Reviewer #1: The authors develop a framework for approximating grain boundary geodesic distance kernels that is more computationally efficient than prior work without significantly impacting the accuracy of property regression on reference datasets. Two ideas central to the efficiency of the framework are 1) the approximation of geodesic GB distances with a Euclidean metric and 2) the construction of a fundamental zone for GBs based on Voronoi tessellations on the 7-sphere. It is argued that the distance approximation errors arising from (2) are small and insignificant for property prediction. Out of several regression methods explored in the paper, the authors recommend Gaussian Process Regression (GPR) for GB property interpolation because of its relative accuracy and built-in uncertainty quantification.   
  
The main impact of this work, in my view, is that it provides a method for GB property interpolation that scales to large GB datasets. The structure of the GB character space, and the way in which properties vary over that space, are complex and not well understood yet, and methods which help advance scientific efforts in this direction are welcome.   
  
I do have several comments and concerns. My primary concern about this paper is its atypical and somewhat confusing organization, addressed in points 1-4 below. One theoretical comment is given in point 5. Finally, several typos are identified in points 6.   
  
1. Several sections seem to overlap in content and should be merged or reorganized, such as "3.1.3 Our Four Interpolation Methods" and "3.2.1 Comparison of Models".

2. The narrative is difficult to read sequentially: sections do not clearly transition from one to another, sections do not have topic sentences summarizing main points.

3. The Introduction starts with an outline, where claims about a "highly efficient" model seem premature. Consider moving the outline to the end of the Introduction.

This paragraph has been moved to the end of the Introduction.

4. Consider relegating details of the codebase to the Supplementary Information. This is a suggestion meant to lead to a better reading experience.

We appreciate this suggestion. While we consider the development of our approach and the results of this work meaningful on their own, we believe that the full value is realized through use of the extensive codebase developed around this metric. We wish to emphasize that not only are the distance calculations and interpolation methods efficient, but that these and other methods are available in functions with well-defined purposes. We worry that relegation to the Supplementary Information will cause this important aspect of the work to be overlooked or diminished.

5. In Section 2.1.1, an example is given for VFZ construction on the 2-sphere. It seems that this construction should be related to the standard stereographic triangle, a space with which materials scientists are already familiar. Is there a relationship between these constructions?

With the standard stereographic triangle (SST), the fundamental zone will naturally have three vertices. While the SST bears resemblance to the VFZ construction on the 2-sphere, these are distinct. One difference is that the 2-sphere VFZ has 5 vertices. Another aspect is that O­h cubic point group proper rotations are used as the symmetry operators. A non-crystallographic set of symmetry operators could have been used to illustrate the VFZ. The primary of purpose of the 2-sphere construction was to help with visualization of higher-dimensional hyperspheres and illustrate key similarities (e.g. crossing the border of a VFZ).

6. Typos:   
Figure 1 caption, "fixed reference point": no comma needed  
Figure 1 caption, insert comma before "demonstrating"

This has been updated in the manuscript.

Reviewer #2: See “Summary of Comments on A-21-0401-Review 2 Attachment.pdf”

Reviewer #3: This paper was interesting for this reviewer to read, but I am not sure it will find its readership if published in Acta Mat. It is rather atypical in its focus on algorithmic and computational aspects whereas Acta papers are traditionally more focused on physical aspects of material behavior. I understand quite well where the authors are coming from but, really, will a typical Acta reader go into all or any of the technical details presented here? I am not necessarily recommending doing it, but if I were its author, I'd rather submit a brief but logical summary of this work to Scripta and a full article to some other journal more directly focused on data sciences.   
  
Having wrote all that above, I find the logic and the methods presented here elegant and potentially quite useful. Especially nice is the idea to map all symmetry equivalent representations of a given GB to a single point inside the Voronoi volume. Then, the approximate euclidization of the resulting space is making interpolations so much faster. On its technical content, it is a high quality paper. And very well written. The authors even post their user friendly [matlab] functions that can be used by others to much benefit.  
  
And a couple of miscellaneous comments:  
  
1. Unless I missed it somehow, the authors are not giving a comparison of their, say, GPR predictions for Ni against the original computed data of Olmstead. Instead, in Fig. 6 they only compare their four interpolation functions with the BRK function. But the BRK validation function is an interpolation function itself so, while it is obviously useful for generating synthetic training data, it is smoother by construction than the atomistic data set to which it was fitted. It would be of interest to compare BRK and GPR interpolations directly to Olmstead's atomistic data.

In “Results and Discussion” -> “Literature Datasets” -> “Comparison with Prior Work”, Figure S9 is referenced. This gives a direct comparison between the GPR interpolation results and the interpolation results presented in Figure 6a of Chesser et al. [49]. To better highlight this, this figure has been moved into the main body. Additionally, we add a parity plot for BRK interpolation relative to the atomistic data. Unfortunately, it will be difficult for us to implement leave-one-out cross-validation, which would provide the most comparable results across the board.  
  
2. Is it really important to be able to fit to 50,000 boundaries? Given uncertainties in experimental and computational determination of the GB energy, how much really gained by sampling so many more GBs than 388 in Olmsted's dataset?