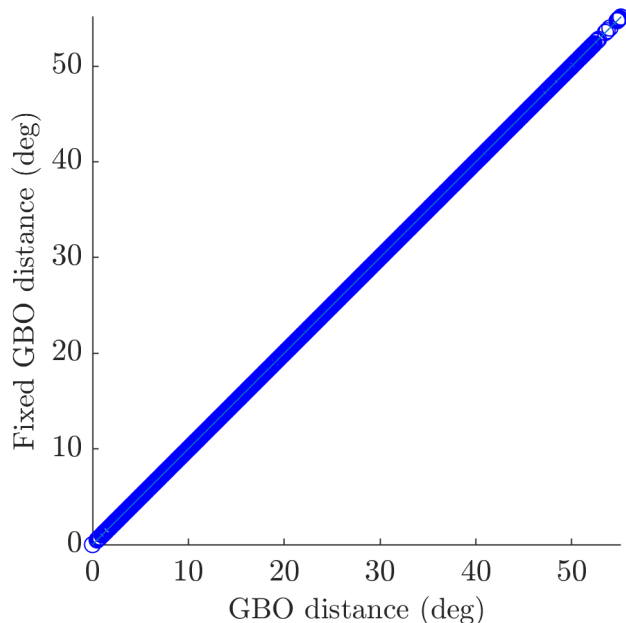


Figure S2: GB distances calculated with one GBO fixed vs. the traditional calculations in Chesser et al. [3] show that the isometry equivalence discussed in Morawiec [4] applies to GBOs. The pairwise-distance matrix for the Olmsted GBs supplied in [5] was used.

S2. Additional Interpolation Results

S2.1. Smaller Set Sizes of Input GBs

Interpolation results for 388 and 10 000 GBs are given in Figure S3 and Figure S4, respectively.

S3. Ensemble Interpolation Results

Ensemble interpolation is a classic technique that can be used to enhance predictive performance of models. Here we describe our methods (Section S3.1), results (Section S3.2), and the potential of integrating ensemble interpolation with a Gaussian process regression mixture scheme (Section S3.3).

S3.1. Methods

VFZ-GBO ensemble¹ interpolation occurs by:

1. generating multiple reference GBOs to define multiple VFZs
2. obtaining multiple VFZ-GBO representations for a set of GBs based on the various reference GBOs
3. performing an interpolation (e.g. GPR) for each of the representations
4. homogenizing the ensemble of models (e.g. by taking the mean or median of the various models)

¹Ours is a “bagging”-esque ensemble scheme because the same interpolation method (GPR) is used but with different representations for the input data.

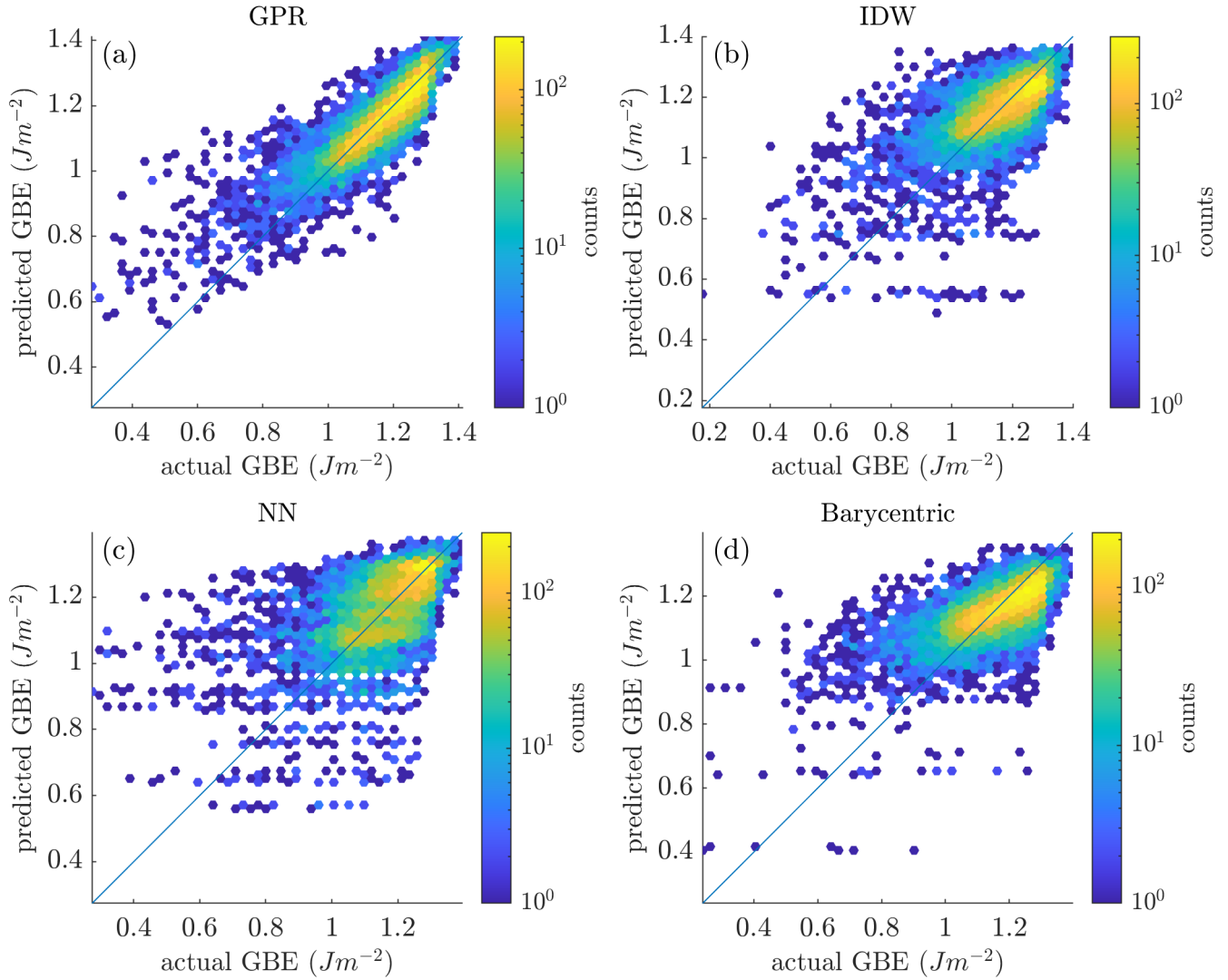


Figure S3: Hexagonally binned parity plots for 388 input and 10 000 prediction GBOs formed via pairs of a random cubochorically sampled quaternion and a spherically sampled random boundary plane normal. Interpolation via (a) GPR, (b) inverse-distance weighting, (c) NN, and (d) barycentric coordinates. Bulatov Reed Kumar GBE function for face-centered cubic Ni [6] was used as the test function.

S3.2. Results

Use of an ensemble interpolation scheme decreases interpolation error for a GPR model with 50 000 input and 10 000 prediction VFZ-GBOs. By using an ensemble size of 10 (i.e. 10 GPR models each with different reference GBOs and therefore different VFZs), root mean square error (RMSE) and mean absolute error (MAE) decreased from $0.0241 J m^{-2}$ and $0.0160 J m^{-2}$ to $0.0187 J m^{-2}$ and $0.0116 J m^{-2}$, respectively, using the median homogenization function (Figure S5).

Figure S6 shows the hexagonally binned parity plots for predictions made using the mean, median, minimum, and maximum predicted values over an ensemble of 10 VFZs. Qualitatively, the ensemble mean and ensemble median parity plots look similar to those from the main text (Figure 6), though the distributions of the ensemble scheme are somewhat tighter. The ensemble minimum produces better predictions of low GBE than any of the other models, but underestimates high GBE as expected. Naturally, the ensemble maximum overestimates in general. Diminishing returns manifest in Figure S5 for mean and median homogenizations. This is to be expected because the original GBO distances [2]

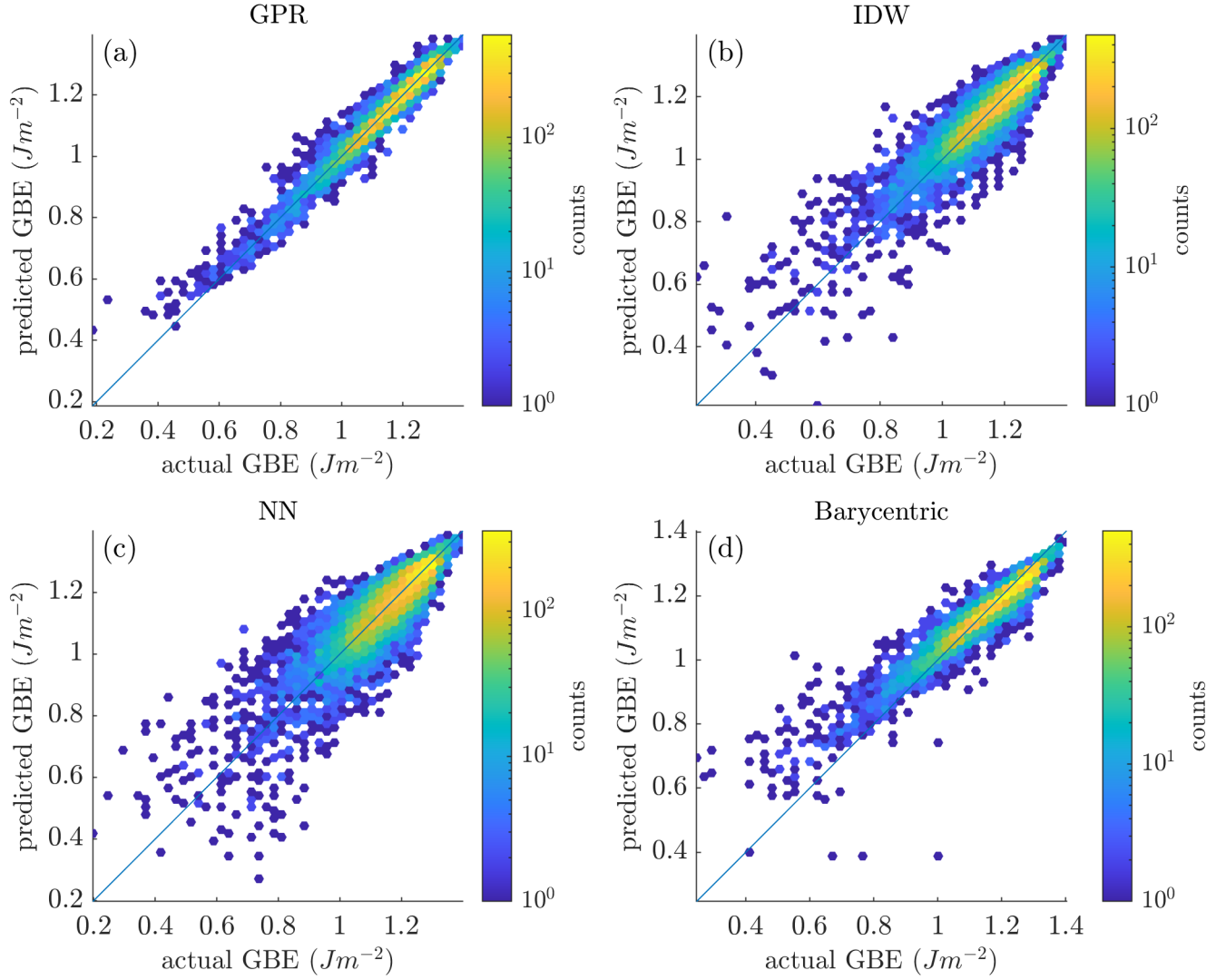


Figure S4: Hexagonally binned parity plots for 10 000 input and 10 000 prediction GBOs formed via pairs of a random cubochorically sampled quaternion and a spherically sampled random boundary plane normal. Interpolation via (a) GPR, (b) inverse-distance weighting, (c) NN, and (d) barycentric coordinates. Bulatov Reed Kumar GBE function for face-centered cubic Ni [6] was used as the test function.

are well-approximated using an ensemble size of 10 (Figure 3c and Figure 4).

S3.3. Possibility: Combining Ensemble with Gaussian Process Regression Mixture

A scheme which preferentially favors the ensemble minimum for low GBE predictions and defaults to ensemble mean or median for all other GBEs may produce even better results across the full range of GBEs. For example, this could be accomplished by combining the ensemble scheme described here with the GPR mixture model described in Section S5.1.

S4. Barycentric Interpolation

S4.1. High-Aspect Ratios

An artifact of the barycentric interpolation method which occurs due to the presence of high-aspect ratio facets is shown in Figure S7. As the dimensionality increases for a constant number of points and

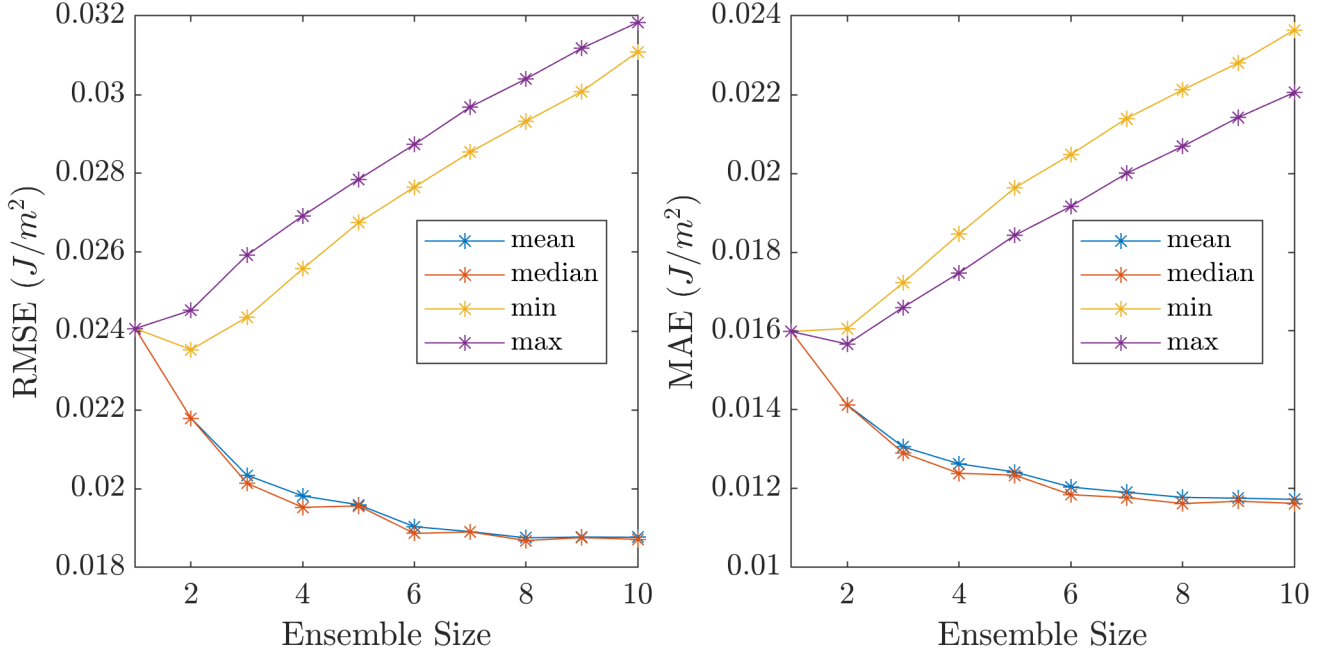


Figure S5: (a) RMSE and (b) MAE vs. ensemble size for mean, median, minimum, and maximum homogenization functions. A GPR model with 50 000 input and 10 000 prediction VFZ-GBOs was used.

from our numerical tests, the rate of missed facet intersections increases. This artifact and our method for addressing it are discussed in [Appendix B.2](#) of the main text.

S5. Kim Interpolation

A GPR mixing model is developed to accommodate the non-uniformly distributed, noisy Fe simulation data [7] and better predict low GBE. Details of the method ([Section S5.1](#)) and an analysis of the input data quality ([Section S5.2](#)) are given. The code implementation is given in `gprmix.m` and `gprmix_test.m` of the VFZ repository [1].

S5.1. Details of Gaussian Process Regression Mixture

As shown in [Figure S8a](#), prediction using the standard approach of the main document (termed the ϵ_1 model) overestimates low GBEs for this dataset. By training the model on only GBs with a GBE less than 1.2 J m^{-2} (termed the ϵ_2 model) and by using an exponential (`KernelFunction='exponential'`) rather than a squared exponential kernel, prediction of low GBEs improves, but naturally underestimation occurs for higher GBEs ([Figure S8b](#)).

A combined, disjoint model ([Figure S8c](#)) is taken (ϵ_3) by replacing ϵ_1 GBE predictions for GBs with GBE less than 1.1 J m^{-2} with the corresponding ϵ_2 predictions. Finally, a weighted average ([Eq. \(S1\)](#)) is taken according to:

$$\epsilon_{mix} = f\epsilon_1 + (f - 1)\epsilon_2 \quad (\text{S1})$$

where ϵ_1 and ϵ_2 represent the standard GPR model and the GPR model trained on the subset of GBs with a GBE less than 1.2 J m^{-2} , respectively, and f is the sigmoid mixing fraction given by:

$$f = \frac{1}{e^{-m(\epsilon_3 - b)} + 1} \quad (\text{S2})$$

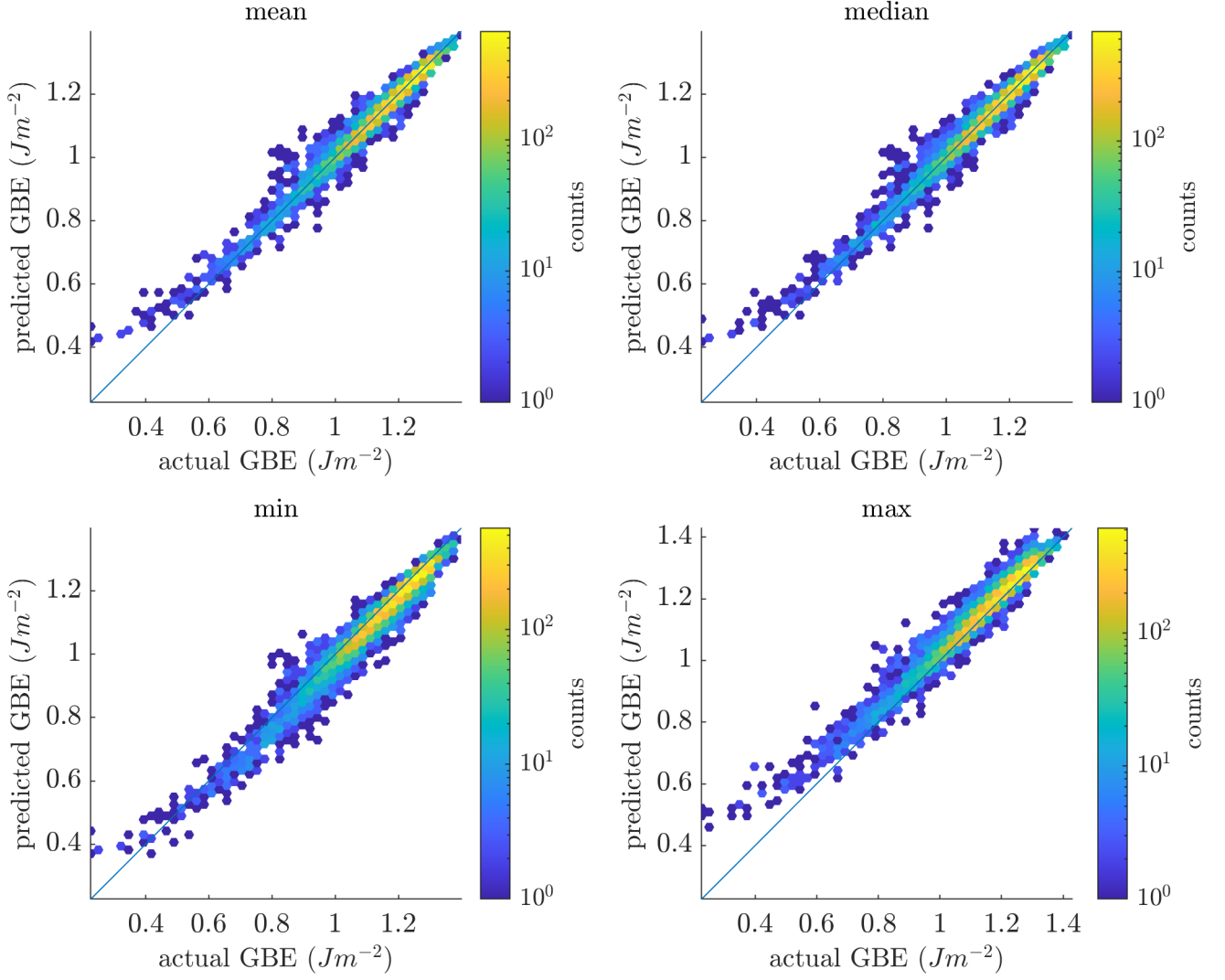


Figure S6: Hexagonally binned parity plots for (a) mean, (b) median, (c) minimum, and (d) maximum ensemble homogenization functions. A GPR model with 50 000 input and 10 000 prediction VFZ-GBOs was used.

and shown in Figure S9 with $m = 30$ and $b = 1.1 J m^{-2}$, as used in this work. Larger values of m yield a steeper sigmoid function and larger values of b shift the sigmoid function further to the right. Specific values for m and b were chosen by visual inspection and trial and error. This results in a GPR mixing model which better predicts low GBEs while retaining overall predictive accuracy (Figure S8d).

Uncertainty of the GPR mixing model is similarly obtained by taking a weighted average of the uncertainties of each model according to:

$$\sigma_{mix} = f\sigma_1 + (f - 1)\sigma_2 \quad (S3)$$

where σ_1 and σ_2 are the corresponding uncertainties of ϵ_1 and ϵ_2 , respectively, and f is given by Eq. (S2).

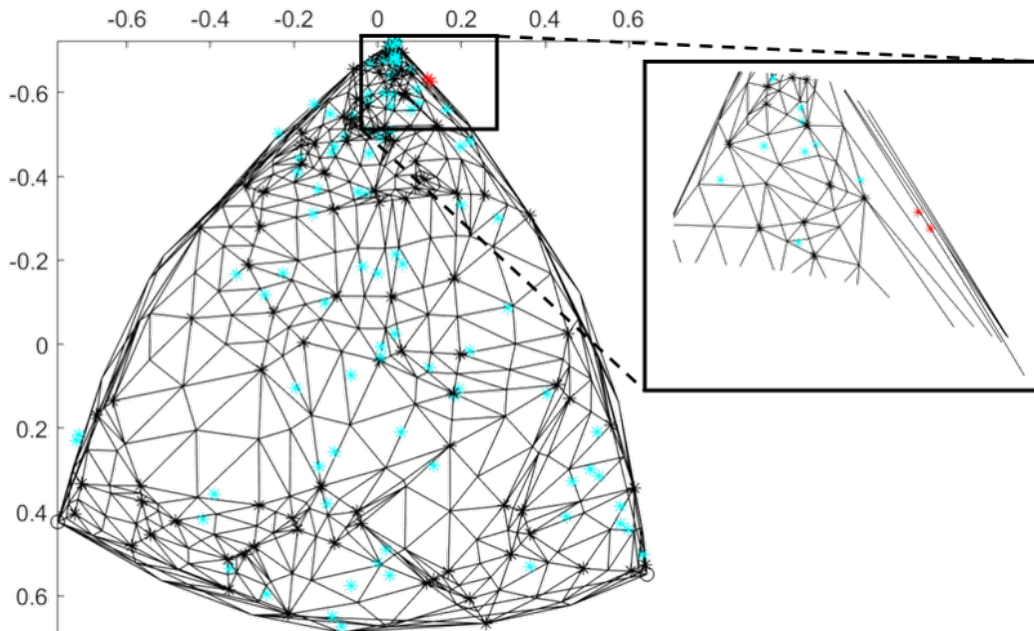


Figure S7: Illustration of two prediction points (red) for which no intersecting facet is found due to being positioned within a high-aspect ratio facet. The inset shows that facets connected to the NN do not contain the prediction point. Many NNs would need to be considered before an intersection is found. Additionally, it is expected that if found, the interpolation will suffer from higher error due to use of facet vertices far from the interpolation point. Proper intersections of prediction points with the mesh are shown in blue.

S5.2. Input Data Quality

Of the $\sim 60\,000$ ² GBs in [7], $\sim 10\,000$ GBs were repeats that were identified by converting to VFZ-GBOs and applying VFZ repository function `avg_repeats.m`. In [7], mechanically selected GBs were those which involved sampling in equally spaced increments³ for each five degree-of-freedom parameter, and a few thousand intentionally selected GBs (i.e. special GBs) were also considered. Of mechanically and intentionally selected GBs, 9170 and 112 are repeats, respectively, with a total of 2496 degenerate sets⁴ (see Figure S10 for a degeneracy histogram). Thus, on average there is a degeneracy of approximately four per set of degenerate GBs.

By comparing GBE values of (unintentionally⁵) repeated GBs in the Fe simulation dataset [7], we can estimate the intrinsic error of the input data. For example, minimum and maximum deviations from the average value of a degenerate set are -0.2625 J m^{-2} and 0.2625 J m^{-2} , respectively, indicating that a repeated Fe GB simulation from [7] can vary by as much as 0.525 J m^{-2} , though rare. Additionally, RMSE and MAE values can be obtained within each degenerate set by comparing against the set mean. Overall RMSE and MAE are then obtained by averaging and weighting by the number of GBs in each degenerate set. Following this procedure, we obtain an average set-wise RMSE and MAE of $0.065\,29\text{ J m}^{-2}$ and $0.061\,90\text{ J m}^{-2}$, respectively, which is an approximate measure of the intrinsic error of the data. Figure S11 shows histograms and parity plots of the intrinsic error. The overestimation of intrinsic error mentioned in the main text (Section 3.4) could stem from bias as to what type of

²The “no-boundary” GBs (i.e. GBs with close to 0 J m^{-2} GBE) were removed before testing for degeneracy.

³In some cases, this was equally spaced increments of the argument of a trigonometric function.

⁴A degenerate “set” is distinct from a VFZ-GBO “set”, the latter of which is often used in the main text.

⁵To our knowledge, the presence of repeat GBs were not mentioned in [7] or [8]

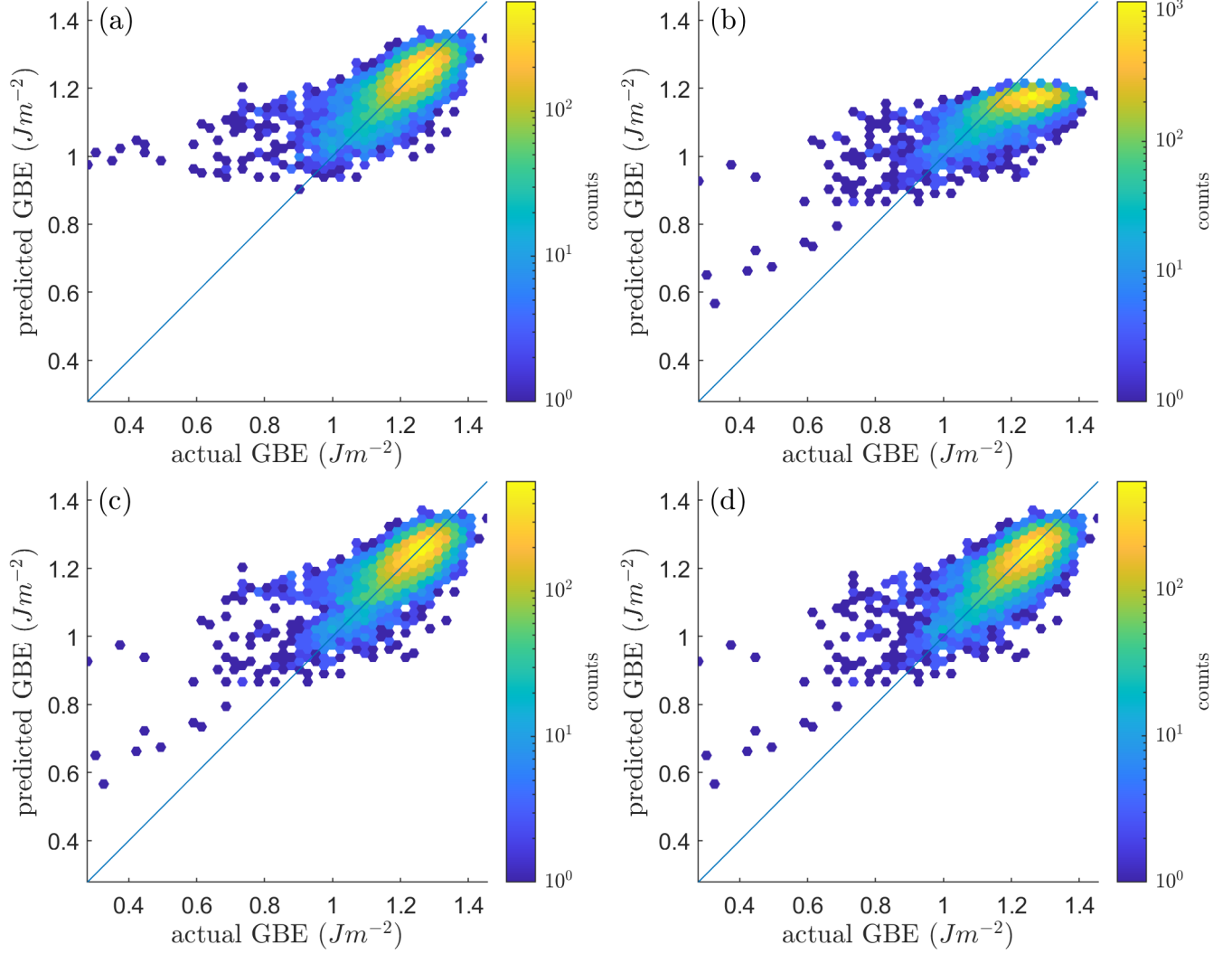


Figure S8: (a) Hexagonally binned parity plot of the standard GPR model. (b) All prediction GBs based on the model using only training GBs with a GBE less than $1.2 Jm^{-2}$. (c) Combined disjoint model as explained in the text. (d) Hexagonally binned parity plots of the final GPR mixing model. Points in (c) are produced by splitting the prediction data into less than and greater than $1.1 Jm^{-2}$. A sigmoid mixing function (Figure S9) is then applied where the predicted GBEs shown in (c) determines the mixing fraction (f) to produce a weighted average of models (a) and (b). A large Fe simulation database [7] using 46 883 training datapoints and 11 721 validation datapoints in an 80%/20% split. The GPR mixture model decreases error for low GBE and changes overall RMSE and MAE from $0.057 932 J m^{-2}$ and $0.039 857 J m^{-2}$ in the original model (shown in (a)) to $0.055 035 J m^{-2}$ and $0.039 185 J m^{-2}$ (shown in (d)), respectively.

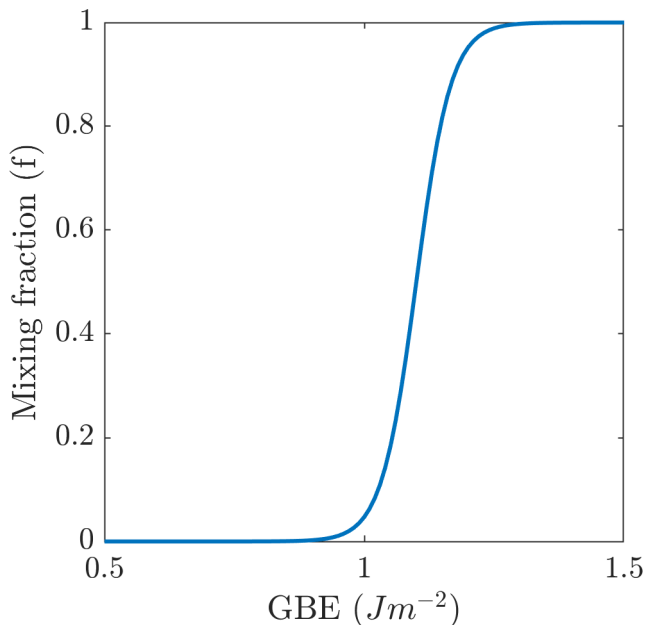


Figure S9: Sigmoid mixing function used in the GPR mixing model with $m = 30$ and $b = 1.1 \text{ J m}^{-2}$ (Eq. (S2)).

GBs exhibit repeats based on the sampling scheme used in [7] and/or that many of the degenerate sets contain a low number of repeats (Figure S10).

Next, we see that by binning GBs into degenerate sets, most degenerate sets have a degeneracy of fewer than 5 Figure S10. We split the repeated data into sets with a degeneracy of fewer than 5 and greater than or equal to 5 and plot the errors (relative to the respective set mean) in both histogram form (Figure S11a and Figure S11c, respectively) and as hexagonally-binned parity plots (Figure S11b and Figure S11d, respectively). While heavily repeated GBs tend to give similar results, occasionally repeated GBs often have larger GBE variability. This could have physical meaning: Certain types of (e.g. high-symmetry) GBs tend to have less variation (i.e. fewer and/or more tightly distributed metastable states). However, it could also be an artifact of the simulation setup that produced this data (e.g. deterministic simulation output for certain types of GBs).

S6. Olmsted Interpolation

As illustrated in Figure S12, leave-one-out cross validation interpolation results for 0 K molecular statics low-noise Ni simulations using the GPR method are similar to Laplacian kernel regression results reported in Figure 6a of Chesser et al. [3] (reproduced on the right of Figure S12 for convenience).

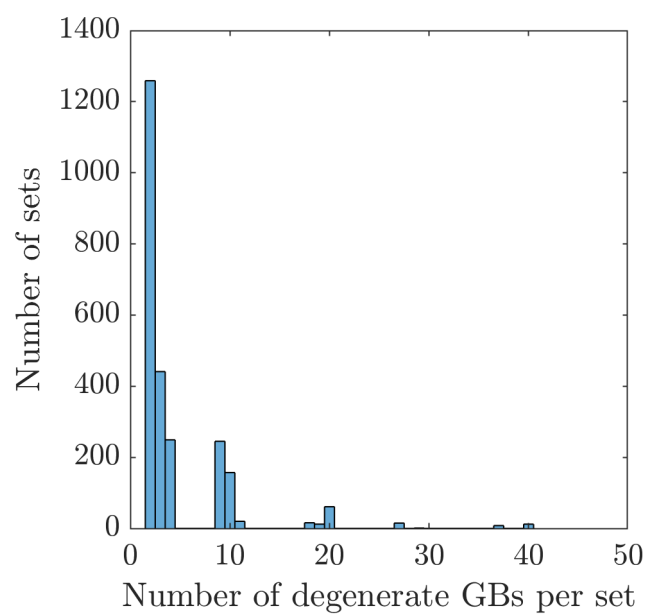


Figure S10: Histogram of number of sets vs. number of degenerate GBs per set for the Fe simulation dataset [7]. Most sets have a degeneracy of fewer than 5.

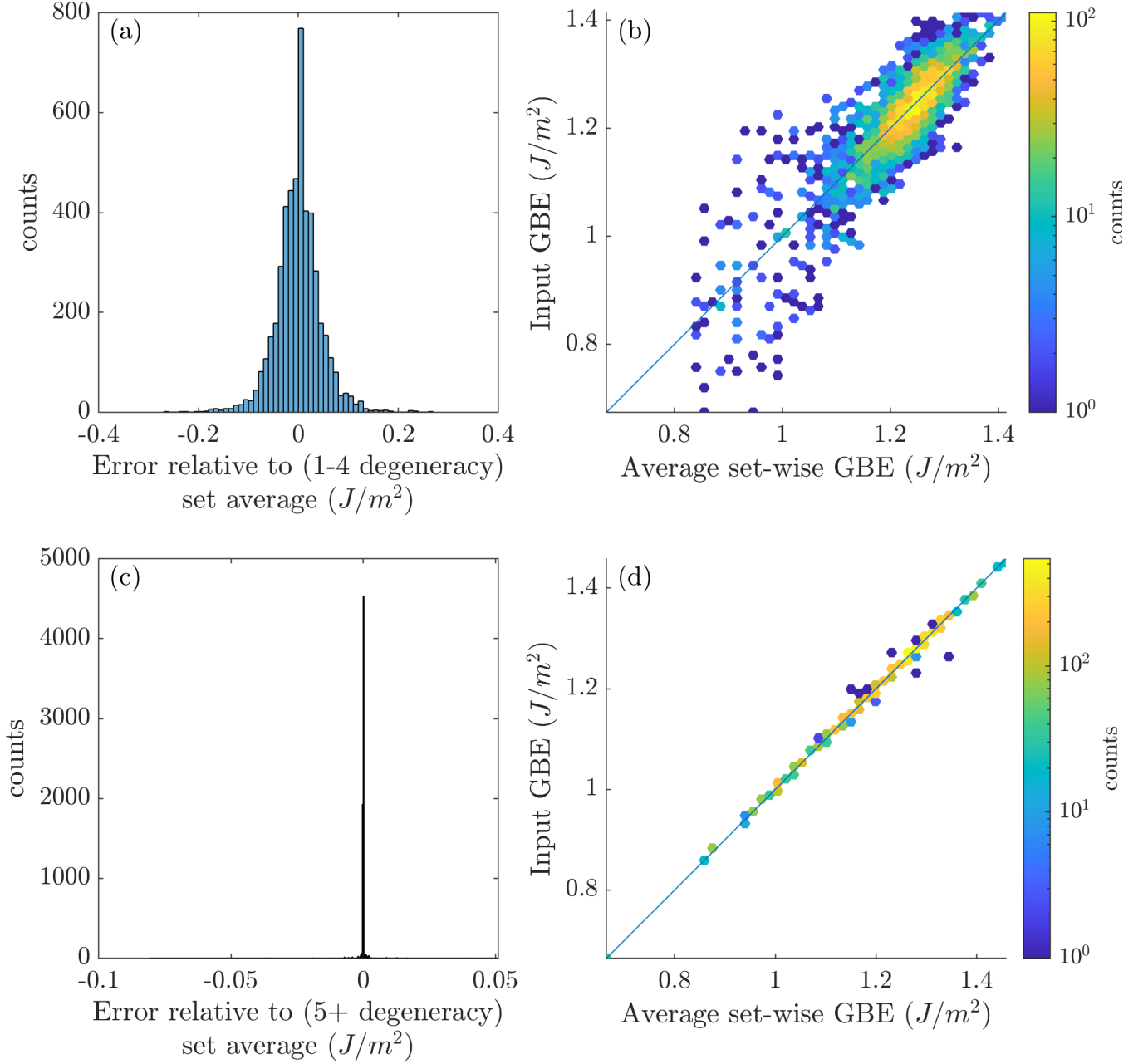


Figure S11: Degenerate GBs sets are split into those with a degeneracy of fewer than 5 and greater than or equal to 5 and plotted as ((a) and (c), respectively) error histograms and ((b) and (d), respectively) hexagonally-binned parity plots. Large degenerate sets tend to have very low error, whereas small degenerate sets tend to have higher error. In other words, GBs that are more likely to be repeated many times based on the sampling scheme in [7] tend to give similar results, whereas GBs that are less likely to be repeated often have larger variability in the simulation output. We do not know if this has physical meaning or is an artifact of the simulation setup.

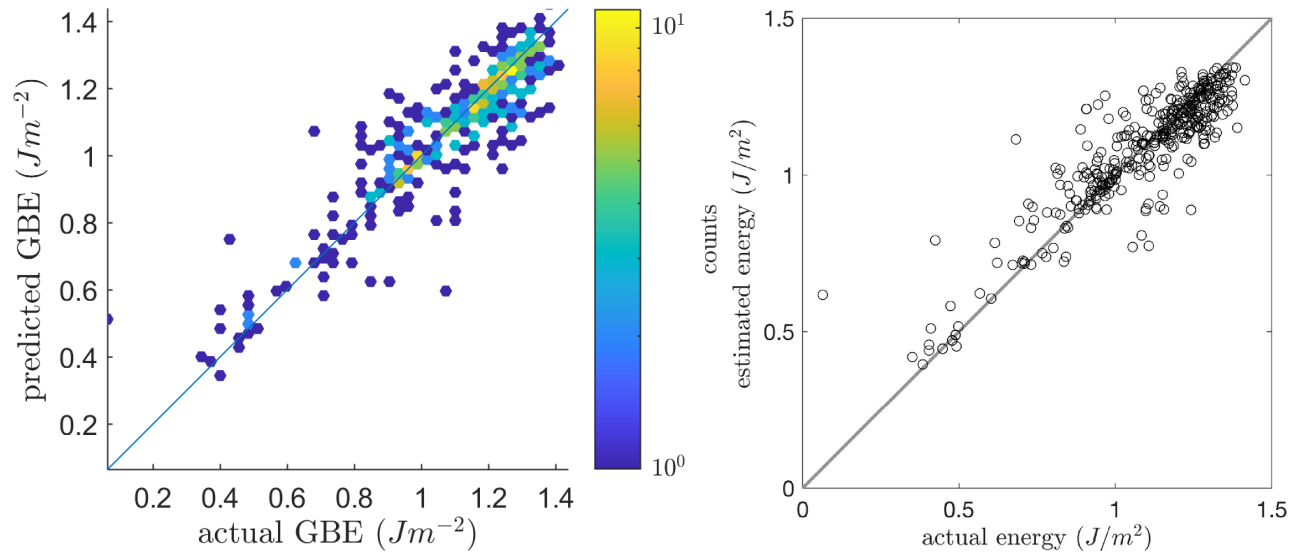


Figure S12: (left) Hexagonally binned parity plot for Ni simulation grain boundary energy (GBE) interpolation using LOOCV. (right) Parity plot for leave-one-out cross validation (LOOCV) interpolation results reproduced from Figure 6a of Chesser et al. [3] under CC-BY Creative Commons license.

Glossary

GB grain boundary [2](#), [3](#), [6](#), [8–12](#)

GBE grain boundary energy [2](#), [4–10](#), [13](#)

GBO grain boundary octonion [2–5](#)

GPR Gaussian process regression [2–7](#), [9](#), [10](#)

LOOCV leave-one-out cross validation [13](#)

MAE mean absolute error [4](#), [6](#), [8](#), [9](#)

NN nearest neighbor [2](#), [4](#), [5](#), [8](#)

RMSE root mean square error [4](#), [6](#), [8](#), [9](#)

VFZ Voronoi fundamental zone [1–4](#), [6](#), [8](#)

VFZ-GBO Voronoi fundamental zone grain boundary octonion [2–4](#), [6–8](#)

References

- [1] S. Baird, O. Johnson, Five Degree-of-Freedom (5DOF) Interpolation, 2020. URL: github.com/sgbaird-5dof/interp.
- [2] T. Francis, I. Chesser, S. Singh, E. A. Holm, M. De Graef, A geodesic octonion metric for grain boundaries, *Acta Materialia* 166 (2019) 135–147. doi:[10.1016/j.actamat.2018.12.034](https://doi.org/10.1016/j.actamat.2018.12.034).
- [3] I. Chesser, T. Francis, M. De Graef, E. Holm, Learning the grain boundary manifold: Tools for visualizing and fitting grain boundary properties, *Acta Materialia* 195 (2020) 209–218. doi:[10.1016/j.actamat.2020.05.024](https://doi.org/10.1016/j.actamat.2020.05.024).
- [4] A. Morawiec, On distances between grain interfaces in macroscopic parameter space, *Acta Materialia* 181 (2019) 399–407. doi:[10.1016/j.actamat.2019.09.032](https://doi.org/10.1016/j.actamat.2019.09.032).
- [5] I. Chesser, GB Octonion Code, 2019. URL: https://github.com/ichesser/GB_octonion_code.
- [6] V. V. Bulatov, B. W. Reed, M. Kumar, Grain boundary energy function for fcc metals, *Acta Materialia* 65 (2014) 161–175. doi:[10.1016/j.actamat.2013.10.057](https://doi.org/10.1016/j.actamat.2013.10.057).
- [7] H.-K. Kim, S. G. Kim, W. Dong, I. Steinbach, B.-J. Lee, Phase-field modeling for 3D grain growth based on a grain boundary energy database, *Modelling and Simulation in Materials Science and Engineering* 22 (2014) 034004. doi:[10.1088/0965-0393/22/3/034004](https://doi.org/10.1088/0965-0393/22/3/034004).
- [8] H. K. Kim, W. S. Ko, H. J. Lee, S. G. Kim, B. J. Lee, An identification scheme of grain boundaries and construction of a grain boundary energy database, *Scripta Materialia* 64 (2011) 1152–1155. doi:[10.1016/j.scriptamat.2011.03.020](https://doi.org/10.1016/j.scriptamat.2011.03.020).