Applied Geochemistry 172 (2024) 106124

Contents lists available at ScienceDirect

Applied Geochemistry

journal homepage: www.elsevier.com/locate/apgeochem

Data generation for exploration geochemistry: Past, present and future

Julie E. Bourdeau [a], Steven E. Zhang [a], Glen T. Nwaila [b,*], Yousef Ghorbani [c]

[a] 39 Kiewiet Street, Helikon Park, Randfontein, 1759, South Africa
[b] School of Geosciences, University of the Witwatersrand, 1 Jan Smuts Ave., Johannesburg, 2000, South Africa
[c] College of Health and Science, School of Natural Science, University of Lincoln, Joseph Banks Laboratories, Green Lane, Lincoln, LN6 7DL, United Kingdom

ARTICLE INFO

Editorial handling by: Dr. O.P. Kreuzer

Keywords:
Exploration
Geochemistry
Sampling
Future
Data generation

ABSTRACT

Geochemical surveys are a cornerstone for data generation in geosciences, facilitating resource exploration. Geochemical data is essential for identifying mineralized areas, understanding local geology and assessing environmental impacts. This paper aims to: (1) review the process of geochemical data generation to gain an appreciation of the characteristics of traditional geochemical data; (2) review recent developments in the usage of geochemical data, particularly the disruption brought forth by those in data science and artificial intelligence; and (3) envision a future specification of geochemical data that would be fit-for-purpose for modern and emerging data users. This review reveals that benefits brought by advancements in automation, analytical technique and computing capability have unlocked unprecedented insights from geochemical data. However, the sustainability of re-purposing small geochemical data for big data methods is intrinsically questionable. The mismatch stems from rapidly changing data requirements in geochemistry, which is brought forth by: (1) a pivot away from scientific reduction through the adoption of system-level methods; (2) developments in geometallurgy; (3) skill gaps in geoscientific education; and (4) a growing demand of raw materials. While traditional methods will likely continue to serve many scientific needs, new strategies and techniques must be developed and implemented to cost effectively and efficiently generate bigger geochemical data. Bigger geochemical data must emerge in response to the already changing landscape of geochemical data usage. Solutions are multidimensional, from evolving geoscientific education, leveraging modern technology, explicating and differentiating data user and generator roles and responsibilities, to modernizing data management.

1. Introduction

Geochemical surveys are a dominant mechanism of geochemical data generation. Surveys provide the essential linkage that connects the physical world with data users. Its specific purpose in exploration geochemistry is to provide data that could potentially lead to the discovery of mineral deposits (Haldar, 2018). Aside from locating mineralization, geochemical data can be used to establish geochemical baselines, resolve bedrock geology, support method development, conduct environmental assessments, among others (Garrett, 1983; Govett, 1983; Demetrides et al., 2018; Hosseini-Dinani et al., 2018). In the broadest terms, geochemical surveys are a systematic and multi-stage process that encompasses: (1) data generation (survey design, sample collection and preparation, element determination); (2) data management (quality assurance and quality control [QA/QC]), data processing and curation; and (3) data usage (modelling and analysis, and outcome verification). Geochemical surveys have been conducted using a variety of media and at regional (or reconnaissance) to deposit scales, depending on the purpose (Demetrides et al., 2018; Hosseini-Dinani et al., 2018). Geochemical data, like all geoscientific data (geodata), continues to grow in abundance and complexity (including synthetic data), aided by an increased capability across the data pipeline (Johnson et al., 2013; Talapatra, 2020; Ma, 2023). For example, improvements in transport and infrastructure have facilitated the mapping and surveying of increasingly remote areas. Simultaneously, improvements in analytical instrument capability have created more voluminous and higher-dimensional data.

Automation, data science, artificial intelligence (AI) and computational methods continue to revolutionize our ability to derive insights from geochemical data (e.g., Zuo, 2020; Lawley et al., 2021; Zuo et al., 2019, 2022; Zhang et al., 2023c; Bourdeau et al., 2023; Zuo and Xu, 2023). It is critical to appreciate that unlike developments in transport technology, AI and computing hardware are exponential technologies, meaning that their growth rate is proportional to their present state

* Corresponding author.
E-mail address: glen.nwaila@wits.ac.za (G.T. Nwaila).

https://doi.org/10.1016/j.apgeochem.2024.106124
Received 7 February 2024; Received in revised form 17 July 2024; Accepted 28 July 2024
Available online 30 July 2024
0883-2927/© 2024 The Authors. Published by Elsevier Ltd. This is an open access article under the CC BY-NC license (http://creativecommons.org/licenses/by-nc/4.0/).

(Grossi et al., 2021). Exponential growth is fundamentally disruptive, because its impact is not humanly or linearly predictable in the long term. There is already an exponential increase in the number of publications in AI, which has a doubling time of 23 months and shows no sign of slowing (Krenn et al., 2023). A similar trend in the number of publications using AI (and related methods, such as statistical learning, hereon broadly referred to as AI) is also applicable in geosciences (Dramsch, 2020) and is also noticed in exploration geochemistry (Fig. 1). Disruption in geochemical data usage is witnessed by geoscientists adopting data science frameworks, transcribing in-discipline activities (e.g., geochemical targeting) into AI tasks (anomaly detection) (Zuo and Xiong, 2018; He et al., 2022; Zhang et al., 2022b, 2023c) and the rise of geodata science pedagogy, formalizing previously ad-hoc techniques (Ma, 2023; Wang et al., 2023). This is unsurprising, because data science is intentionally transdisciplinary (designed for other disciplines) and specializes in inferential modelling, by abstracting the concept of data and standardizing the process to model data using a framework (e.g., a workflow; Hazzan and Mike, 2023). Therefore, applicative data science guides applications of AI and is methodology-centric (whereas computer science guides development of AI algorithms and hardware). Adaptations of data science to exploration has already resulted in a geodata science framework (which is applicable to all geodata types beyond geochemical) that adds geospatial analysis to a data science framework (e.g., Zuo, 2020; Zhang et al., 2024b; Yousefi and Carranza, 2015a, b). The disruption to data generation is significant but less obvious given that AI demands specific data characteristics, primarily volume (abundance), quality (e.g., machine readable) and velocity (e.g., fast enough to refine or validate models during deployment). Such characteristics are very different than those provided by traditional geochemical data, which is difficult to address because of the growing expertise gap between data generators and users.

Most historic improvements to geochemical survey methodology were centered around scientific reduction (Kleinhans et al., 2005). The impact has been an increasingly more granular data in its depiction of reality, becoming more accurate and precise (among other properties; Cohen et al., 2010), such that physical reality is more reduced to its components (van Riel and Van Gulick, 2024). Reductive data is designed to meet the needs of hypothesis-driven science, which is a key domain aside from data-driven science. More reductive data lends to higher discriminatory power per sample, which means that hypotheses or models are more easily tested or differentiable. Adherence to reductionism kept geochemical survey practices resource- and time-intensive, making geochemical data costly (e.g., Demetrides et al., 2018). It may already be the case that the most voluminous uses of geochemical data are no longer hypothesis-driven, but rather data-driven, since data science methods are increasingly used in publications (He et al., 2022; Ma, 2023; Fig. 1). It must be recognized that data-driven science is philosophically complementary but diametrically opposite of hypothesis-driven science. The fundamental goal of data-driven science is to create inferential (rather than causal) models of reality. It requires different data characteristics. Notions of quality are no longer focused on per-sample characteristics, but rather on the system aspects of data. The disruption on characteristics of geochemical data is a result of a change in the purpose of geochemical data, which is increasingly to suit the use of big data-modelling techniques that are exemplified by AI. This is a sweeping disruption to many scientific domains that is only recently and slowly receiving due attention (e.g., Karpatne et al., 2018; Keane and Wilson, 2018; Bergen et al., 2019; He et al., 2022; Ma, 2023).

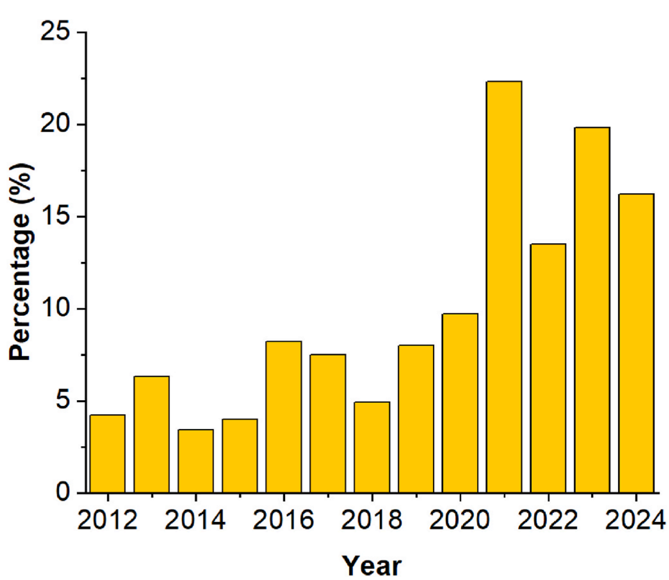

Fig. 1. Percentage of AI-related articles published in the topic of exploration geochemistry from 2012 to 2024. A total of six journals (Applied Geochemistry, Journal of Geochemical Exploration, Ore Geology Reviews, Natural Resources Research, Geochemistry: Exploration, Environment, Analysis and Computers and Geosciences) were investigated. Yearly volumes were randomly sampled, and articles (i.e., original research, review) featuring AI methods (including statistical learning but excluding solely traditional statistics) were counted.

Frameworks govern the use of big-data modelling methods (e.g., Cross Industry Standard Process for Data Mining [CRISP-DM]; Wirth and Hipp, 2000; Chapman et al., 2000), because Blackbox algorithms are unexplainable and unvalidated models are worthless (e.g., Hastie et al., 2017; Hazzan and Mike, 2023; Xu et al., 2019; Hawkins et al., 2023). Hence, data science emphasizes validation, which must occur in the form of additional data. In realms with sustained data science application, the tendency is to rely on big data, particularly voluminous data for initial modelling and sufficient velocity for deployment and validation (Chen and Lin, 2014; Zhang and Han, 2021). Therefore, it is unsustainable to deploy big data methods on traditional geochemical data, which can be voluminous, but is costly and of low velocity. Thus, transdisciplinary methods and traditional geochemical survey requirements are a major source of data specification discrepancy. Currently, the use of AI on geochemical data constitutes a transient or periodic phenomenon due to an abundance of accumulated data (see He et al., 2022 and references therein). Streams of geochemical data, e.g., 'big geochemical data', are generally unheard of (Zhang et al., 2023b, 2023c). High data velocity is conceptually foreign to exploration geochemistry because geochemical data generation, such as sampling and analysis, in their current state, are only linearly scalable.

If data-driven science continues to flourish, then traditional exploration geochemistry is faced with an evolutionary barrier. There has been no requirement analysis of geochemical data, its management, usage and education in light of rapidly evolving data usage patterns and societal demands (e.g., green energy transition). A re-examination of survey-scale geochemical data generation serves to ensure that future data is specified for modern activities. To accomplish this task, we provide a summary review of the progress of geochemical data generation in surveys, mainly at the regional scale for reconnaissance. Specifically, we examine the role that exploration geochemistry has had over the years and provide a vision of how it could progress into the future. We first provide a historical overview (historical – 1990s), detailing survey design, limitations and challenges. We then review modern advancements (1990s–2020s), present current practices, and discuss pertinent and outstanding issues. Lastly, we present a vision for the future (2020s – beyond), including practices that will remain relevant and those that will experience changes. To ensure continuity across the data pipeline, practicality and relevance for the academies and industry, we provide a brief analysis of data management, education and downstream requirements, although the focus remains on data generation. We expect that the review and discussion are applicable to all spatial scales of exploration. In essence, we hope to drive and facilitate collaborations between geoscientists, engineers, data scientists and statisticians (among others), such that the field of geochemical exploration can continue to evolve.

2. Past (historical – 1990s)

2.1. Historical overview

The first geochemical 'surveys' (*sensu lato*) were undertaken by prospectors panning for heavy minerals centuries ago (Fig. 2). In the late 1800s to early 1900s, the first method used to chemically identify minerals (qualitatively) retained in the prospecting pan is thought to have been a carbon block and blowpipe (Prider, 1981; Fisher, 1988). The first documented exploration geochemical surveys considered as true geochemical surveys on a basis of their systematic sampling procedures, were developed in the Soviet Union in the 1920s and 1930s (Fig. 2; Garrett et al., 2008). A total of 17 field projects were documented with the intention to prospect for bedrock tin deposits, aided by the development of new chemical determination methods (especially optical spectroscopy; Flerov, 1935; Sergeyev, 1941; Garrett et al., 2008). Thereafter, geochemical surveys quickly spread to Europe, North America, and Africa (e.g., Hawkes, 1954; James, 1957). Notably, from 1961 to 1975, the first continental-scale geochemical mapping project was conducted by the U.S. Geological Survey (USGS) by Hansford T. Shacklette and colleagues (Fig. 2; Shacklette et al., 1971; Boerngen and Shacklette, 1981; Shacklette and Boerngen, 1984; Gustavsson et al., 2001). This began the era of modern regional-to continental-scale geochemical surveys. By the late 1970s–1980s, many countries such as Canada (Friske and Hornbrook, 1991), United Kingdom (McGrath and Loveland, 1992), Finland (Elo et al., 1992), and China (Xie and Ren, 1993), among others, subsequently followed suit with their own regional-to country-wide geochemical surveys.

Early geochemical surveys generally had a high sample density (multiple samples per km^2) with only one media, preferentially stream sediments (Smith et al., 2013; Garrett et al., 2008; Hosseini-Dinani et al., 2018). Subsequently, other types of media were added (mainly soil, lake sediment, vegetation, and rock), and by the 1970s–1980s, sampling density varied greatly, ranging from 1 per 1 km^2 for exploration projects to 1 per 10,000 km^2 for international-scale mapping projects (Reimann et al., 2001; Garrett et al., 2008; Smith et al., 2013). Results from early geochemical surveys demonstrated that stream and lake sediment samples are overall representatives of the entire basin catchment area, whereas soil and rock samples are representative of the immediate area (Webb et al., 1964; Coker and Nichol, 1975).

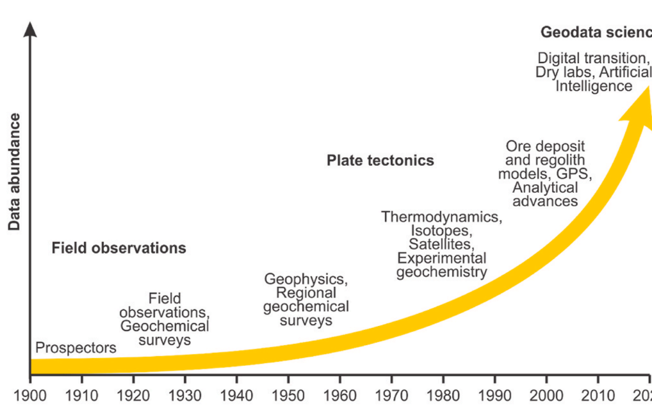

Fig. 2. Overview of major changes in exploration geochemistry with respect to data abundance. Figure inspired by a lecture presented by Dr. John F.H. Thompson (Cornell University; PetraScience Consultants).

2.2. Laying the foundations for the design of geochemical exploration surveys

Design constraints on geochemical surveys are more onerous at larger scales because, the cost per increment in sampling resolution is quadratic for two-dimensional surveys. Geochemical surveys must optimize perceived benefits against costs, which are mainly economic, but can also include non-economic considerations, such as incursions into protected areas. The benefits depend on the goal of the survey, and in the case of mineral reconnaissance missions, they are mainly the discovery of deposits. Surveyors therefore optimize survey design under two constraints that vary by sample media: (1) maximizing sampling density to match the survey budget; and (2) maximize the probability to capture foreseeable geochemical anomalies (of a characteristic size). Since the inception of true exploration geochemical surveys (1920s–1930s), numerous prospectors, exploration managers, agencies, and companies sought to optimize surveys, however, little quantitative work was done on the subject until the 1970s–1980s (e.g., Garrett, 1983; De Geoffroy and Wignall, 1985). The earliest quantitative survey designs employed probability and statistics.

Knowledge is built iteratively. The complete absence of any a-priori knowledge of an area and mineral deposit(s) implies that it would be impossible to optimize any survey. Consequently, the notion of greenfield to brownfield is a continuum. To engineer an ideal sampling density, it is important to have some pre-existing knowledge of the occurrence being sought (e.g., size and shape of occurrence, size and shape of alteration halo, dispersion mechanisms, drainage characteristics, geological environment, topographical relief, mineralization type, climate, etc.; Hosseini-Dinani et al., 2018 and references therein). Such knowledge is usually collected during a reconnaissance survey, whose purpose is to assess an area in anticipation of further surveys (Govett, 1983; Demetrides et al., 2018; Talapatra, 2020). It is important to consider the survey scale (e.g., to find a mineralized vein versus geochemical anomalies at a regional scale). Thus, both occurrence size and survey scale are important in survey design (Garrett, 1983; Hosseini-Dinani et al., 2018).

Using data collected during the reconnaissance survey, the average distance from the target (orebody) to where the geochemical signal is lost (background) can be quantified (Fig. 3; Cohen et al., 2010). In essence, the boundary of the expected geochemical halo and/or dispersion pattern can be taken to represent the limit of the local geochemical background (Fig. 3a; Garrett, 1983; Reimann and Garrett, 2005). Of course, the determination of this threshold is objectively difficult, as it is spatially variable depending on the type of orebody (e.g., vein, disseminated, lode deposits), bedrock geology, climate (weathering), physical environment, past environmental events (e.g., glaciation), among others (Fig. 3b; Hosseini-Dinani et al., 2018; Talapatra, 2020). A baseline method to determine this threshold is to sample an area lacking mineral deposits (background) and an area with known mineral deposits (anomaly). For baselining, areas with similar bedrock geology are preferable to minimize geochemical variance. However, samples from different areas will inevitably exhibit differences (e.g., contrasting geochemical variance) as samples typically capture different lithologies (e.g., bedrock geology; Matheron, 1963; Miesch, 1975). To minimize such issues, samples collected should be maximally related (e.g., a single stream segment), thus improving sample stationarity. However, such a solution is rarely available and samples must be collected from several populations. To mitigate this issue, several auto-correlation techniques were initially developed by Dijkstra and Kubik (1975) and Howarth and Martin (1979). Ultimately, the quantification of the threshold relies on the determination of the spatial semi- and co-variance (Garrett, 1983), which can be determined following the calculation of the mean and variance:

$$\text{Mean } \bar{z} = \frac{1}{n}\sum_{i=1}^{n} z_i \quad (1)$$

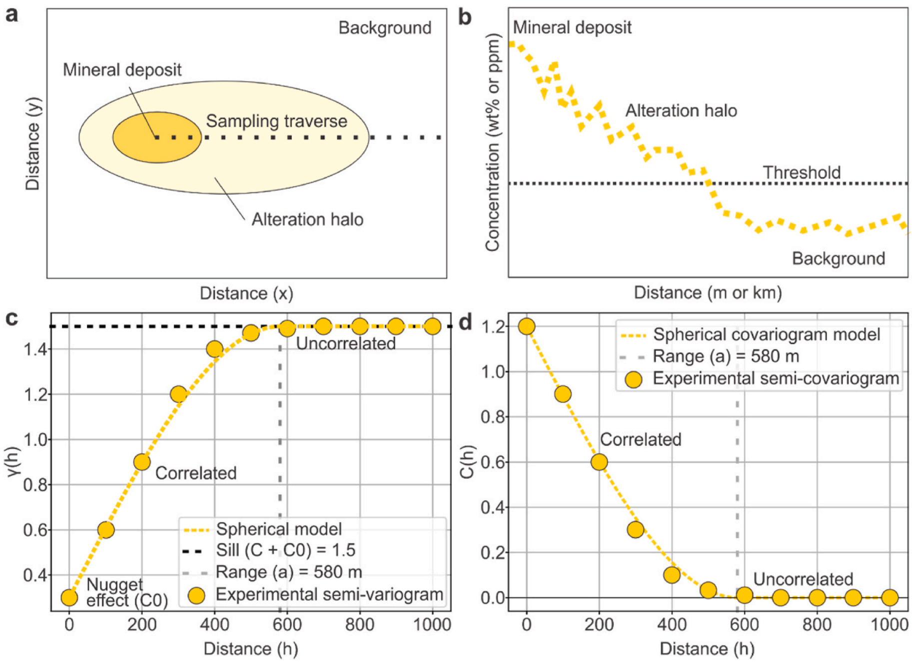

Fig. 3. (a) Mineral deposit in an area of interest. Here, a targeted sampling traverse (dashed line) is conducted to determine the sampling interval in the region. (b) Concentration of a chemical in samples obtained from the traverse in (a) progressing away from the mineral deposit and into the regional background. (c) Semi-variogram plot of the semi-variance (γ(h); Equation (3)) versus distance. Sample locations separated by distances closer than the range are correlated (mineral deposit and alteration halo), whereas sample locations above the range are uncorrelated (background). (d) Covariogram plot of the covariance (C(h)) versus distance. Note that the covariogram plot is the inverse of the semi-variogram and is another method of visualizing and determining the range.

$$\text{Variance } s^2 = \frac{1}{n-1}\sum_{i=1}^{n}(z_i - \bar{z})^2 \tag{2}$$

$$\text{Semi}-\text{variance } \gamma(h) = \frac{1}{2N(h)}\sum_{a=1}^{N(h)}[z(u_\alpha) - z(u_\alpha + h)]^2 \tag{3}$$

$$\text{Covariance } C(h) = \frac{1}{N(h)}\sum_{a=1}^{N(h)}(z(u_\alpha) - \bar{z})(z(u_\alpha + h) - \bar{z}) \tag{4}$$

Where z_i is a regionalized variable (a spatial measurement), N is the number of samples, h is the lag distance per interval, u_α is a location vector, $N(h)$ is number of pairs of points separated by the lag distance h, and $z(u_\alpha)$ and $z(u_\alpha + h)$ are values of a variable z at locations u_α and $(u_\alpha + h)$, respectively.

A semi-variogram is constructed by computing the variance of the differences between pairs of samples as a function of their separation distance or lag (Equation (3)). This is done by first calculating the semi-variance, which is half the average squared difference between all pairs of points separated by a certain distance. The semi-variance is then plotted against the lag distance to create a graph called a semi-variogram, which shows how the degree of spatial dependence of the variable varies with distance (Fig. 3c). An inverse of the semi-variogram in a form of $C(h) = \sigma^2 - \gamma(h)$, can be used to plot the co-variogram (Fig. 3d). The covariance is related to the semi-variance and measures the degree of correlation between pairs of sample points as a function of their separation distance or lag. It is calculated by computing the covariance of the variable at all pairs of sample points separated by a certain distance or lag (Equation (4)). Spatial continuity significantly impacts sampling interval and subsequent modelling of regionalized variables. Directional semi-variograms and semi-covariograms help to determine spatial anisotropy, which may guide sampling design. These two methods illustrated in Fig. 3c and d enable quantification of optimal sample closeness, by assessing the spatial correlation between samples and unknown locations, and the approximation of redundancy, which is proxied by spatial correlation between samples. Specifically, Fig. 3c shows that after about 600 m, the semi-variogram attains and maintains a high value, known as the sill. From the sill, the range is determined and here corresponds to 580 m. At the deposit scale, this value is a measure of the size of the target (mineral deposit and alteration halo; Fig. 3a). Therefore, applying the Nyquist-Shannon sampling theorem, the sampling interval will be set at a maximum of ½ (Whittaker, 1915) to a discipline-specific heuristic minimum of 1/3 (Garrett, 1983) of the estimated target size, to best detect possible anomalies (ore bodies) at the chosen scale. Other, more modern approaches are summarised in Hosseini-Dinani et al. (2018). Both the semi-variogram and co-variogram can be used to help determine the optimum sample spacing or interval for spatial estimation using methods such as kriging by examining the shape of the graph (Fig. 3c and d). The stability of the sill would suggest that there is little spatial dependence at larger sample spacings. If the sill is not stable or is cyclical, there is variable spatial dependence at greater distances.

Unfortunately, even to date, there is no consensus on how to spatially design a survey to best capture geochemical anomalies. This is because targets of reconnaissance surveys are unknown unknowns (e.g., mineral deposits), created by complex systems, which exhibit fractal dynamics, nonlinear distance to spatial proxies and a spatial and size distribution exhibiting power law behaviour (Zuo, 2016; Hobbs et al., 2022). Consequently, completely unknown mineral deposits are likely black swan events (Gelman, 2007). Instead, targeting criterion is usually formulated as knowledge-guided heuristics that attempt to locate proxies of known unknowns. Reconnaissance surveys in greenfield settings are generally more knowledge- or heuristics-driven rather than data-driven, based on operator expertise and past experience. Another key difference at the reconnaissance stage relative to the finer scales of geochemical surveys is that the metal accounting frameworks (including legal practices, e.g., SAMREC code and NI 43–101 standards of disclosure; SAMREC, 2016; OSC, 2011, respectively) do not apply. Hence, there is more variability in geochemical survey design, and because there are usually more parameters in a survey design than availability of knowledge to constrain them (surveys problems are generally underdetermined), some choices must be subjective or experiential. Nonetheless, survey designs can be iteratively refined using successive reconnaissance surveys to better monitor sampling and analytical variability. At the dawn of modern geochemical surveys, the systematic grid sampling pattern was canon (Garrett, 1983, Fig. 4a). However, this pattern faces a number of limitations (e.g., scale, topography, climate) and is therefore restricted to more accessible media types, such as soil, rocks and till (where possible). A large number of studies have demonstrated that stream and lake sediment samples are also useful (Hale and Plant, 1994; Hosseini-Dinani et al., 2018 and references therein). However, streams and lakes are not uniformly distributed and are rare in some parts of the world. As a result, the stratified random sample design was developed (Fig. 4b), but does not allow for monitoring of sample variance. Another survey design is the balanced nested hierarchical (Fig. 4c), which is preferred to analyze the variance between samples (for the purpose of testing the robustness of results), but requires many samples (Miesch, 1976; Garrett, 1983). Thus, to minimize the number of samples and preserve the ability to analyze sample variance, the unbalanced nested sampling design (Fig. 4d) was developed as a compromise (Tidball and Severson, 1976; Garrett and Goss, 1980).

2.3. Limitations and challenges faced by legacy geochemical exploration surveys

Early exploration geochemical surveys could only determine a limited number of elements (up to 20 elements for Canadian surveys conducted in the 1960s–1970s; Garrett et al., 2008; Shacklette et al., 1971; Smith et al., 2013). The most complete example documented is from the first continental-scale geochemical mapping project conducted by the USGS (1961–1975; Smith et al., 2013), which reported analytical concentrations for 50 elements (Ag, Al, As, B, Ba, Be, Bi, Br, C, Ca, Cd, Ce, Co, Cr, Cu, F, Fe, Ga, Ge, Hg, I, K, La, Li, Mg, Mn, Mo, Na, Nb, Nd, Ni, P, Pb, Pr, Rb, S, Sb, Sc, Se, Si, Sn, Sr, Ti, Th, U, V, Y, Yb, Zn and Zr) that were analyzed using a variety of analytical methods. However, 30 of these elements (Ag, Al, B, Ba, Be, Bi, Cd, Ce, Co, Cr, Cu, Fe, Ga, La, Mg, Mn, Mo, Na, Nb, Nd, Ni, Pb, Pr, Sc, Sr, Ti, V, Y, Yb and Zr) were semi-quantitatively determined, and cannot be used in rigorous data analysis (Myers et al., 1961; Grimes and Marranzino, 1968; Neiman, 1976; Smith et al., 2013).

Early analytical methods (circa 1960s–1980s) included: X-ray fluorescence (XRF); optical spectroscopy; flame photometry; colorimetry; delayed neutron activation analysis; instrumental neutron activation analysis (INAA); atomic absorption spectrometry (AAS); specific-ion electrode; EDTA titration; and combustion, among others (Shacklette et al., 1971; Friske and Hornbrook, 1991; Garrett et al., 2008; Smith et al., 2013). Early regional-to national-scale geochemical surveys (1960s–1970s) are often criticized for a lack of standardized sampling and analytical protocols, casting doubt over the obtained results. Thus, numerous government-funded regional-to national-scale geochemical surveys designed and implemented standardized sampling, analytical techniques and QA/QC programs (e.g., Friske and Hornbrook, 1991 for stream and lake sediment sampling under the National Geochemical Reconnaissance programme at the Geological Survey of Canada). Furthermore, detection limits for many of the analytical methods aforementioned were high (considerably above crustal abundance for some elements), further limiting the achievable quality of geochemical data. Nonetheless, these early geochemical surveys (1960s–1970s) pioneered our understanding of the Earth and its physical processes, as well as how to generate and analyze data for mineral discovery.

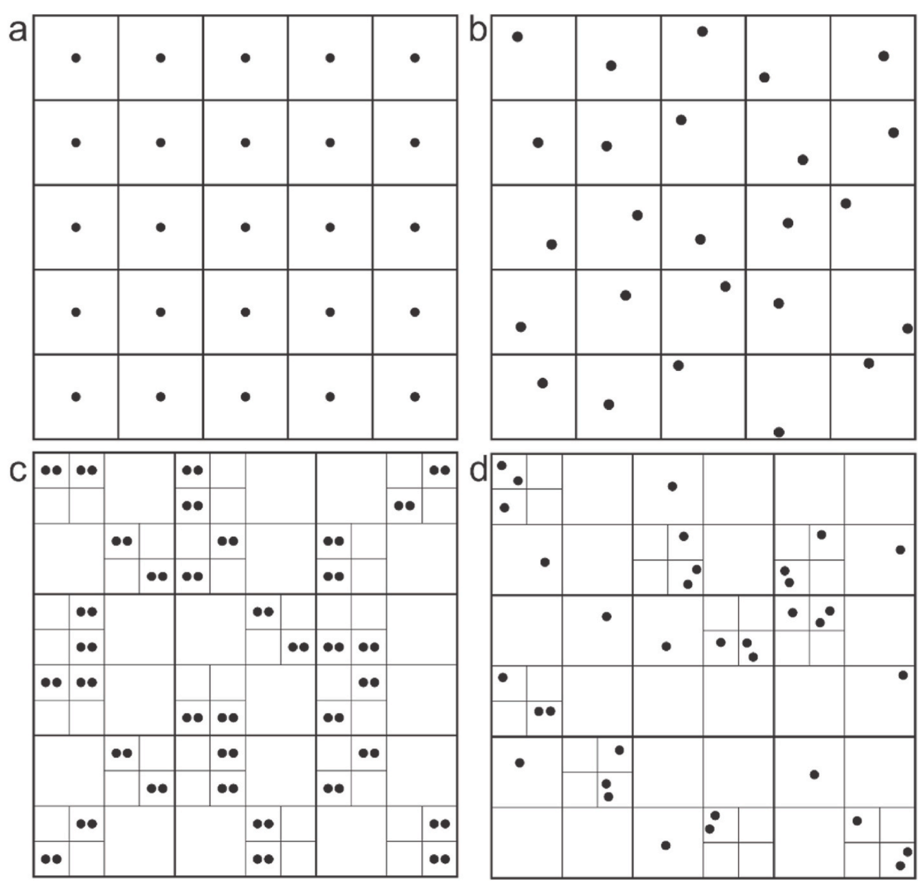

Fig. 4. Examples of survey designs. (a) Systematic grid sampling, preferentially used in early geochemical surveys. (b) Stratified random sampling, which allows for sampling of material which is not uniformly available (e.g., stream, lake, till). (c) Stratified random sampling, used to monitor variance between samples. (d) Unbalanced nested sampling, preserves the ability to monitor for sampling variance, while reducing the number of samples to be analyzed.

3. Present (1990s–2020s)

3.1. Modern advancements

The 1990s were denoted by stagnant to depressed commodity prices that led mining companies to de-emphasize mineral exploration (Cohen et al., 2010; Connolly and Orsmond, 2011). Of the exploration that took place, mining companies focused it nearby or on their own properties. Another consequence of the depressed commodity market was that the re-use of older datasets became popular. This period experienced probably the first major occurrence of data re-use and re-purposing. By the 1970s–1980s, most mineral deposits exposed at the surface had been discovered (Blain, 2000; Ghorbani et al., 2023a). Thus, the 1990–2000s marked a significant change in research direction to understand the regolith in order to discover buried mineral deposits using geochemistry (Fig. 2; e.g., Butt et al., 2000; Cameron et al., 2004, Coker, 2010 and references therein; Winterburn et al., 2020). Understanding the regolith ultimately led to the incorporation of physics-based modelling methods (e.g., elemental diffusion and transport) to aid mineral deposit targeting (Cohen et al., 2010; Johnson et al., 2013). To discover buried mineral deposits, new sample mediums and methods were developed and added to the repertoire, including: ground water (Giblin and Mazzuchelli, 1997; Pirlo and Giblin, 2004; Leybourne and Cameron, 2010), soil gas (Plet and Noble, 2023), vegetation (Cohen et al., 1998; Dunn, 2007), and more recently isotopes (Kirste et al., 2004; White, 2015). Hence, this period also witnessed the development of non-conventional geochemical surveys and the start of the post-conventional period of geochemical surveying.

Notable developments in analytical methods include: (1) development and uses of selective extraction techniques; (2) improvement in detection limits; (3) decrease in analytical costs; (4) improvement in the intelligibility of analytical devices (e.g., inductively coupled plasma-mass spectrometry [ICP-MS]); (5) additional elements that can be determined; and (6) instrument miniaturization and the rise of portable analytical devices (e.g., portable XRF [pXRF], portable X-ray diffraction [pXRD] and laser-induced breakdown spectroscopy [LIBS] instruments) (Fig. 2). The development of selective extraction techniques occurred in response to the development of new exploration models (e.g., elemental diffusion and transport, and development of alteration haloes). Selective extraction techniques are preferentially used to separate signals emanating from secondary processes (e.g., alteration of minerals in the regolith originating from a buried mineral deposit), versus elements from the transported regolith (e.g., glaciation; Hall and Bonham-Carter, 1998). Overall, the lower detection limits for routine elements have decreased by one half to an order of magnitude through improvements in analytical instrumentation-technology (e.g., sensors and algorithms; Cohen et al., 2010). This has allowed the proliferation of ultra-trace stable isotope analytical devices such as the multi-collector ICP-MS instrument (Caughlin, 2010), or elemental determinations on a micro-to nano-scale (e.g., scanning electron microprobe [SEM], focused-ion beam SEM and laser-ablation ICP-MS). The reduction of analytical costs is observed in both sample preparation (e.g., widespread use of microwave ovens in the late 1990s for sample digestion) and analytical procedures (e.g., reduction in the number of procedural steps, miniaturization and automation of instruments/procedures; de la Guardia and Garrigues, 2019). A direct effect of the reduction of analytical costs has been a drastic increase in sample throughput. Consequently, QA/QC programs, which can monitor for potential contamination, as well as the accuracy, precision and fitness-for-purpose of the data, started becoming commonplace (Piercey, 2014; Grunsky and de Caritat, 2017; de Caritat, 2022). Modern exploration geochemical datasets contain determinations for more than 40 elements, and include some previously difficult-to-detect elements such as Re and Te (e.g., McCurdy et al., 2016; de Caritat, 2022). Since 1994, the XRF instrument became sufficiently miniaturized to become fully portable (hand-held; pXRF; Cuffari, 2016). Additionally, in 1996 a portable LIBS instrument for elemental determination was developed (Yamamoto et al., 1996). Although portable instruments are not as accurate and precise as the much larger, laboratory-sized instruments, they have been proven to be advantageous in field settings and for rapid data generation (e.g., Winterburn et al., 2020; Lemière and Uvarova, 2020; Knight et al., 2021). Certainly, in a quantitative sense, these portable instruments provide a breadth of high-quality data that is in many cases comparable to analytical data from the decades prior.

Another notable, yet often overlooked advancement in geochemistry, is the near-exponential increase in computer availability and performance. This permitted the: (1) improvement in analytical instrumentation (e.g., Lemière and Uvarova, 2020); (2) use of more accurate algorithms during sample analysis (e.g., Tuoriniemi et al., 2015); (3) instrument automation and control infrastructure (e.g., Ghorbani et al., 2022, 2023b); and (4) creation of increasingly user-friendly software, making it possible to construct, handle, process and interpret larger datasets (Johnson et al., 2013; Talapatra, 2020; Ma, 2023). Evolution of computer hardware and software have also aided in-field and out-of-laboratory work (Lemière and Uvarova, 2020). Aside from portable instruments, digital devices (e.g., tablets, cellphones, laptops) are commonly used during field activities, because they contain a multitude of useful miniaturized sensors that include: cameras; compasses; altimeter and barometers; GPS receivers; and distance-measurement sensors such as lasers and detectors. These sensors permit the collection of field notes that accompany sampling activities, which become ancillary or metadata that permits the construction of modern datasets and databases, and facilitates QA/QC. Data can now be easily spatially visualized and interpreted using software, largely in the GIS domain (e.g., ESRI's ArcGIS, Google Earth Pro, QGIS). Additionally, user-friendly, and in some cases, free, data analysis and visualization packages are available (e.g., Python, MATLAB, Origin Pro, R Project, Orange Data Mining, Data Mosaic). Furthermore, statistical methods and programs allowed for the integration of all types of geodata (Grunsky, 2010; Hill et al., 2021). The dramatic increase in data volume and the desire to re-use and re-purpose data has led to the development of data governance and management frameworks to ensure the effective and efficient use of data (see USGS, 2020 for additional information). However, such frameworks are still in their infancy, with more work remaining to be done (Chamberlain et al., 2021; Sun et al., 2022; Ma, 2023).

3.2. Current practices

Multi-element geochemical surveys using stream, soil, lake, water, vegetation, rock and till samples continue to be the main staple of regional geochemical exploration (Cohen et al., 2010; Smith et al., 2013; Hosseini-Dinani et al., 2018; Talapatra, 2020). As an example of current practices, we present a regional geochemical survey performed at the Geological Survey of Canada that uses lake sediments. Much of the survey in terms of planning and design follows what had been developed from the 1960s–1990s. Specifically, regional-scale lake sediment geochemical surveys are typically planned at the scale of a NTS 1:250, 000 sheet (Friske and Hornbrook, 1991). Sampling is typically conducted at a density of 1 sample per 13 km^2 (or 1 per 5 mi^2), but can reach a density of 1 sample per 6 km^2 (Friske, 1991; Friske and Hornbrook, 1991). Because lake sediment samples cannot be collected in a systematic manner (Fig. 4a), a stratified random grid is used (Fig. 4b). Lakes measuring between 1 and 5 km in length and reaching depths >3 m are preferred as they contain more organic-rich sediments (gyttja; desired for elemental determinations). Lake sediment samples are collected from a helicopter on floats, which lands in the middle of the select lake. A total of 1.2 kg of lake sediment material (gyttja) is collected using a gravity torpedo sampler (Garrett et al., 2008), which is attached by a rope to a helicopter. The sample is then placed in a bag, labeled, dried (<40 °C for 10 days), sieved and milled (−80 mesh or 177 μm) before being placed in a labeled container for longer-term storage. In anticipation of improvements in analytical methods, large survey programmes collect excess mass per sample, which permits re-analysis of archived samples, thus cyclically modernizing some datasets (Smith et al., 2013). Re-analysis is more cost-effective than re-surveys, with the former at approximately 5% of the cost of the latter (McCurdy et al., 2014).

Select samples are analyzed by commercial laboratories via ICP-MS (65 elements) and occasionally INAA (35 elements) (McCurdy et al., 2016). Data is then analyzed for contamination, accuracy, precision and fitness-for-purpose (QA/QC; McCurdy and Garrett, 2016) and released to the public. In essence, the entire survey activity, from inception to the publishing of data/findings can take between years to decades (de Caritat, 2022). Unanticipated delays, including due to social, political and environmental factors, are increasingly commonplace, wreaking havoc on survey timelines. Such large-scale regional surveys are also resource intensive (equipment, personnel, infrastructure), running at a cost of ~430,000 USD (in 2022) for a lake sediment sampling survey on the scale of 1:250,000, over an area of ~14,000 km^2 (which is about the size of an NTS zone in Canada), with a sampling density of 1 sample per 13 km^2.

The problem of identifying deposits using geochemical data, similar to that using geophysical data, is an inverse problem, which means that solutions are not unique. This is because although mechanisms of material transport and mixing are necessary for chemical dispersion, which creates detectable geochemical signatures (Fig. 2), they increase entropy, which is irreversible. Consequently, there is no "best" method to process and analyze geochemical data (Winterburn et al., 2020; Zuo, 2020; Zuo et al., 2021; Yousefi et al., 2021). To date, the question of what constitutes a geochemical anomaly and how to best detect it remains an active research topic (Fig. 2; Cohen et al., 2010; Winterburn et al., 2020). Known methods include: data visualization (e.g., GIS), statistical techniques, geostatistics, fractal and multifractal models, and more recently data science (AI-based) methods. Below we provide a summary, for details see the comprehensive works of Carranza (2008), Zuo et al. (2016, 2019, 2021), Yousefi et al. (2021), Zhang et al. (2022c) and He et al. (2022).

Visualization methods delegate analysis to the human visual cognition system and include, for example, mapping of elemental concentrations. Visualization is limited by the two-dimensional nature of human vision and therefore, works better for low-dimensional data (e.g., Friske and Hornbrook, 1991; Zuo et al., 2016). Univariate statistical methods (including classical statistics and exploratory data analysis; Tukey, 1977) focus on leveraging distribution properties to identify outliers and patterns, and include analyzing the mean, standard deviation, quantiles, percentiles, cumulative frequency distributions, conducting regression analysis, Q-Q plots, extreme value statistics, among others (e.g., Chung and Agterberg, 1980; Reimann and Garrett, 2005; Reimann, 2005; Carranza, 2008; Zuo et al., 2021 and references therein). Multivariate statistical methods model interdependence between variables to analyze trends, such as the dispersal of minerals. Classic multivariate methods include principal component analysis (Grunsky, 2010; Grunsky and de Caritat, 2017, 2020), factor analysis (Fabrigar and Wegener, 2011), cluster analysis (Wang et al., 2017a), and others (Porwal et al., 2006; Parsa et al., 2017; Carranza, 2008; Zuo et al., 2021 and references therein). Geostatistical methods include stationarity tests, geodomaining, variography analysis, kriging and stochastic simulation, to integrate data, estimate values at unsampled locations and detect anomalies (Fig. 3; Matheron, 1963; Isaaks and Srivastava, 1989; Liu et al., 2019). In particular, geostatistical interpolation is unavoidable for rigorous gridding of sparse data, which is necessary for spatial indexing and integration of datasets for combined analyses, such as in mineral prospectivity mapping (MPM). Fractal and multi-fractal models leverage the fractal spatial distribution of deposits to detect them, using methods such as concentration-area (Cheng et al., 1994), concentration-distance (Li et al., 2003) and concentration-volume methods (Afzal et al., 2011). AI methods used for geochemical data analysis fall into two broad categories, supervised and unsupervised learning. Supervised learning for deposit predictions makes use of algorithms to model relationships between known mineral deposits and their covariate data (Rodriguez-Galiano et al., 2015; He et al., 2022). Supervised learning can also be used indirectly for synthetic data generation, which can be used for exploration (e.g., Zhang et al., 2022b; Bourdeau et al., 2023). Unsupervised methods formulate geochemical anomaly detection as AI-based anomaly detection tasks, whose algorithms used to date include shallow learning algorithms (Zhang et al., 2022a), autoencoders (Zhang et al., 2024a; Zuo et al., 2019, 2022) and convolutional neural networks (Zhang et al., 2021b).

3.3. Outstanding issues

3.3.1. Changes in philosophy, especially with regards to data generation and data usage

Geodata will continue to accumulate, which drives the expansion of data-driven geoscience (Figs. 1 and 2; e.g., Dramsch, 2020; Yousefi et al., 2021; He et al., 2022; Sun et al., 2022; Ma, 2023). This is motivated by the desire to: (1) extract additional value from data through data re-purposing; (2) expedite mineral deposit discovery; and (3) attract funding and evolve geoscientific practices. Data-driven geoscience consists of theory and applications of transdisciplinary methods, which include: (1) statistical methods such as regression analysis; (2) statistical learning, data mining and AI methods; and (3) generation of synthetic datasets. This implies that there will be increasing adoption and adaptation of data science into geosciences, to capitalize on more powerful and rigorous methods (Hastie et al., 2017; Grossi et al., 2021; Hazzan and Mike, 2023; Ma, 2023). This trend is being formalized through increasingly formal geodata science training, as witnessed by the arrival of the first pedagogical materials (Wang et al., 2023; Stanford University, 2024; Berkeley, 2024).

The pace of AI development and application results from the interaction of a diversity of disciplines. Therefore, no single discipline can compete in terms of method development in data usage. Consequently, AI methods can handle any combination of data, from monodisciplinary data, such as in geochemical anomaly detection (e.g., Parsa et al., 2017; Grunsky and de Caritat, 2017, 2020; Zhang et al., 2022a), to multidisciplinary data in MPM (e.g., Yousefi et al., 2021; Lawley et al., 2021, 2022). In effect, AI displaces and subsumes a large portion of traditional inferential data modelling. For example, statistical data analysis techniques such as principal component analysis and regression analysis, which are staples of geochemical data analysis have become subsumed under linear dimensionality reduction and regression modelling in AI. This is especially true of deep learning algorithms in geodata analysis, because deep learning methods are the most general and sometimes the most powerful (Zuo et al., 2019; Dramsch, 2020; Yousefi et al., 2021; He et al., 2022; Sun et al., 2022; Ma, 2023), to the extent that visual or unstructured geodata can be modelled using computer vision or language processing methods (e.g., Yang et al., 2022; Lawley et al., 2023). In fact, a key distinction between deep and shallow learning is that while the latter requires geoscientists to engineer effective features (e.g., elemental ratios and indicators), the former eschews discipline-specific expertise through automation of feature extraction and is far more disruptive.

Disruption occurs for in-discipline data users because: (1) data science-specific training is broader than typical geoscientific training for data usage (e.g., mathematics, statistics, programming, automation and computation); and (2) a shift in required aptitude from physical/observational and discipline-specific to abstract and multidisciplinary. However, the disruption extends beyond the data usage stage of the data pipeline into data generation. New algorithms and frameworks (e.g., deep learning and data science) demand new specific characteristics of data, which can be very different than characteristics of traditional geochemical data. For example, the number of degrees of freedom in non-parametric, nonlinear models can be exceptionally high compared with those encountered in principal component or parametric regression analysis. This necessitates a higher volume of data to support modern modelling methods. Insufficient geochemical data abundance inadvertently facilitates the displacement of geochemical exploration in favour of integrated uses of more voluminous types of geodata, such as using MPM for mineral reconnaissance (Lawley et al., 2021, 2022).

These disruptions can be understood by contextualizing the entire history of geochemical surveys into a modern data perspective, which is that (all) data is engineered to be fit-for-purpose (Figs. 3 and 4). This is particularly true for small data, because such data is targeted in intent and costly. The purpose of traditional geochemical data was functionally defined as data that is suitable for traditional data analysis methods in exploration (e.g., spatial analysis, statistical methods) (Fig. 5a). Guided by this purpose, modern advancements have focused on the quality of each data point (increasingly higher dimensional data that is more precise and accurate; Caughlin, 2010; Cohen et al., 2010; Johnson et al., 2013), which aligns with scientific reductionism. This philosophy resulted in a vast increase in the resolution of geochemical data in the variable domain (e.g., better analytical precision and elemental coverage), but generally not in the spatial domain (e.g., sampling resolution has not generally increased). For surveys at large scales, the probability of making a truly greenfield discovery is foremost controlled by the spatial resolution (Section 2.2; Fig. 3). This is because deposits are generally smaller than typical greenfield sample spacing, and therefore, their discovery depends on whether their spatial proxies have been adequately sampled.

It is important to appreciate that data science methods such as modelling using AI are mostly system-level techniques that do not adhere to scientific reduction. Consequently, increasing use of system-oriented methods constitutes a definitive change in the functional and philosophical purpose of survey data. One implication that is obvious is that whereas clean, precise data is important for scientific reduction, it is unnecessary for big data techniques. In fact, making inference on noisy data is a strength of AI methods (e.g., to identify a face in noisy images; Ding et al., 2017), and model performance can sometimes be improved using noisier data (e.g., Reed and Marks, 1999; Bishop, 1995a, b; An, 1996). Thus, the recent increase in variable domain resolution is not as beneficial to modern data users. In particular, AI models are inferential, and therefore, whether data captures an element that is causally indicative of a deposit, is functionally unimportant. Instead, there must be some high-dimensional covariates of at least one targetable characteristic of a deposit being sought. This could vary from a simple linear combination of elements (e.g., Grunsky, 2010; Grunsky and de Caritat, 2020), or a complex and uninterpretable transformation of elements (e.g., Zuo et al., 2019; Zuo and Xu, 2023). In particular, because AI-based methods are specifically capable of ignoring noise in data, traditional notions of static and stationary geochemical baselines no longer carry the same significance (e.g., Luo et al., 2020; Zhang et al., 2021a, 2024a; Feng et al., 2022).

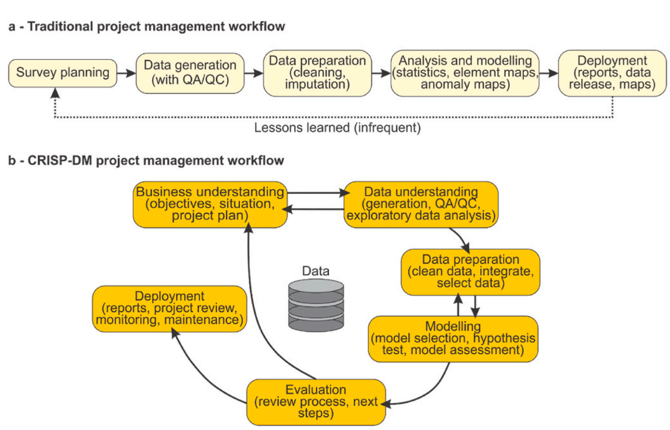

Fig. 5. Project management workflows for exploration geochemistry. Included with the major phases are some of the tasks between parentheses. (a) Traditional project management workflow currently practiced. (b) Data science workflow for future integrated survey practices. Here, the Cross Industry Standard Process for Data Mining (CRISP-DM) workflow model is presented as an example.

Application of AI to traditional geochemical data is considered data re-purposing, because such data was never specified to suit AI algorithms. Data re-purposing is very risky, because it is unobvious whether any discoveries could be made until after at least one complete execution of a data science framework (Woodall, 2017; e.g., the CRISP-DM model; Wirth and Hipp, 2000; Chapman et al., 2000, Fig. 5b). This makes it tenuous to sustain the generation of traditional geochemical data to supply both in-discipline and transdisciplinary data users. The situation is compounded by the fact that exploration is becoming more challenging, because easily discoverable, near-surface and geologically-simple mineral deposits are essentially exhausted (Blain, 2000; Ghorbani et al., 2023a). Therefore, geochemical signatures emanating from remaining deposits can be more complex, fainter and smaller (Fig. 3a), which also overlap with brownfield signatures. Thus, to lower exploration risk and increase the probability of discovery, geochemical exploration will need to increasingly adopt data modelling methods that are more noise tolerant, scalable to bigger data and able to extract high-dimensional relationships. This essentially assures the continued rise of geodata science as applied to exploration geochemistry, which secures a new purpose for geochemical data – to be used specifically with system-oriented (e.g., AI) methods. A solution is to engineer at least one more type of geochemical data that is fit-for-purpose for system-oriented data users (Woodall, 2017; Ma, 2023). This issue has been recognized more generally, to support modern data analytics (Woodall and Wainman, 2015).

3.3.2. *Evolution of roles and responsibilities in exploration*

Traditionally, geochemical data analysis is commonly carried out by geoscientists who also generate data (Fig. 5a), meaning that the traditional data pipeline is implicit. Poor delineation of roles and responsibilities is becoming a limiting factor in exploration productivity. This is because data generation and usage have become independent, because they are both complexifying, and geochemical data is increasingly used out of discipline, e.g., in MPM (Yousefi et al., 2021). It is impractical to increase responsibility for delivery of the entire data pipeline implicitly, because this practice requires that relatively small groups of people must specialize in data generation, management and data science to be competitive and sustainable. In geosciences and related fields (e.g., mining engineering, metallurgy), the canonical approach is to specialize into focused sub-disciplines (Keane, 2001). This results a reduced overlap in terms of expertise and aptitude from individuals in different sub-disciplines (e.g., geochemistry and geophysics). In the modern data pipeline, knowledge of data generation and usage are becoming foreign to data users and generators, respectively (e.g., Chamberlain et al., 2021; Sun et al., 2022). Hence, there is an increasing gap between data generators and users due to a divergence in aptitudes, interests and skills (Fig. 5b). Proposed solutions are unsatisfactory because they still revolve around cross-training data users and generators (e.g., He et al., 2022; Zhang et al., 2022c), which is antithetical to natural functional specialization and therefore, counter productive. An explicit pedagogical recognition must be made that the interest of the data generator likely lies in the interaction with nature, observations and metrology. This is in stark contrast to the interest of data users, which center around abstraction, algorithms, math, statistics and computation. Cross-training divergent interests is also impractical, because the talent pool that would exhibit both sets of interests is diminishingly small. In other disciplines, there is a clear role-distinction for example, between data generators and users (e.g., experimentalists and theorists). Hence, there is now a need for expertise to bridge data generators and data users, which does not currently exist in geochemistry (Sun et al., 2022).

3.3.3. Geometallurgical and other downstream requirements

Geochemical exploration is integrated with downstream considerations increasingly through geometallurgy (integration of geosciences, mining and metallurgy) and closely related social, environmental and governance aspects (e.g., ESG; Lawley et al., 2024). Geometallurgy is a response to the time-consuming and resource-intensive nature of the segregated traditional mineral value chain, from deposit discovery to mineral processing (Fig. 7a). Geometallurgy specifically tries to integrate data and knowledge across the chain, and therefore, increase its agility (Fig. 7b; Van der Boogaart and Tolosana-Delgado, 2018; Lishchuk et al., 2020; Michaux and O'Connor, 2020; Frenzel et al., 2023). A fundamental goal of geometallurgy is to gather exploration data and information early in reconnaissance and exploration stages, such that engineering constraints downstream of exploration can be known as early as possible. In simple terms, geometallurgy addresses the issue of: that which is explorable is not necessarily extractable, that which is extractable is not necessarily processable, and that which is processable is not necessarily economical. Insufficient or untimely information passed down the mineral value chain not only hinders resource utilization, but can decrease economic performance, increase environmental/social impact and evaporate investor confidence. Geometallurgy is expected to produce greater system-level insights and potential for control, relative to gated or siloed stages (Michaux and O'Connor, 2020; Ghorbani et al., 2022). Ideally, data from each stage of the mineral value chain would become interoperable, big and online (live generation, processing and usage) (Ghorbani et al., 2022, 2023b).

The integration of data can also be viewed along financial, operational, scientific and engineering dimensions. The integration of financial, scientific and engineering dimensions can be easily understood from the perspective of by-products and co-products. Traditionally, geochemical exploration was dominantly focused on a single commodity (e.g., uranium, gold, copper, zinc). The extraction of mono-mineralic or mono-elemental ores are essentially obsolete today. Instead, the current focus is on designing maximally economical extraction to beneficiation processes. This means that by-products or co-products of an ore deposit are now key considerations for the feasibility to extract and process (e.g., Whitworth et al., 2022). For example, many currently designated critical raw materials (for national self-sufficiency or green energy transition), such as the rare earth elements were not considered during historical exploration, and hence, reside in either mine or metallurgical waste. The implications of geometallurgy on surveys is profound, as information regarding observed mineral associations, textures, mineralogy and chemistry can be used to constrain downstream feasibility, even during reconnaissance. For example, the presence of pyrite with other economically valuable minerals such as sphalerite, chalcopyrite, galena and gold, will determine the economic feasibility of mineral separation (Chandra and Gerson, 2010). An integrated survey combining geochemistry and mineralogy is possible (e.g., of the latter using automated mineralogy or element to mineral conversion; Gu et al., 2014; Parian et al., 2015) and the most feasible using hard rock geochemistry. Aside from impacting the type of data generated, geometallurgical constraints added to surveys will alter the cost-to-benefit ratio of various media types, and therefore, affect data generation in general.

3.3.4. Education and careers

Geochemical exploration is increasingly faced with a diminished quantity of graduates with geoscientific degrees at the tertiary level (Cohen et al., 2010). Geoscientific education is also increasingly recognized as insufficient in its present form in university curriculums (Mikeš, 2015 and references therein). In particular, deficiencies are spread across key themes, including: (1) complex systems and geological time; (2) quantitative skills such as mathematics, physics, statistics, logic and programming; (3) system thinking; (4) spatial reasoning; and (5) data-driven science (Kastens et al., 2009; Mikeš, 2015; Medunić et al., 2022; Sun et al., 2022; Zhang et al., 2022c). Geochemical surveys encompass all of the above themes directly or indirectly. The lack of quantitative and/or data-centric education (and much less data-driven) (Mikeš, 2015; Chamberlain et al., 2021; Sun et al., 2022; Zhang et al., 2022c) is a thorny issue because of: (1) the increasing complexity and volume of geodata (including synthetic/simulated datasets); (2) emergence of integrated surveys (Coker, 2010; Johnson et al., 2013; Winterburn et al., 2020; Ma, 2023); and (3) transdisciplinary uses of geodata (e.g., Lawley et al., 2021, 2022; Zhang et al., 2021b). This problem is worsened by the rapid retirement of existing scientists (Coker, 2010; Wilson, 2014; Winterburn et al., 2020). There is therefore an urgent need to rejuvenate university curriculums such that new talent can both replace traditional roles and face modern challenges.

Major issues also exist for career data users, or geodata scientists. Notably, although exploration geochemical data can be voluminous, it is low velocity (accumulated slowly). This results in a gradual accumulation and a saturation, which is marked by an onset of significant data analysis (outside of QA/QC). It is unclear how careers of geodata scientists working with primarily geochemical data would develop, if data analysis is an occasional event. Thereafter, data becomes dormant until substantial improvements in either data (e.g., a re-analysis or re-survey) or analysis techniques have occurred. At the moment, a growing category of geodata scientists are transdisciplinary (Ma, 2023). Geosciences is following other domains (e.g., Karpatne et al., 2018; Hazzan and Mike, 2023) with a growing number of traditional geoscientists reskilling to become geodata scientists (e.g., for MPM). Unfortunately, traditional sampling techniques cannot scale up in data velocity (Sun et al., 2022). This truly hampers the deployment of AI and data science techniques because: (1) meaningful models require continuous data input (e.g., to compensate for covariate shift, Diethe et al., 2019; Hoffimann et al., 2021); and (2) model deployment and post-deployment validation can only be met through a sustainable stream of data (e.g., Hawkins et al., 2023; Zhang et al., 2022c). This partly explains why in geosciences, almost all publications involving AI and data science are proof-of-concept studies, for which model deployment is restricted to a singular occurrence within the study (see He et al., 2022 and references therein) and there continues to be a paucity of validation studies and a positive outcome bias. Whereas in other domains (e.g., finance, commerce, entertainment, medicine and marketing), models built by data scientists are passed to engineers for deployment, which permits definitive validation (Hawkins et al., 2023). These are career problems for geodata scientists, as data is easily exhaustible and professional output (e.g., models or maps) remains unvalidated for prolonged durations, which makes it challenging to assess career achievement.

3.3.5. Material needs for the modern society and the future

The demand for raw materials evolves with time. For example, rare earth elements were an uninteresting commodity historically (Castor and Hendrick, 2006), but are highly desirable today due to their role in high-tech, military and energy technologies (Balaram, 2019). There is presently a strong drive at the national level to: (1) pursue resource self-sufficiency; (2) reduce environmental/social impacts; and (3) maximize or counter geopolitical gains and leverage (Zhang et al., 2023a; Ghorbani et al., 2024). This has resulted in national to super-national policy frameworks that designate various minerals and metals as "critical" (e.g., Wang et al., 2017b; Burton, 2022; Government of Canada, 2022; European Commission, 2023). Energy is a modern issue for many countries, because disruptions in the global supply chain and geopolitics have greatly sensitized the need for energy abundance and stability. Part of the solution of sustainable energy generation is an enormous and unprecedented (up to many orders of magnitudes higher rate of extraction compared to pre-COVID levels) increase in the extraction of minerals and metals (e.g., cobalt, lithium, rare earth elements) (Kleijn et al., 2011; Watari et al., 2019; Michaux, 2021).

The insatiable material desire is met with the immovable timescale of the mineral value chain. Global estimates on the average time to the onset of mineral extraction from initial reconnaissance is about 17 years (IEA, 2023). Whether the urgency for sustainable energy and environmental/social impact reduction can be met using traditional methods alone is questionable. However, it is certain that for many countries that have pledged to reach the Agenda 2030 or the Net-Zero 2050 goals (UN, 2015), the amount of time remaining is exceedingly short. This also assumes that new deposits would be discovered at rates comparable to the past, which is unlikely (Blain, 2000). Nevertheless, there is hope that data-driven reconnaissance and exploration using geodata science-based methods will expedite exploration (e.g., MPM, Lawley et al., 2021; 2022). As data generation is the rate-limiting process in geochemical exploration, it would be efficacious to examine how to best alleviate or mitigate this bottleneck, to suit foreseeable downstream requirements (Fig. 8).

4. Future (2020s – beyond)

4.1. Stream 1 - relevance of traditional practices

Traditional geochemical practices will continue in the future. Notably, there will always be a need for highly accurate and precise data, such that conjectures, hypotheses, models and theory can be tested. This requires modern laboratories (XRF, ICP-MS, SEM, etc.) and their continued evolution. Scientifically reductive data is necessary for: (1) studies of the timelines and processes of ore deposit formation (e.g., elemental diffusion and transport); (2) creation of exploration or metallogenetic models (for deep or in unconventional settings); (3) integration of multi-resolution data (i.e., km to µm); (4) development and incorporation of new sample mediums; and (5) refinement of analytical methodologies. Lastly, it is possible that high quality data will support data-driven practices by validating findings (e.g., verification of a geochemical anomaly or a prospectivity map).

4.2. Stream 2 - anticipated changes in data requirements

Key drivers of the evolution of geosciences as a whole are directly affecting geochemical data generation, which are: (1) a rise in data-driven philosophy of science, as a counterpart of hypothesis-driven science; (2) evolving roles and responsibilities (e.g., differentiation between geodata generators and geodata scientists); (3) geometallurgical integration of the mineral value chain from exploration to beneficiation; (4) a need to modernize geoscientific education, to incorporate more abstract training and to maximize the value of future geoscientific skillsets; and (5) a need to expedite exploration (Fig. 8 and Table 1). It is time to re-think key aspects of how geochemical data is generated, not to replace traditional data generation, because hypothesis-driven science will always remain, but to fit to the purpose of data-driven exploration and its antecedent activities in geodata science. Notably, data requirements for the purpose of data-driven exploration, relative to existing geochemical data include: (1) higher throughput (velocity); (2) higher spatial resolution (density); (3) more integrated with other measurements (dimensionality); (4) better managed and accessible to non-geochemists; and (5) more agile and less grid-based. Geochemical data already engineered to satisfy all the aforementioned requirements do not exist today, although it is an emerging research domain (e.g., Kyle and Ketcham, 2015; Lemière and Uvarova, 2020; Wang et al., 2021; Zhang et al., 2023b, c; Fontana et al., 2023, MineSense, 2024). Below, we discuss solutions with a summary of key drivers, issues and proposed solutions presented in Table 1.

Table 1
Summary of outstanding issues and proposed solutions.

Aspects	Issues	Proposed solutions
Scientific reduction philosophy	- Maladapted for geodata science - Costly data, with cost dedicated to benefits that are primarily extrinsic to geodata science	- Adoption of a systemic scientific philosophy - Culture change with regards to data management - Adoption of 'Universal Reference Materials' (URM) to increase data interoperability - Foster multidisciplinary understanding
Evolving Data Culture, changing roles and responsibilities	- Implicit data pipeline - Poor delineation of roles - Lack of consideration for divergent interests	- Creation of new roles: data managers and data engineers - Train data generators and users separately - Use of project management models that are compatible with data-centric disciplines (e.g., CRISP-DM) - foster multidisciplinary understanding
Geometallurgy and downstream processes	- Data are siloed with respect to the mineral value chain - Lack of consideration for downstream processes	- Adoption of portable instruments to increase data velocity and density (applicable as early as mineral exploration) - Adoption of out-of-discipline instruments to capture additional data dimensions - Better data integration across disciplines
Education and careers	- Educational deficiencies - Lack of data-centric education and learning opportunities - Lack of career opportunities for data-centric experts	- New curriculum that increases transdisciplinary training and mutual understanding - Adoption of portable instruments to increase data velocity and density - Adoption of out-of-discipline instruments to capture additional data dimensions - Culture change with regards to data management - Train talents according to their aptitude and modernized roles
Material needs for the future	- Unprecedented need for raw materials - Traditional Data generation is not time-effective	- Adoption of portable instruments to increase data velocity and density - Adoption of out-of-discipline instruments to capture additional data dimensions - Better data integration across disciplines for rapid discoveries and new insights

4.2.1. Adoption and integration of new tools and sensor technologies

Portable instruments (e.g., pXRF, pXRD, LIBS) seem inevitable because technologically they are along an unstoppable trend towards instrument (and sensor) miniaturization. Their desirability is multi-fold as they: (1) are fast at data generation (a few seconds to minutes for a pXRF; Knight et al., 2021); (2) can be used in-situ with non-destructive sampling methods; (3) require less total sample preparation; and (4) are low in cost (Lemière and Uvarova, 2020). A common concern guided by scientific reduction is that data generated using portable instruments is of lesser quality (less accurate and precise) as compared to traditional laboratory instruments (e.g., XRF, ICP-MS, INAA; Lemière and Uvarova, 2020). However, this trade-off is desirable as data-driven methods are mostly designed for big data analysis and can therefore handle noisy data, which can even improve model performance (e.g., Ding et al., 2017). Portable instruments are beginning to be adopted into the mineral value chain, such as for geochemical analysis of drill hole samples (e.g., Fontana et al., 2023) and bulk material characterization in mining and deposit-scale profiling (e.g., MineSense, 2024). Its use in reconnaissance and exploration constitutes a type of data-driven geometallurgical integration (Table 1). Portable instruments have the advantage that they are maximally compatible with traditional data usage methods, without displacing geochemical exploration using other methods (e.g., through remote sensing). Consequently, protocols (including QA/QC) and workflows applied to traditional geochemical data can be easily adapted to portable or miniaturized instruments.

An emerging trend in exploration geochemistry is secondary data or predictive assays, through the inversion of hyperspectral remote sensing-data, which can provide rapid, systematic, dense and non-destructive data generation both on a large- (e.g., Kurz et al., 2013; Booysen et al., 2019) and per-sample scale (e.g., Schodlok et al., 2016; Simán, 2020; Semereab, 2021; Linton et al., 2024). Drone-based remote sensing can perform 3D mapping at resolutions down to the centimeter-scale (Ghosh et al., 2017; Ullo et al., 2020; Baek et al., 2022) and are employable for ore delineation, mine planning and hazard monitoring. Technically, remote sensing derives a spectral response from surficial materials, in a manner that is similar to fluorescence-based chemical analysis in geochemistry (Kurz et al., 2013). However, because in remote sensing, the photon energies involved are usually insufficient to cause ionization, the information contained in remote sensing data is not directly reducible to chemical concentrations, but rather mineral composition (Marghany, 2022). This means that chemical information can be inverted from remote sensing data, which creates secondary geochemical data that inherits the velocity and volume of remote sensing data (e.g., Wang et al., 2021; Zhang et al., 2023b, c).

Online multi-sensor-based technologies for drill core characterization are being developed to integrate chemical, mineralogical and structural data (e.g., Schodlok et al., 2016; Simán, 2020; Semereab, 2021; Linton et al., 2024). A 3D analytical technique using X-ray Computed Tomography (XCT) is also incorporated to allow machine-vision based approaches for material characterization (e.g., Kyle and Ketcham, 2015; Guntoro et al., 2019; Delgado Yanez, 2020). However, XCT methods would still require sampling, because such instruments are not yet portable.

Other known avenues for the rapid generation of geochemical data include inferential data generation and augmentation (e.g., synthetic data via generative methods; Liu et al., 2019; Zhang et al., 2022b, 2023c; Bourdeau et al., 2023; Ferreira and Koeshidayatullah, 2024). There are physical and abstract limitations with this class of approaches. Physically, despite the accumulated geochemical data, the spatial and depth coverage of geochemical data is still a trifle fraction of the Earth, which means that generative methods trained on geochemical data cannot create a diversity of geochemical data that approaches the lithodiversity of the Earth. Abstractly, the use of inferentially generated data to support geodata science leads to model collapse through the curse of recursion (Shumailov et al., 2023). This is due to compounding errors incurred through recursion and include: functional approximation errors, sampling errors and learning errors (e.g., Liu et al., 2019; Zhang et al., 2022b, 2023c; Bourdeau et al., 2023; Ferreira and Koeshidayatullah, 2024). It affects models roughly commensurate with their complexity (e.g., complex models collapse faster). Detection and remediation for model collapse in geodata is an unresearched topic presently. It is not clear if novel mitigation strategies in other domains, such as using a human- or AI-based discriminators to reject synthetic data, or trade model degradation for more data through data blending (Gerstgrasser et al., 2024) can be applied to exploration geochemistry.

4.2.2. *Evolving data culture, new roles and responsibilities*

Bigger geochemical data and its use by non-geochemists (e.g., geodata scientists) will necessitate formal data management (e.g., DAMA DMBOK; Henderson et al., 2017, Table 1). Others have lamented that geochemical data is generally not FAIR (findable, accessible, interoperable and re-useable) due to the lack of international standards and best-practices (Chamberlain et al., 2021; Sun et al., 2022; Ma, 2023). This lack of a data sharing culture is hindering data quality assessments, interoperation and re-use of data, and outcomes (Ma, 2023). Formalized data management and curation entails comprehensive, well-structured and machine-readable metadata, the use of international geo-sample numbers (IGSN), standardized data processing workflows and data publications (Chamberlain et al., 2021; Sun et al., 2022; Ma, 2023). There is a need to designate and adopt what we term ‘universal reference materials’ or URMs, which permit integrated data usage of geochemical data from different generators. At present, no study/-committee has yet designed URMs for any sample media. Ultimately, the long-term survival of geochemical data will depend on how useable they will be in the future.

Managing and using diverse geodata encompassing geochemical data would require dedicated roles with new responsibilities (Ghorbani et al., 2022; 2023b, Table 1). An increasing separation of responsibilities driven by increasing task complexity in each role means that the future role of a geodata scientist is strictly data analysis, and of a surveyor, data generation. It is unreasonable to expect future users of geochemical data will revert to solely geochemists. Functional specialization along the data pipeline (Fig. 6) would create new roles to bridge the gap between data users and generators: which are: (1) data managers; and (2) data engineers (Sun et al., 2022, Fig. 6c). Consequently, data managers must exist to ensure that data quality and standards are met for use for geodata scientists, by synthesizing data requirements and transmitting them to data engineers (e.g., DAMA-DMBOK; Henderson et al., 2017, Fig. 6c). Data engineers bridge the gap between traditional geochemists and transdisciplinary talents, by performing geoscience-specific data processing and steering data generation. For example, data engineers would perform QA/QC, data annotation (e.g., deposit presence), interpolation, compositing, levelling and imputation. As an example, Guntoro et al. (2019, 2020) demonstrated that engineering the appropriate AI classification methods for specific ore types improves 3D data processing. The roles between data generators, managers, engineers and users are detailed in Fig. 9.

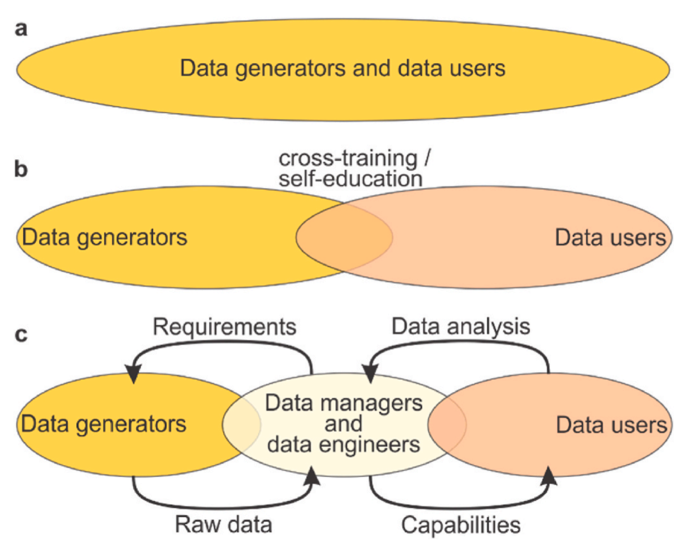

Fig. 6. Schematic representations of the specialization of geodata generators and users. (a) Historically, geodata generators were also data users. (b) Currently, a divergence exists between geodata generators and modern data users (e.g., AI experts). Numerous advocates (e.g., He et al., 2022; Zhang et al., 2022c) have proposed cross-training to bridge the gap. (c) Proposed future integration can occur through a formal implementation of the data pipeline through functional specialization instead of cross-training.

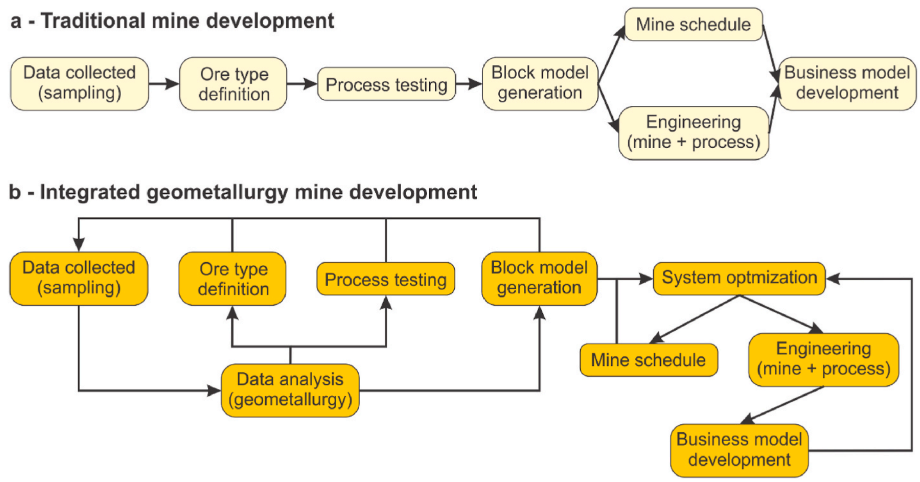

Fig. 7. Traditional (a) and integrated geometallurgical (b) mine development flow chart. Figure modified from Michaux and O'Connor (2020).

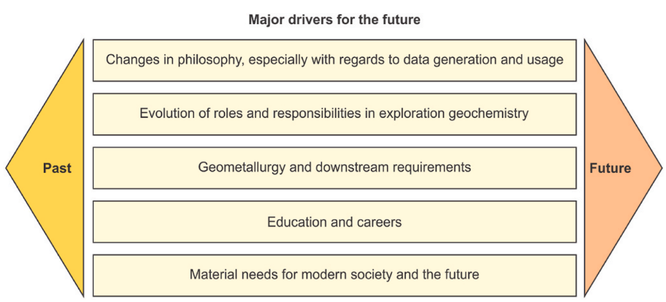

Fig. 8. Major drivers of future geochemical data generation. These drivers emerged as a result of legacy and current practices and will undoubtably change how geochemical data is generated so that the field will remain relevant in the future.

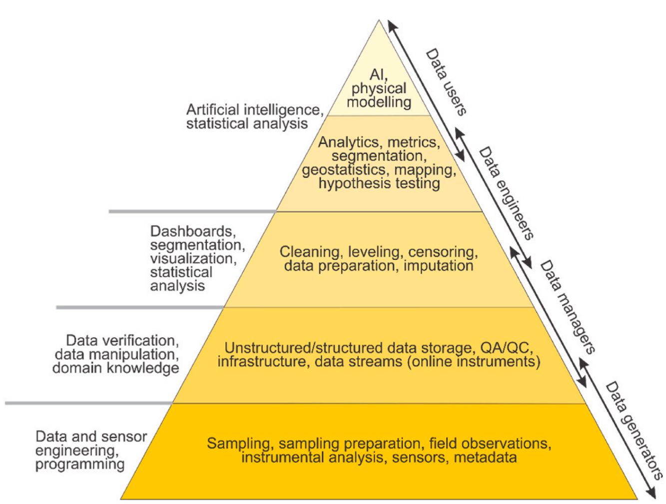

Fig. 9. Future roles between data generators, managers, engineers and users to accomplish data-driven geochemical exploration. Diagram modified from Rogati (2017) and adapted for data generation for exploration geochemistry, and broadly geodata generation.

4.2.3. Re-thinking education

Geodata science is nascent (Fig. 1). Very few universities offer programs at any level (e.g., Stanford University, 2024; Berkeley, 2024), which is unfortunate for exploration geochemistry, because the overlap between our vision of the future of exploration geochemistry and geodata science is substantial (e.g., statistics and programming). The dichotomy of the desirability of geodata science (Ma, 2023) and its immaturity in the academies results from transdisciplinary methods evolving rapidly because of exponential technologies (e.g., computing hardware; Roser et al., 2023). Exponential technologies rapidly ascend to a state of maturity, followed by a long period of quiescence as they reach a physical or practical limitation, similar a biological concept known as punctuated equilibrium. Therefore, in the growth phase, geodata science is highly desirable and disruptive, but its impacts are not exactly foreseeable, since breakthroughs occur regularly. It is challenging to design an applicative, future-proof curriculum in this environment; most universities subsume data science under computer science instead (Grossi et al., 2021; Ma, 2023). Currently, geodata science pedagogy is mostly taught ad hoc (e.g., conferences, workshops, online courses, or through discipline-bridging or educational software, such as Orange Data Mining, Data Mosaic or ArcMPM, see Zuo et al., 2024). However, informal education is insufficient to sustainably supply talents, especially those with a Whitebox understanding of data science (e.g., mathematics of algorithms).

A future curriculum for the exploration geoscientist should contain early exposure to data science (e.g., statistics, programming, data lifecycle, project management), which would foster cross-domain understanding and scientific communication (Table 1). Exposure can be a combination of theoretical and practical classes, with the latter substituting for laboratory and fieldwork for people on the geodata scientist track (Fig. 9). Observational geosciences should be mainly reserved for future geodata generators and those desiring field work. Future exploration geochemists would therefore, specialize in data usage or generation, but not both simultaneously. A new pedagogical direction for exploration geosciences should focus on broadening knowledge (along the data pipeline, not across geoscientific subdisciplines in counter to functional specialization) in preparation for multidisciplinary data usage and role-integration along the data pipeline, as opposed to solely becoming increasingly narrow (increasing reduction). This improves cross-discipline understanding using data as an abstraction and the data pipeline as an integrator, enhancing meaningful collaboration and knowledge transfer, and making geodata careers more appealing given the wider range of data accessible. This heralds the rise of 'integrative geoscientists', specifically unification using data and its treatise.

More generalist training presented by geodata science means that future exploration geodata scientists will work less like client-centric data analysts (Ma, 2023) and more towards leading exploration projects. This is because although traditional geoscientists appreciate data analysis advancements brought forth by data science, few are capable of training their own, which results in a recruitment pattern of general-purpose data scientists into ad-hoc projects. It is therefore a constant struggle to attract, sustain and retain geodata talent because of interest mismatch (data generators versus users), small data exhaustion and the dichotomous nature of knowledge-versus data-driven science. Early exposure between disciplines (e.g., geodata usage) would help attract talent and foster interest, but sustaining their careers would require scalable data generation methods (Table 1). Education of future geodata generators, such as geochemical surveyors, should explicitly incorporate modern and emerging data requirements and be specific to their roles and careers in the data lifecycle or pipeline (e.g., Figs. 6 and 9). Fundamental data management as well as data-centric project management concepts (e.g., CRISP-DM; Fig. 7b) are also important, similar to how education of data modelling methods has been standardized in data science (Henderson et al., 2017 and e.g., Ghorbani et al., 2022). This ensures that talents of the future can appreciate the full value of data in relation to data usage and the inefficiency of ad-hoc practices (Ma, 2023).

5. Summary and outlook

Our review highlights the potential for geochemical exploration to evolve and ensure that it stays relevant-for-purpose. The future of exploration geochemistry is dependent on the full recognition of the effects of modern drivers on a core component of geochemistry – data generation. Engineering a future type of geochemical data to suit modern data usage methods is important and provides a direction of research towards the development of a new type of geochemical data (and its accompanying methodology) that is far more scalable in terms of volume, velocity and dimensionality. This is because traditional processes in geochemical data generation are unscalable to suit modern data usage methods. In particular, there is a need to focus on designing data-driven surveys that enable the efficient collection of denser geochemical data, using predictive assays, secondary data and portable instruments, with an integration of other types of analyses in a geometallurgical sense to reduce costs and increase agility.

This review has demonstrated that data generation in exploration geochemistry is on the cusp of a significant shift. We acknowledge that there are broader challenges associated with implementing our vision of the future of exploration geochemistry, including the need for rigorous data management and the development of appropriate training programs. However, the potential benefits of our vision, including reduced exploration costs, improved survey agility, and enhanced sustainability, make the concept of "bigger" and cheaper geochemical data an area of research that should be pursued with enthusiasm. The review has also shown that exploration geochemistry should continue to evolve, because its data usage portion of the data pipeline has already experienced a substantial evolution. Its implications on data generation are profoundly philosophical but is only beginning to be appreciated. Only by embracing the change that is already occurring in geosciences, geometallurgy and modern resource requirements, will exploration geochemistry remain relevant and flourish in the future.

CRediT authorship contribution statement

Julie E. Bourdeau: Writing – review & editing, Writing – original draft, Visualization, Investigation, Conceptualization. **Steven E. Zhang:** Writing – review & editing, Writing – original draft, Validation, Conceptualization. **Glen T. Nwaila:** Writing – review & editing, Writing – original draft, Validation. **Yousef Ghorbani:** Writing – review & editing, Validation.

Declaration of competing interest

The authors declare that they have no known competing financial interests of personal relationships that could have appeared to influence the work reported in this paper.

Data availability

No data was used for the research described in the article.

Acknowledgements

The authors would like to thank Stephen J.A. Day (Geological Survey of Canada) for providing cost and time estimates to accomplish a regional-scale lake sediment survey. We thank Dr. Christopher J.M. Lawley (Geological Survey of Canada) for insightful discussions and constructive comments provided to this review. The authors are also thankful to MineSense who shared their latest technological advancements in ore characterization and extraction process routing. Two anonymous reviewers are thanked for their highly constructive comments which have helped shape this review study. Finally, Drs. Oliver Kreuzer and Zimeng Wang are thanked for editorial handling.

References

Afzal, P., Fadakar, A.Y., Khakzad, A., Moarefvand, P., Rashidnejad, O.N., 2011. Delineation of mineralization zones in porphyry Cu deposits by fractal concentration–volume modeling. J. Geochem. Explor. 18, 220–232. https://doi.org/10.1016/j.gexplo.2011.03.005.

An, G., 1996. The effects of adding noise during backpropagation training on a generalization performance. Neural Comput. 8 (3), 643–674. https://doi.org/10.1162/neco.1996.8.3.643.

Baek, J., Junhyeok, P., Seongjun, C., Changwon, L., 2022. 3D global localization in the underground mine environment using mobile LiDAR mapping and point cloud registration. J. Sens. 22 (8), 2873. https://doi.org/10.3390/s22082873.

Balaram, V., 2019. Rare earth elements: a review of applications, occurrence, exploration, analysis, recycling, and environmental impact. Geosci. Front. 10 (4), 1285–1303. https://doi.org/10.1016/j.gsf.2018.12.005.

Bergen, K.A., Johnson, P.A., de Hoop, M.V., Beroza, G.C., 2019. Machine learning for data-driven discoveries in solid Earth geoscience. Science 363 (6433), eaau0323. https://doi.org/10.1126/science.aau0323.

Berkeley, 2024. Data 8: Foundations of Data Science. University of, Berkeley. https://data.berkeley.edu/education/courses/data-8. (Accessed 1 January 2024).

Bishop, C.M., 1995a. Neural Networks for Pattern Recognition. Carendon Press.

Bishop, C.M., 1995b. Training with noise is equivalent to Tikhonov Regularization. Neural Comput. 7 (1), 108–116. https://doi.org/10.1162/neco.1995.7.1.108.

Blain, C., 2000. Fifty-year trends in mineral discovery – commodity and ore-type targets. Explor. Min. Geol. 9 (1), 1–11. https://doi.org/10.2113/0090001.

Boerngen, J.G., Shacklette, H.T., 1981. Chemical Analyses of Soils and Other Surficial Materials of the Conterminous United States. U.S. Geological Survey, Open-File Report, pp. 81–197.

Booysen, R., Gloaguen, R., Lorenz, S., Zimmermann, R., Andreani, L., Nex, P.A., 2019. The potential of multi-sensor remote sensing mineral exploration: examples from southern Africa. In: IGARSS 2019-2019 IEEE International Geoscience and Remote Sensing Symposium. IEEE, pp. 6027–6030.

Bourdeau, J.E., Zhang, S.E., Lawley, C.J., Parsa, M., Nwaila, G.T., Ghorbani, Y., 2023. Predictive geochemical exploration: inferential generation of modern geochemical data, anomaly detection and application to Northern Manitoba. Nat. Resour. Res. 32, 2355–2386. https://doi.org/10.1007/s11053-023-10273-6.

Burton, J., 2022. U.S. Geological Survey Releases 2022 List of Critical Minerals. U.S. Geological Survey. https://www.usgs.gov/news/national-news-release/us-geological-survey-releases-2022-list-critical-minerals. (Accessed 25 November 2022).

Butt, C.R.M., Lintern, M.J., Anand, R.R., 2000. Evolution of regolith and landscapes in deeply weathered terrain-implications for geochemical exploration. Ore Geol. Rev. 16, 167–183. https://doi.org/10.1016/S0169-1368(99)00029-3.

Cameron, E.M., Hamilton, S.M., Leybourne, M.I., Hall, G.E.M., McClenaghan, M.B., 2004. Finding deeply buried deposits using geochemistry. Geochem. Explor. Environ. Anal. 4, 7–32. https://doi.org/10.1144/1467-7873/03-019.

Carranza, E.J.M., 2008. Geochemical Anomaly and Mineral Prospectivity Mapping in GIS. Elsevier.

Castor, S.B., Hedrick, J.B., 2006. Rare earth elements. In: Industrial Minerals Volume, seventh ed., pp. 769–792 Society for Mining Metallurgy and Exploration, Colorado.

Caughlin, B., 2010. Developments in analytical technology. Geochem-Explor. Env. A. 10 https://doi.org/10.1144/1467-7873/09-214.

Chamberlain, K.J., Lehnert, K.A., McIntosh, I.M., Morgan, D.J., Wörner, G., 2021. Time to change the data culture in geochemistry. Nat. Rev. Earth Environ. 2 (11), 737–739. https://doi.org/10.1038/s43017-021-00237-w.

Chandra, A.P., Gerson, A.R., 2010. The mechanisms of pyrite oxidation and leaching: A fundamental perspective. Surf. Sci. Rep. 65 (9), 293–315. https://doi.org/10.1016/j.surfrep.2010.08.003.

Chapman, P., Clinton, J., Kerber, R., Khabaza, T., Reinartz, T.P., Shearer, C., Wirth, R., 2000. CRISP-DM 1.0: Step-by-step Data Mining Guide.

Chen, X.W., Lin, X., 2014. Big data deep learning: challenges and perspectives, 2. IEEE access, pp. 514–525. https://doi.org/10.1109/ACCESS.2014.2325029.

Cheng, Q., Agterberg, F.P., Ballantyne, S.B., 1994. The separation of geochemical anomalies from background by fractal methods. J. Geochem. Explor. 51, 109–130. https://doi.org/10.1016/0375-6742(94)90013-2.

Chung, C.F., Agterberg, F.P., 1980. Regression models for estimating mineral resources from geological map data. Math. Geol. 12, 473–488. https://doi.org/10.1007/BF01028881.

Cohen, D.R., Kelley, D.L., Anand, R., Coker, W.B., 2010. Major advances in exploration geochemistry, 1998-2007. Geochem. Explor. Environ. Anal. 10, 3–16. https://doi.org/10.1144/1467-7873/09-21.

Cohen, D.R., Shen, X.C., Dunlop, A.C., Rutherford, N.F., 1998. A comparison of selective extraction soil geochemistry and biogeochemistry in the Cobar Area, NSW. J. Geochem. Explor. 61 (1–3), 173–190. https://doi.org/10.1016/S0375-6742(97)00052-6.

Coker, W.B., 2010. Future research direction in exploration geochemistry. Geochem. Explor. Environ. Anal. 10, 75–80. https://doi.org/10.1144/1467-7873/09-218.

Coker, W.B., Nichol, I., 1975. The relation of lake sediment geochemistry to mineralization in the northwestern Ontario region of the Canadian Shield. Econ. Geol. 70 (1), 202–218. https://doi.org/10.2113/gsecongeo.70.1.202.

Connolly, E., Orsmond, D., 2011. The mining industry: from bust to boom. Reserve Bank of Australia Conference 2011.

Cuffari, B., 2016. Evolution of handheld XRF analysis in mining. AZO Mining. https://www.azomining.com/Article.aspx?ArticleID=1344. (Accessed 20 September 2022).

de Caritat, P.D., 2022. The national geochemical survey of Australia: review and impact. Geochem. Explor. Environ. Anal. 22 (4) https://doi.org/10.1144/geochem2022-032 geochem2022-032.

De Geoffroy, J.G., Wignall, T.K., 1985. Designing Optimal Strategies for Mineral Exploration. Plenum Press, New York.

de la Guardia, M., Garrigues, S., 2019. Analytical research based on the use of low-cost instrumentation. J. Pharmaceut. Sci. 25, 82–84. https://doi.org/10.15171/PS.2019.13.

Delgado Yanez, D., 2020. Geochemical and Mineralogical Comparison of XRT-XRF Scanning Technology versus Traditional Analysis of Drill Core from the Assarel Porphyry Copper-Gold Deposit (Panagyurishte District, Bulgaria). M.Sc. thesis, Uppsala University.

Demetrides, A., Smith, D.B., Wang, X., 2018. General concepts of geochemical mapping at global, regional, and local scales for mineral exploration and environmental purposes. Geochim. Bras. 32 (2), 136–179. https://doi.org/10.21715/GB2358-2812.2018322136.

Diethe, T., Borchert, T., Thereska, E., Balle, B., Lawrence, N., 2019. Continual learning in practice. ArXiv preprint arXiv:1903.05202. https://doi.org/10.48550/arXiv.1903.05202.

Dijkstra, S., Kubik, K., 1975. Autocorrelation studies in the analysis of stream sediment data. In: Elliott, I.L., Fletcher, W.K. (Eds.), Geochemical Exploration 1974. Elsevier, Amsterdam, pp. 141–161.

Ding, Y., Cheng, Y., Cheng, X., Li, B., You, X., Yuan, X., 2017. Noise-resistant network: a deep-learning method for face recognition under noise. J. Image Video Proc. 43 https://doi.org/10.1186/s13640-017-0188-z.

Dramsch, J.S., 2020. 70 years of machine learning in geoscience in review. Adv. Geophys. 61, 1–55. https://doi.org/10.1016/bs.agph.2020.08.002.

Dunn, C.E., 2007. Biogeochemistry in mineral exploration. In: Handbook of Exploration and Environmental Geochemistry, 9. Elsevier.

Elo, S., Gustavsson, N., Huhma, H., Kauranne, L.K., Koljonen, T., Naras, P., Pesonen, L.J., Ruotoistenmäki, T., Saltikoff, B., Sillanpä, M., Tanskanen, H., Vaasjoki, M., Vuorela, P., 1992. The Geochemical Atlas of Finland, Part 2: till. Geological Survey of Finland, Espoo.

European Commission, 2023. Establishing a Framework for Ensuring a Secure and Sustainable Supply of Critical Raw Materials and Amending Regulations (EU) 168/2013, (EU) 2018/858, 2018/1724 and (EU) 2019/1020. European Commission, COM(2023) 160 final.

Fabrigar, L.R., Wegener, D.T., 2011. Exploratory Factor Analysis. Oxford University Press.

Feng, B., Chen, L., Xu, Y., Zhang, Y., 2022. Comparative study on three autoencoder-based deep learning algorithms for geochemical anomaly identification. Earth Space Sci. 9 (11), e2022EA002626 https://doi.org/10.1029/2022EA002626.

Ferreira, I., Koeshidayatullah, A., 2024. Synthetic data in geosciences: challenges and opportunities. 85th EAGE Annual Conference & Exhibition. European Association of Geoscientists & Engineers 2024 (1), 1–5. https://doi.org/10.3997/2214-4609.2024101691.

Fisher, I., 1988. Blowpipe analysis. In: General Geology. Encyclopedia of Earth Science. Springer, Boston, MA. https://doi.org/10.1007/0-387-30844-X_11.

Flerov, B.M., 1935. The application of the stannometric survey for the exploration of primary tin deposits. Redkie Metally (1), 31–39 (in Russian).

Fontana, F.F., van der Hoek, B., Tassios, S., Tiddy, C., Stromberg, J., Francis, N., Uvarova, Y.A., Lancaster, D.G., 2023. Laser induced breakdown spectroscopy (LIBS) for whole rock geochemistry. J. Geochem. Explor. 246, 107160 https://doi.org/10.1016/j.gexplo.2023.107160.

Frenzel, M., Baumgartner, R., Tolosana-Delgado, R., Gutzmer, J., 2023. Geometallurgy: present and future. Elements 19 (6), 345–351. https://doi.org/10.2138/gselements.19.6.345.

Friske, P.W.B., 1991. The application of lake sediment geochemistry in mineral exploration. In: Franklin, J.M., Duke, J.M., Shilts, W.W., Coker, W.B., Friske, P.W.B., Maurice, Y.T., Ballantyne, S.B., Dunn, C.E., Hall, G.E.M., G, R., Garrett, R.G. (Eds.), Exploration Geochemistry Workshop. Geological Survey of Canada, pp. 157–180. https://doi.org/10.4095/132392. Open File 2390.

Friske, P.W.B., Hornbrook, E.H.W., 1991. Canada's national geochemical reconnaissance program. Trans. Inst. Min. Metall-Sec. B. 100, B47–B56.

Garrett, R.G., 1983. Sampling methodology. In: Howarth, R.J. (Ed.), Statistics and Data Analysis in Geochemical Prospecting, 2. Elsevier Science, pp. 83–100. https://doi.org/10.1016/B978-0-444-42038-1.50010-1.

Garrett, R.G., Goss, T.I., 1980. The appraisal of survey effectiveness in regional geochemical surveys of Canada's Uranium Reconnaissance Program. Math. Geol. 12, 443–458.

Garrett, R.G., Reimann, C., Smith, D.B., Xie, X., 2008. From geochemical prospecting to international geochemical mapping: a historical overview. Geochem. Explor. Environ. Anal. (8), 205–217. https://doi.org/10.1144/1467-7873/08-174.

Gelman, A., 2007. Nassim Taleb's "The Black Swan". Statistical Modeling, Causal Inference and Social Science. Columbia University. https://statmodeling.stat.columbia.edu/2007/04/09/nassim_talebs_t/. (Accessed 19 January 2024).

Gerstgrasser, M., Schaeffer, R., Dey, A., Rafailov, R., Sleight, H., Hughes, J., Korbak, T., Agrawal, R., Pai, D., Gromov, A., Roberts, D.A., 2024. Is model collapse inevitable? breaking the curse of recursion by accumulating real and synthetic data. ArXiv preprint arXiv:2404.01413. https://doi.org/10.48550/arXiv.2404.01413.

Ghorbani, Y., Nwaila, G.T., Zhang, S.E., Bourdeau, J.E., Cánovas, M., Arzua, J., Nikadat, N., 2023a. Moving towards deep underground mineral resources: drivers, challenges and potential solutions. Res. Pol. 80, 103222 https://doi.org/10.1016/j.resourpol.2022.103222.

Ghorbani, Y., Zhang, S.E., Bourdeau, J.E., Chipangamate, N.S., Rose, D.H., Valodia, I., Nwaila, G.T., 2024. The strategic role of lithium in the green energy transition: towards and OPEC-style framework for Green Energy-Mineral Exporting Countries (GEMEC). Res. Pol. 90, 104737 https://doi.org/10.1016/j.resourpol.2024.104737.

Ghorbani, Y., Zhang, S.E., Nwaila, G.T., Bourdeau, J.E., 2022. Framework components for data-centric dry laboratories in the minerals industry: a path to science-and-technology-led innovation. Extr. Ind. Soc. 10, 101089 https://doi.org/10.1016/j.exis.2022.101089.

Ghorbani, Y., Zhang, S.E., Nwaila, G.T., Bourdeau, J.E., Safari, M., Hoseinie, S.H., Nwaila, P., Ruuska, J., 2023b. Dry laboratories–Mapping the required instrumentation and infrastructure for online monitoring, analysis, and characterization in the mineral industry. Miner. Eng. 191, 107971 https://doi.org/10.1016/j.mineng.2022.107971.

Ghosh, D., Samanta, B., Chakravarty, D., 2017. Multi sensor data fusion for 6D pose estimation and 3D underground mine mapping using autonomous mobile robot. Int. J. Image Data Fusion 8 (2), 173–187. https://doi.org/10.1080/19479832.2016.1226966.

Giblin, A., Mazzuchelli, R., 1997. Groundwater geochemistry in exploration. an investigation in the Black Flag district, Western Australia. Aust. J. Earth Sci. 44, 433–443. https://doi.org/10.1080/08120099708728324.

Government of Canada, 2022. Critical Minerals Centre of Excellence. Government of Canada. https://www.canada.ca/en/campaign/critical-minerals-in-canada.html. (Accessed 25 November 2022).

Govett, G.J.S., 1983. Geochemistry in the exploration sequence. In: Govett, G.J.S. (Ed.), Handbook of Exploration Geochemistry, 3. Elsevier, pp. 7–15. https://doi.org/10.1016/B978-0-444-42021-3.50006-9.

Grimes, D.J., Marranzino, A.P., 1968. Direct-current arc and alternating-current spark emission spectrographic field methods for the semiquantitative analysis of geologic materials. U.S. Geol. Surv. Circ. 591, 1–6. https://doi.org/10.3133/cir591.

Grossi, V., Giannotti, F., Pedreschi, D., Manghi, P., Pagano, P., Assante, M., 2021. Data science: a game changer for science and innovation. Int. J. Data Sci. Anal. 11, 263–278. https://doi.org/10.1007/s41060-020-00240-2.

Grunsky, E.C., 2010. The interpretation of regional geochemical survey data: proceedings of exploration 2007. Geochem. Explor. Environ. Anal. 10, 27–74.

Grunsky, E.C., de Caritat, P., 2017. Advances in the use of geochemical data for mineral exploration. In: Tschirhart, V., Thomas, M.D. (Eds.), Proceedings of Exploration 17: Sixth Decennial International Conference on Mineral Exploration, Decennial Minerals Exploration Conferences, pp. 441–456.

Grunsky, E.C., de Caritat, P., 2020. State-of-the-art analysis of geochemical data for mineral exploration. Geochem. Explor. Environ. Anal. 20 (2), 217–232. https://doi.org/10.1144/geochem2019-031.

Gu, Y., Schouwstra, R.P., Rule, C., 2014. The value of automated mineralogy. Miner. Eng. 58, 100–103. https://doi.org/10.1016/j.mineng.2014.01.020.

Guntoro, P.I., Ghorbani, Y., Butcher, A.R., Kuva, J., Rosenkranz, J., 2020. Textural quantification and classification of drill cores for geometallurgy: moving toward 3D with X-ray Microcomputed Tomography (μCT). Nat. Resour. Res. 29 (6), 3547–3565. https://doi.org/10.1007/s11053-020-09685-5.

Guntoro, P.I., Tiu, G., Ghorbani, Y., Lund, C., Rosenkranz, J., 2019. Application of machine learning techniques in mineral phase segmentation for X-ray microcomputed tomography (μCT) data. Miner. Eng. 142, 105882 https://doi.org/10.1016/j.mineng.2019.105882.

Gustavsson, N., Bølviken, B., Smith, D.B., Severson, R.C., 2001. Geochemical Landscapes of the Conterminous United States: New Map Presentations for 22 Elements. U.S. Geological Survey Professional Paper 1648.

Haldar, S.K., 2018. Exploration geochemistry. In: Haldar, S.K. (Ed.), Mineral Exploration, second ed. Elsevier, pp. 85–101. https://doi.org/10.1016/B978-0-12-814022-2.00005-8.

Hale, M., Plant, J.A., 1994. Drainage geochemistry. In: Govett, G.J.S. (Ed.), Handbook of Exploration Geochemistry, 6. Elsevier, Amsterdam.

Hall, G.E.M., Bonham-Carter, G.F., 1998. Selective extractions - foreword. J. Geochem. Explor. 61 (1–3) (VII-VIII).

Hastie, T., Tibshirani, R., Friedman, J.H., Friedman, J.H., 2017. The Elements of Statistical Learning: Data Mining, Inference, and Prediction, second ed. Springer, New York.

Hawkes, H.E., 1954. Geochemical Prospecting Investigations in the Nyeba Lead-Zinc District, Nigeria, No. 1000. US Government Printing Office, Washington.
Hawkins, J.R., Olson, M.P., Harouni, A., Qin, M.M., Hess, C.P., Majumdar, S., Crane, J.C., 2023. Implementation and prospective real-time evaluation of a generalized system for in-clinic deployment and validation of machine learning models in radiology. PLOS Digit. Health 2 (8), e0000227. https://doi.org/10.1371/journal.pdig.0000227.
Hazzan, O., Mike, K., 2023. The data science workflow. In: Guide to Teaching Data Science. Springer, Cham. https://doi.org/10.1007/978-3-031-24758-3_10.
He, Y., Zhou, Y., Wen, T., Zhang, S., Huang, F., Zou, X., Ma, X., Zhu, Y., 2022. A review of machine learning in geochemistry and cosmochemistry: method improvements and applications. J. App. Geochem. 140, 105273 https://doi.org/10.1016/j.apgeochem.2022.105273.
Henderson, D., Earley, S., Sebastian-Coleman, L., Sykora, E., Smith, E., 2017. DAMA-DMBOK: Data Management Body of Knowledge, second ed. DAMA International, Technics Publications, New Jersey.
Hill, E.J., Pearce, M.A., Stromberg, J.M., 2021. Improving automated geological logging of drill holes by incorporating multiscale spatial methods. Math. Geosci. 53 (1), 21–53. https://doi.org/10.1007/s11004-020-09859-0.
Hobbs, B.E., Ord, A., Blenkinsop, T., 2022. The spatial distributions of mineralisation. J. Struct. Geol. 156, 104529 https://doi.org/10.1016/j.jsg.2022.104529.
Hoffimann, J., Zortea, M., de Carvalho, B., Zadrozny, B., 2021. Geostatistical learning: challenges and opportunities. Front. App. Math. Stat. 7, 689393 https://doi.org/10.3389/fams.2021.689393.
Hosseini-Dinani, H., Mokhtari, A.R., Shahrestani, S., De Vivo, B., 2018. Sampling density in regional exploration and environmental geochemical studies: a review. Nat. Resour. Res. 28, 967–994. https://doi.org/10.1007/s11053-018-9431-2.
Howarth, R.J., Martin, L., 1979. Computer based techniques in the compilation, mapping and interpretation of exploration geochemical data. In: Hood, J.P. (Ed.), Geophysics and Geochemistry in the Search for Metallic Ores, 31. Geological Survey of Canada, Economic Geology Report, pp. 545–574. https://doi.org/10.4095/106037.
IEA, 2023. Energy Technology Perspectives 2023. International Energy Agency.
Isaaks, E.H., Srivastava, M.R., 1989. Applied Geostatistics. Oxford University Press.
James, C.H., 1957. Applied Geochemical Studies in Southern Rhodesia and Great Britain. Ph.D. Thesis. University of London.
Johnson, C.M., McLennan, S.M., McSween, H.Y., Summons, R.E., 2013. Smaller, better, more: five decades of advances in geochemistry. In: Bickford, M.E. (Ed.), The Web of Geological Sciences: Advances, Impacts and Interactions, 500. The Geological Society of America, p. 529. https://doi.org/10.1130/2013.2500(08, 302.
Karpatne, A., Ebert-Uphoff, I., Ravela, S., Babaie, H.A., Kumar, V., 2018. Machine learning for the geosciences: challenges and opportunities. IEEE Trans. Knowl. Data Eng. 31 (8), 1544–1554.
Kastens, K., Manduca, C., Cervato, C., Frodeman, R., Goodwin, C., Liben, L., Mogk, D., Spangler, T., Stillings, N., Titus, S., 2009. How geoscientists think and learn. Eos 90 (31), 265–272. https://doi.org/10.1029/2009EO310001.
Keane, C.M., 2001. Technical and Soft Skills Expectations During the Transition from Recent Graduate to New Hire. In: *AGU Fall Meeting Abstracts*, 2001, pp. ED22B–02.
Keane, C.M., Wilson, C.E., 2018. The mid-21st century geophysics workforce: how today's trends across geoscience impact geophysics human resources of the future. 2018 SEG International Exposition and Annual Meeting (OnePetro).
Kirste, D., de Caritat, P., Dann, R., 2004. The application of the stable isotopes of sulfur and oxygen in groundwater sulfate to mineral exploration in the Broken Hill region of Australia. J. Geochem. Explor. 78/79, 81–84. https://doi.org/10.1016/S0375-6742(03)00076-1.
Kleijn, R., Van der Voet, E., Kramer, G.J., Van Oers, L., Van der Giesen, C., 2011. Metal requirements of low-carbon power generation. Energy 36 (9), 5640–5648. https://doi.org/10.1016/j.energy.2011.07.003.
Kleinhans, M.G., Buskes, C.J., de Regt, H.W., 2005. Terra Incognita: explanation and reduction in earth science. Int. Stud. Philos. Sci. 19 (3), 289–317. https://doi.org/10.1080/02698590500462356.
Knight, R.D., Kjarsgaard, B.A., Russell, H.A., 2021. An analytical protocol for determining the elemental chemistry of Quaternary sediments using a portable X-ray fluorescence spectrometer. Appl. Geochem. 131, 105026 https://doi.org/10.1016/j.apgeochem.2021.105026.
Krenn, M., Buffoni, L., Coutinho, B., Eppel, S., Foster, J.G., Gritsevskiy, A., Lee, H., Lu, Y., Moutinho, J.P., Sanjabi, N., Sonthalia, R., 2023. Forecasting the future of artificial intelligence with machine learning-based link prediction in an exponentially growing knowledge network. Nat. Mach. Intell. 5, 1326–1335. https://doi.org/10.1038/s42256-023-00735-0.
Kurz, T.H., Buckley, S.J., Howell, J.A., 2013. Close-range hyperspectral imaging for geological field studies: workflow and methods. Int. J. Rem. Sens. 34 (5), 1798–1822. https://doi.org/10.1080/01431161.2012.727039.
Kyle, J.R., Ketcham, R.A., 2015. Application of high-resolution X-ray computed tomography to mineral deposit origin, evaluation, and processing. Ore Geol. Rev. 65, 821–839. https://doi.org/10.1016/j.oregeorev.2014.09.034.
Lawley, C.J., Gadd, M.G., Parsa, M., Lederer, G.W., Graham, G.E., Ford, A., 2023. Applications of natural language processing to geoscience text data and prospectivity modeling. Nat. Resour. Res. 32, 1503–1527. https://doi.org/10.1007/s11053-023-10216-1.
Lawley, C.J., Haynes, M., Chudasama, B., Goodenough, K., Eerola, T., Golev, A., Zhang, S.E., Park, J., Lèbre, E., 2024. Geospatial data and deep learning expose ESG risks to critical raw materials supply: the case of lithium. ES3 4, 10109. https://doi.org/10.3389/esss.2024.10109.
Lawley, C.J., McCafferty, A.E., Graham, G.E., Huston, D.L., Kelley, K.D., Czarnota, K., Paradis, S., Peter, J.M., Hayward, N., Barlow, M., Emsbo, P., 2022. Data–driven prospectivity modelling of sediment–hosted Zn–Pb mineral systems and their critical raw materials. Ore Geol. Rev. 141, 104635 https://doi.org/10.1016/j.oregeorev.2021.104635.
Lawley, C.J.M., Tschirhart, V., Smith, J.W., Pehrsson, S.J., Schetselaar, E.M., Schaeffer, A.J., Houlé, M.G., Eglington, B.M., 2021. Prospectivity modeling of Canadian magmatic Ni (± Cu ± Co ± PGE) sulphide mineral systems. Ore Geol. Rev. 132, 103985 https://doi.org/10.1016/j.oregeorev.2021.103985.
Lemière, B., Uvarova, Y.A., 2020. New developments in field-portable geochemical techniques and on-site technologies and their place in mineral exploration. Geochem. Explor. Environ. Anal. 20 (2), 205–216. https://doi.org/10.1144/geochem2019-044.
Leybourne, M.I., Cameron, E.M., 2010. Groundwaters in geochemical exploration. Geochem. Explor. Environ. Anal. 10 (2), 99–118. https://doi.org/10.1144/1467-7873/09-222.
Li, C., Ma, T., Shi, J., 2003. Application of a fractal method relating concentrations and distances for separation of geochemical anomalies from background. J. Geochem. Explor. 77, 167–175. https://doi.org/10.1016/S0375-6742(02)00276-5.
Linton, P., Kosanke, T., Greene, J., Porter, B., 2024. The application of hyperspectral core imaging for oil and gas. Geol. Soc. London Spec. Pub. 527 (1), 95–119. https://doi.org/10.1144/SP527-2022-2.
Lishchuk, V., Koch, P.H., Ghorbani, Y., Butcher, A.R., 2020. Towards integrated geometallurgical approach: critical review of current practices and future trends. Miner. Eng. 145, 106072 https://doi.org/10.1016/j.mineng.2019.106072.
Liu, Y., Cheng, Q., Carranza, E.J.M., Zhou, K., 2019. Assessment of geochemical anomaly uncertainty through geostatistical simulation and singularity analysis. Nat. Resour. Res. 28, 199–212. https://doi.org/10.1007/s11053-018-9388-1.
Luo, Z., Xiong, Y., Zuo, R., 2020. Recognition of geochemical anomalies using a deep variational autoencoder network. Appl. Geochem. 122, 104710 https://doi.org/10.1016/j.cageo.2021.104890.
Ma, X., 2023. Data science for geoscience: recent progress and future trends from the perspective of a data life cycle. In: Ma, X., Mookerjee, M., Hsu, L., Hills, D. (Eds.), Recent Advancement in Geoinformatics and Data Science, 558. Geological Society of America special paper, pp. 57–69. https://doi.org/10.1130/2022.2558(05.
Marghany, M., 2022. Remote Sensing and Image Processing in Mineralogy. CRC Press.
Matheron, G., 1963. Principles of geostatistics. Econ. Geol. 58 (8), 1246–1266. https://doi.org/10.2113/gsecongeo.58.8.1246.
McCurdy, M.W., Amor, S.D., Finch, C., 2016. Regional lake sediment and water geochemical data, western and central Labrador (NTS 13-L, 13-M, 14-D, 23-I and 23-J). Geological Survey of Canada. https://doi.org/10.4095/298834. Open file 8026.
McCurdy, M.W., Garrett, R.G., 2016. Geochemical data quality control for soil, till and lake and stream sediment samples. Geological Survey of Canada, Open File 7944. https://doi.org/10.4095/297562.
McCurdy, M.W., Spirito, W.A., Grunsky, E.C., Day, S.J.A., McNeil, R.J., Coker, W.B., 2014. The evolution of the geological survey of Canada's regional reconnaissance geochemical drainage sediment and water surveys. In: EXPLORE Newsletter, 163. Association of Applied Geochemists, pp. 1–10.
McGrath, S.P., Loveland, P.J., 1992. The Soil Geochemical Atlas of England and Wales. Blackie Academic and Professional, Glasgow.
Medunić, G., Chakravarty, S., Kundu, R., 2022. Computational skills in geosciences higher education system for the 21st century. Mathematical Methods and Terminology in Geology.
Michaux, S., 2021. Assessment of the extra capacity required of alternative energy electrical power systems to completely replace fossil fuels. Geological Survey of Finland (GTK), report 42/2021, pp. 1–1000. https://doi.org/10.13140/RG.2.2.34895.00160.
Michaux, S., O'Connor, L., 2020. How to Set up and Develop a Geometallurgical Program. Geological Survey of Finland. GTK Open File Work Report 72/2019.
Miesch, A.T., 1975. Variograms and variance components in geochemistry and ore evaluation. In: Whitten, E.H.T. (Ed.), Quantitative Studies in the Earth Sciences, 142. Geological Society of America, Memoir, pp. 333–340.
Miesch, A.T., 1976. Geochemical Survey of Missouri – Methods of Sampling, Laboratory Analysis, and Statistical Reduction of Data, 954-A. U.S. Geological Survey professional paper. https://doi.org/10.3133/pp954A.
Mikeš, D., 2015. Geoscience education is outdated. South Afr. J. Geol. 118 (3), 299–301. https://doi.org/10.2113/gssajg.118.3.299.
MineSense, 2024. Unlocking New Orebody Knowledge. MineSense. https://minesense.com/. (Accessed 10 January 2024).
Myers, A.T., Havens, R.G., Dunton, P.J., 1961. A spectrochemical method for the semiquantitative analysis of rocks, minerals, and ores. U.S. Geological Survey Bulletin, 1084-I, pp. 207–229. https://doi.org/10.3133/b1084I.
Neiman, H.G., 1976. Analysis of rocks, soils, and plant ashes by emission spectroscopy. In: Miesch, A.T. (Ed.), Geochemical Survey of Missouri - Methods of Sampling, Laboratory Analysis, and Statistical Reduction of Data. U.S. Geological Survey Professional Paper 954-A, A14-A15. https://doi.org/10.3133/pp954A.
OSC, 2011. Ontario securities commission bulletin. In: The Ontario Securities Commission, 34, pp. 7043–7086. Chap. Rules and Policies: NI 43-101 standards of disclosure for mineral projects, form 43-101F1 Technical Report and related consequential amendments.
Parian, M., Lamberg, P., Möckel, R., Rosenkranz, J., 2015. Analysis of mineral grades for geometallurgy: combined element-to-mineral conversion and quantitative X-ray diffraction. Miner. Eng. 82, 25–35. https://doi.org/10.1016/j.mineng.2015.04.023.
Parsa, M., Maghsoudi, A., Yousefi, M., 2017. An improved data-driven fuzzy mineral prospectivity mapping procedure; cosine amplitude-based similarity approach to delineate exploration targets. Int. J. Appl. Earth Obs. Geoinf. 58, 157–167. https://doi.org/10.1016/j.jag.2017.02.006.

Piercey, S.J., 2014. Modern analytical facilities 2. A review of quality assurance and quality control (QA/QC) procedures for lithogeochemical data. Geosci. Can. 41, 75–88. https://doi.org/10.12789/geocanj.2014.41.035.
Pirlo, M.C., Giblin, A.M., 2004. Application of groundwater-mineral equilibrium calculations to geochemical exploration for sediment-hosted uranium. observations from the Frome Embayment, South Australia. Geochem-Explor. Env. A. 4, 113–127.
Plet, C., Noble, R.R., 2023. Soil gases in mineral exploration: a review and the potential for future developments. Geochem. Explor. Environ. Anal. 23 (2) https://doi.org/10.1144/geochem2023-008 geochem2023-008.
Porwal, A., Carranza, E.J.M., Hale, M., 2006. A hybrid fuzzy weights-of-evidence model for mineral potential mapping. Nat. Resour. Res. 8, 27–35. https://doi.org/10.1007/s11053-006-9012-7.
Prider, R.T., 1981. Blowpipe analysis of minerals. In: Mineralogy. Encyclopedia of Earth Science. Springer, Boston. https://doi.org/10.1007/0-387-30720-6_19.
Reed, R., Marks, R.J., 1999. Neural Smithing: Supervised Learning in Feedforward Artificial Neural Networks. MIT Press.
Reimann, C., Garrett, R.G., 2005. Geochemical background – concept and reality. Sci. Total Environ. 350, 12–27. https://doi.org/10.1016/j.scitotenv.2005.01.047.
Reimann, C., 2005. Sub-continental-scale geochemical mapping: sampling, quality control and data analysis issues. Geochem. Explor. Environ. Anal. 5, 311–323. https://doi.org/10.1144/1467-7873/03-065.
Reimann, C., Kashulina, G., de Caritat, P., Niskavaara, H., 2001. Multi-element, multi-medium regional geochemistry in the European Arctic: element concentration, variation and correlation. Appl. Geochem. 16 (7–8), 759–780. https://doi.org/10.1016/S0883-2927(00)00070-6.
Rodriguez-Galiano, V., Sanchez-Castillo, M., Chica-Olmo, M., Chica-Rivas, M.J.O.G.R., 2015. Machine learning predictive models for mineral prospectivity: an evaluation of neural networks, random forest, regression trees and support vector machines. Ore Geol. Rev. 71, 804–818. https://doi.org/10.1016/j.oregeorev.2015.01.001.
Rogati, M., 2017. The artificial intelligence hierarchy of needs. Hackernoon. https://hackernoon.com/the-ai-hierarchy-of-needs-18f111fcc007. (Accessed 23 October 2022).
Roser, M., Ritchie, H., Mathieu, E., 2023. What is moore's law. Our world in data. https://ourworldindata.org/moores-law. (Accessed 10 January 2024).
SAMREC, 2016. The South African code for the reporting of exploration results, mineral resources and mineral reserves (the SAMREC code). South African Mineral Codes, pp. 1–98.
Schodlok, M.C., Whitbourn, L., Huntington, J., Mason, P., Green, A., Berman, M., Coward, D., Connor, P., Wright, W., Jolivet, M., Martinez, R., 2016. HyLogger-3, a visible to shortwave and thermal infrared reflectance spectrometer system for drill core logging: functional description. Aust. J. Earth Sci. 63, 929–940. https://doi.org/10.1080/08120099.2016.1231133.
Semereab, M., 2021. XRF Spectroscopy in Analytical Multi-Sensor Drill Core-Scanning: Evaluating the Potential of Qualitative and Quantitative Analysis for Aguas Teñidas Deposit, Iberian Pyrite Belt, Spain. M.Sc. thesis, Université de Liègey.
Sergeyev, Y.E.A., 1941. Geochemical Methods of Prospecting for Ore Deposits. U.S. Geological Survey. Open File Report 48-21-B.
Shacklette, H.T., Boerngen, J.G., 1984. Element Concentrations in Soils and Other Surficial Aaterials of the Conterminous United States. U.S. Geological Survey Professional. Paper 1270.
Shacklette, H.T., Hamilton, J.C., Boerngen, J.G., Bowles, J.M., 1971. Elemental Composition of Surficial Materials in the Conterminous United States. U.S. Geological Survey Professional Paper 574-D.
Shumailov, I., Shumaylov, Z., Zhao, Y., Gal, Y., Papernot, N., Anderson, R., 2023. The curse of recursion: training on generated data makes models forget. ArXiv preprint arXiv:2305.17493. https://doi.org/10.48550/arXiv.2305.17493.
Simán, F., 2020. Assessment of Machine Learning Applied to X-Ray Fluorescence Core Scan Data from the Zinkgruvan Zn-Pb-Ag Deposit, Bergslagen, Sweden. M.Sc. Thesis, Luleå University of Technology.
Smith, D.B., Smith, S.M., Horton, J.D., 2013. History and evaluation of national-scale geochemical data sets for the United States. Geosci. Front. 4, 167–183. https://doi.org/10.1016/j.gsf.2012.07.002.
Stanford University, 2024. Data Science for Geoscience. Stanford University. https://eps ci.stanford.edu/academics-admissions/why-major-earth-and-planetary-sciences/data-science-geoscience. (Accessed 1 January 2024).
Sun, Z., Sandoval, L., Crystal-Ornelas, R., Mousavi, S.M., Wang, J., Lin, C., Cristea, N., Tong, D., Carande, W.H., Ma, X., Rao, Y., 2022. A review of earth artificial intelligence. Comput. Geosci. 159, 105034 https://doi.org/10.1016/j.cageo.2022.105034.
Talapatra, A.K., 2020. Geochemical Exploration and Modelling of Concealed Mineral Deposits. Springer. https://doi.org/10.1007/978-3-030-48756-0.
Tidball, R.R., Severson, R.C., 1976. Chemistry of northern great plains soils. U.S. Geological Survey, Open-File Report 76–729, 57–81.
Tukey, J.W., 1977. Exploratory Data Analysis. Addison-Wesley, Reading, MS.
Tuoriniemi, J., Cornelis, G., Hassellöv, M., 2015. A new peak recognition algorithm for detection of ultra-small nano-particles by single particle ICP-MS using rapid time resolved data acquisition on a sector-field mass spectrometer. J. Anal. At. Spectrom. 30 (8), 1723–1729. https://doi.org/10.1039/C5JA00113G.
Ullo, S.L., Zarro, C., Wojtowicz, K., Meoli, G., Focareta, M., 2020. LiDAR-based system and optical VHR data for building detection and mapping. J. Sens. 20 (5), 1285. https://doi.org/10.3390/s20051285.
UN, 2015. Paris agreement. United nations. Clim. Change. In: https://unfccc.int/process-and-meetings/the-paris-agreement. (Accessed 10 July 2024).
USGS, 2020. Data Management Website. U.S. Geological Survey. https://doi.org/10.5066/F7MW2G15.
Van der Boogaart, K.G., Tolosana-Delgado, R., 2018. Predictive geometallurgy: an interdisciplinary key challenge for mathematical geosciences. In: Handbook of Mathematical Geosciences. Fifty Years of IAMG, pp. 673–686.
van Riel, R., Van Gulick, R., 2024. Scientific Reduction. In: Zalta, E.N., Nodelman, U. (Eds.), The Stanford Encyclopedia of Philosophy. https://plato.stanford.edu/archives/spr2024/entries/scientific-reduction/. accessed 10 July 2024.
Wang, J., Zuo, R., Caers, J., 2017a. Discovering geochemical patterns by factor-based cluster analysis. J. Geochem. Explor. 181, 106–115. https://doi.org/10.1016/j.gexplo.2017.07.006.
Wang, L., Yin, D.Z., Caers, J., 2023. Data Science for the Geosciences. Cambridge University Press.
Wang, X., Yao, M., Li, J., Zhang, K., Zhu, H., Zheng, M., 2017b. China's rare earths production forecasting and sustainable development policy implications. Sustainability 9 (6), 1–14. https://doi.org/10.3390/su9061003.
Wang, Z., Zuo, R., Jing, L., 2021. Fusion of geochemical and remote-sensing data for lithological mapping using random forest metric learning. Math. Geosci. 53, 1125–1145. https://doi.org/10.1007/s11004-020-09897-8.
Watari, T., McLellan, B.C., Giurco, D., Dominish, E., Yamasue, E., Nansai, K., 2019. Total material requirement for the global energy transition to 2050: a focus on transport and electricity. Resour. Conserv. Recycl. 148, 91–103. https://doi.org/10.1016/j.resconrec.2019.05.015.
Webb, J.S., Fortescue, J., Nichol, I., Tooms, J.S., 1964. Regional Geochemical Reconnaissance in the Namwala Concession Area, Zambia, 47. Geochemical Prospecting Research Centre, Technical Communication.
White, W.M., 2015. Isotope Geochemistry. Wiley Blackwell.
Whittaker, E.T., 1915. XVIII. – on the functions which are represented by the expansions of the interpolation-theory. P. Roy. Soc. Edin 35, 181–194. https://doi.org/10.1017/S0370164600017806.
Whitworth, A.J., Vaughan, J., Southam, G., van der Ent, A., Nkrumah, P.N., Ma, X., Parbhakar-Fox, A., 2022. Review on metal extraction technologies suitable for critical metal recovery from mining and processing wastes. Miner. Eng. 182, 107537 https://doi.org/10.1016/j.mineng.2022.107537.
Wilson, C.E., 2014. Explanation of the Predicted Geoscience Workforce Shortage. American Geosciences Institute. Brief 2014-011. https://www.americangeosciences.org/geoscience-currents/explanation-predicted-geoscience-workforce-shortage. (Accessed 1 January 2024).
Winterburn, P.A., Noble, R.R., Lawie, D., 2020. Advances in exploration geochemistry, 2007 to 2017 and beyond. Geochem. Explor. Environ. Anal. 20 (2), 157–166. https://doi.org/10.1144/geochem2019-030.
Wirth, R., Hipp, J., 2000. CRISP-DM: towards a standard process model for data mining. Proceedings of the 4th international conference on the practical applications of knowledge discovery and data mining 1, 29–39.
Woodall, P., 2017. The data repurposing challenge: new pressures from data analytics. J. Data Inf. Qual. 8 (3–4), 1–4. https://doi.org/10.1145/3022698.
Woodall, P., Wainman, A., 2015. Data quality in analytics: key problems arising from the repurposing of manufacturing data. In: International Conference on Information Quality (ICIQ), pp. 174–184. Cambridge.
Xie, X., Ren, T., 1993. National geochemical mapping and environmental geochemistry–Progress in China. In: Davenport, P.H. (Ed.), Geochemical Mapping. J. Geochem. Explor. 49, 15–34. https://doi.org/10.1016/0375-6742(93)90037-M.
Xu, F., Uszkoreit, H., Du, Y., Fan, W., Zhao, D., Zhu, J., 2019. Explainable AI: a brief survey on history, research areas, approaches and challenges, in: natural language processing and Chinese computing. In: Proceedings of the 8th cCF International Conference, NLPCC 2019, Part II, 8. Springer, pp. 563–574.
Yamamoto, K.Y., Cremers, D.A., Ferris, M.J., Foster, L.E., 1996. Detection of metals in the environment using a portable laser-induced breakdown spectroscopy instrument. Appl. Spectroc. 50, 222–233.
Yang, N., Zhang, Z., Yang, J., Hong, Z., 2022. Applications of data augmentation in mineral prospectivity prediction based on convolutional neural networks. Comput. Geosci. 161, 105075 https://doi.org/10.1016/j.cageo.2022.105075.
Yousefi, M., Carranza, E.J.M., 2015a. Fuzzification of continuous-value spatial evidence for mineral prospectivity mapping. Comput. Geosci. 74, 97–109. https://doi.org/10.1016/j.cageo.2014.10.014.
Yousefi, M., Carranza, E.J.M., 2015b. Prediction–area (P–A) plot and C–A fractal analysis to classify and evaluate evidential maps for mineral prospectivity modeling. Comput. Geosci. 79, 69–81. https://doi.org/10.1016/j.cageo.2015.03.007.
Yousefi, M., Carranza, E.J.M., Kreuzer, O.P., Nykänen, V., Hronsky, J.M., Mihalasky, M. J., 2021. Data analysis methods for prospectivity modelling as applied to mineral exploration targeting: state-of-the-art and outlook. J. Geochem. Explor. 229, 106839 https://doi.org/10.1016/j.gexplo.2021.106839.
Zhang, C., Han, J., 2021. Data mining and knowledge discovery. In: Shi, W., Goodchild, M.F., Batty, M., Kwan, M.P., Zhang, A. (Eds.), Urban Informatics. The Urban Book Series. Springer, Singapore. https://doi.org/10.1007/978-981-15-8983-6_42.
Zhang, C., Zuo, R., Xiong, Y., 2021b. Detection of the multivariate geochemical anomalies associated with mineralization using a deep convolutional neural network and a pixel-pair feature method. Appl. Geochem. 130, 104994 https://doi.org/10.1016/j.apgeochem.2021.104994.
Zhang, S.E., Bourdeau, J.E., Nwaila, G.T., Corrigan, D., 2022a. Towards a fully data-driven prospectivity mapping methodology: case study of the Southeastern Churchill Province, Quebec and Labrador. Artif. Intell. Geosci. 2, 128–147. https://doi.org/10.1016/j.aiig.2022.02.002.
Zhang, S.E., Bourdeau, J.E., Nwaila, G.T., Ghorbani, Y., 2022b. Advanced geochemical exploration knowledge using machine learning: prediction of unknown elemental concentrations and operational prioritization of re-analysis campaigns. Artif. Intell. Geosci. 3, 86–100. https://doi.org/10.1016/j.aiig.2022.10.003.

Zhang, S.E., Bourdeau, J.E., Nwaila, G.T., Ghorbani, Y., 2023a. Emerging criticality: unraveling shifting dynamics of the EU's critical raw materials and their implications on Canada and South Africa. Res. Pol. 86, 104247 https://doi.org/10.1016/j.resourpol.2023.104247.

Zhang, S.E., Bourdeau, J.E., Nwaila, G.T., Parsa, M., Ghorbani, Y., 2024a. Denoising of geochemical data using deep learning – implications for regional surveys. Nat. Resour. Res. 33 (2), 495–520. https://doi.org/10.1007/s11053-024-10317-5.

Zhang, S.E., Lawley, C.J., Bourdeau, J.E., Nwaila, G.T., Ghorbani, Y., 2024b. Workflow-induced uncertainty in data-driven mineral prospectivity mapping. Nat. Resour. Res. 33, 995–1023. https://doi.org/10.1007/s11053-024-10322-8.

Zhang, S.E., Nwaila, G.T., Agard, S., Bourdeau, J.E., Carranza, E.J.M., Ghorbani, Y., 2023b. Big geochemical data through remote sensing for dynamic mineral resource monitoring in tailing storage facilities. Artif. Intell. Geosci. 4, 137–149. https://doi.org/10.1016/j.aiig.2023.09.002.

Zhang, S.E., Nwaila, G.T., Bourdeau, J.E., Ashwal, L.D., 2021a. Machine learning based prediction of trace element concentrations using data from the Karoo large igneous province and its application in prospectivity mapping. Artif. Intell. Geosci. 2, 60–75. https://doi.org/10.1016/j.aiig.2021.11.002.

Zhang, S.E., Nwaila, G.T., Bourdeau, J.E., Ghorbani, Y., Carranza, E.J.M., 2023c. Deriving big geochemical data from high-resolution remote sensing data via machine learning: application to a tailing storage facility in the Witwatersrand goldfields. Artif. Intell. Geosci. 4, 9–21. https://doi.org/10.1016/j.aiig.2023.01.005.

Zhang, W., Gu, X., Tang, L., Yin, Y., Liu, D., Zhang, Y., 2022c. Application of machine learning, deep learning and optimization algorithms in geoengineering and geoscience: comprehensive review and future challenge. Gondwana Res. 109, 1–17. https://doi.org/10.1016/j.gr.2022.03.015.

Zuo, R., 2016. A nonlinear controlling function of geological features on magmatic–hydrothermal mineralization. Sci. Rep. 6 (1), 27127 https://doi.org/10.1038/srep27127.

Zuo, R., 2020. Geodata science-based mineral prospectivity mapping: a review. Nat. Resour. Res. 29, 3415–3424. https://doi.org/10.1007/s11053-020-09700-9.

Zuo, R., Carranza, E.J.M., Wang, J., 2016. Spatial analysis and visualization of exploration geochemical data. Earth Sci. Rev. 158, 9–18. https://doi.org/10.1016/j.earscirev.2016.04.006.

Zuo, R., Luo, Z., Xiong, Y., Yin, B., 2022. A geologically constrained variational autoencoder for mineral prospectivity mapping. Nat. Resour. Res. 31 (3), 1121–1133. https://doi.org/10.1007/s11053-022-10050-x.

Zuo, R., Shi, L., Yang, F., Xu, Y., Xiong, Y., 2024. ArcMPM: an ArcEngine-based software for mineral prospectivity mapping via artificial intelligence algorithms. Nat. Resour. Res. 33 (1), 1–21. https://doi.org/10.1007/s11053-023-10286-1.

Zuo, R., Wang, J., Xiong, Y., Wang, Z., 2021. The processing methods of geochemical exploration data: past, present, and future. Appl. Geochem. 132, 105072. https://doi.org/10.1016/j.apgeochem.2021.105072.

Zuo, R., Xiong, Y., 2018. Big data analytics of identifying geochemical anomalies supported by machine learning methods. Nat. Resour. Res. 27, 5–13. https://doi.org/10.1007/s11053-017-9357-0.

Zuo, R., Xiong, Y., Wang, J., Carranza, E.J.M., 2019. Deep learning and its application in geochemical mapping. Earth Sci. Rev. 192, 1–14. https://doi.org/10.1016/j.earscirev.2019.02.023.

Zuo, R., Xu, Y., 2023. Graph deep learning model for mapping mineral prospectivity. Math. Geosci. 55 (1), 1–21. https://doi.org/10.1007/s11004-022-10015-z.