A High-Throughput Computational Dataset of Halide Perovskite Alloys[†]

Jiaqi Yang[a], Panayotis Manganaris[a], and Arun Mannodi-Kanakkithodi[a]

Abstract

Novel halide Perovskites with improved stability and optoelectronic properties can be designed via composition engineering at cation and/or anion sites. Data-driven methods, especially high-throughput first principles computations and subsequent analysis based on unique materials descriptors, are key to achieving this goal. In this work, we report a Density Functional Theory (DFT) based dataset of 495 ABX_3 halide Perovskite compounds, with various atomic and molecular species considered at A, B and X sites, and different amounts of mixing considered at each site generated using the Special Quasirandom Structures (SQS) algorithm for alloys. We perform GGA-PBE calculations on pseudo-cubic Perovskite structures to determine their lattice constants, stability in terms of formation and decomposition energies, electronic band gaps, and properties extracted from optical absorption spectra. To elucidate the importance of the level of theory used, we further perform 299 calculations using the more expensive HSE06 functional and determine lattice constant, stability and band gap, and compare PBE and HSE06 properties with some experimentally measured results. Trends in the datasets are unraveled in terms of the effects of mixing at different sites, the composition in terms of specific atomic or molecular species, and averaged elemental properties of species at different sites. This work presents the most comprehensive DFT perovskite alloy dataset to date and the data, which is open-source, can be exploited to train predictive and optimization models for accelerating the design of completely new compositions that may yield large solar cell efficiencies and improved performance across many optoelectronic applications.

TODO Introduction

- State ”DONE” from [2022-06-05 Sun 18:16]

The challenge of optimizing Perovskite performance is one with many facets. Almost every detail of a Perovskite crystal’s structure and chemistry effects its performance as a semiconductor. The size of the unit cell effects its substrate affinity and in turn its carrier concentrations[1]. The crystal’s phase effects many aspects of the electronic structure, including the band gap and optical response. Of course, these qualities of a structure are largely dependent on the specification, proportions and arrangements of the constituent elements.

We report a synthetic dataset collected for 495 chemically distinct, pseudo cubic Halide Perovskites. This dataset builds on that of 229 samples which formed the foundation of the prior work by Mannodi-Kanakkithodi and Chan[2]. The DFT computed properties we collected and the levels of theory used are discussed in ??. The relatively large size of this dataset is intended to provide an initial sampling suitable for guiding exploration of the alloy space. Structural information is considered constant to better focus on obtaining physically meaningful interpretations of models dealing only with information derived from a sample composition.

TODO Methodology

NEXT Building Perovskite Dataset

- State ”NEXT” from ”TODO” [2022-06-05 Sun 22:11]

The dataset we report is based on standard cubic phase ABX_3 Perovskite structures obtained from public databases. Fourteen common Perovskite constituents are selected to form our Halide Perovskite composition space 1. Five constituents including Methylammonium and Formamidinium cations represent the possible A-site occupants. Six metals represent the possible B-site occupants. Three halides represent the possible X-site occupants. In total, these component vectors form a constrained 14 dimensional space (Figure 2) within which all Perovskite compounds consisting of the elements in Figure 1 (a) must exist.

The pure (non-alloyed) possibilities are exhaustively sampled using $5 * 6 * 3 = 90$ Perovskites. Based on these pure Perovskite structures, we mix candidates for A, B, and X sites systematically. The alloy space sees combinatorial scaling and must be sparsely sampled ??.

To generate site-mixing structures with high representativeness, special quasi-random structures (SQS) are applied. The SQS method is to build a special periodic structure and make the first nearest-neighbor shells as similar to

[a] School of Materials Engineering, Purdue University, West Lafayette, IN 47907, USA; E-mail: amannodi@purdue.edu

[†] Electronic Supplementary Information (ESI) available: https://www.github.com/PanayotisManganaris/REPO$_{T}$*ODO.SeeDOI* : 00.0000/00000000.

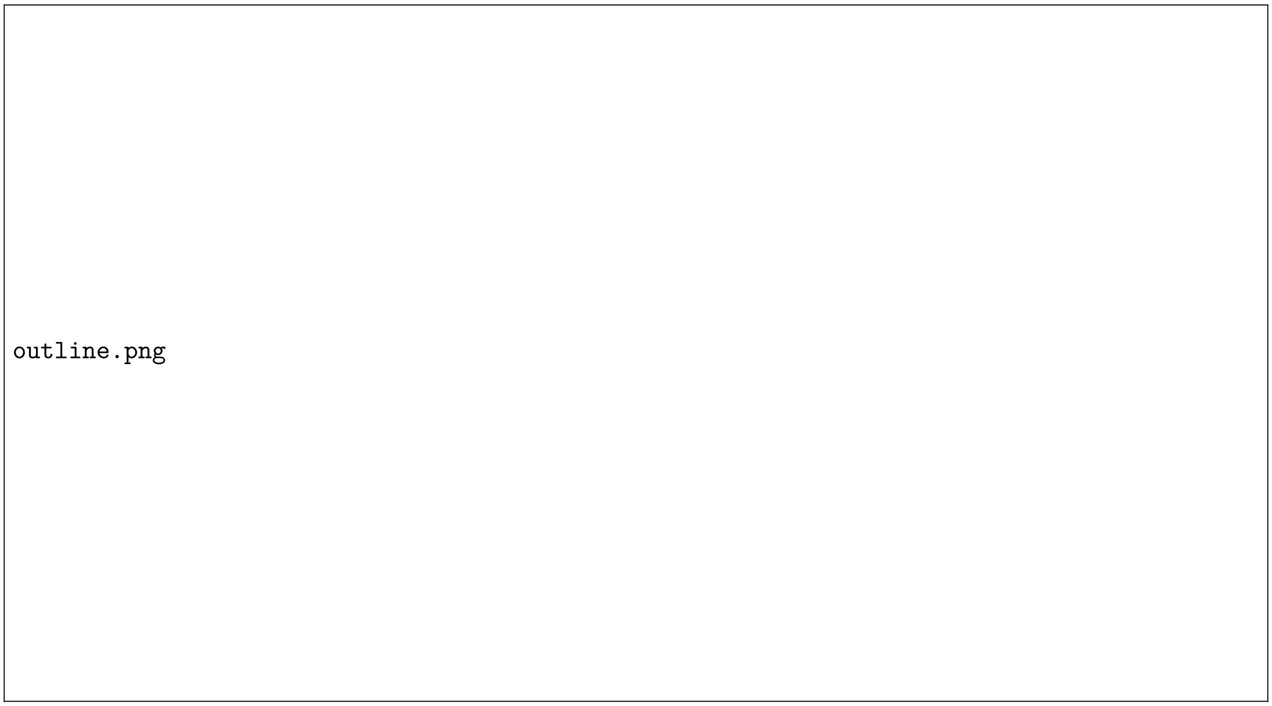

Fig. 1 (a) Chemical space of ABX_3 perovskites. (b) Number of samples representing each kind of primary alloy. (c) Detailed outline.

the target random alloy as possible. The SQS can be considered as the best possible periodic unit cell representing a given random alloy.

Each computational run is performed using a 2x2x2 supercell, this allows A and B site doping to be modeled in discrete $1/8^{\mathrm{th}}$ fractions of the total site occupancy, and it allows X site doping to be modeled in $1/24^{\mathrm{th}}$ fractions. At these mixing levels, it is appropriate to call all these Perovskites alloys.

following the process above, 126 A-site mixing samples, 151 B-site mixing samples and 127 X-site mixing samples are generated. All resulting structures are optimized using a DFT variable-cell relaxation under (pbe). These same initial structures also underwent a full (hse) relaxation to help ensure the validity of the PBE relaxations, however only 299 were computationally tractable.

DONE Calculation Details

- State "DONE" from "TODO" [2022-06-06 Mon 17:17]

DFT calculations are performed with ?? version 6.2. The paw potentials were used as pseudopotentials. The generalized gradient approximation gga of pbe and the hybrid hse (α=0.25 and ω=0.2) functionals are used as exchange-correlation energy. The energy cutoff for the plane-wave basis is set to 500 eV. For pbe the Brillouin zone was sampled by 6x6x6 reciprocal mesh using the Monkhorst-Pack k-point mesh. For hse the Brillouin zone was sampled by 2x2x2 reciprocal mesh using the Monkhorst-Pack k-point mesh. The structural force convergence threshold is set to be 0.005 eV/Å.

TODO Discussion of DFT Computed Properties

- State "DONE" from "TODO" [2022-06-06 Mon 17:19]

DONE Decomposition Energy

- State "DONE" from "TODO" [2022-06-06 Mon 00:43]

The decomposition energy indicates the stability of a compound. To calculate the decomposition energy for ABX3 perovskite, we assume it will decompose to two phases, AX and BX2. Using DFT calculations, we can get the optimized energy of a Perovskite and that of its constituent phases. The decomposition energy is calculated using equation (1). This calculation is performed separately for each level of theory.

$$E_{decomp} = E_{opt}(ABX_3) - E_{opt}(AX) - E_{opt}(BX2) \qquad (1)$$

TODO ?? calculations

- State "DONE" from "TODO" [2022-06-06 Mon 17:30]

The ?? metric developed by Yu and Zunger[3] is used as the primary criterion screening perovskites for their photovoltaic merits. The SLME value is computed considering a 5μm absorption layer for every Perovskite according to ().

Fig. 2 Plots showing number of Perovskites representing a constituent at a certain atomic fraction of a complete Perovskite.

(2)

This calculation is performed using Logan William's SL3ME[4]. Based on absorption spectra obtained from from VASP. This calculation is performed separately for the pbe band gaps and the hse band gaps, resulting in two synthetic efficiencies for each record.

TODO Results and discussion

TODO Visualization of DFT Data

DONE Lattice constant PBE vs HSE

- State "DONE" from [2022-06-06 Mon 01:02]

Fig 4 presents the lattice parameter comparison of ?? calculation and HSE calculations. Full relaxations at both levels of theory mostly agree on the lattice parameters. At least, any deviation appears to be very linearly explained. This suggests the accuracy of PBE relaxation is enough to optimize most Perovskite samples.

Note this also served as a reasonable validation for our results. A few samples did have significantly differing lattice parameters. This prompted checking the optimized structures. We found those Perovskite structures were substantially deformed and no longer had obvious octahedral structures. Thus, we exclude these outliers from any analysis concerned with the dominant pseudo cubic structures which are the focus of this report.

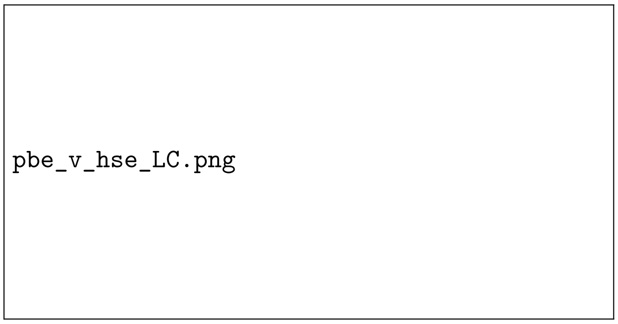

Fig. 4 Comparing lattice constants obtain by full PBE and HSE relax calculations

DONE Comparing synthetic to physical data

- State "DONE" from [2022-06-06 Mon 01:01]

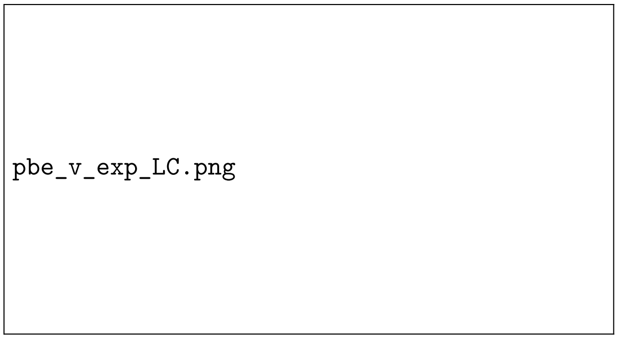

Figure2.png

Fig. 3 DFT Results: PBE and HSE properties; lattice constants, decomposition energies, band gaps, and converter efficiencies.

Fig. 5 Comparison of PBE and HSE computed pseudo-cubic lattice constants with crystallographic measures for lattice constants

hse_v_exp_LC.png

Fig. 6 Comparison of PBE and HSE computed pseudo-cubic lattice constants with crystallographic measures for lattice constants

The physical data used in comparison is collected from the works of Jiang et al.[1], Briones et al.[5], Chen et al.[6].

TODO Decomposition energy vs band gap PBE and HSE

Fig 2(b) is showing the PBE band gap compared to the PBE decomposition energy. It presents the diversity of our perovskite dataset. The data sets cover a large range of band gap and decomposition energy. For example, we have perovskite with low decomposition energy (good stability) and suitable band gap value (between 1 eV to 2.5 eV for PBE calculations). And we can also find samples with low stability and large band gap. The distribution of band gap and decomposition energy shows a great diversity of all perovskite samples and indicates that our data set can statistically represent a sufficient perovskite space. Fig 2(c) shows the plot of HSE band gaps and HSE decomposition energy. Since we applied HSE calculations for part of the samples, it shows some grouping on low decomposition energy. Compared to PBE plot, more data should be added in high decomposition energy region and suitable band gap region.

TODO Spectroscopic Limited Maximum Efficiency (SLME) vs PBE Band gap

Fig 2(d) presents the Spectroscopic Limited Maximum Efficiency (SLME) values related to the PBE band gap. Spectroscopic Limited Maximum Efficiency (SLME) is a very important properties for photovoltaic performance. SLME measures the absorption efficiency of light for the perovskite. As Fig 2(d) showing, a peak around 1.5 eV is obvious. The peak indicates that these samples with 1.5 eV PBE band gap will also have best absorption efficiency as photovoltaic materials. As the band gap increases, the SLME value decreases and eventually goes to zero due to the high band gap values.

PBE_v_comp_pearson2.png

Fig. 7 Pearson linear correlation coefficients between 50 composition and elemental descriptors and (a) 6 PBE computed properties, and (b) 4 HSE computed properties.

PBE_v_site_prop_pearson.png

Fig. 8 Pearson linear correlation coefficients between 50 composition and elemental descriptors and (a) 6 PBE computed properties, and (b) 4 HSE computed properties.

TODO Pearson Correlation Results

It is unlikely that any of the targets is fully explained by a single composition or composition derived axis. But there are helpful relations that aid in obtaining a physical understanding.

as in a Pearson correlation map is produced to check for strong relations. Those that exist, when plotted in detail show some trending, but always with extensive variability. Evidently, an accurate model will have to be formed on a multidimensional domain.

TODO PCA

Principal Component Analysis is a method of projecting high dimensional data onto a plane defined by the two linear combinations of axes that explain as much of the variance as possible.

The method of PCA is the Singular Value Decomposition, a Unitary Transform which generalizes the familiar eigendecomposition. PCA will "rotate" the N data points in M-D space until their widest 2D cross section is displayed.

At this point it is readily apparent that this dataset is highly topological. The data exists on a mostly bounded domain in high dimensions, so there is some geometry the features constitute.

Our models will prefer to use this this geometric structure in their explanation for why perovskite properties vary, this can be useful for accuracy, it can also be a bias-inducing hindrance.

TODO t-SNE visualization of stability clustering

DONE Screening process of all samples in database

- State "DONE" from [2022-05-24 Tue 13:31]

The synthetic Perovskite data is collected for sampling variety and without regard for each compound's viability. In order to extract potentially high-performing candidates for synthesis, we screen all samples according to certain constraints, thereby obtaining some Perovskites worth examining with higher Level of Theory DFT calculations and future physical experiments by brute force. This dataset can only confidently claim to sample cubic phase perovskites. Therefore, a promising candidate should have low deformation. Additionally, a lower decomposition energy combined with a maximal SLME value would be ideal for physical testing. Our screening procedure consists of cutting based on the Deviation from Cubicity metric, octahedral factor, Goldschmidt tolerance factor, Bartel tolerance factor, decomposition energy, and band gap – acting as less stringent proxy for the SLME spectrum. The following sections will discuss the details of each constraint.

Deviation from Cubicity

For all the compositions we tested in our data set, some of them will have large strain and deformation because of the combination of elements. For these largely deformed samples, they are no longer remain a cubic perovskite structure. In this section, we want to analyst how the structure is apart from the cubic perovskite structure and rule out these largely deformed samples. Firstly, we need to define deviation of cubicity. For each perovskite sample, we measure the difference between b, c lattice parameter against a lattice, showing in Equation X. If the lattice deviation of cubicity is larger than 10%, we will consider this perovskite is no longer remain a cubic perovskite and exclude it. Similarly, we also need to consider 3 angles, , and , to make sure the angle also remain 90 degrees. We calculated the difference of , and versus 90 degrees, showing in Equation X, and take this as angular deviation of cubicity. We also consider the samples that have more than 5% of angular deviation of cubicity are not cubic perovskites and exclude them. Fig 7 (d) shows the screening results for angle deviation of cubicity. SI Fig X shows the screening results for b, c lattice and , angle. Most of the samples with non-cubic perovskites structures are organic-inorganic hybrid perovskites. A large part of the excluded samples are A-mixing hybrid perovskites. It indicates that organic ligands in A site sometimes increase the lattice along some direction and make the perovskite deformed.

Octahedral factor, tolerance factor, and Bartel tolerance factor

The stability of perovskite can be predicted by using the atom radius of all components. There are 3 types of factors are usually considered, Octahedral factor, tolerance factor, and Bartel tolerance factor. The formula of these three factors are shown in equation XXX. In our screening process, we set the criteria for Octahedral factor as 0.442 – 0.895. The criteria for tolerance factor is set to be 0.813 – 1.107. The criteria for Bartel tolerance factor is set to be less than 4.18. Fig 7 (a) shows the Octahedral factor versus decomposition energy plot. Fig 7 (b) shows the tolerance factor versus decomposition energy plot. The tolerance factor shows a trend that as the tolerance factor increases, the decomposition energy decreases. Fig 7 (c) shows the Bartel tolerance factor versus decomposition energy plot. Within the criteria of Bartel tolerance factor, the decomposition energy is rapidly decreased.

Screening Results

...

comp_ratio_projection_annot.png

Fig. 9 PCA

tsne_comp_DecoE_clusters.png

Fig. 10 t-SNE

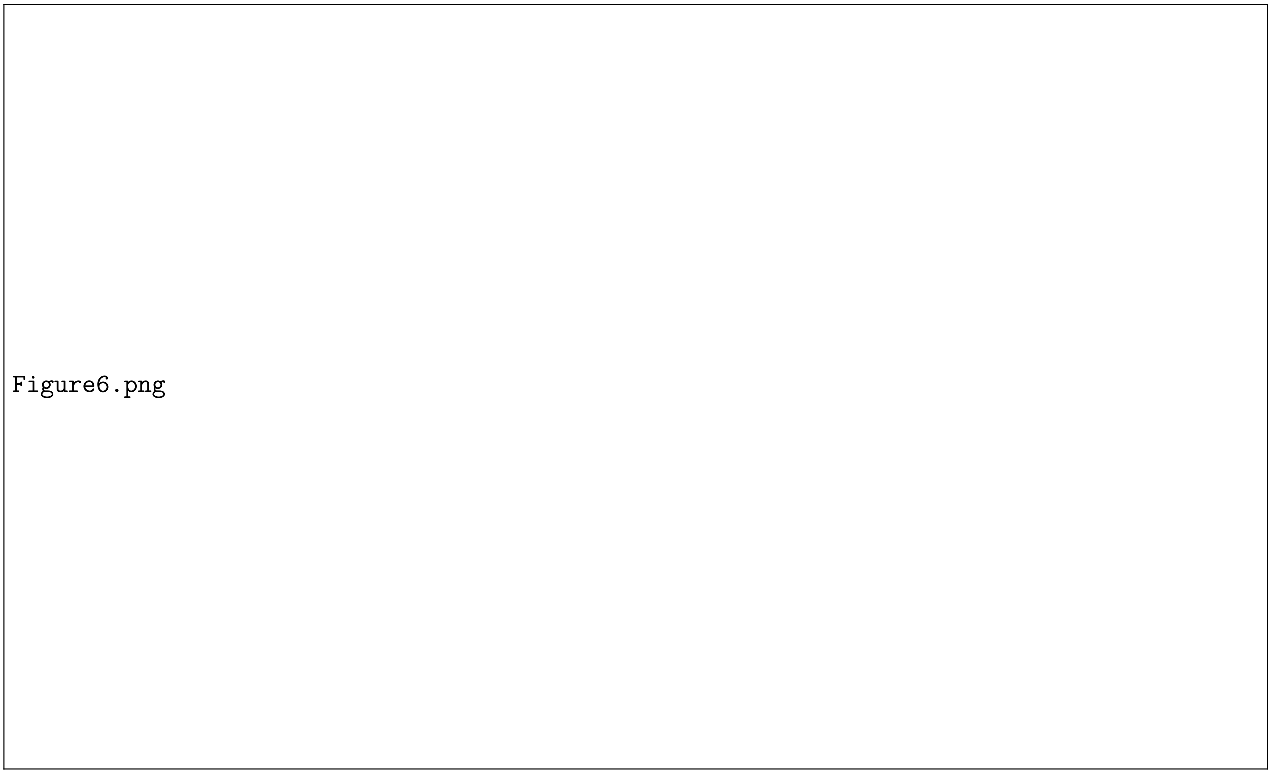

Fig. 11 placing limits on properties governing viability results in a subset of the dataset potentially suitable for physical testing.

TODO Perspective and Future Work

The design of this dataset is uniquely suited to the exploration of alloying effects on Perovskite properties. The combinatorial space of possible alloys has been sparsely but systematically sampled along four primary alloy schemes. This sample space affords the opportunity for a QM/ML surrogate model to form the basis of an active learning strategy which can begin selecting potentially high performing multi-site alloy candidates based on the current sample set.

For the calculation of future data points obtained by a surrogate optimizer, we will follow a strategy of performing full structural optimization at a PBE level of theory unless circumstances demand otherwise. This is justified by 4.

Also, we will explore methods for combining the insights of the PBE and HSE datasets in training surrogates, e.g. in section ?? the two SLME spectra could be used together to converge on a physically accurate PCE value.

We anticipate our principal challenge will be in extracting useful predictor variables from the composition information. The basic feature sets examined here are highly correlated, but nonetheless show promise both as a basic screening criterion, and as good classifier features under t-SNE transformation.

Modeling pipelines capable of predicting Perovskite decomposition energy will likely be very achievable using a transductive and invertible equivalent of the t-SNE algorithm, potentially SONG.

For this reason, kernel learning methods appear to be particularly promising for high speed optimization of the space.

Conclusions

...

Conflicts of Interest

There are no conflicts to declare.

Acknowledgements

Extensive discussions with and scientific feedback from UC San Diego researchers David Fenning and Rishi Kumar and Argonne National Lab scientist Maria Chan are acknowledged. This work was performed at Purdue University, under startup account F.10023800.05.002 from the Materials Engineering department. This research used resources of the National Energy Research Scientific Computing Center, the Laboratory Computing Resource Center at Argonne National Laboratory, and the RCAC clusters at Purdue.

References

1 L. Jiang, J. Guo, H. Liu, M. Zhu, X. Zhou, P. Wu and C. Li, Journal of Physics and Chemistry of Solids, 2006,

67, 1531–1536.

2 A. Mannodi-Kanakkithodi and M. K. Y. Chan, Energy and Environmental Science, 2021.

3 L. Yu and A. Zunger, Physical Review Letters, 2012, 108, year.

4 L. Williams, Sl3me – a Python3 Implementation of the Spectroscopic Limited Maximum Efficiency (SLME) Analysis of Solar Absorbers, https://github.com/ldwillia/SL3ME.

5 J. Briones, M. C. Guinto and C. M. Pelicano, Materials Letters, 2021, 298, 130040.

6 Q. Chen, N. D. Marco, Y. M. Yang, T.-B. Song, C.-C. Chen, H. Zhao, Z. Hong, H. Zhou and Y. Yang, Nano Today, 2015, 10, 355–396.

Supplemental Material

Glossary

This document is incomplete. The external file associated with the glossary 'main' (which should be called `main.gls`) hasn't been created.

This has probably happened because there are no entries defined in this glossary. If you don't want this glossary, add `nomain` to your package option list when you load `glossaries-extra.sty`. For example:

```
\usepackage[nomain,acronym]{glossaries-extra}
```

This message will be removed once the problem has been fixed.

Acronyms

This document is incomplete. The external file associated with the glossary 'acronym' (which should be called `main.acr`) hasn't been created.

This has probably happened because there are no entries defined in this glossary. Did you forget to use `type=acronym` when you defined your entries? If you tried to load entries into this glossary with `\loadglsentries` did you remember to use `[acronym]` as the optional argument? If you did, check that the definitions in the file you loaded all had the type set to `\glsdefaulttype`.

This message will be removed once the problem has been fixed.

HSE_v_comp_pearson2.png

HSE_v_site_prop_pearson.png