American Mineralogist—In Press—June 2015

Earth's "missing" minerals

Robert M. Hazen[1*], Grethe Hystad[2], Robert T. Downs[3],
Joshua Golden[3], Alex J. Pires[3], and Edward S. Grew[4]

[1]*Geophysical Laboratory, Carnegie Institution, 5251 Broad Branch Road NW, Washington, D. C. 20015, USA.*
[2]*Department of Mathematics, University of Arizona, Tucson, Arizona 85721-0089, USA.*
[3]*Department of Geosciences, University of Arizona, 1040 East 4th Street, Tucson, Arizona 85721-0077, USA.*
[4]*School of Earth and Climate Sciences, University of Maine, Orono, Maine 04469, USA.*

ABSTRACT

Recent studies of mineral diversity and distribution lead to the prediction of >1563 mineral species on Earth today that have yet to be described—approximately one fourth of the 6394 estimated total mineralogical diversity. The distribution of these "missing" minerals is not uniform with respect to their essential chemical elements. Of 15 geochemically diverse elements (Al, B, C, Cr, Cu, Mg, Na, Ni, P, S, Si, Ta, Te, U, and V), we predict that approximately 25% of the minerals of Al, B, C, Cr, P, Si, and Ta remain to be described—a percentage similar to that predicted for all minerals. Almost 35% of the minerals of Na are predicted to be undiscovered, a situation resulting from more than 50% of Na minerals being white, poorly crystallized, and/or water soluble, and thus easily overlooked. In contrast, we predict that fewer than 20% of the minerals of Cu, Mg, Ni, S, Te, U, and V remain to be discovered. In addition to the economic value of most of these elements, their minerals tend to be brightly colored and/or well crystallized, and thus likely to draw attention and interest. These disparities in percentages of undiscovered minerals reflect not only natural processes, but also sociological factors in the search, discovery, and description of mineral species.

E-mail: rhazen@ciw.edu

Keywords: Mineral diversity, nickel, sodium, tellurium, mineral ecology, mineral evolution, chance versus necessity, philosophy of mineralogy, sociology of mineralogy

INTRODUCTION

Earth's near-surface mineralogy has diversified significantly over 4.5 billion years as a consequence of physical, chemical, and biological processes (Hazen et al. 2008, 2011, 2012; Hazen and Ferry 2010). Hazen (2013) estimated that 420 mineral species may have been present in the Hadean Eon, whereas as many as 1500 species arose from physical and chemical events prior to 3 billion years ago. Biological processes, most notably near-surface environmental changes following the Great Oxidation Event at 2.4 to 2.2 Ga and Phanerozoic biomineralization subsequent to ~540 Ma, have led to numerous additional phases (Hazen et al. 2008, 2013a, 2013b; Sverjensky and Lee 2010; Dove 2010), including an estimated 70 percent of the ~5000 minerals approved by the International Mineralogical Association (rruff.info/ima; Downs 2006).

Earth's mineralogical diversity and distribution arise from both deterministic processes and frozen accidents (Grew and Hazen 2014; Hazen et al. 2015). For example, chance and necessity both play a role in the distribution of mineral species among the 72 essential mineral-forming chemical elements. Previous authors have identified a correlation between the crustal abundance of an element and the number of mineral species incorporating that element (Fig. 1; Yaroshevsky and Bulakh 1994; Higgins and Smith 2010; Christy 2015; Hazen et al. 2015).

Deviations from this trend arise from several factors. On the one hand, fewer mineral species than predicted by the general trend occur for rare elements that mimic more abundant elements (*e.g.*, Ga for Al, Hf for Zr, and rare earth elements for Ce and Y). On the other hand, significantly more mineral species tend to occur for elements with multiple oxidation states (*e.g.*, Cu, Te, or U), unique coordination geometries (*i.e.*, elements with lone-pair electrons), and/or multiple geochemical roles (*e.g.*, Cu forms chalcogenides, oxides, and halides).

In this contribution we examine another potential contribution in the scatter of data in Figure 1—the unequal distribution of Earth's "missing" minerals among different elements. Our statistical studies of 15 diverse mineral-forming chemical elements suggest that documentation of Earth's mineralogy is more complete for some elements than for others. This analysis not only reveals a new factor in the study of "anomalous mineral diversity" in the Periodic Table (Christy 2015), but it also points to as yet rarely explored sociological influences on the scientific study of mineral diversity and distribution—influences in addition to the problem of the characterization of new mineral species (*e.g.*, Bulakh et al. 2003).

Mineral Diversity-Distribution Relationships

The relative roles of chance versus necessity are notably reflected in the distribution of rare minerals—those known from only a few localities. More than half of all mineral species are recorded from 5 or fewer localities, with 22% of species known from only 1 locality (data recorded in mindat.org as of 1 February 2014). Hystad et al. (2015a) discovered that the relationship between mineral diversity and distribution is analogous to the frequency distribution of words in a book—i.e., a few words such as "a", "and", and "the" are very common, but most words are used infrequently. Idiosyncratic combinations of rare words and phrases can be used to identify the genre and authorship of an unsigned text, and they usually conform to a Large Number of Rare Events (LNRE) frequency distribution (Baayen 2001; Evert and Baron 2008).

Here we explore how LNRE models can also be exploited in mineralogy to estimate the minimum abundances, as well as some chemical characteristics, of as-yet-undiscovered mineral species. In this work we employ two types of Large Number of Rare Events models—both Generalized Inverse Gauss-Poisson (GIGP) and finite Zipf-Mandelbrot (fZM) distribution functions (Hystad et al. 2015a, 2015b; see Supplementary Materials). These frequency distributions use an empirical 3-parameter fit to model the number of mineral species found at m localities, typically for $1 \leq m \leq 15$. The GIGP model is the more robust formulation; it works well for all minerals, as well as for the 15 subsets of minerals containing diverse elements. The fZM model is more sensitive to scatter in mineral species-locality data; hence, we were unable to fit the data for C, Cr, Na, and Si to fZM models. We employed the fZM model when possible, because it facilitates calculation of individual probabilities in the population, which can be used to run simulations of alternative "Earth-like" planets (Hystad et al. 2015b). By contrast, direct calculation of probability distributions are not yet possible for GIGP models, though efforts are now underway to solve this problem.

Our initial analysis considered 4831 Earth minerals from 135,415 localities with 652,856 individual mineral-locality data (Fig. 2*a*; Hazen et al. 2015; Hystad et al. 2015a). Large number of rare events models are used to calculate "accumulation curve," which document the rates of discovery of new mineral species as more mineral species-locality data are recorded. Accumulation curves extrapolate to the predicted total number of mineral species on Earth today (Fig. 2*b*). These methods, which are widely employed in ecological studies of species diversity, lead to the prediction of a minimum of 6394 minerals, implying that at least 1563 mineral species exist on Earth today, but have yet to be documented. This prediction is a robust minimum estimate, in spite of the many inaccuracies and incompleteness of the mindat.org crowd-sourced data resource. Because the mindat resource constitutes such a large dataset, we suggest that database errors will not introduce significant bias if used only for extrapolating large-scale patterns—a conclusion reached by Adrain and Westrop (2000) for a comparable global database of fossil genera "rife" with error. Our predicted numbers of missing species represent minimum values because new search strategies (*e.g.*, field deployment of hand-held Raman spectrometers or targeted field studies based on the findings reported herein) and analytical tools (*e.g.*, transmission electron microscopic or confocal micro-Raman spectroscopic identification of micro/nanophases) are continuously being introduced and thus expand the range of detectable species.

ELEMENTS DISPLAY DIFFERENT PERCENTAGES OF MISSING MINERALS

Formalisms employed to estimate the total number of Earth's missing minerals can also be applied to subsets of the 652,856 species-locality data based on geographic locality, age of formation, lithological context, chemistry, or other objective criteria. Here we employ LNRE frequency models to analyze the numbers and percentages of missing minerals for 15 diverse chemical elements (Goldschmidt 1937): Al, B, C, Cr, Cu, Mg, Na, Ni, P, S, Si, Ta, Te, U, and V. Table 1 lists the total number of known minerals, predicted mineral diversity, and percent missing minerals for each of these elements.

The 15 elements for which we have obtained LNRE frequency distributions vary in terms of the percentages of their missing minerals. Seven elements—Al (27.1%), B (26.1%), C (26.5%), Cr (22.3%), P (25.5%), Si (28.4%), and Ta (28.6%)—have percentages similar to that for all minerals (24.4%). However, fewer than 20% of minerals are predicted to be missing for 7 elements—Cu (17.4%), Mg (18.4%), Ni (15.6%), S (17.8%), Te (9.9%), U (17.6%), and V (19.6%). By contrast, 34.7% of Earth's sodium minerals are predicted to be as yet undescribed.

Figures 2, 3, and 4 display frequency spectra in the form of bar plots (*a*) for m localities ($1 \leq m \leq 15$), and mineral species accumulation functions (*b*), for all minerals (Fig. 2), nickel minerals (Fig. 3), and sodium minerals (Fig. 4). These plots reveal important characteristics of mineral diversity-distribution systematics. Of special importance in Figures 2*b*, 3*b*, and 4*b* are ratios of minerals from 1 ($m = 1$) versus 2 ($m = 2$) localities, as indicated in curves labeled 1 and 2, respectively. Initially, the number of minerals known from only 1 locality rapidly outpaces minerals known from exactly 2 localities. However, as the number of data increases, minerals with $m = 1$ reach a maximum and then start to decrease, but those with $m = 2$ continue to increase. Eventually, when most mineral species have been discovered, both the $m = 1$ and $m = 2$ curves decrease; ultimately, the number of minerals known from 2 localities surpasses the number known from only 1 locality. Thus, sodium (Fig. 4), with the highest percentage of missing minerals, also has the highest ratio of minerals with $m = 1$ (273 species) compared to $m = 2$ (122 species; 273/122 = 2.24). All minerals (Fig. 2) display an intermediate value (1062/569 = 1.87), whereas nickel (Fig. 3) has a smaller ratio (30/25 = 1.20), and tellurium (Table 1) has the smallest ratio (25/25 = 1.00) and, accordingly, the smallest percentage of missing minerals (9.9%). Thus, for any subset of minerals, the ratio of minerals with $m = 1$ versus $m = 2$ correlates with the percentage of missing minerals.

MINERAL SPECIMEN APPEARANCE INFLUENCES THE PERCENTAGE OF MISSING MINERALS

The distinctive appearance of mineral specimens appears to be the most important factor in the varied percentages of missing minerals (Table 1; Fig. 5). On the one hand, sodium has the highest predicted percentage of missing minerals (34.7%), while it also has the lowest percentage of minerals that are easy to detect in hand specimen. Based on a systematic survey of mineral photographs on mindat.org, we find that 49.5% of sodium minerals are colorful and/or occur in distinctive crystals, whereas more than half of Na species are white or grey in color and fine-grained or poorly crystalized. In addition, with the exception of some salts, sodium minerals tend not to be economically important and thus may receive less scrutiny than minerals of some other elements. Furthermore, many Na mineral species are water soluble and therefore unstable and ephemeral in most crustal environments.

By contrast, the minerals of Cu, Mg, Ni, S, Te, U, and V—all elements for which smaller than average percentages of minerals are predicted to be missing—also tend to produce minerals with distinctive color, luster, and/or crystal form that are easily recognized (and thus popular with mineral collectors). Of the 15 elements studied, copper (68.5%), magnesium (68%), nickel (72%), and uranium (76%) display the greatest percentages of distinctive specimens, while sulfur (62.5%) and vanadium (61%) are closer to the average value for all minerals (63%). In addition, 6 of these elements (Cu, Ni, S, Te, U, and V) are concentrated in economically important deposits that have thus been disproportionately studied.

The case of tellurium minerals is of special interest. Of the 72 mineral-forming elements Te, with more than 10 times the number of species than might be predicted from tellurium's crustal abundance alone, displays the greatest positive deviation from the general trends in Figures 1 and 5. The deviation in Figure 1 results, in part, from tellurium's multiple oxidation states and varied crystal chemical roles (Christy 2015). An additional factor is that Te has the smallest predicted percentage of missing minerals (9.9%) of the elements studied (16 of 178 predicted species total). As a corollary, Te is also the only element studied for which the numbers of minerals known from 1 and from 2 localities are equal. Tellurium is also an extreme outlier in Figure 5, having far fewer predicted missing species than would be expected from the relatively low percentage of easily recognized Te mineral species (51%). We suggest that an important contributing factor to the relative completeness of Te mineralogy, as well as the low percentage of species that are easily recognized in hand specimens, is the intense targeted study of Te minerals by a few distinguished research teams, notably S.A. Williams (21 new species, 1974 to 1982), E.M. Spiridonov and coworkers (8 new species, 1978 to 1989), A.C. Roberts and coworkers (14 new species, 1994 to 2010), and the ongoing research of A.R. Kampf, S.J. Mills, J. Marty, and coworkers (13 new species, 2010 to present; *e.g.*, Housely et al. 2011, and references therein). The effectiveness of their efforts in identifying microscopic phases in thin section is reflected in the predicted relative completeness of Te mineral inventories, despite the relatively low percentage of species easily recognized in hand specimen.

IMPLICATIONS

Several factors influence the differing percentages of missing minerals for different chemical elements. Physical and chemical characteristics play obvious roles. White and/or poorly crystallized phases are more likely to be overlooked. In addition, elements that form highly soluble salts, including halogens or alkali metals, or that form phases that are otherwise unstable in a near-surface environment, are less likely to be found and catalogued.

Economic factors also play a significant role, as minerals of valuable elements have received special scrutiny by mineralogists—intense study that must have biased the observed distribution. Patterns of mineral discovery also reflect the sociology of mineralogy, particularly the sensibilities of the mineral collecting community. We obtain mineral locality information from the crowd-sourced database mindat.org. It is not surprising, therefore, that brightly colored, lustrous, and/or morphologically distinct minerals are disproportionately represented owing to observational bias.

The effects of different percentages of missing minerals for different elements would have a modest but significant impact on the positions of points in Figure 1. If we considered all minerals, including Earth's predicted missing minerals, for each element, then all points in Figure 1 would shift upward. However, points for elements with relatively well-documented minerals, including Cu, Mg, Ni, S, Te, U, and V, would shift less relative to the average, whereas the point for underdescribed Na minerals would shift more.

This study illustrates the great promise of exploiting ever growing mineral data resources, coupled with the application of powerful statistical methods. We anticipate that the discovery of new minerals, which has traditionlly been based on chance finds, can transformed to a predictive science. To bring this opportunity into reality, we have commenced a series of contributions in which we will identify likely compositions and localities of Earth's "missing" minerals.

ACKNOWLEDGEMENTS

We are grateful to A.G. Christy, S.J. Mills, M.J. Origlieri, and H. Yang for valuable discussions, and to J.M. Hughes, I.P. Swainson, and an anonymous reviewer for thoughtful and constructive analysis and suggestions. This work was supported in part by the NASA Astrobiology Institute, the Deep Carbon Observatory, the Alfred P. Sloan Foundation, the W.M.Keck Foundation, a private foundation, and the Carnegie Institution for Science for support of mineral evolution research.

REFERENCES CITED

Adrain, J.M. and Westrop, S.R. (2000) An empirical assessment of taxic paleobiology. Science, 289, 110-112.

Baayen, R.H. (2001) Word Frequency Distributions. Dordrecht, The Netherlands: Kluwer.

Bulakh, A.G., Zolotarev, A., and Britvin, S.N. (2003) A retrospect of discovery of minerals (1775−2000) and a look into the future. Neues Jahrbuch für Mineralogie. Monatshefte, 2003, 446-460.

Christy, A.G. (2015) Causes of anomalous mineralogical diversity in the Periodic Table. Mineralogical Magazine, 79, 33-49.

Dove, P.L. (2010) The rise of skeletal biomineralization. Elements, 6, 37-42.

Downs, R.T. (2006) The RRUFF Project: an integrated study of the chemistry, crystallography, Raman and infrared spectroscopy of minerals. Program and Abstracts of the 19th General Meeting of the International Mineralogical Association in Kobe, Japan. O03-13.

Evert, S. and Baron, M. (2008) Statistical Models for Word Frequency Distributions, Package zipfR.

Goldschmidt, V.M. (1937) The principles of distribution of chemical elements in minerals and rocks. Journal of the Chemical Society, 1937, 655-673.

Grew, E.S. and Hazen, R.M. (2014) Beryllium mineral evolution. American Mineralogist, 99, 999-1021.

Hazen, R.M. (2013) Paleomineralogy of the Hadean Eon: A preliminary list. American Journal of Science, 313, 807-843.

Hazen, R.M. and Ferry, J.M. (2010) Mineral evolution: mineralogy in the fourth dimension. Elements, 10, 9-12.

Hazen, R. M., Papineau, D., Bleeker, W., Downs, R.T., Ferry, J., McCoy, T., Sverjensky, D., and Yang, H. (2008) Mineral evolution. American Mineralogist, 93, 1693-1720.

Hazen R.M., Bekker A., Bish D.L., Bleeker W., Downs R.T., Farquhar J., Ferry J.M., Grew E.S., Knoll A.H., Papineau D., Ralph J.P., Sverjensky D.A., and Valley J.W. (2011) Needs and opportunities in mineral evolution research. American Mineralogist, 96, 953-963.

Hazen, R.M., Golden, J., Downs, R.T., Hystad, G., Grew, E.S., Azzolini, D., and Sverjensky, D.A. (2012) Mercury (Hg) mineral evolution: A mineralogical record of supercontinent assembly, changing ocean geochemistry, and the emerging terrestrial biosphere. American Mineralogist, 97, 1013-1042.

Hazen, R.M., Downs, R.T., Kah, L., and Sverjensky, D.A. (2013a) Carbon mineral evolution. Reviews in Mineralogy & Geochemistry, 75, 79-107.

Hazen, R.M., Sverjensky, D.A., Azzolini, Bish, D.L., Elmore, S., Hinnov, L., and Milliken, R.E. (2013b) Clay mineral evolution. American Mineralogist, 98, 2007-2029.

Hazen, R.M., Grew, E.S, Downs, R.T., Golden, J., and Hystad, G. (2015) Mineral ecology: chance and necessity in the mineral diversity of terrestrial planets. Canadian Mineralogist, in press. DOI: 10.3749/canmin.1400086

Higgins, M.D. and Smith, D.G.W. (2010) A census of mineral species in 2010. Elements, 6, 346.

Housely, R.M., Kampf, A.R., Mills, S.J., Marty, J., and Thorne, B. (2011) The remarkable occurrence of rare secondary tellurium minerals at Otto Mountain, near Baker, California. Rocks & Minerals, 86, 132-142.

Hystad, G., Downs, R.T., and Hazen, R.M. (2015a) Mineral frequency distribution conforms to a Large Number of Rare Events model: Prediction of Earth's missing minerals. Mathematical Geosciences, DOI 10.1007/s1 1004-015-9600-3.

Hystad, G., Downs, R.T., Grew, E.S., and Hazen, R.M. (2015b) Statistical analysis of mineral diversity and distribution: Earth's mineralogy is unique. Earth & Planetary Science Letters, in press.

Sverjensky, D.A. and Lee, N. (2010) The Great Oxidation Event and mineral diversification. Elements, 6, 31-36.

Yaroshevsky, A.A. and Bulakh, A.G. (1994) The mineral composition of the Earth's crust, mantle, meteorites, moon and planets. In Marfunin, A.S. (ed) Advanced Mineralogy, Volume 1: Composition, Structure, and Properties of Mineral Matter: Concepts, Results and Problems, p. 27-36. Springer-Verlag, Berlin, Heidelberg.

Table 1. Numbers of known mineral species; predicted number of species; inferred number of missing species; percentage of missing species; and percentage of species easily recognized in hand specimens owing to color and/or crystal form (see text).

Element	**# Known Species**	**# Predicted**	**# Missing**	**% Missing**	**% Easily Recognized**
All species	4831	6394	1563	24.4	63
Al	1005	1378	373	27.1	65
B	269	364	95	26.1	62
C	403	548	145	26.5	62
Cr	94	121	27	22.3	56
Cu	658	797	139	17.4	68.5
Mg	628	770	142	18.4	68
Na	933	1429	496	34.7	49.5
Ni	151	179	28	15.6	72
P	579	777	198	25.5	62
S	1028	1250	222	17.8	62.5
Si	1436	2002	568	28.4	61.5
Ta	60	84	24	28.6	57
Te	162	178	16	9.9	51
U	252	306	54	17.6	76
V	218	271	53	19.6	61

Figures and Captions

Figure 1. The number of known mineral species in Earth's upper crust is plotted versus crustal abundance (in atom percent) for 72 essential mineral-forming elements (based on the data in Hazen et al. 2015). Most elements plot close to the linear trend defined by all elements on this log-log plot. Several rare elements that lie below the trend mimic more common elements (e.g., Ga for Al; Hf for Zr; REE for Ce or Y). Several elements that lie above the trend have multiple oxidation states and/or varied crystal chemical roles. The percentage of as yet undescribed minerals also plays a role in these deviations.

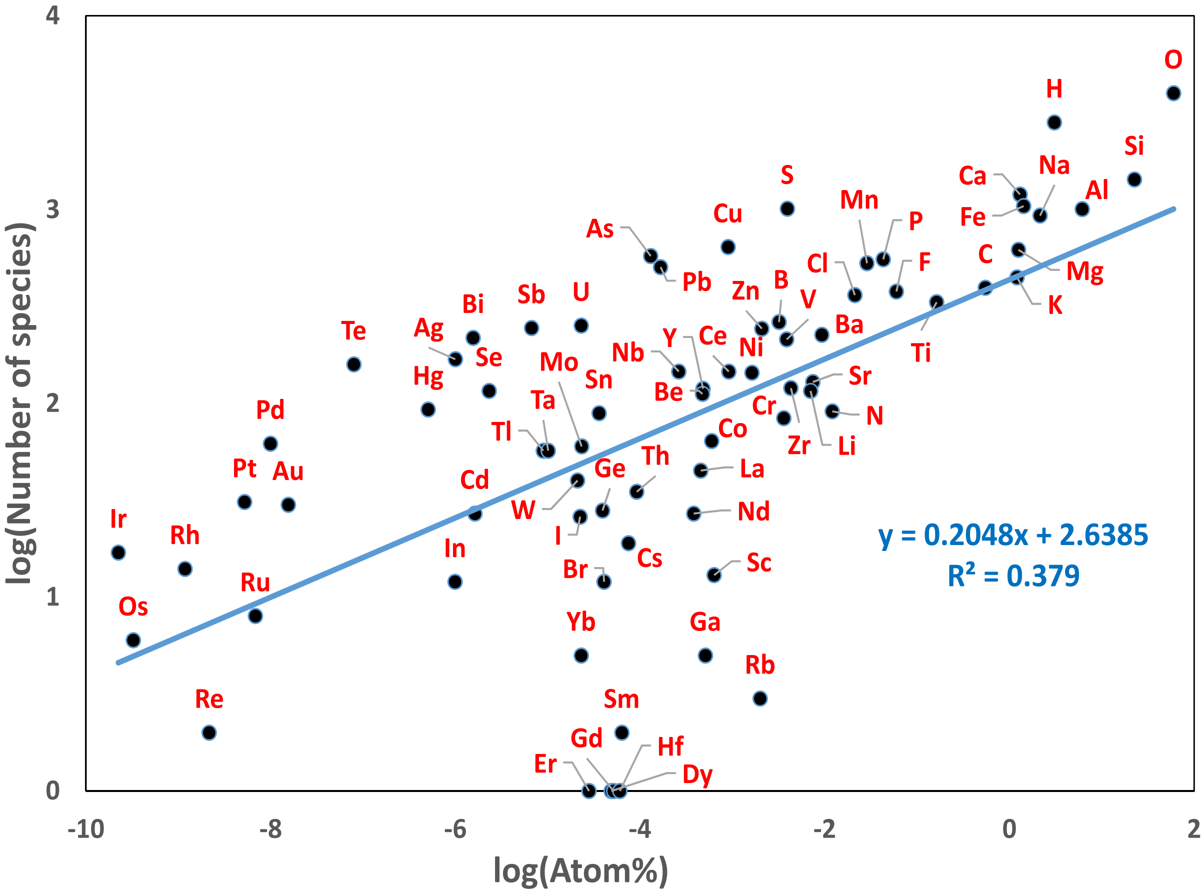

Figure 2. (*a*) Frequency spectrum analysis of 4831 Earth minerals, with 652,856 individual mineral-locality data (from mindat.org as of February 2014), employed a Generalized Inverse Gauss-Poisson (GIGP) function to model the number of mineral species for minerals found at from 1 to 15 localities (after Hazen et al. 2015). (*b*) This model facilitates the prediction of the mineral species accumulation curve (upper curve, "All"), which plots the number of expected mineral species (*y*-axis) as additional mineral species/locality data (*x*-axis) are discovered. The vertical dashed line indicates data recorded as of 1 February 2014 in mindat.org. The model also predicts the varying numbers of mineral species known from exactly 1 locality (curve 1) or from exactly 2 localities (curve 2). Note that the number of mineral species from only 1 locality is now decreasing, whereas the number from 2 localities is now increasing, though it will eventually decrease. We predict that the number of minerals known from 2 localities will surpass those from 1 locality when the number of species-locality data exceeds ~3 x 10^6.

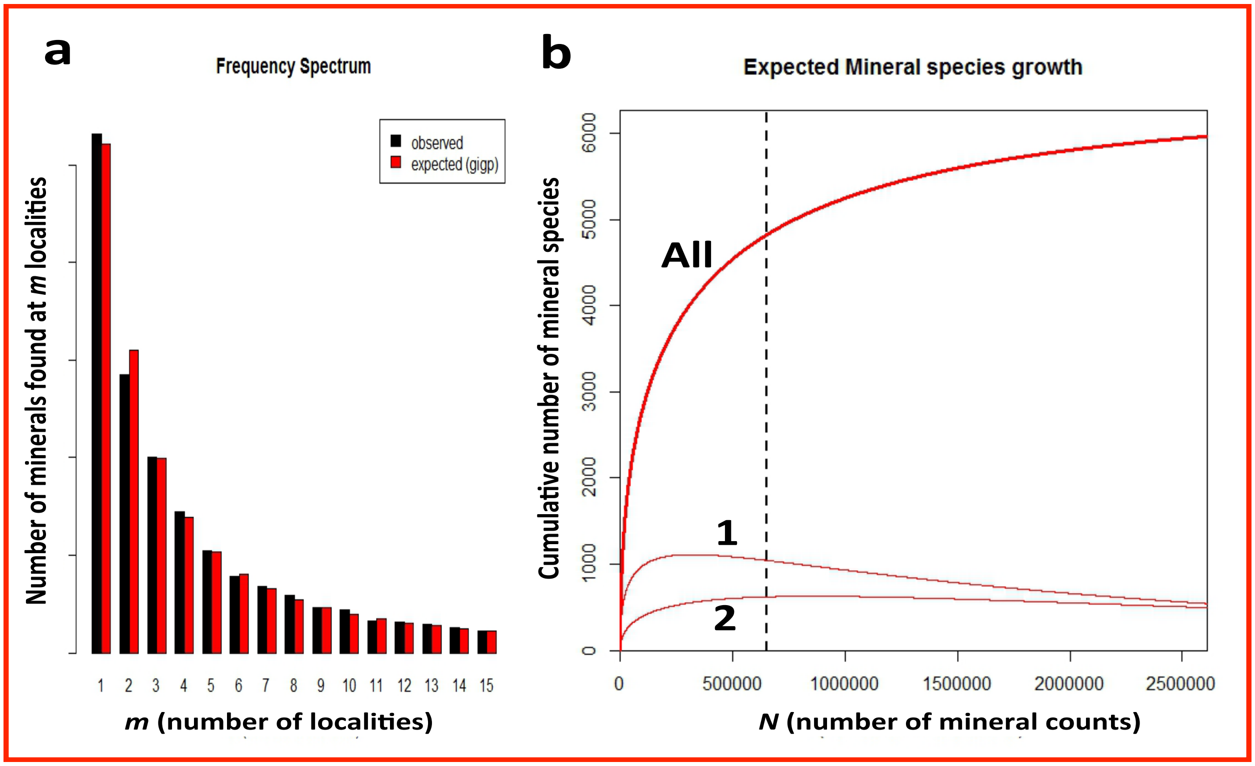

Figure 3. (*a*) Frequency spectrum analysis of 151 nickel minerals with 7567 individual mineral-locality data (from mindat.org as of 1 February 2014). These data conform to a finite Zipf-Mandelbrot (fZM) function that models the number of Ni mineral species found at from 1 to 15 localities. (*b*) Mineral species accumulation curve for nickel minerals: Note that the number of mineral species from only 1 locality (curve 1) is now decreasing, whereas the number of minerals from 2 localities (curve 2) is close to its maximum. We predict that the number of Ni minerals known from 2 localities will surpass those from 1 locality when the number of species-locality data exceeds 15,000.

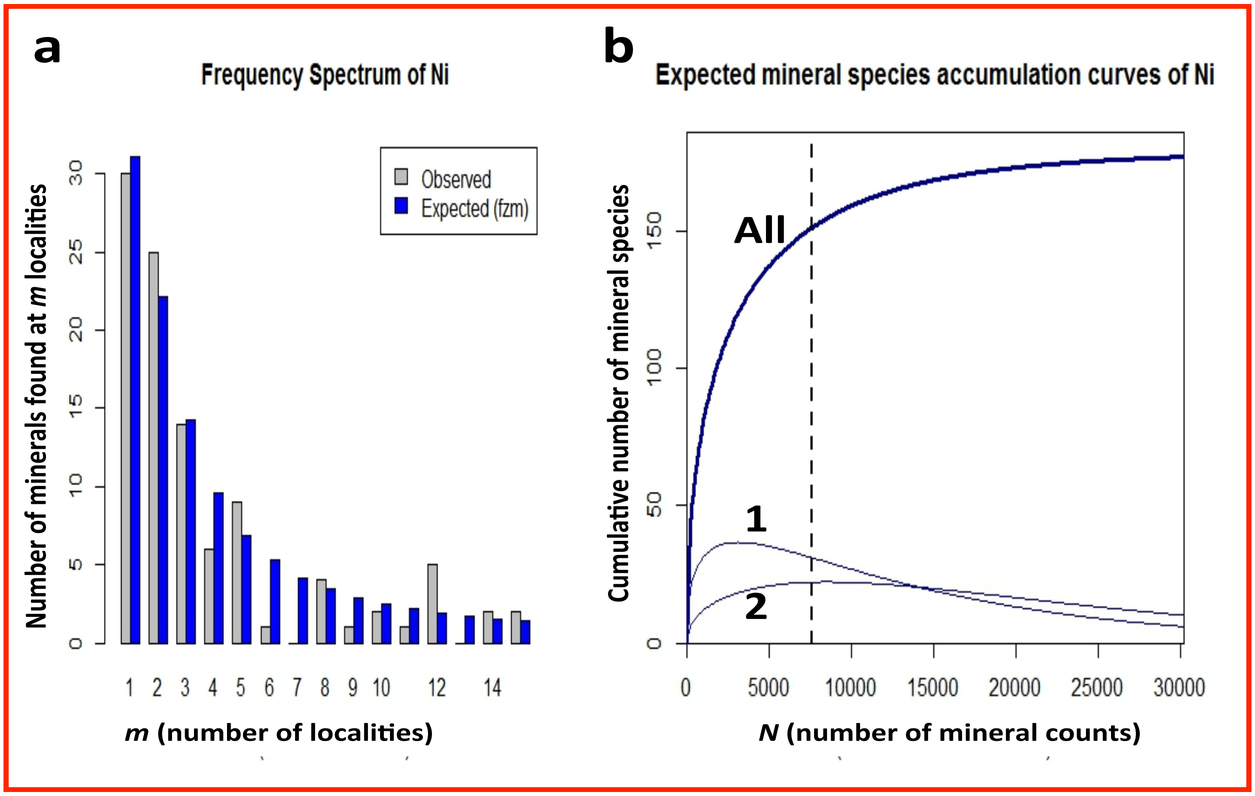

Figure 4. (*a*) Frequency spectrum analysis of 933 sodium minerals with 35,651 individual mineral-locality data (from mindat.org as of 1 February 2014). These data conform to a finite Zipf-Mandelbrot (fZM) function that models the number of Na mineral species found at from 1 to 15 localities. (*b*) Mineral species accumulation curve for sodium minerals: Note that the number of mineral species from only 1 locality (curve 1) is more than twice that of the number of minerals known from 2 localities (curve 2), in contrast to values in Figures 2 and 3.

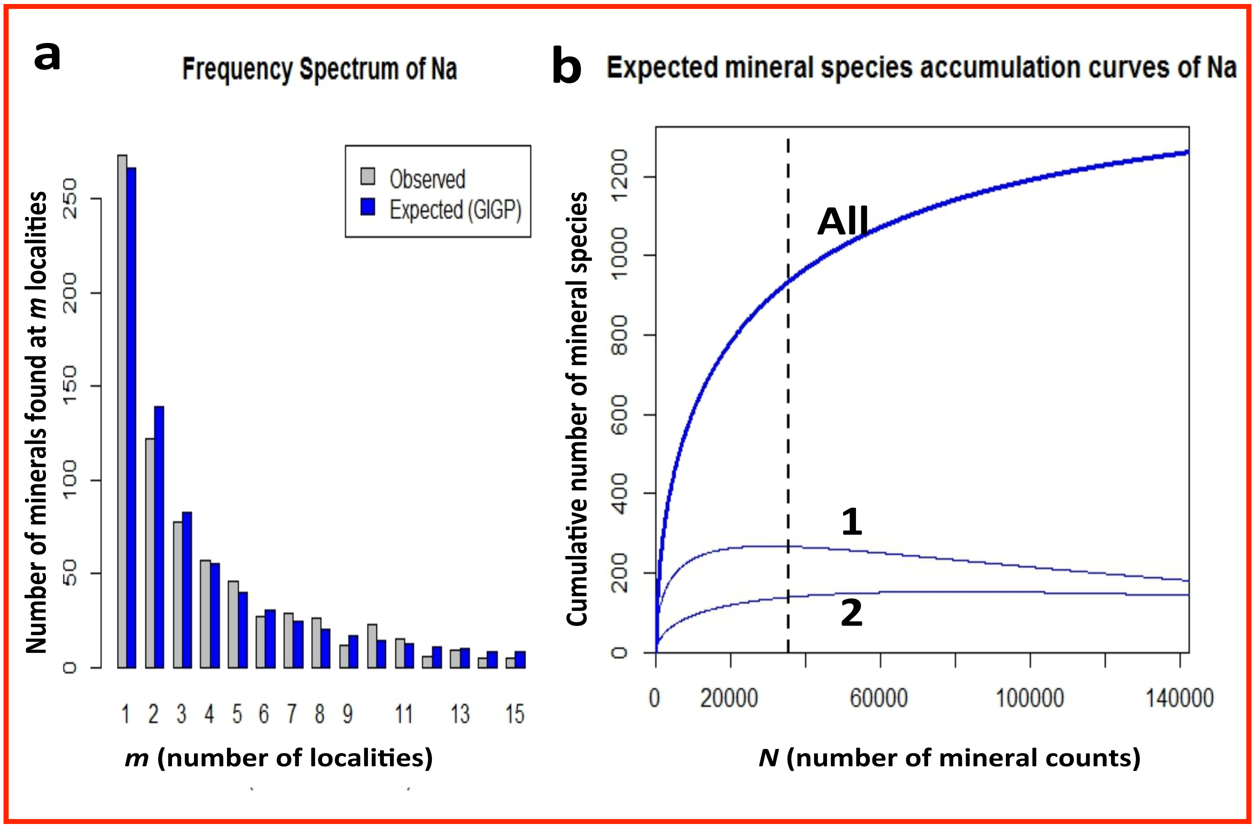

Figure 5. The predicted percentage of missing minerals is inversely correlated with the observed percent of specimens easily recognized in hand samples owing to their color and/or crystal form (see text). The linear regression line excludes tellurium, which may be an outlier because of the intense focus on discovering microscopic phases in thin section.