Evolution of symmetry index in minerals

Bermanec, Marko; Vidović, Noa; Gavryliv, Liubomyr; Morrison, Shaunna M.; Hazen, Robert M.

DATA ARTICLE

Geoscience Data Journal RMetS WILEY

Evolution of symmetry index in minerals

Marko Bermanec[1] | **Noa Vidović[2]** | **Liubomyr Gavryliv[3]** | **Shaunna M. Morrison[4]** | **Robert M. Hazen[5]**

[1]Institute of Geological Sciences, Universität Bern, Bern, Switzerland

[2]Department of Physics, Faculty of Science, University of Zagreb, Zagreb, Croatia

[3]Department of Mineralogy, Petrology and Economic Geology, Faculty of Natural Sciences, Comenius University, Bratislava, Slovakia

[4]Earth and Planets Laboratory, Carnegie Institution of Science, Washington, District of Columbia, USA

[5]Earth and Planets Laboratory, Carnegie Institution of Science, Washington, District of Columbia, USA

Correspondence
Marko Bermanec, Institute of Geological Sciences, Universität Bern, Baltzerstrasse 1+3, Bern 3012, Switzerland.
Email: marko.bermanec@gmail.com

Abstract

Crystal structures of minerals are defined by a specific atomic arrangement within the unit-cell, which follows the laws of symmetry specific to each crystal system. The causes for a mineral to crystallize in a given crystal system have been the subject of many studies showing their dependency on different formation conditions, such as the presence of aqueous fluids, biotic activity and many others. Different attempts have been made to quantify and interpret the information that we can gather from studying crystal symmetry and its distribution in the mineral kingdom. However, these methods are mostly outdated or at least not compatible for use on large datasets available today. Therefore, a revision of symmetry index calculation has been made in accordance with the growing understanding of mineral species and their characteristics. In the gathered data, we observe a gradual but significant decrease in crystal symmetry through the stages of mineral evolution, from the formation of the solar system to modern day. However, this decrease is neither uniform nor linear, which provides further implications for mineral evolution from the viewpoint of crystal symmetry. The temporal distribution of minerals based on the number of essential elements in their chemical formulae and their symmetry index has been calculated and compared to explore their behaviour. Minerals with four to eight essential elements have the lowest average symmetry index, while being the most abundant throughout all stages of mineral evolution. There are many open questions, including those pertaining to whether or not biological activity on Earth has influenced the observed decrease in mineral symmetry through

time and whether or not the trajectory of planetary evolution of a geologically active body is one of decreasing mineral symmetry/increasing complexity.

1 | INTRODUCTION

Minerals are characterized by their crystal structures, specifically the internal, repeating arrangement of atoms within a crystal. This repeating crystallographic unit is referred to as the *unit-cell*, the symmetry of which can be most broadly categorized by *crystal system*. Each of the seven crystal systems is characterized by the axes and associated angles of the three-dimensional structure of the unit-cell. Different symmetry operators lead to a different *multiplicity* of the unit-cell, which represents a maximum number of times a single spot can be multiplied within one unit-cell by the present symmetry operators. If we take multiplicity as a metric of symmetry, the seven crystals systems, in order of decreasing symmetry with maximum multiplicity in parenthesis are as follows: cubic (48), hexagonal (24), tetragonal (16), trigonal (12), orthorhombic (8), monoclinic (4) and triclinic (2). While trigonal is a subsystem of the hexagonal system, here, it is separated because of the lower multiplicity.

Naumann (1855), who gathered information on crystal systems of 546 mineral species, was the first to discuss the distribution of mineral species among different symmetries. After this pioneering research, several authors discussed the same topic (Kostov & Kostov, 1999; Lebedev, 1891; Nowacki, 1942; Povarennykh, 1966; Shafranovsky & Feklichev, 1982; Vernadsky, 1903 etc.). From the original 546 mineral species discussed, the number increased over seven times to 3,958 at the brink of the 21st century (Nikolaev, 2000).

However, despite the attention given to this topic, early researchers all came to the consensus that the distribution of minerals among crystal symmetry classes remains constant despite the increasing number of newly discovered, rare minerals (e.g. Shafranovsky, 1983; Vernadsky, 1988).

In the 1980s, however, the first signs of doubt occurred, questioning the validity of this theory. Yushkin et al. (1987) brought up the problem of equalizing abundant minerals with rare ones. The same 'weight' was given to quartz, an abundant and widespread mineral, and some minerals discovered only at one locality. Furthermore, Dolivo-Dobrovol'sky (1988) demonstrated that the crystal structures of minerals discovered between 1980 and 1984 have a higher percentage of monoclinic

and lower portion of cubic minerals than the complete dataset available.

The lowering of the symmetry index in the newly discovered mineral species was explained by Urusov (2002) due to the increased proportion of rare minerals. While this trend is present and the symmetry index drops with the discovery of new species, this trend does not provide a conclusive explanation for the nature of this 'dissymmetrization'.

Urusov (2007) provided an overview of the previous work done in the study of symmetry statistics in the mineral world and expanded the ideas of mineral dissymmetrization. Several examples of reactions were given in which the products' symmetry is lower than those of the reactants. Furthermore, these reactions are characterized by a decrease in the thermodynamic entropy and an increase in the informational entropy. This thesis is further confirmed by Krivovichev, Krivovichev, et al. Hazen (2018) who showed that minerals' chemical and structural complexity both increase with the progression of mineral evolution.

With the change in the overall symmetry index of discovered mineral species and those present on Earth being more accepted in the scientific community, the reasons for this change have become the object of further investigations. (Filatov, 2021), who explored the symmetry statistics in different thermodynamic environments (notably, different depths of the Earth's crust and mantle), noticed that the symmetry index significantly grows with the increase in temperature and with the increase in depth from Earth's surface to the lower mantle. Furthermore, in this work, the 'monoclinic anomaly' – the dominance of monoclinic compounds among the mineral kingdom – is discussed. It is suggested that two main factors affect the distribution of mineral species among the symmetry groups: lattice dynamics and site multiplicity in the given group. The lattice dynamics are characterized by the number of unit-cell parameters that are unfixed according to symmetry, which increases from 1 for cubic to 6 for triclinic crystals (Filatov, 2021). With the increase in the number of unfixed parameters of the unit-cell, the fitting of coordination polyhedra in the crystal lattice is more easily achieved. However, this trend implies that triclinic minerals should be more abundant than monoclinic ones, when in reality, triclinic minerals are much less abundant. This situation

is explained by the fact that a decrease in symmetry is accompanied by decreased site multiplicity and therefore fails with regard to the principle of the economy of a crystal structure (Pauling's rule 5; Pauling, 1929). Following this rule, the maximum site multiplicity is much smaller in the triclinic system than in the monoclinic system.

Krivovichev, Krivovichev, et al. (2018) discussed the evolution of chemical and structural complexity of minerals through the first three stages of mineral evolution according to (Hazen et al., 2008). A brief overview of different stages of mineral evolution is given in Table 1. Unfortunately, at that time, comprehensive data on mineral crystal system distributions between all of the stages of mineral evolution were not available. Therefore, this paper presents a more comprehensive dataset of mineral distribution among different crystal systems with respect to different stages of mineral evolution.

Also, as there are several ways of calculating the symmetry index of a given dataset, all of the methods have been discussed, and an adjustment has been made for a better and more indicative calculation of the symmetry index.

2 | DATA DESCRIPTION AND DEVELOPMENT

The dataset was developed by assimilating data from the resources listed in Section 2.1. During the creation of the initial dataset, all the mineral species present in each stage were separated based on the crystal system(s) they crystallize in. In addition, the analysis was made for all the minerals first appearing in each given stage to see whether there was a difference from the changes in the complete set of minerals thought to be present on Earth's surface at a given time and those that first appeared in each stage. The data for the number of all mineral species that first appeared in each stage, segregated by the crystal system that they crystallize in are accessible in Tables 2 and 3, as well as Figure 1 in the form of pie charts. The same data for all species that appeared in each stage are accessible in Tables 4 and 5. Stages 6 'Anoxic biological world', 8 'Intermediate Ocean' and 9 'Snowball Earth Events' were excluded from consideration because during these stages of mineral evolution, few new minerals have been reported.

TABLE 1 A short description of stages of mineral evolution

Era/stage	Age	Cumulative number of species
0. Prenebular Ur-Mineralogy	>4.6 Ga	24
Era of Planetary Accretion (>4.55 Ga)		
1. Primary Chondrite minerals	>4.56 Ga	99
2. Achondrite and Planetesimal alteration	>4.56 to 4.55 Ga	314
Era of Crust and Mantle Reworking (4.55 to 2.5 Ga)		
3a. Igneous Rock Evolution (fractionation)	4.55 to 4.0 Ga	663
3b. Igneous Rock Evolution (volcanism, outgassing, surface hydration)	4.55 to 4.0 Ga	899
4a. Granite formation (granitoids)	4.5 to 3.5 Ga	2,909
4b. Granite formation (pegmatites)	4.5 to 3.5 Ga	3,741
5. Plate tectonics	> 3.0 Ga	3,816
Era of Biologically Mediated Mineralogy (>2.5 Ga to Present)		
6. Anoxic biological world	3.9 to 2.5 Ga	3,816
7. Great Oxidation Event	2.5 to 1.9 Ga	5,538
8. Intermediate Ocean	1.9 to 1.0 Ga	5,538
9. Snowball Earth Events	1.0 to 0.542 Ga	5,538
10a. Phanerozoic Era (Biomineralization)	0.542 Ga to Present	5,664
10b. Phanerozoic Era (Bioweathering)	0.542 Ga to Present	5,674

Adapted from (Hazen et al., 2008) with current known cumulative number of mineral species in each stage.

TABLE 2 The number of new species that appeared in each stage, segregated by crystal system

	Amorphous	Triclinic	Monoclinic	Orthorhombic	Trigonal	Tetragonal	Hexagonal	Cubic	Total
0	0	0	0	4	0	2	9	9	24
1	1	7	19	11	0	6	16	15	75
2	0	17	50	39	1	15	51	42	215
3a	1	18	113	73	3	29	64	48	349
3b	1	15	75	31	0	39	38	37	236
4a	5	183	652	440	6	163	370	191	2,010
4b	4	91	315	159	3	31	177	52	832
5	0	4	31	16	0	1	17	6	75
7	7	274	636	339	1	109	267	89	1,722
10a	0	11	40	32	0	8	28	7	126
10b	0	2	6	0	0	0	2	0	10

TABLE 3 The percentage of new species that appeared in each stage that crystallize in each crystal system

	Amorphous	Triclinic	Monoclinic	Orthorhombic	Trigonal	Tetragonal	Hexagonal	Cubic
0	0	0	0	16.7	0	8.3	37.5	37.5
1	1.3	9.3	25.3	14.7	0	8	21.3	20
2	0	7.9	23.3	18.1	0.5	7	23.7	19.5
3a	0.3	5.2	32.4	20.9	0.9	8.3	18.3	13.8
3b	0.4	6.4	31.8	13.1	0	16.5	16.1	15.7
4a	0.2	9.1	32.4	21.9	0.3	8.1	18.4	9.5
4b	0.5	10.9	37.9	19.1	0.4	3.7	21.3	6.2
5	0	5.3	41.3	21.3	0	1.3	22.7	8
7	0.4	15.9	36.9	19.7	0.1	6.3	15.5	5.2
10a	0	8.7	31.7	25.4	0	6.3	22.2	5.6
10b	0	20	60	0	0	0	20	0

2.1 | Data sources

The core of the data analytics consists of International Mineralogical Association (IMA) approved species along with mineral formulas, crystal systems and mineral evolution stages described by Hazen et al. (2008) and Hazen and Ferry (2010). The temporary local data warehouse is compiled from several Web resources and includes extraction from RRUFF databases and from peer-reviewed scientific publications (e.g. Canadian Mineralogist and American Mineralogist).

2.1.1 | The RRUFF project

The RRUFF Project, available at https://rruff.info (accessed 1st February 2022), is a set of mineral libraries and relational databases that allow interactive access to systematic chemical, X-ray powder diffraction and Raman spectroscopic data for IMA-approved mineral species. The project began with a goal to provide freely available Raman spectroscopic data and subsequently grew into one of the most widely used Raman and X-ray diffraction libraries in the world. The project is maintained by Prof. Robert T. Downs at the Department of Geosciences, The University of Arizona (Lafuente et al., 2016). Currently, the database contains 3,729 total mineral species with samples and 9,813 total RRUFF samples (as of 9th March 2022). The RRUFF database, with its relational databases, enables many sorting options based on the maximum or minimum age, locality name and many other attributes. These data are then available for download directly from the Website with various user-defined sort, display and file format options. This project and its associated database have become invaluable for data-driven exploration of the mineral world.

2.1.2 | The International Mineralogical Association (IMA) list of mineral species

The complete list of IMA-approved minerals is accessible via the RRUFF Project (https://RRUFF.info/IMA).

FIGURE 1 The distribution of crystal systems among newly appeared mineral species in each stage.

TABLE 4 The number of all the species present in each stage, segregated by crystal system

	Amorphous	Triclinic	Monoclinic	Orthorhombic	Trigonal	Tetragonal	Hexagonal	Cubic	Total
0	0	0	0	4	0	2	9	9	24
1	1	7	19	14	0	8	22	22	93
2	0	20	57	49	1	22	67	59	275
3a	1	28	130	90	4	37	81	68	439
3b	2	29	121	62	4	47	70	58	393
4a	7	212	785	513	10	207	457	269	2,460
4b	8	136	506	283	6	96	326	162	1,523
5	0	39	161	80	5	27	113	72	497
7	8	328	835	460	3	142	369	129	2,274
10a	4	59	197	126	3	43	130	71	633
10b	4	82	227	110	2	44	102	59	630

The IMA list allows users to search through nearly 5,800 species (date accessed: 30th January 2022) approved by the Commission on New Minerals, Nomenclature and Classification (CNMNC) and apply filtering by composition, crystal system, space group, point group, unit-cell parameters, origins, paragenetic mode, IMA status and other properties. Additionally, cross-references to other valuable Web resources are provided: the Mineral Evolution Database (MED; (Golden et al., 2016), American Mineralogist Crystal Structure Database (AMCSD; [Downs & Hall-Wallace, 2003]), webminerals (Barthelmy, 2007), Mindat, and downloadable links to essential literature references, hosted under the RRUFF Project. The IMA list of mineral species also offers links to information pages on each mineral in the Handbook of Mineralogy ([Anthony et al., 1990-2003]; http://www.handbookofmineralogy.org/) and crystal structure files in the American Mineralogist Crystal Structure Database (AMCSD; [Downs & Hall-Wallace, 2003]).

2.1.3 | The Mineral Evolution Database

The Mineral Evolution Database is another essential resource, designed initially for mineral evolution and ecology studies accessible through RRUFF Project (https://RRUFF.info/Evolution). The MED integrates mineral-locality data from the crowd-sourced mindat.org with the official IMA list of approved mineral species and age data from geologic literature. As of the 30th of January 2022, these data provide a sample size of 810,907 total observations of which 210,037 are dated, where each observation is a unique mineral species-locality pair. This database provides data on specific mineral formations, mineralization events, element concentrations and/or deposit formations which maximize the accuracy of age associations between the locality and the mineralization.

2.2 | Symmetry index calculation

During the investigation of symmetry statistics, three different ways of calculating the symmetry index of a set of minerals have been proposed. In this paper, all three methods were used, with some adjustments, and the results were compared visually and statistically to test the value of each proposed method.

The first two methods, proposed by Yushkin et al. (1987) and Dolivo-Dobrovol'sky (1988), are similar. In both methods, a number was assigned to each crystal system based on its properties of symmetry, and then, this number was multiplied by the number of mineral phases in a given crystal system. This approach produced the non-normalized symmetry index $I_N = \sum_i n_i s_i$, where n_i is the number of mineral species that crystallize in crystal system i, and s_i is the number assigned to said crystal system. The difference between the two approaches was the number assigned to each crystal system. In Yushkin's work, the numbers s_i assigned to each crystal system were 0 for triclinic, 1 for monoclinic, 2 for orthorhombic, 3 for trigonal, 4 for tetragonal, 5 for hexagonal and 6 for cubic. In contrast, for the Dolivo-Dobrovolsky's approach, it was 2 for triclinic, 4 for monoclinic, 8 for orthorhombic, 12 for trigonal, 16 for tetragonal, 24 for hexagonal and 48 for cubic, based on the maximum multiplicity of the holohedral symmetry class for each group.

To compare the symmetry indices of different datasets, the data have to be normalized, dividing the sum of symmetry indices by the total number of mineral species in the set. Furthermore, to obtain a neat number between 0 and 1, the result is also divided by the maximum value of s_i, which is that of the cubic system, giving us the final formula:

$$I = \frac{\sum_i n_i s_i}{\max\{s_j\} \cdot \sum_i n_i}.$$

TABLE 5 The percentage of all the species present in each stage that crystallize in each crystal system

	Amorphous	Triclinic	Monoclinic	Orthorhombic	Trigonal	Tetragonal	Hexagonal	Cubic
0	0	0	0	16.7	0	8.3	37.5	37.5
1	1.1	7.5	20.4	15.1	0	8.6	23.7	23.7
2	0	7.3	20.7	17.8	0.4	8	24.4	21.5
3a	0.2	6.4	29.6	20.5	0.9	8.4	18.5	15.5
3b	0.5	7.4	30.8	15.8	1	12	17.8	14.8
4a	0.3	8.6	31.9	20.9	0.4	8.4	18.6	10.9
4b	0.5	8.9	33.2	18.6	0.4	6.3	21.4	10.6
5	0	7.8	32.4	16.1	1	5.4	22.7	14.5
7	0.4	14.4	36.7	20.2	0.1	6.2	16.2	5.7
10a	0.6	9.3	31.1	19.9	0.5	6.8	20.5	11.2
10b	0.6	13	36	17.5	0.3	7	16.2	9.4

In Urusov's complementary approach, the number of mineral phases in a given dataset that crystallizes in a higher symmetry system (cubic, hexagonal, tetragonal and trigonal) is counted and divided by the number of phases that crystallize in a lower symmetry system (orthorhombic, monoclinic and triclinic). This method is not normalized, so its values can vary up to greater than 1 in specific conditions where there are more mineral phases of higher symmetry.

An additional adjustment must be made to apply these three methods to the data available from the RRUFF project and the MED. Since, in the MED, there are several amorphous phases also present, they were also taken into account. In Urusov's calculation, these phases were simply added to the count of the lower symmetry phases. For the Yushkin-inspired approach, the numbers s_i assigned to each crystal system had to be changed to 0 for amorphous, 1 for triclinic, 2 for monoclinic, 3 for orthorhombic, 4 for trigonal, 5 for tetragonal, 6 for hexagonal and 7 for cubic. For the Dolivo-Dobrovolsky approach, value 1 was simply added to amorphous phases, while the given value for other crystal systems was not changed. Once these adjustments in the methods were made, all three ways of calculations were applied to each stage of mineral evolution separately.

2.3 | Data processing

The data processing included parsing the initial dataset, cleaning, transforming, treating missing values and calculating the symmetry indices along with the visual output of the dependencies within the data. The calculations were performed using Python 3.10 in the Jupyter Notebook environment, adjusted with several standard data science computing libraries – Pandas, NumPy, SciPy and Matplotlib. During the first step of the analysis, all of the local text files in .csv format, accessed from https://rruff.info, were uploaded into the local environment and

further checked for consistency of data headers, data formats and general data quality. The mineral species without an assigned crystal system were further identified and exported into a machine-readable file to fill missing values before running the rest of the analytics pipeline. The second step was to determine the list of minerals that are present in each stage of mineral evolution and those that appear in each stage the first time only. The third step included calculating the symmetry indices using different methodologies for every mineral evolution stage based on: (1) minerals that first appeared in that stage, (2) all minerals present in the respective stage and (3) all minerals that appeared in the respective stage and those present in the previous ones – a cumulative metric. Accordingly, a separate .ipynb file is provided in the GitHub repository for calculating the symmetry index using a different approach where a file name designates the taken approach:

1. `data_analysis_dobrovolski.ipynb` – for Dolivo-Dobrovolsky's approach.
2. `data_analysis_urusov.ipynb` – for Urusov's approach.
3. `data_analysis_yushkin.ipynb` – for Yushkin's approach.

Furthermore, the data were visualized using the bar chart, pie chart and line chart graph representations to better understand the patterns within the symmetry evolution, discover the general trends of the crystal systems distribution through mineral evolution periods, and choose the data interpretation strategy. Afterwards, the number of elements that dominate one or more unique sites in the crystal structure of a mineral, called essential elements, data provided by https://rruff.info, was used to discover how the number of essential elements present in the mineral stoichiometric formula affects its symmetry index and how it fluctuates during each mineral evolution

stage. When working with text (string) data, regular expression matching operations were used. For instance, the essential elements data provided by https://rruff.info is a string concatenation of unique elements present in minerals' IMA-approved formula using a space character, which is a distinct separator for normalizing this data. Accordingly, the essential elements dataset was used to compile a visual representation of (1) the symmetry index of newly appeared mineral species through each evolution stage divided into 14 categories, designating the number of unique essential elements present in the formula; (2) a frequency of the symmetry index depending on the number of unique essential elements without relation to evolution stages; (3) a proportion of minerals with a certain number of unique essential elements through each evolution stage; (4) an average number of unique essential elements present in minerals through each evolution stage. Selected figures are provided here (Figures 2 and 3); all of the plots listed above are available in the GitHub repository.

For each calculation approach, the symmetry index was first calculated based only on the mineral species that first appeared in a given stage (Figures 4, 5 and 6), as well as the symmetry index based on all of the mineral species that were present in a given stage (Figures 7, 8 and 9). While there are noticeable differences in the results on first inspection, namely the differences between the ratio of symmetry index between stages in each approach, little can be gathered from this rudimentary look at the data apart from the fact that there is an obvious downward trend in the symmetry index in all approaches. When considering only the mineral species that first appeared in a stage (Figures 4, 5 and 6), we see noticeably large drops between stages 0 'Prenebular Ur-Mineralogy' and 1 'Primary Chondrite minerals', 3b 'Igneous Rock Evolution (volcanism, outgassing,

surface hydration)' and 4a 'Granite formation (granitoids)', as well as 5 'Plate tectonics' and 7 'Great Oxidation Event', and then a large spike upward in stage 10a 'Phanerozoic Era (Bioweathering)'. However, motivated by the difference in sample size between stages (namely, stages 4a and 7 comprising the vast majority of all mineral species and 10a comprising a comparably minuscule number of species), a cumulative symmetry index was also calculated for each stage by taking into account all of the mineral species that appeared in a given stage, as well as all prior stages (Figures 10, 11 and 12). In these data, once again a strong downward trend can be seen (this time without upward spikes, implying a continuous drop in symmetry index), as well as a major drop in the symmetry index between stages 0 'Prenebular Ur-Mineralogy' and 1 'Primary Chondrite minerals', 3b 'Igneous Rock Evolution (volcanism, outgassing, surface hydration)' and 4a 'Granite formation (granitoids)', and 5 'Plate tectonics' and 7 'Great Oxidation Event'. However, given the downward trend and the fact that the vast majority of species originated in stages 4a 'Granite formation (granitoids)' and 7 'Great Oxidation Event', these large decreases, on their own, are unremarkable. To take a closer look at which stages had the largest impact on symmetry evolution, the difference in symmetry index between each stage was calculated, normalized per newly appeared species in a stage:

$$\left(\frac{\Delta I}{\Delta n} \right)_{j,j+1} = \frac{I_{j+1} - I_j}{n_i},$$

where I_j is the symmetry index of stage j. The results are shown in Figures 13, 14 and 15 and illustrate the decrease per species, $\left(-\frac{\Delta I}{\Delta n} \right)$. In all three graphs, the major difference can be seen between stages: stage 1 'Primary Chondrite minerals'
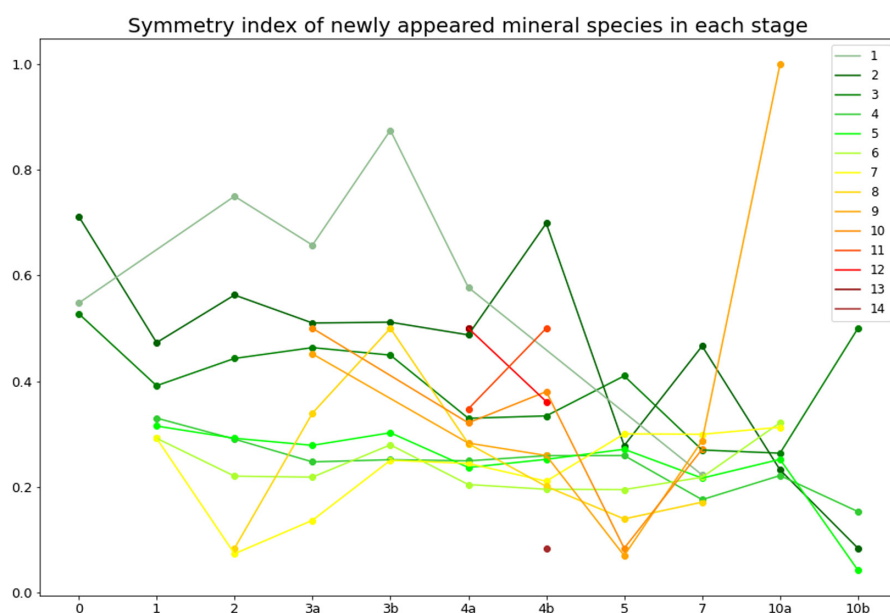

FIGURE 2 Dolivo-Dobrovolsky symmetry index of different stages, segregated by number of essential elements of a mineral.

FIGURE 3 Average Dolivo-Dobrovol'sky symmetry index of minerals depending on the number of essential elements therein.
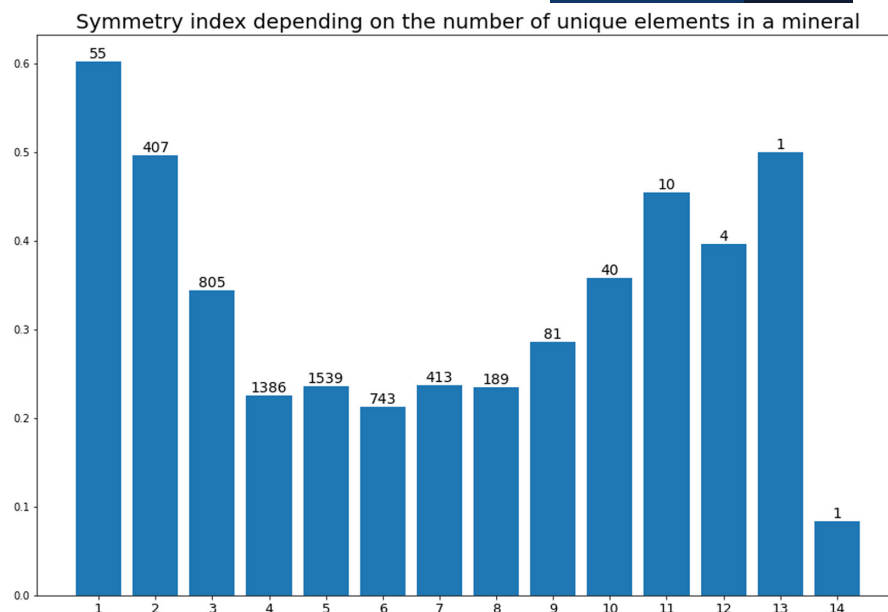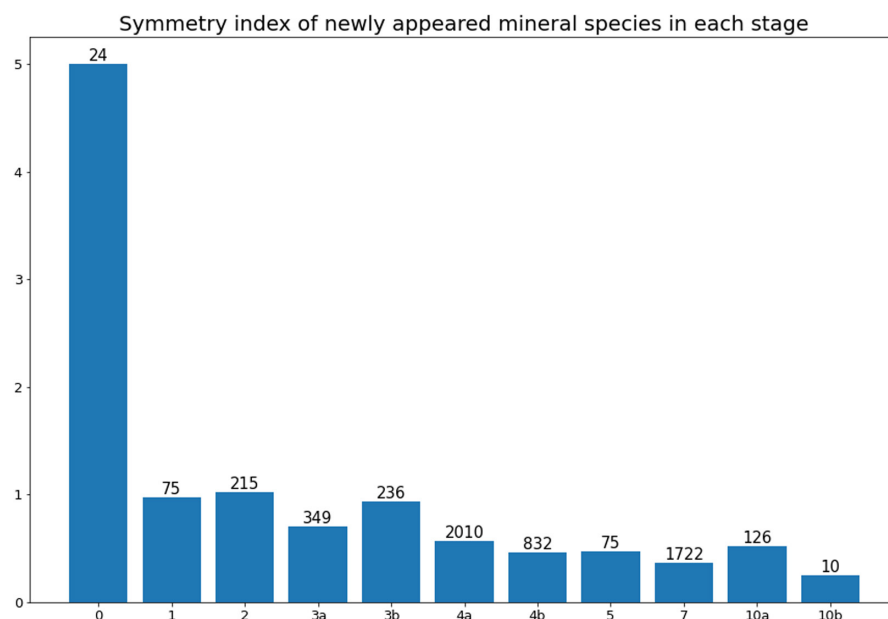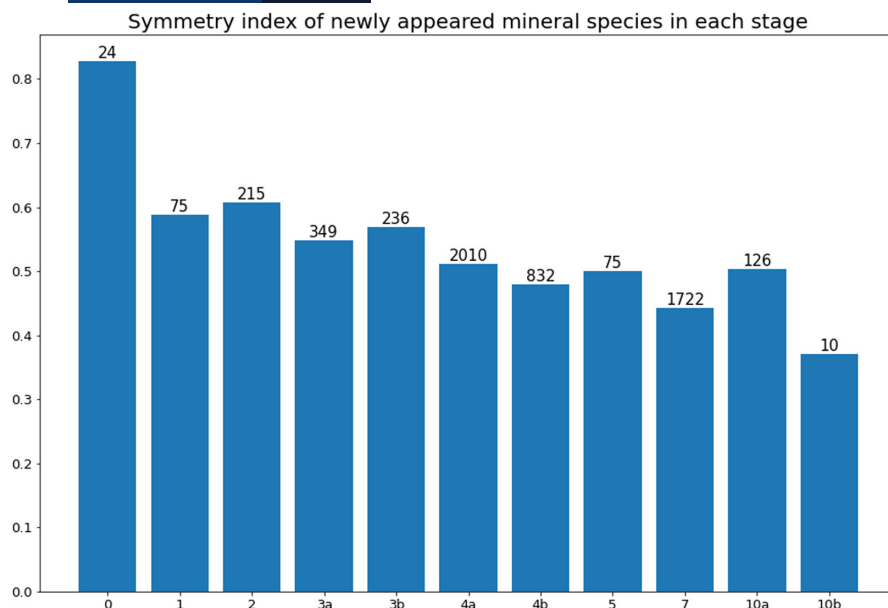

FIGURE 4 The Urusov symmetry index of each stage, calculated by only considering the mineral species that first appeared in each stage. Above the bars is the number of newly appeared species in each stage.


has a very large impact on the symmetry index, changing it drastically from that of stage 0 'Prenebular Ur-Mineralogy'. Stages 2 'Achondrite and Planetesimal alteration', 3a 'Igneous Rock Evolution (fractionation)', 4a 'Granite formation (granitoids)' and 7 'Great Oxidation Event' also have an impact at least one order of magnitude larger than the other remaining stages. Note that using Yushkin's approach, a small increase in the symmetry index can be observed in the transition from stage 7 'Great Oxidation Event' to stage 10a 'Phanerozoic Era (Bioweathering)', instead of the decrease observed in all other stages, but the amount is not significant owing to a very small sample size.

It is interesting to note that despite Urusov's index not being normalized, in contrast to Yushkin's and Dolivo-Dobrovolsky's, it still shows very similar trends when applied to different stages of mineral evolution. This similarity shows that, despite the index values being different, all three approaches yield similar results when used to compare different datasets, or in these examples, different stages of mineral evolution.

The average Dolivo-Dobrovolsky symmetry index across different stages of mineral evolution was segregated by number of essential elements in a mineral in Figure 2. It can be seen that minerals consisting of the largest number of unique elements (12, 13, 14) form exclusively in stage 4 'Granite formation (granitoids)' of mineral evolution, with the mineral eveslogite $((Na,K,Ca,Sr,Ba)_{48} [(Ti,Nb,Mn,Fe^{2+})_{12}Si_{48}O_{144}(OH)_{12}] (F,OH,Cl)_{14})$ being the most chemically complex mineral with respect to the number of suggested essential
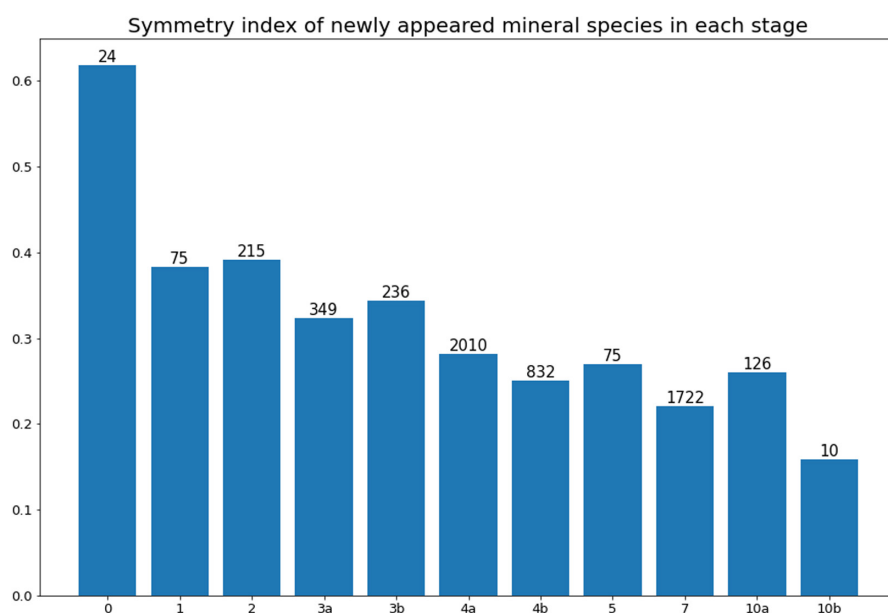
Symmetry index of newly appeared mineral species in each stage

Symmetry index of newly appeared mineral species in each stage

elements. It can also be seen that almost all of the first minerals to appear in the first stages of mineral evolution form with 1 to 3 essential elements and they tend to have somewhat higher symmetry index than minerals with more essential elements; however, this trend is not observed in the later stages of mineral evolution (stage 5 'Plate tectonics' and beyond).

To better understand the distribution of mineral symmetry through different stages of mineral evolution, Figure 3 was created to show the number of minerals and their average Dolivo-Dobrovolsky symmetry index depending on the number of unique essential elements therein. The plurality of minerals contains four or five different chemical elements (Krivovichev & Charykova, 2013); (Krivovichev, Charykova, et al., 2018). This feature can also be seen in 14,

which shows an almost reverse dependency of symmetry index on the number of unique elements in a mineral with minerals having between four and eight different essential elements having the lowest symmetry index. (Krivovichev et al., 2022) discusses the similar, lognormal distribution of the number of atoms per formula or per unit-cell as a result of balance between the need to accommodate different elements in the same cell and the tendency of crystal structures to be as simple as possible.

3 | DATASET ACCESS

All of the data are available in the public repository on Github at the following link. The raw data were pulled from

FIGURE 7 The Urusov symmetry index of each stage, calculated by considering all of the mineral species present in each stage. Above the bars is the number of all species present in each stage.
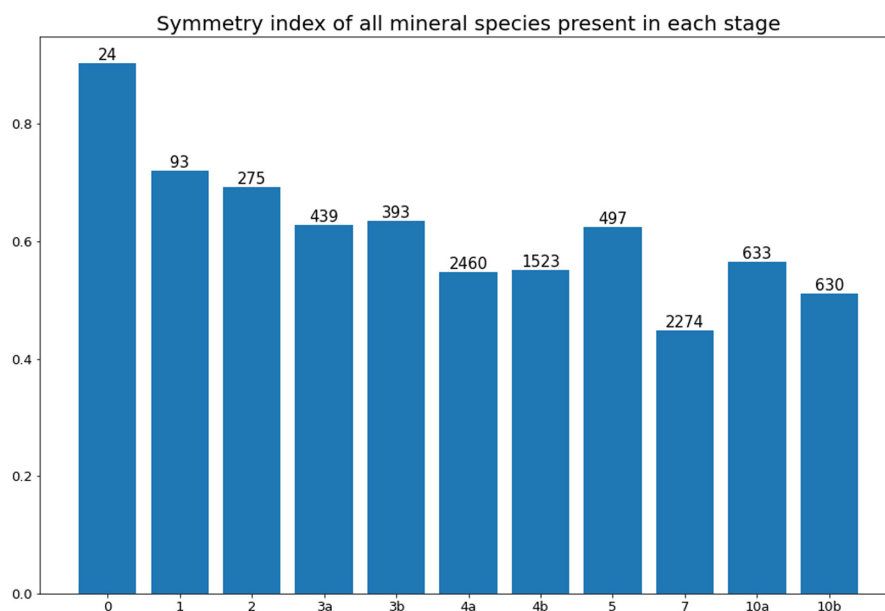

Symmetry index of all mineral species present in each stage
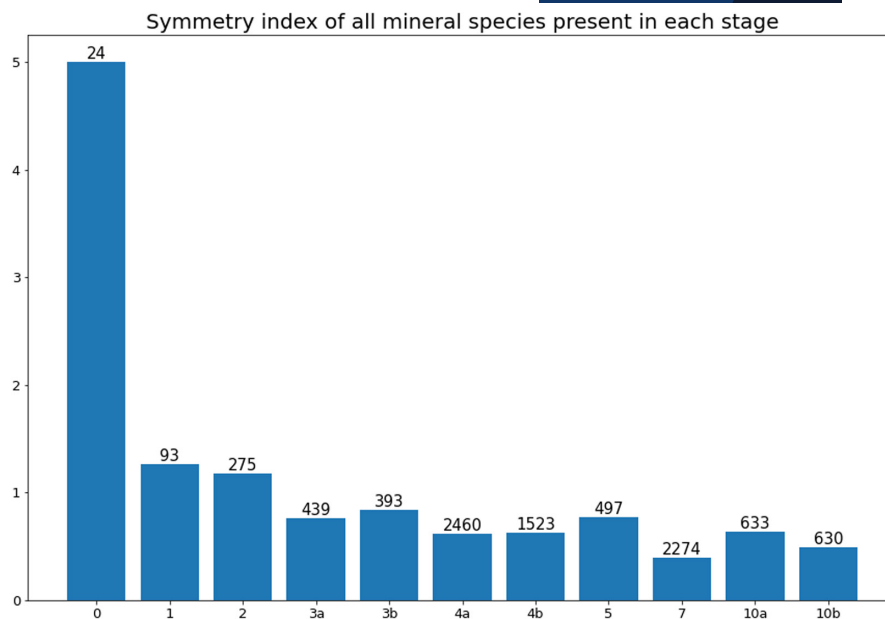
FIGURE 8 The Yushkin symmetry index of each stage, calculated by considering all of the mineral species present in each stage. Above the bars is the number of all species present in each stage.

Symmetry index of all mineral species present in each stage

the publicly available database of minerals created and maintained by the RRUFF Project at the following link. For use in this paper, the data were pulled and stored in the form of CSV files, which are up to date as of 25 December 2021. For ease of access for other projects, the data have been compiled from these CSV files into a MongoDB database that can be reconstructed from the dump directory of the previously mentioned Github repository using the instructions provided in the README file of said repository.

The code in the repository uses Python 3 and Jupyter Notebook, along with many commonly used data processing libraries. In processing the data, we used a virtual environment managed using the Anaconda platform. The README file of the repository includes instructions on setting up a Python virtual environment with the help of

Anaconda's conda CLI utility, installing all the necessary packages in the process. Once this is set up, one can open the notebooks in the repository to see and modify the code that was used to process the data.

4 | POTENTIAL DATASET USE

Data-driven discovery in mineralogy is a new approach to analysing the ever-growing large volumes of data available on chemical composition, crystal structure, physical properties and geological origins of minerals (Hazen, 2014; Hazen et al., 2019; Hazen & Morrison, 2021; Hystad et al., 2019; Morrison et al., 2017; Prabhu et al., 2020; Prabhu et al., 2022).
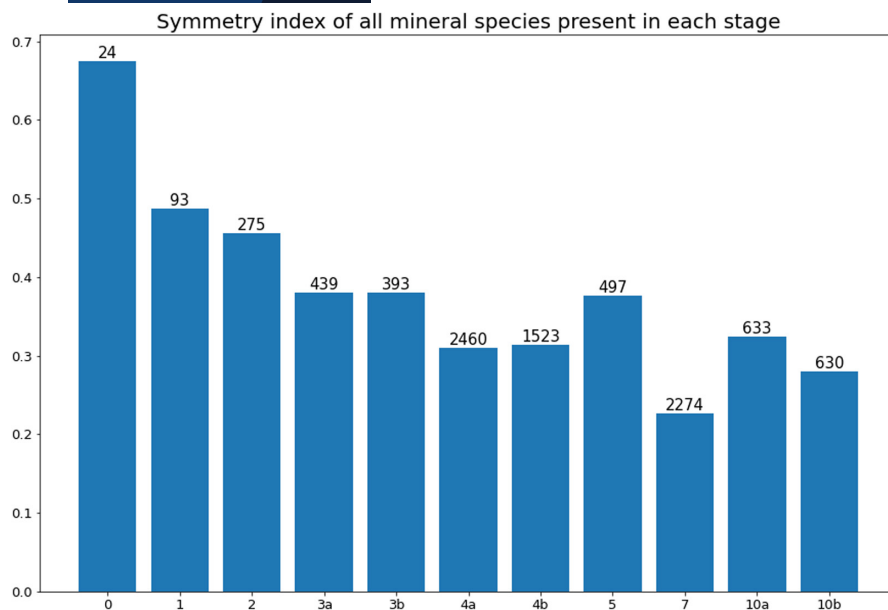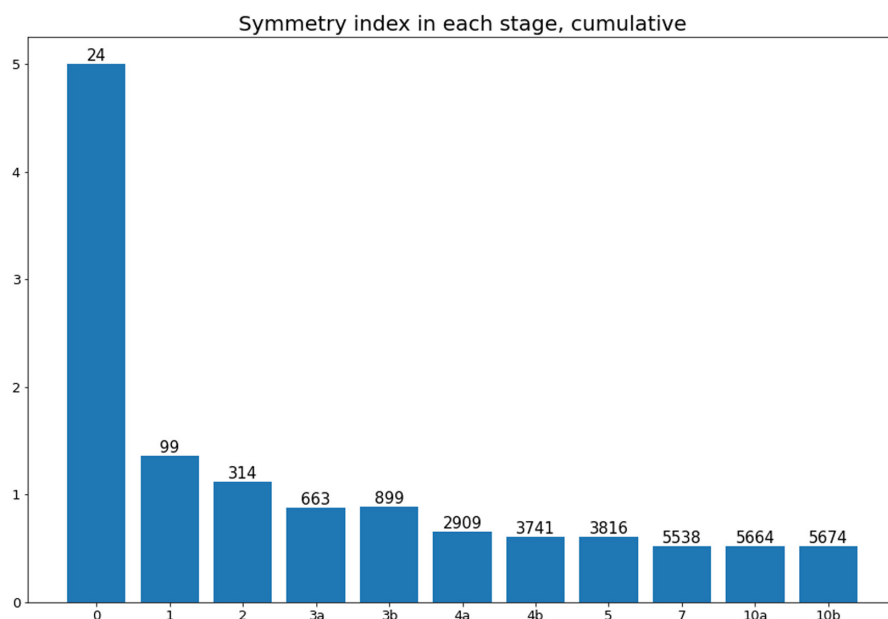
Large, open-access databases, such as the ones available through the RRUFF Project, open doors for many statistical and numerical analyses of the mineral kingdom which were not possible before.

This approach aspires to combine the statistical methods used in big-data science with the available knowledge of mineralogy and crystal chemistry of minerals. The dataset herein will serve as a basis for further research of mineral symmetry distribution with respect to other mineral properties such as complexity, rarity, spatial distribution, paragenetic modes, chemical composition, physical attributes and more. Likewise, these techniques can be applied to mineral data from other planetary bodies, including Mars and the moon (Morrison et al., 2017; Morrison et al., 2018; Rampe et al., 2020).

5 | FUTURE WORK

Currently, only the crystal systems are defined within the dataset (i.e. cubic and hexagonal), while point groups and symmetry classes are not taken into account. This is mostly due to the fact that it would significantly overcomplicate the dataset without providing much additional information about the symmetry evolution trends. (Hummer, 2021) investigated the distribution of mineral species among 32 point groups and noticed that it seems like minerals prefer higher symmetry when only one crystal system is studied (majority of minerals form in holohedral classes of seven crystal systems) while preferring lower symmetry when observing all crystal systems at once (with monoclinic being most abundant). Furthermore,

FIGURE 11 The Yushkin symmetry index of each stage, calculated by considering all the mineral species that appeared up to each stage. Above the bars is the number of species that appeared before or during each stage.
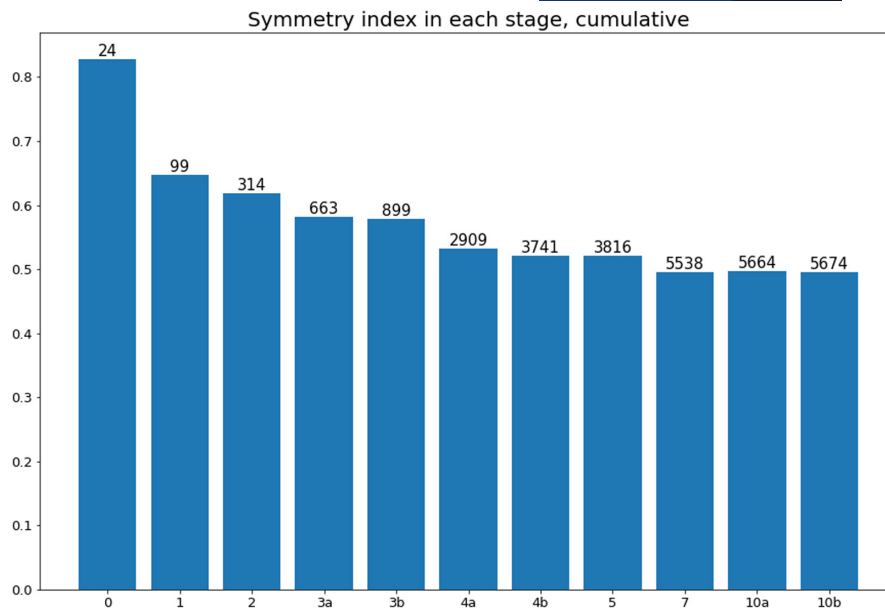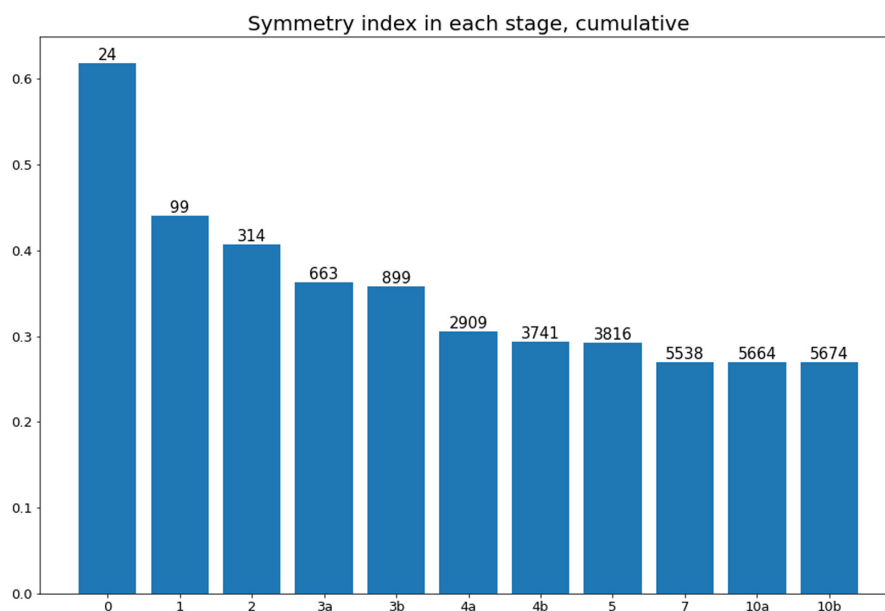

FIGURE 12 The Dolivo-Dobrovolsky symmetry index of each stage, calculated by considering all the mineral species that appeared up to each stage. Above the bars is the number of species that appeared before or during each stage.

point groups are not known for each mineral species and some of them can crystallize in multiple point groups, and even crystal systems (i.e. monoclinic and triclinic mica polytypes) further complicating the dataset.

Additionally, the dataset is currently limited to the IMA definitions of mineral species which have some internal inconsistency and bias towards end-member compositions.

In most cases, IMA formulas are idealized end-member compositions and do not reflect the complete range of natural chemical variation observed in mineral specimens. Mineralogical nomenclature of binary systems such as solid solutions still follows a 50% rule, or the dominant-constituent rule, while it should be extended with the dominant-valency rule as proposed by (Hatert

& Burke, 2008). Forsterite is Mg_2SiO_4, even though forsterite always has significant Fe (up to 49 atom % in some samples). Orthoclase is $KAlSi_3O_8$, even though it always has significant Na (up to 10s of atom %). Another example provided by (Hatert & Burke, 2008) concerns structural order involving the ions that define the end members. For instance, the ordering of Ca and Mg in dolomite, $CaMg(CO_3)_2$, results in a crystal structure different from the end members of the $(Ca,Mg)CO_3$ series – calcite, $CaCO_3$ and magnesite, $MgCO_3$. However, it gets even more complicated with other mineral groups and other elements. For example, rare earth elements (REEs) and platinum group elements (PGEs) always occur collectively. The IMA divides specimens with very close REE compositions into different species when they display no
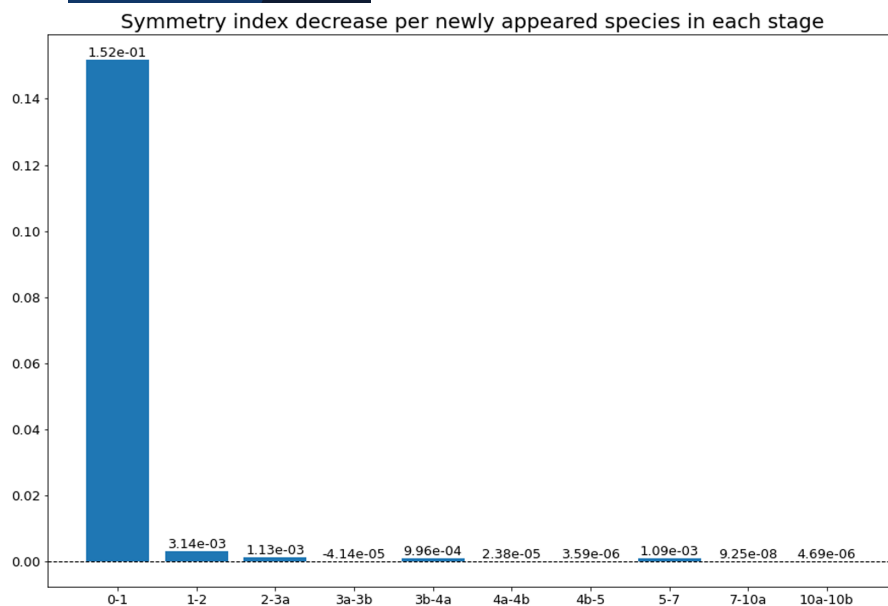
FIGURE 13   The decrease in the Urusov symmetry index of each stage, normalized per newly appeared species in each stage. The number above the bar is the magnitude of the decrease. The dashed line represents the point of no change.
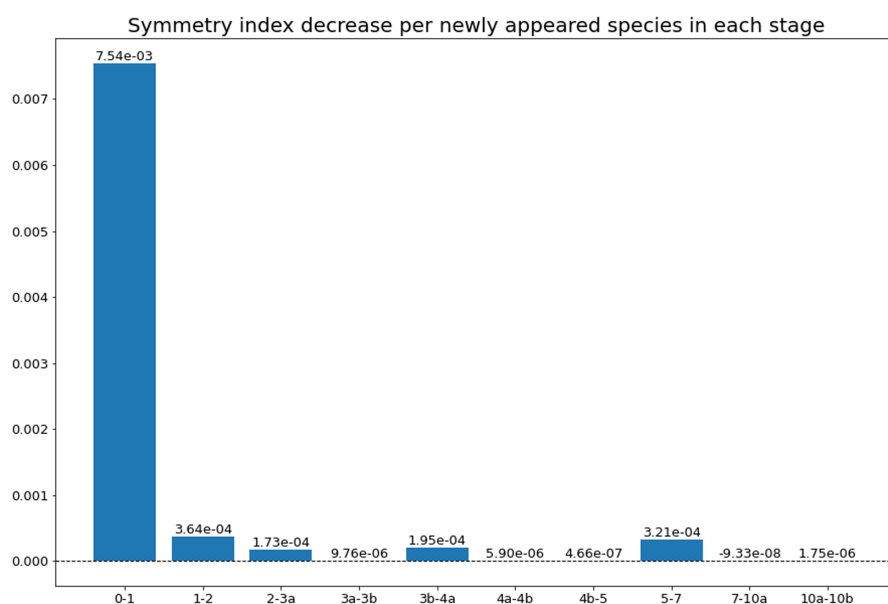

FIGURE 14   The decrease in the Yushkin symmetry index of each stage, normalized per newly appeared species in each stage. The number above the bar is the magnitude of the decrease. The dashed line represents the point of no change.

difference in paragenesis or formation conditions except slight shifts in ratios of elements. However, in some cases, even those slight shifts in chemistry, such as adding a cation (or an anion) with different dimensions than the one present in the original crystal structure, leads to desymmetrization. Therefore, it is still important to consider minor and trace elements when possible. For example, substituting one SiO_4 tetrahedra with a $[(OH)_4]$ functional group in henritermierite and holtstamite leads to tetragonal symmetry, instead of cubic present in 'regular' garnets. These subtleties are important when computing the numerical representation model of mineral formulas for data analysis and descriptive statistics with no or substantially less impact on mineralogical nomenclature.

Future investigations could use mineral natural kinds (Hazen et al., 2019; Hazen et al., 2020; Hazen & Morrison, 2020; Hazen & Morrison, 2021; Morrison & Hazen, 2020; Morrison & Hazen, 2021), which better illustrate natural chemical variations observed in mineral specimens, rather than mineral species; however, these data are not yet fully assembled and are therefore not yet available for study. Another solution is to define the term 'chemical formula', which differs from the IMA definition of idealized mineralogical formula and reflects the content of the impurities present in species. The latter could potentially increase the quality of the dataset and provide a consistent chemical context.

The data from this paper will be further combined with new data on the chemical and structural complexity

FIGURE 15 The decrease in the Dolivo-Dobrovolsky symmetry index of each stage, normalized per newly appeared species in each stage. The number above the bar is the magnitude of the decrease. The dashed line represents the point of no change.
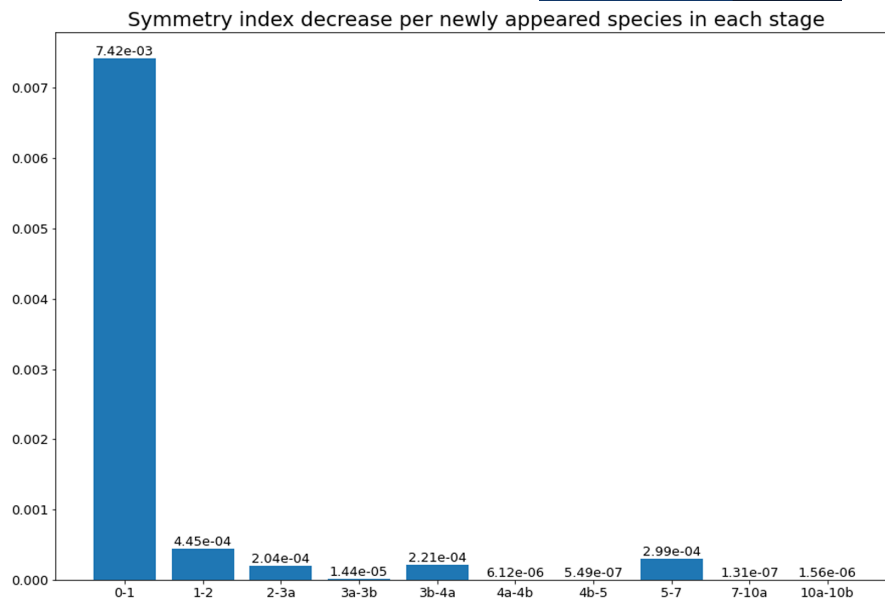
of minerals (Krivovichev et al., 2022), which have partially been incorporated into the Global Earth Mineral Inventory (GEMI) online database (Prabhu et al., 2020). With this approach, new insights into the complexities, rarity, symmetry distribution and other mineral properties will be explored.

6 | CONCLUSIONS

The work presented in this paper deals with the creation of a dataset demonstrating evolutionary trends of symmetry index and distribution in the mineral kingdom across the stages of mineral evolution.

This study has adapted the Yushkin, Dolivo-Dobrovolsky and Urusov approaches to symmetry index calculations to take into consideration amorphous minerals. All three methods of symmetry index calculation gave similar results; however, the authors recommend using the Dolivo-Dobrovolsky symmetry index because it is based on multiplicity of symmetry classes and is standardized, unlike the other approaches.

In addition to the symmetry index, the decrease in symmetry index of each stage, normalized per newly appeared species in each stage, was calculated as a metric of impact of each stage on the symmetry evolutionary trends. The most significant change in overall symmetry index occurred in Stage 1 'Primary Chondrite minerals' of mineral evolution with other notable drops of symmetry index occurring in stages 2 'Achondrite and Planetesimal alteration', 3a 'Igneous Rock Evolution (fractionation)', 4a 'Granite formation (granitoids)' and 7 'Great Oxidation Event'. It is obvious that the largest drops in symmetry indices occurred as a result of major changes in the Earth's history. Apart from the initial changes during early Solar

system, stages 3a 'Igneous Rock Evolution (fractionation)' and 4a 'Granite formation (granitoids)' influenced the drop by increasing the temperature and pressure ranges of mineral formation, as well as the degree of elemental fractionation. The drop in symmetry indices associated with stage 7 'Great Oxidation Event' can be associated with the increase in number of oxidative states in which different elements can occur on Earth's surface, giving the opportunity for more oxidized minerals to form as well.

This work also calculated the average Dolivo-Dobrovolsky symmetry index across different stages of mineral evolution as well as the distribution of minerals by the number of essential elements. The distribution of minerals based on number of essential elements shows a lognormal dependency, with minerals with four to five essential elements being most abundant. The symmetry index follows an opposite trend to the mineral distribution with minerals with four to eight essential elements having the lowest symmetry index. The symmetry index rises from three to one essential elements as well as from five to 13. There are only six minerals with 12 or more essential elements, so a small dataset may account for the slight deviation from this rule. These results are in accordance with Fedorov–Groth law which shows that the symmetry is correlated to the chemical complexity of minerals and that this correlation is statistically meaningful (Krivovichev & Krivovichev, 2020).

AUTHOR CONTRIBUTIONS
Marko Bermanec: Conceptualization (lead); project administration (equal); writing – original draft (lead); writing – review and editing (equal). **Noa Vidović:** Data curation (equal); visualization (lead). **Liubomyr Gavryliv:** Data curation (supporting); methodology (equal); visualization (equal); writing – review and editing (equal). **Shaunna**

CONFLICTS OF INTEREST
SM Morrison is one of the guest editors for this special issue of Geoscience Data Journal.

OPEN RESEARCH BADGES
This article has earned an Open Data badge for making publicly available the digitally-shareable data necessary to reproduce the reported results. The data is available at [insert provided URL from Open Research Disclosure Form]. Learn more about the Open Practices badges from the Center for Open Science: https://osf.io/tvyxz/wiki

ORCID
Marko Bermanec https://orcid.org/0000-0001-6165-3417

REFERENCES
Anthony JW, Bideaux RA, Bladh KW and Nichols MC eds. 1990-2003. Handbook of mineralogy. Chantilly, VA: Mineralogical Society of America.

Barthelmy D. 2007. *Webmineral mineralogy database*.

Dolivo-Dobrovol'sky, V.V. (1988) On so-called rules of statistic mineralogy. *Zap. Vseross. Mineral. O–Va*, 117(3), 387–393.

Downs, R.T. & Hall-Wallace, M. (2003) The American mineralogist crystal structure database. *American Mineralogist*, 88, 247–250.

Filatov, S.K. (2021) The symmetry statistics of mineral species in various thermodynamic environments. *Geology of Ore Deposits*, 62, 547–553. https://doi.org/10.1134/s1075701520070041

Golden JJ, Pires AJ, Hazen RM, Downs RT, Ralph J and Meyer M. (2016). "*Building the mineral evolution database: implications for future big data analysis. In mineralogical evidence for the co-evolution of the geosphere and biosphere: in honor of Robert M. Hazen.* Proceedings from the GSA annual meeting in Denver, Colorado, USA.

Hatert, F. & Burke, E.A.J. (2008) The IMA-CNMNC dominant-constituent rule revisited and extended. *The Canadian Mineralogist*, 46(3), 717–728. https://doi.org/10.3749/canmin.46.3.717

Hazen, R.M. (2014) Data-driven abductive discovery in mineralogy. *American Mineralogist*, 99(11–12), 2165–2170. https://doi.org/10.2138/am-2014-4895

Hazen, R.M. & Ferry, J.M. (2010) Mineral evolution: Mineralogy in the fourth dimension. *Elements*, 6(1), 9–12. https://doi.org/10.2113/gselements.6.1.9

Hazen, R.M. & Morrison, S.M. (2020) An evolutionary system of mineralogy, part I: Stellar mineralogy (>13 to 4.6 Ga). *American Mineralogist*, 105(5), 627–651. https://doi.org/10.2138/am-2020-7173

Hazen, R.M. & Morrison, S.M. (2021) An evolutionary system of mineralogy, part V: Aqueous and thermal alteration of planetesimals (4565 to 4550 Ma). *American Mineralogist*, 106(9), 1388–1419. https://doi.org/10.2138/am-2021-7760

Hazen, R.M., Papineau, D., Bleeker, W., Downs, R.T., Ferry, J.M., McCoy, T.J. et al. (2008) Mineral evolution. *American Mineralogist*, 93(11–12), 1693–1720. https://doi.org/10.2138/am.2008.2955

Hazen, R.M., Downs, R.T., Eleish, A., Fox, P., Gagné, O.C., Golden, J.J. et al. (2019) Data-driven discovery in mineralogy: Recent advances in data resources, analysis, and visualization. *Engineering*, 5, 397–405. https://doi.org/10.1016/j.eng.2019.03.006

Hazen, R.M., Morrison, S.M. & Prabhu, A. (2020) An evolutionary system of mineralogy, part III: Primary chondrule mineralogy (4566 to 4561 Ma). *American Mineralogist*, 106(3), 325–350. https://doi.org/10.2138/am-2020-7564

Hummer, D.R. (2021) Fractal distribution of mineral species among the crystallographic point groups. *American Mineralogist*, 106, 1574–1579. https://doi.org/10.2138/am-2021-7698

Hystad, G., Morrison, S.M. & Hazen, R.M. (2019) Statistical analysis of mineral evolution and mineral ecology: the current state and a vision for the future. *Applied Computing and Geosciences*, 1, 100005. https://doi.org/10.1016/j.acags.2019.100005

Kostov, I. & Kostov, R.I. (1999) Crystal habits of minerals. *Bulgarian Academic Monographs*, 1, 3–56

Krivovichev, V.G. & Charykova, M.V. (2013) *Classification of mineral systems*. St. Petersburg, Russia: St. Petersburg University Press, p. 196.

Krivovichev, S.V. & Krivovichev, V.G. (2020) The Fedorov-Groth law revisited: complexity analysis using mineralogical data. *Acta Crystallographica Section A Foundations and Advances*, 76(3), 429–431. https://doi.org/10.1107/s2053273320004209

Krivovichev, S.V., Krivovichev, V.G. & Hazen, R.M. (2018) Structural and chemical complexity of minerals: correlations and time evolution. *European Journal of Mineralogy*, 30(2), 321–326. https://doi.org/10.1127/ejm/2018/0030-2694

Krivovichev, V.G., Charykova, M.V. & Krivovichev, S.V. (2018) The concept of mineral systems and its application to the study of mineral diversity and evolution. *European Journal of Mineralogy*, 30, 219–230. https://doi.org/10.1127/ejm/2018/0030-2699

Krivovichev, S.V., Krivovichev, V.G., Hazen, R.M., Aksenov, S.M., Avdontceva, M.S., Banaru, A.M. et al. (2022) Structural and chemical complexity of minerals: An update. *Mineralogical Magazine*, 1–76, 183–204. https://doi.org/10.1180/mgm.2022.23

Lafuente, B., Downs, R.T., Yang, H. & Stone, N. (2016) The power of databases: the RRUFF project. In: Bindi, L., Churakov, S.V., Downs, R.T., Gfeller, F., Krivovichev, S.V., Lafuente, B. et al. (Eds.) *Highlights in mineralogical crystallography*. Berlin, Boston: Walter de Gruyter GmbH, pp. 1–29.

Lebedev GG. 1891. *Course of mineralogy. Descriptive Part*. St. Petersburg, Russia: St. Petersburg University Press.

Morrison, S.M. & Hazen, R.M. (2020) An evolutionary system of mineralogy, part II: interstellar and solar nebula primary condensation mineralogy (>4.565 Ga). *American Mineralogist*, 105(10), 1508–1535. https://doi.org/10.2138/am-2020-7447

Morrison, S.M. & Hazen, R.M. (2021) An evolutionary system of mineralogy, part IV: planetesimal differentiation and impact mineralization (4566 to 4560 Ma). *American Mineralogist*, 106(5), 730–761. https://doi.org/10.2138/am-2021-7632

Morrison, S.M., Liu, C., Eleish, A., Prabhu, A., Li, C., Ralph, J. et al. (2017) Network analysis of mineralogical systems. *American Mineralogist*, 102, 1588–1596. https://doi.org/10.2138/am-2017-6104CCBYNCND

Morrison, S.M., Downs, R.T., Blake, D.F., Vaniman, D.T., Ming, D.W., Rampe, E.B. et al. (2018) Crystal chemistry of martian minerals from Bradbury landing through Naukluft plateau, Gale crater, Mars. *American Mineralogist*, 103(6), 857–871. https://doi.org/10.2138/am-2018-6124

Naumann CF. 1855. *Elemente Der Mineralogie*. St. Petersburg, Russia: St. Petersburg University Press.

Nikolaev SM. (2000). *Statistics of modern mineralogical information*. Novosibirsk, Russia: GEO.

Nowacki, W. (1942) Symmetry and physical–chemical properties of crystallized compounds. The distribution development of the crystal structures over the 219 space groups. *Helvetica Chimica Acta*, 25, 863–880.

Pauling, L. (1929) The principles determining the structure of complex ionic crystals. *Journal of American Chemical Society*, 51(4), 1010–1026. https://doi.org/10.1021/ja01379a006

Povarennykh, A.S. (1966) On regularity in distribution of mineral species by systems, symmetry classes, and spatial groups. *Mineral. Zbornik Lvov. Gos. Univ.*, 3(20), 341–351.

Prabhu, A., Morrison, S.M., Eleish, A., Zhong, H., Huang, F., Golden, J.J. et al. (2020) Global earth mineral inventory: a deep carbon observatory data legacy. *Geoscience Data Journal*. DCO Special Issue, 8, 74–89. https://doi.org/10.1002/gdj3.106

Prabhu A, Morrison SM, Fox PA, Ma X, Wong ML, Williams J, McGuinness KN, Krivovichev S, Lehnert K, Ralph J, Lafuente B, Downs RT, Walter MJ and Hazen RM. 2022. *What is mineral informatics?*. Earth-Science Reviews submitted. https://doi.org/10.1002/essoar.10511201.2

Rampe, E.B., Blake, D.F., Bristow, T.F., Ming, D.W., Vaniman, D.T., Morris, R.V. et al. (2020) Mineralogy and geochemistry of sedimentary rocks and eolian sediments in gale crater, Mars: A review after six earth years of exploration with curiosity. *Chemie Der Erde Geochemistry*, 80(2), 125605. https://doi.org/10.1016/j.chemer.2020.125605

Shafranovsky, I.I. (1983) Statistical regularities and generalizing law in mineral distribution by symmetry systems. *Zap. Vsesoyuzn. Mineral O–Va*, 112(2), 177–184.

Shafranovsky, I.I. & Feklichev, V.G. (1982) Symmetry statistics of minerals. *Mineralogicheskii Zhurnal*, 4(6), 31–36.

Urusov, V.S. (2002) Principle of structural Dissymmetrization minimum and its disruption by scarce new minerals. *Doklady Earth Sciences*, 386(3), 838–842.

Urusov, V.S. (2007) Symmetry statistics of mineral species and the evolutionary Dissymmetrization of mineral matter. *Geology of Ore Deposits*, 49, 497–504. https://doi.org/10.1134/S1075701150707001X

Vernadsky VI. 1903. *The fundamentals of crystallography*. Moscow, Russia: Moskva Gosuniversitet

Vernadsky, V.I. (1988) Crystallography. Selected works, ed. by Urusov, VS. In: Frank-Kamenetsky, V.A., Koptsik, V.A. & Urusov, V.S. (Eds.) *Comments of Shafranovsky, II*. Moscow, Russia: Nauka, pp. 66–181.

Yushkin, N.P., Shafranovsky, I.I. & Yanulov, K.P. (1987) *The Laws of symmetry in mineralogy*. Moscow, Russia: Nauka.