Journal of Cleaner Production 540 (2026) 147437

Contents lists available at ScienceDirect

Journal of Cleaner Production

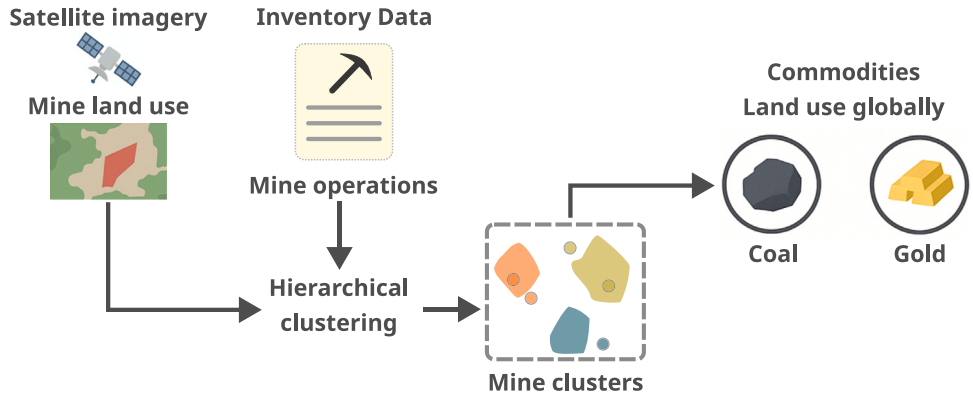
journal homepage: www.elsevier.com/locate/jclepro

A data-driven approach to mapping global commodity-specific mining land-use

Victor Maus *

Institute for Ecological Economics, Vienna University of Economics and Business (WU), Welthandelsplatz 1, Vienna, 1020, Austria
Advancing Systems Analysis Program, International Institute for Applied Systems Analysis (IIASA), Schlossplatz 1, Laxenburg, A-2361, Austria

GRAPHICAL ABSTRACT



HIGHLIGHTS

- New approach to link satellite-derived mine land-use and mine operation data.
- Unsupervised clustering reduces subjectivity and increases the linking accuracy.
- Integrated data sources reveal a total of 145,738.1 km^2 of extractive land globally.
- Coal and gold each account for over 21% of global mine land area.
- 26.8% of mining land could not be linked to a specific commodity.

ARTICLE INFO

Dataset link: https://maps.minethegap.eu, https://github.com/vwmaus/mining-spatial-data-integration

Keywords:
Minerals supply chain
Land use change
LCI
Environmental assessment
Geospatial data integration

ABSTRACT

Mineral extraction is a key driver of environmental change globally, yet geospatial data on mining operations remains fragmented and incomplete across data sources. Datasets with complementary information, such as mining project inventories (points) and satellite-derived land use (polygons), are often disconnected due to spatial mismatches and the complex distribution of infrastructures, such as open pits, tailings, and processing facilities, which are frequently scattered. Integrating these geographic features is critical for enhancing mining data availability and leveraging data complementarity, thereby advancing the understanding of mining impacts globally. This study proposes a scalable approach to link heterogeneous mining datasets and demonstrates its applicability by quantifying the global mine land associated with specific commodities. The new approach introduces data-driven *mine clusters*, grouping geographic features through hierarchical clustering with locally

* Correspondence to: Institute for Ecological Economics, Vienna University of Economics and Business (WU), Welthandelsplatz 1, Vienna, 1020, Austria.
E-mail address: victor.maus@wu.ac.at.
URL: https://victormaus.github.io.

https://doi.org/10.1016/j.jclepro.2025.147437
Received 17 June 2025; Received in revised form 5 December 2025; Accepted 28 December 2025
Available online 8 January 2026

Hierarchical clustering

optimised distance thresholds. This method enables associating information from inventory data with land-use polygons covering mine infrastructure derived from satellite data. To test the approach, data from various sources were integrated. The resulting integrated dataset covers over 145,000 km^2 and offers the most comprehensive overview of global mine land use linked to mineral commodities. Validation of the clusters against expert-labelled mines shows a high level of agreement, with 95 % of the clusters sharing at least one primary commodity. Results revealed that coal (22.5 %) and gold (21.1 %) dominate global mining land footprints. 26.8 % of the area could not be assigned to a commodity. This methodology provides a reproducible approach to enhancing the integration of spatial data on mining activities, supporting more robust global assessments of mining impacts.

1. Introduction

Global demand for minerals has accelerated rapidly over recent decades (OECD, 2019), fuelling an unprecedented expansion of mineral extraction activities across diverse geographic contexts (Giljum et al., 2025). This surge in mining is closely linked to a wide array of environmental and social impacts, including deforestation, biodiversity loss, community displacement, and resource-related conflicts (Giljum et al., 2022; Owen et al., 2022; Siqueira-Gay et al., 2022; Luckeneder et al., 2021; Siqueira-Gay et al., 2020; Sonter et al., 2018; Conde, 2017; Bebbington and Williams, 2008; Lagos et al., 2018; Sonter et al., 2014). These impacts are mainly driven by complex global supply chains that spatially decouple resource consumption from the regions that bear the environmental and social costs (Berthet et al., 2024). A single commodity or economic sector can therefore drive ecological degradation across multiple geographically dispersed locations, each with distinct socio-ecological vulnerabilities.

As mining pressures intensify, there is growing recognition of the need to study mining globally, while accounting for the local heterogeneity of its impacts (Sun et al., 2025; Lèbre et al., 2024; Crona et al., 2023). The expansion of geospatial datasets, particularly those based on remote sensing technologies, has created new opportunities to map, monitor, and assess the impacts of mining at a planetary scale (Guo et al., 2024; Tang and Werner, 2023a; Maus et al., 2022a; Werner et al., 2020b). Datasets derived from satellite imagery can capture various features of mining infrastructure, such as pits, tailings, storage facilities, overburden piles, waste rock dumps, and processing plants. Although derived from satellite images, these features are usually represented as vector data, specifically as polygons. However, these polygons currently lack a direct link to mine-specific operation data reported by companies and governments, which is necessary to associate their impacts with commodities and economic flows. In contrast, mining inventories, compiled from company reports and sustainability disclosures, provide detailed information on individual mines, including mined commodities, production volumes, ore grades, ownership, and in some cases, waste volumes, tailings management practices, and land and water use. Despite these details, such inventories can be incomplete and often do not include georeferenced delineations of the areas physically occupied by mining operations; at best, they provide location coordinates as point data (Maus and Werner, 2024; Fonseca et al., 2014).

Despite the incompleteness and fragmentation of available datasets, they can complement each other to enable more comprehensive analyses. However, attempts to integrate multiple sources often fall short due to spatial mismatches and ambiguities in the association between land-use features derived from satellite imagery and mine-site information in inventories. Even time-consuming, manually conducted approaches are compromised by spatial mismatches and ambiguities (Tang and Werner, 2023a; Maus et al., 2022a; Werner et al., 2020b), and remain vulnerable to the subjectivity of the experts conducting the task. Furthermore, it is impractical to gather local knowledge for each mine worldwide. These challenges lead to the adoption of rather simplistic approaches, for example, based on the nearest ground feature (Cabernard and Pfister, 2022), which produce unrealistic relationships between land use and sites in inventories. Thus, resolving all point-to-polygon ambiguities in mining datasets remains an open methodological challenge.

This study proposes a pragmatic solution by introducing *mine clusters*, which can automatically and reproducibly derive associations between spatial entities solely from the information available in the mining datasets. The proposed approach to construct mine clusters leverages the geographic proximity of georeferenced features (polygons and points) and the mined commodities to infer relationships between features using a hierarchical clustering algorithm with locally optimal distance thresholds. The resulting mine clusters are groups of nearby inventory points and land-use polygons that, taken together, likely represent coherent mining operations, even when including multiple properties or operators. Visual assessments of inferred mine clusters and comparisons with existing independent data collections indicate that the methodology can robustly integrate diverse data sources. While this approach does not eliminate all uncertainty, it provides a consistent and scalable framework for approximating associations between spatial features, enabling the systematic integration of heterogeneous datasets and providing a more comprehensive foundation for global socio-ecological assessments of mining.

2. Methods

2.1. Mine data definition

This study integrates two complementary types of geospatial datasets to characterise mining activity: (i) mines from inventory data, which includes reported locations of mining operations (points), and (ii) land-use data derived from satellite imagery, which delineates the physical areas occupied by mining infrastructures (polygons). Each dataset contributes distinct but essential information to support a wide range of analyses related to mining impacts. Common to both data types is their georeferencing, i.e., they are assigned to a geographic coordinate system corresponding to a real-world location on the Earth's surface. Therefore, entries in these datasets (points and polygons) will be generically referred to as *geographic feature*.

(i) Mine locations (points): These data are extracted from mining inventories compiled from corporate sustainability reports, government disclosures, and third-party datasets. Each entry corresponds to a mine or mining project and includes geographic point coordinates representing its reported or approximate inferred location. The inventory data also contains non-spatial attributes, such as mined commodities, production volume, and ownership, among others. These points may represent centroids of mining concessions, administrative headquarters, or approximate locations.

(ii) Mining land-use (polygons): These data are generated through visual interpretation of high-resolution satellite imagery. The resulting polygons delineate the surface footprint of mining infrastructures, including extraction pits, tailings ponds, waste rock dumps, and related facilities. While these features offer detailed spatial coverage and a consistent representation of land disturbance, they do not contain direct references to the mine names, companies, commodities, or production attributes.

2.2. Configuration and association problem of geographic features

Combining the two primary mine data types to increase their utility is desirable but nontrivial due to spatial mismatches and relational

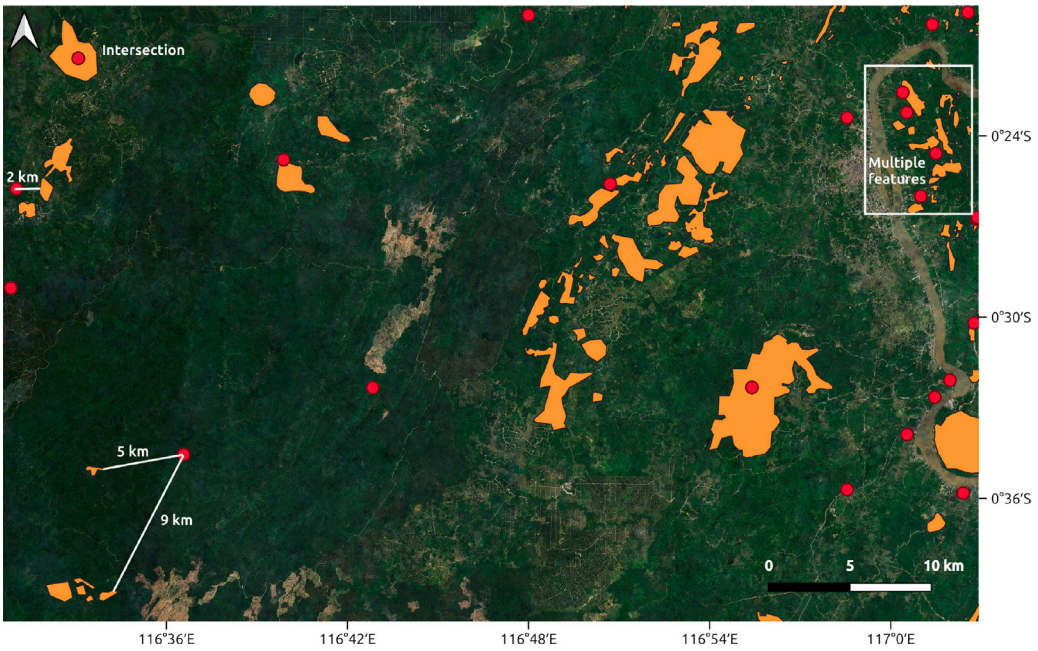


Fig. 1. Illustration of the complex spatial distribution of data points and polygons in the Kutai Kartanegara Regency, East Kalimantan, Indonesia. Image: Sentinel-2 cloudless https://s2maps.eu by EOX IT Services GmbH (Contains modified Copernicus Sentinel data 2019).

ambiguities. In the inventories, a mine property represented by a single pair of coordinates can have several disconnected infrastructures, e.g. tailings, waste rock, or multiple pits, each represented by a polygon, (Tang and Werner, 2023a; Maus et al., 2022a; Werner et al., 2020b). The pair of coordinates, therefore, only provides a general approximation of the location of each mine with varying precision (Murguía et al., 2016; Kobayashi et al., 2014) and the infrastructures linked to a single mine can be displaced several kilometres away from each other (Maus et al., 2020; Werner et al., 2020a).

Fig. 1 illustrates the spatial distribution of geographic features (points and polygons) related to mining. In some cases, the association is simple, for example, a point located directly within a single polygon that is far from other features (top left of Fig. 1). However, other configurations are more problematic: polygons may appear far from any associated point (bottom left), or features may be densely clustered in ways that obscure individual relationships (top right), i.e. many properties are often reported near each other, so it is impossible to precisely create a point-to-polygon association without ground knowledge on the specific cases.

2.3. Mine data integration

To tackle this problem, this study proposes grouping spatial features into *mine clusters* that are small enough to preserve local heterogeneity yet large enough to capture complementary information from both datasets–inventory points and land-use polygons. The processing workflow illustrated in Fig. 2 was designed to systematically build spatial relationships between the two data types while maintaining traceability, consistency, and enabling scalability. The process includes data harmonisation, spatial aggregation, integration, batching for parallel processing, clustering, and distance threshold optimisation. The implementation is in R (R Core Team, 2024) and is openly available from Maus (2025).

2.3.1. Pre-processing

The process starts with harmonisation steps applied separately to each dataset type. Polygon features are harmonised by aligning attribute schemas and coordinate reference systems (CRS), followed by a spatial aggregation step where overlapping or adjacent polygons are merged to eliminate internal borders. This ensures that spatially overlapping areas are not counted multiple times, allowing for the integration of polygons from various complementary data sources, even if they partially overlap. Similarly, the inventory point data undergoes harmonisation to standardise attributes and CRS alignment.

Once harmonised, the point and polygon datasets are integrated into a unified spatial database. For efficient processing, the features are split into equal-sized batches and spatial groups to reduce computational costs and account for the geographic independence of distant features. For each batch, a sparse pairwise geographical distance matrix is computed, including only the distances between features within the same spatial group, i.e. features that are chain-connected within a maximum distance threshold (usually a large distance). The maximum threshold for pre-grouping is not a clustering parameter itself, but rather a parameter that defines the maximum extent of a local cluster. It should be set to a conservatively high value (e.g., 20 km), which is larger than any reasonably expected cluster diameter but should not be excessively large, to avoid unnecessary additional computational costs. Since this proximity threshold avoids computing all pairwise distances between geographic features, the memory and processing costs are substantially reduced, even for large global datasets.

2.3.2. Clustering

The distance matrices are then used as input to a fast hierarchical agglomerative clustering algorithm with single linkage (Müllner, 2013). This method iteratively merges the two closest clusters based on their pairwise distances, continuing until all features are grouped into a single cluster. The resulting structure can be visualised as a dendrogram, as illustrated in Fig. 3a, where the vertical axis represents distance and the horizontal axis represents individual features and clusters.

Setting a very low maximum distance threshold (tree depth) at the base of the dendrogram allows each feature to become an individual cluster. Increasing the distance threshold will merge adjacent clusters, forming progressively larger groups. This structure intuitively illustrates how increasing the distance threshold results in fewer, larger clusters. A few common strategies are often used to determine the dendrogram cutoff: (i) specifying a maximum number of features per cluster, (ii) specifying the number of clusters, or (iii) defining a maximum tree depth, which in this case corresponds to a geographical

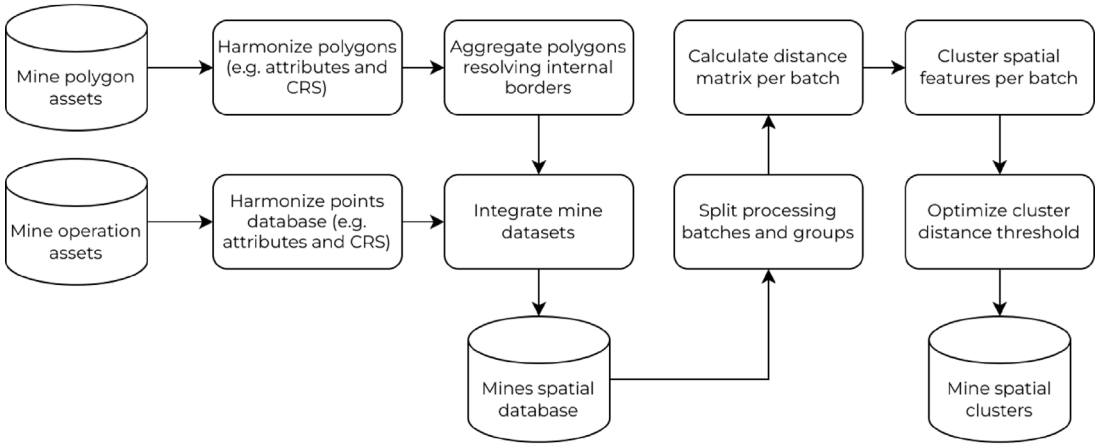


Fig. 2. Workflow for integrating mine polygon and point datasets to produce spatially coherent mine clusters.

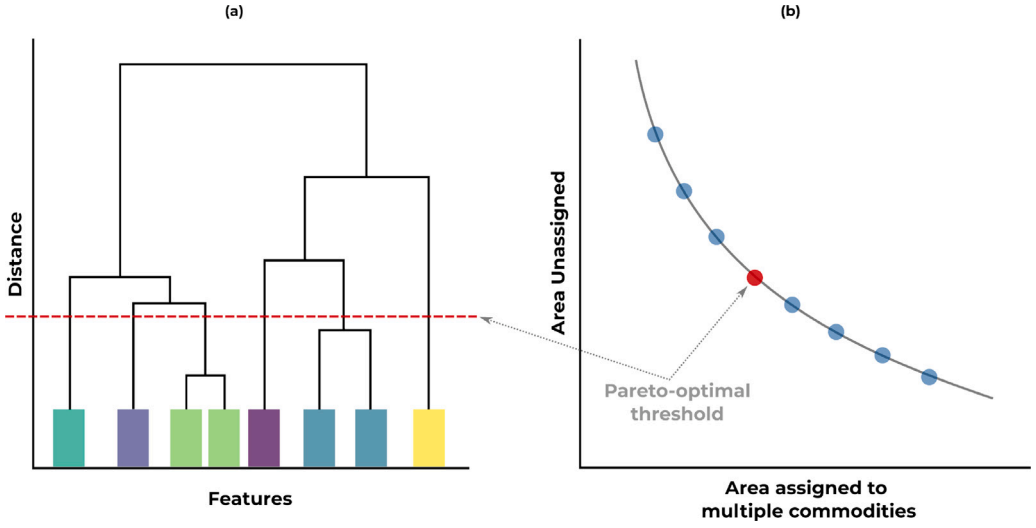


Fig. 3. Illustration of the optimal threshold distance based on Pareto efficiency. (a) Shows the dendrogram with the optimal distance (dashed red line) and feature clusters in different colours. (b) Shows the Pareto frontier between the competing objectives derived from multiple discrete tree depths (distances). The sub-optimal thresholds are represented in blue, and the optimal threshold minimising the objectives is red. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

distance threshold. Since the number of features or clusters can vary widely across regions, a distance-based threshold provides a more suitable approach, as it directly reflects the maximum allowable geographic separation between grouped features. This is more intuitive and flexible when handling the variability of mining features.

To accurately cluster the mining data, the algorithm must also be adaptable to small and large clusters spreading over long distances. To that end, a hierarchical clustering with single linkage is used as a flexible grouping engine. Single linkage defines the distance between two clusters as the minimum distance between any single point/polygon in one cluster and any single point/polygon in the other cluster. This property is suited to overcome the spatial mismatches and relational ambiguities inherent in the integrated mining data. Mining operations often do not form compact, uniform shapes; instead, a single mine property (represented by a single inventory point) can have several disconnected infrastructures (pits, tailings, or waste dumps) that are dispersed several kilometres apart. Single linkage produces the desirable chaining effect, which is essential here to allow the algorithm to bridge these spatial gaps between sparse geographic features, ensuring that all remotely sensed infrastructure (polygons) belonging to a vast, fragmented, or elongated operation is grouped into a single cluster.

2.3.3. *Optimal distance threshold*

Although effective for linking scattered infrastructure, the chaining effect inherent in the single linkage hierarchical clustering method can lead to incorrect groupings if not carefully managed. Mining features exhibit complex spatial distributions: on the one hand, mine infrastructure often forms sparse, chain-like formations that extend over vast regions, necessitating a large distance threshold to link all components. On the other hand, the landscape can contain adjacent but unrelated geographical features belonging to independent mining operations that extract different materials, requiring a small threshold to keep them separate. A threshold that is too small prevents related features in a chain-like formation from clustering, leading to data fragmentation. Conversely, a threshold that is too large incorrectly groups unrelated features, introducing commodity ambiguity. The key task, therefore, is identifying the optimal cut on the dendrogram—the maximum distance at which clustering should stop.

To find this optimal grouping threshold, this study formalises the problem as a multi-objective optimisation: finding an optimal distance threshold for hierarchical clustering that balances two competing objectives:

- Minimising A_u, i.e. the total area of polygons that remain unassigned to any commodity (i.e., with an unknown primary commodity).
- Minimising A_m, i.e. the area of polygons assigned to multiple primary commodities.

The reasoning for using these two metrics rests on the spatial autocorrelation of geological occurrences and extracted minerals. This method acknowledges the reality that while related mine infrastructure can be spatially dispersed across long distances, proximity to an inventory point with known commodity data is critical for attribution. Therefore, minimising the unassigned area is essential to ensure the spatial completeness and linkage of all mine-related geographical features. Simultaneously, minimising multi-commodity assignment is necessary to maximise thematic precision and prevent the incorrect merging of distinct, unrelated operations, e.g. merging adjacent mine properties (points) that report different primary commodities.

This optimisation problem enables the construction of a Pareto efficiency frontier (Gunantara, 2018), which is used to identify the threshold distance that yields the best trade-off between reducing unlinked polygons and avoiding ambiguous multiple primary commodity assignments. A key component of this optimisation algorithm is the definition of the Pareto distance metric, $E(A_u, A_m)$, which mathematically defines the best trade-off between the two error types. The metric used in this study is given by

$$E(A_u, A_m) = \sqrt{((1-w)\cdot \mathrm{norm}(A_u))^2 + (w \cdot \mathrm{norm}(A_m))^2}; \quad 0 \le w \le 1 \quad (1)$$

Setting $w = 0.5$ is equivalent to a standard Euclidean distance that assigns equal importance to minimising unassigned area (A_u) and multi-assigned area (A_m). This balanced approach is ideal for a general-purpose dataset. However, the algorithm can be tuned to different research goals by adjusting the weighting of this metric. For applications where minimising multi-assignments is the primary concern, a weight $w < 0.5$ could be used to penalise multi-assignments more heavily, biasing the algorithm to select a more conservative optimal threshold. Since this study has a more general purpose, the parameter w is fixed at 0.5 for all experiments.

2.3.4. Group optimal distance threshold

As described in Section 2.3.1, the data is first partitioned into independent spatial groups. This partitioning strategy enables local optimisation, adapting to regional variations in geographic feature density and yielding more robust results than a single global threshold.

The core of the method is a two-fold optimisation. A meta-optimisation is applied to find the best value for the maximum limit. This stage systematically tests a range of different limits (e.g., 5 km, 6 km, 7 km...). For each tested limit, it runs the entire local optimisation for all groups and calculates the total aggregate error for the dataset. This process generates a sensitivity curve, and the meta-optimisation identifies the single limit (H^*_{max}) that provides the best trade-off between reducing unlinked features and avoiding ambiguous clusters on a global scale. Constrained by the maximum threshold H^*_{max}, the algorithm determines a locally-optimal distance cutoff by minimising a trade-off between unassigned features (A_u) and multi-commodity assignments (A_m) for each group.

The core of the method involves a **two-stage optimisation** process applied to find the best maximum distance limit (H^*_{max}) for clustering the entire dataset. The algorithm systematically tests a range of different maximum limits (e.g., 5 km, 6 km, 7 km), runs the local optimisation for all groups at each limit, and then calculates the total aggregate error for the entire dataset. This process generates a sensitivity curve, allowing the meta-optimisation to identify the single limit (H^*_{max}) that provides the best trade-off between reducing unlinked features (maximising completeness) and avoiding ambiguous clusters (maximising thematic precision) on a global scale. Constrained by different maximum thresholds H^*_{max}, the algorithm determines a locally optimal distance cutoff by minimising a trade-off between the unassigned area (A_u) and the multi-commodity assigned area (A_m) for each individual group. The complete logic of this algorithm is detailed in the pseudo-algorithm in Appendix A.1.

2.4. Test data sets

The above-described processing workflow is most effective when extensive data collections of points and polygons are available. Ideally, the combined data collections should provide quasi-complete coverage of known mining activities. Therefore, to test the approach, three sources of mine properties (points) were selected, including the mining production dataset from Jasansky et al. (2023), the Mine and Metals Database from S&P (2024), and the Global Coal Mine Tracker from Global Energy Monitor (2023), which were complemented by four sources for mining land use (polygons), included Maus et al. (2022a), Tang and Werner (2023a), OpenStreetMap contributors (2017), and an additional polygons data collection derived using the same methodology proposed in Maus et al. (2020) and Maus et al. (2022a). The Table 1 presents the complete list of data sets.

A supplementary data collection was conducted to complement existing data sources. It includes 1153 mining land-use polygons, without distinguishing different parts of the mine. No land use class is specified in the polygon, as they all belong to the same class "mining". The polygons were delineated using the same visual interpretation methodology described in Maus et al. (2020, 2022b) and the same baseline Sentinel 2 image mosaic from 2019. This dataset was included as an input dataset for the integration along with other datasets listed in Table 1. The new polygons are provided as a supplementary data file (see details in subsection Appendix A.3).

2.5. Validation

In addition to datasets listed in Table 1, a validation was conducted using an independent data source from Werner et al. (2020b). This data source provides detailed mine-specific information linked to polygons covering all their infrastructures delineated from high-resolution satellite images. The dataset comprises 295 mines and includes 8859 manually derived multi-polygon features (groups of one or more individual polygons). Since Werner et al. (2020b) data collection was carefully conducted to ascertain the association between a property and the sparse polygons covering its infrastructure, this source provides excellent grounds for validating the automated clustering approach. A set of standard similarity metrics (details in Appendix A.2) was used to compare the primary commodities assigned by the automated clustering with the primary commodities in the reference data.

2.6. Mine land footprints

This study also exemplifies the application of the resulting spatial clusters by presenting new insights on the global mine land use. The spatial clusters were assessed to provide the areal extent of mining-related land use by primary commodity, country, and terrestrial biome. For the latter, we used the Ecoregions 2017 classification (Dinerstein et al., 2017) to assign each mine cluster to a biome category based on spatial intersection. The results offer the first globally consistent estimates of mine land footprints disaggregated by commodity groups and ecological regions.

3. Results and discussion

Together, the integrated datasets after harmonisation comprise 259,999 geometries, including 42,799 points and 217,200 polygons,

Table 1
List of mining data integrated using the clustering methods.

Number of features	Type	Number of primary commodities	Source
1894	Points	28	Jasansky et al. (2022)
34,820	Points	38	S&P (2024)
6564	Points	1 (Coal)	Global Energy Monitor (2023)
44,929	Polygons	Not specified	Maus et al. (2022b)
74,548	Polygons	Not specified	Tang and Werner (2023b)
168,601	Polygons	Not specified	OpenStreetMap contributors (2017)
1153	Polygons	Not specified	Additional Polygons in Supplementary Materials

covering a total area of 145,738.1 km^2. The spatial distribution of these features is shown in Fig. 4a. The clustering procedure was applied to the complete harmonised dataset, with the distance threshold optimised for the entire world and locally for comparison. Fig. 4b illustrates the optimisation trade-off between reducing the area of unknown commodity (y-axis) and limiting the assignment of multiple commodities to the same cluster (x-axis).

Fig. 4b presents two Pareto fronts based on discrete threshold distances of 1 km steps. One curve seeks a single threshold applicable across all mine features worldwide (gray line), and the second curve presents the Pareto front of different maximum local thresholds used to constrain the optimisation of the threshold of each group of local geografic features. As we can see in the figure, this optimal global threshold was 7 km, which left 25% of the total mine area unassigned to any commodity and resulted in 39% of the area being associated with multiple commodities (i.e., multi-count area). The optimum maximum threshold for the local optimisation was also at 7 km approach, however, substantially reducing the multi-count area significantly to only 8%, while modestly increasing the unassigned area to 26.8%. For local optimisation, a separate Pareto front was constructed for each processing group–defined as a set of spatial features connected by an optimum maximum distance of 7 km. For interpretability, Fig. 4b presents only the aggregated results (red dashed line) obtained after applying the hierarchical clustering tree cut-off at each locally optimal threshold. These results confirm that accounting for spatial heterogeneity is essential to minimise ambiguity in commodity attribution.

For all subsequent land-use analyses and final reported statistics, including the global mine land footprints, the clustering results were generated using the local distance threshold optimisation. This approach was selected because it successfully adapted to the spatial heterogeneity of the mine features and significantly reduced the area assigned to multiple commodities (to 8%) while only modestly increasing the unassigned area (to 26.8%), demonstrating a more robust and thematically precise outcome compared to a single global threshold.

The convergence of both methods on 7 km as the critical distance is a noteworthy finding, suggesting this value represents a fundamental spatial scale for the specific integrated dataset. In contrast, the Local optimisation approach confirmed that allowing individual sub-clusters to seek connections beyond this threshold (e.g., 8 km or 9 km) begins to yield diminishing returns. Specifically, at distances greater than 7 km, the slight gain from resolving a few remaining unassigned polygons is outweighed by the penalty of creating many more ambiguous multi-primary commodity assignments, which worsens the overall result. Thus, even without a direct mathematical link between the global and aggregated local results, the findings from these two distinct optimisation approaches collectively validate 7 km as the critical distance threshold for the integrated data, beyond which the inferred connections between mining features are likely to become noisy, inaccurate, or overly complex. This contradicts the generally adopted 10 km distance threshold adopted in previous mine mapping studies (Maus et al., 2020, 2022b), which lacked a robust data-driven definition, such as then one presented in this work.

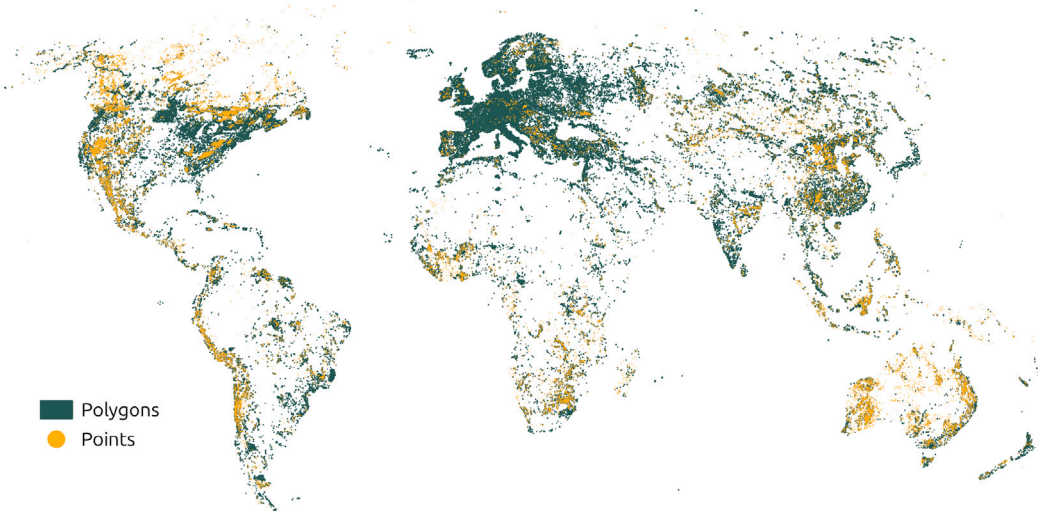
The superiority of the local approach, combined with meta-optimisation, is confirmed by the final distribution of the optimal thresholds chosen across all sub-clusters (Figure A.11). The distribution of these thresholds is strongly bimodal, with a median of 1.0 km and a mean of 2.7 km. This bimodality demonstrates that the method successfully adapted to two distinct spatial patterns: a prominent peak at 1 km indicates that the majority (at least 50%) of features are grouped into small, dense, and tightly clustered operations. Concurrently, a second peak at the 7 km maximum distance shows the method allowed more isolated features to expand and connect with their sparse, distant neighbours up to the defined limit. A rigid application of a global 7 km rule would fail by incorrectly forcing the numerous dense 1 km groups to merge, introducing unnecessary multi-commodity ambiguity, which is evident in Fig. 4. In contrast, the adaptive local approach successfully identified both dense and sparse spatial structures, validating its use over a single global threshold.

3.1. Mine clusters overview

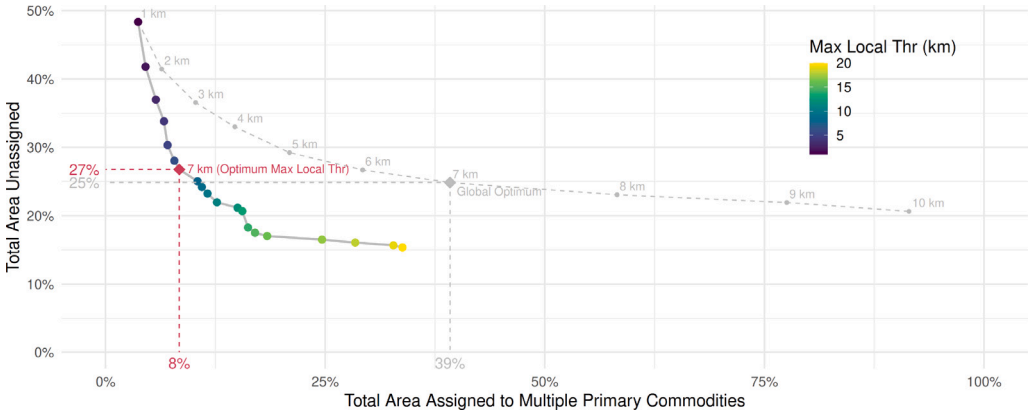
Table 2 summarises the clusters and commodity assignment, and Fig. 5 illustrates the spatial distribution of assigned and unassigned polygons. Of the total 145,738.1 km^2 of mine land, 106,733.4 km^2 (73.2%) could be assigned to at least one primary commodity. These assigned areas are concentrated in 28,453 clusters (26.4% of the total number of clusters), which comprise 86,196 polygons (39.7%) and nearly all points (99.9%). In contrast, unassigned areas account for 39,004.7 km^2 (26.8%), spanning 79,325 clusters (73.6%), 131,004 polygons (60.3%), and only 36 point features. Assigned clusters are on average larger (3.8 km^2) and composed of more spatial features (3.0 polygons and 1.5 points per cluster) compared to unassigned ones, which are typically smaller (0.5 km^2) and less complex (1.7 polygons and virtually no point features since when a point is included in a cluster, it will almost always provide the primary commodity). Moreover, the average area per polygon is markedly higher in the assigned group (1.2 km^2 versus 0.3 km^2). This contrast suggests that polygons with unassigned primary commodities are primarily small and isolated.

The unassigned land is mainly composed of small, isolated polygons that primarily represent land use from quarrying sites identified in the OpenStreetMap database and small-scale artisanal mining activities. Geographically, the unassigned areas are widely distributed but notably concentrated in regions well covered by OpenStreetMap. This pattern highlights persistent gaps in inventory data for small-scale and quarrying operations, underscoring that the lack of commodity attribution is primarily a data incompleteness issue rather than a fundamental flaw in the spatial integration methodology.

The breakdown of polygons with assigned and unassigned primary commodity by data source (Table 3) provides insights to explain the apparent imbalance in their profiling. For traceability and transparency, the table presents the polygons grouped by a combination of the original data sources. Since the integrated data has polygons merged from multiple sources, they may be derived from one or more data sources. Therefore, from Table 3 it is possible to identify the primary sources of unassigned polygons. Notably, polygons sourced from combinations of datasets tend to have higher assignment rates. For example, geometries derived from the integration of Maus et al. (2022a), Tang and Werner (2023a), and OpenStreetMap contributors (2017) jointly account for 49,144.0 km^2 of assigned area, even though they represent only 2.0%



(a)



(b)

Fig. 4. Spatial distribution of data points and clustering optimisation. (a) Integrated global mine land-use (green) and inventory point data (yellow). (b) Threshold optimisation for clustering based on commodity attribution. The gray curve indicates the Pareto frontier using a single global threshold, and the coloured curve indicates the Pareto frontier considering different local thresholds for each maximum local threshold. Robinson coordinate reference system. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

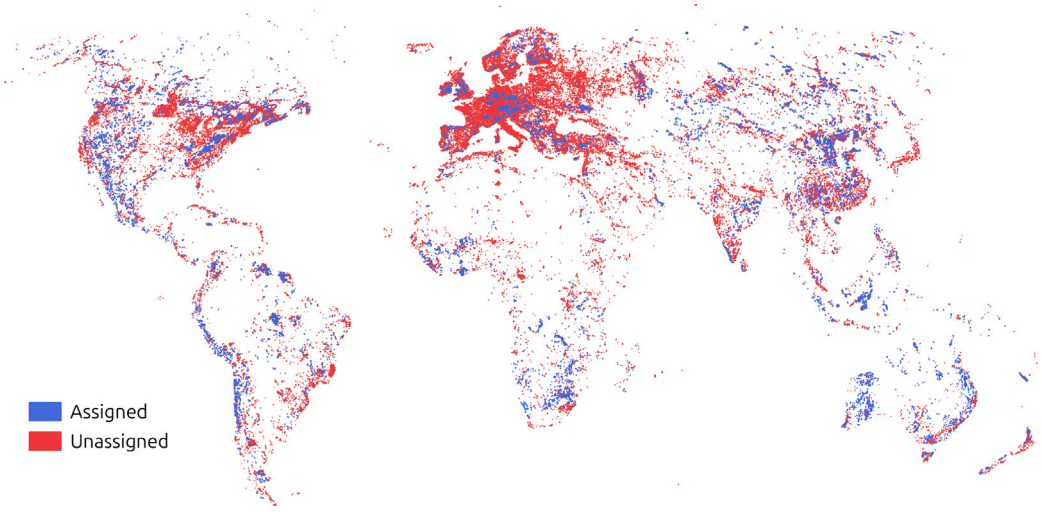


Fig. 5. Distribution of assigned and unassigned polygons worldwide. The coordinate reference system is Robinson.

Table 2
Summary of cluster-level features by commodity assignment.

Commodity	Total area (km²)	Number of			Average			
		Clusters	Polygons	Points	Cluster area (km²)	Polygons/cluster	Points/cluster	Area/polygon (km²)
Assigned	106,733.4 (73.2%)	28,453 (26.4%)	86,196 (39.7%)	42,763 (99.9%)	3.8	3.0	1.5	1.2
Unassigned	39,004.7 (26.8%)	79,325 (73.6%)	131,004 (60.3%)	36 (0.1%)	0.5	1.7	0.0	0.3

Table 3
Polygon counts and areas by commodity assignment and data source.

Status	Data source	Total area (km²)	Number of polygons	Average polygon area (km²)
Assigned	Maus/Tang/OSM	49,144.0 (33.7%)	4,371 (2.0%)	11.2
Assigned	Maus/Tang	30,361.1 (20.8%)	14,229 (6.6%)	2.1
Assigned	Maus	12,782.4 (8.8%)	14,936 (6.9%)	0.9
Assigned	Tang	5731.3 (3.9%)	26,416 (12.2%)	0.2
Assigned	Maus/OSM	5,096.7 (3.5%)	2,320 (1.1%)	2.2
Assigned	OSM	2550.4 (1.8%)	22,962 (10.6%)	0.1
Assigned	Tang/OSM	1,067.5 (0.7%)	962 (0.4%)	1.1
Unassigned	OSM	16,681.4 (11.4%)	113,651 (52.3%)	0.1
Unassigned	Maus	6,739.1 (4.6%)	5,117 (2.4%)	1.3
Unassigned	Maus/OSM	5450.5 (3.7%)	887 (0.4%)	6.1
Unassigned	Tang	3,780.8 (2.6%)	8,538 (3.9%)	0.4
Unassigned	Maus/Tang	2485.9 (1.7%)	1177 (0.5%)	2.1
Unassigned	Maus/Tang/OSM	1,951.3 (1.3%)	544 (0.3%)	3.6
Unassigned	Tang/OSM	1915.8 (1.3%)	1090 (0.5%)	1.8

of the total number of polygons. This suggests that areas commonly known for mining are well-represented in all land use data sources, as well as in the inventory data.

In contrast, OpenStreetMap contributors (2017) alone contributes the highest number of polygons with no overlap with other data sources, 22,962 assigned and 113,651 unassigned. These polygons are generally small (0.1 km² in average) and reflect the crowd-sourced nature of the dataset, with an emphasis on general land use, including mining and quarrying. This data source is also known to be biased towards high-income countries in the global north (Thebault-Spieker et al., 2018). Maus et al. (2022a) and Tang and Werner (2023a), in turn, were developed specifically for the impact assessment of the global mining sector with a focus on metals and coal. These sources integrated information from inventories compiled from company reports to identify and delineate mining infrastructures. However, even these expert-driven sources include a proportion of unassigned polygons, particularly in regions with small-scale and artisanal mining or illegal mining activities that may have been overlooked in the mapping process.

3.2. Examples of mine clusters

As an illustrative example of the optimisation strategy, Fig. 6 demonstrates the method's ability to prevent the incorrect merging of features associated with different commodities. In the Tarkwa-Nsuaem Municipal District in Ghana, gold and manganese mining features exist in close spatial proximity. Despite the short geographical distances between them, the local threshold optimisation successfully maintains distinct clusters, with each one correctly associated with a single primary commodity. This robust delineation prevents the introduction of mixed-commodity clusters, which would otherwise compromise the accuracy and utility of commodity-specific land-use statistics. The example further highlights the strength of the clustering approach in both capturing chain-like features (e.g., the red clusters) while simultaneously keeping self-contained features (like those in purple and blue) separate, thereby ensuring high thematic precision. An illustration of how the clusters progress in this region with increasing distance thresholds is shown in Figure A.12.

Fig. 7 presents two contrasting regions of East Kalimantan, Indonesia. The top map shows a large, contiguous cluster on the right-hand side, where coal is the only commodity reported in the underlying inventory data. This homogeneity allows the clustering algorithm to merge nearby features without introducing ambiguity in commodity assignment, resulting in relatively large and coherent clusters. In contrast, the bottom map displays smaller clusters in an area where both coal and gold are present. In this case, increasing the distance threshold to merging farther away features would have increased the area associated with multiple commodities, which the optimisation process is designed to avoid. As a result, in this area the clustering algorithm yields more fragmented but thematically consistent clusters.

3.3. Validation and limitations

A total of 747 polygons from the harmonised dataset intersect with mine sites delineated by Werner et al. (2020b). Among these, 40.3% exhibit a perfect match with the reference list of primary commodities, 53.9% show partial agreement (i.e., at least one primary commodity is shared between the assigned and reference sets), and only 5.8% show complete disagreement. These results indicate that the clustering-based commodity assignment aligns well with expert-based delineations. The mean value of the Identifier Overlap metric is 0.95, meaning that 95% of the evaluated polygon pairs have at least one commodity in common. This high level of overlap reinforces the robustness of the automated method, although some discrepancies are expected due to differences in how primary commodities are defined and reported across datasets.

Other similarity metrics in Table 4 also support this interpretation. The mean Jaccard Index of 0.65 and Dice Coefficient of 0.74 suggest substantial commodity overlap, even if not always exact. The Overlap Coefficient shows a mean of 0.92 and a median of 1.00, confirming that most clusters share at least one commodity with the reference data. The Jaccard median of 0.5 suggests that for at least half of the intersecting polygons, the set of assigned commodities shares at least 50% of their elements with the set of commodities in the reference. The higher mean than median Jaccard also reflects a right-skewed distribution: while most polygons have partial matches, many achieve high similarity close (Jaccard close to 1.0). These results collectively demonstrate strong agreement between independent data sources.

It is essential to emphasise that all datasets used here–both the harmonised and the reference–are subject to uncertainty coming from subjectivities in their data collection process. Therefore, these similarity metrics should not be interpreted as absolute measures of accuracy, but rather as an indicators of alignment between independently derived sources. High agreement increases confidence in the data, while disagreements signal where uncertainty remains. The example in Fig. 8

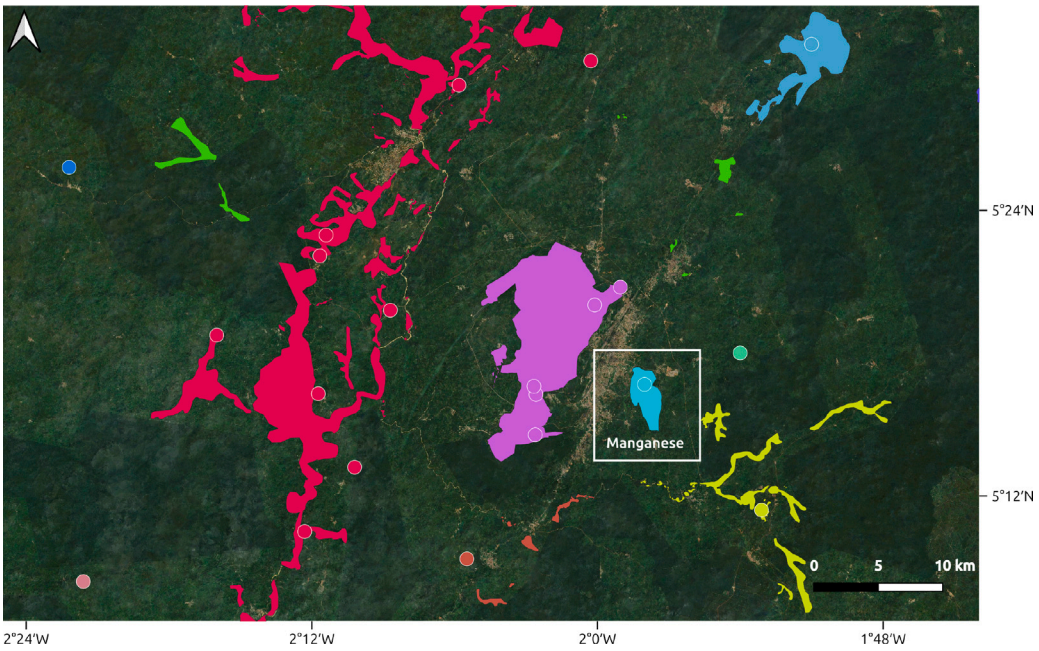


Fig. 6. Manganese mine correctly separated from the surrounding Gold mines in Tarkwa-Nsuaem Municipal District, Ghana. Colours indicate different clusters with distance threshold of 1 km. Image: Sentinel-2 cloudless https://s2maps.eu by EOX IT Services GmbH (Contains modified Copernicus Sentinel data 2019). (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

Table 4
Summary of similarity metrics and match statistics based on the assigned primary commodities and the reference dataset by Werner et al. (2020b).

Statistic	Similarity metrics				Matching counts		
	Jaccard	Dice	Overlap coefficient	ID Overlap	Full match	Partial match	No match
Mean	0.65	0.74	0.93	0.95			
SD	0.32	0.27	0.23	0.21			
Median	0.50	0.67	1.00	1.00			
Count					301	403	43
Proportion					40.3%	53.9%	5.8%

illustrates a specific case of disagreement between the automated data-driven commodity assignment and the reference data, offering insight into the sources of ambiguity in expert-labelled datasets.

In Fig. 8a, two distinct clusters are shown, with yellow polygons representing coal extraction and purple polygons associated with nickel and platinum mining. The clustering algorithm accurately distinguishes the coal complex from the adjacent metal mine, separating them based on their spatial distribution and reported commodities. In contrast, the reference dataset from Werner et al. (2020b) groups all polygons as part of a single nickel complex (Fig. 8b). Closer inspection of high-resolution satellite imagery (Figs. 8c and 8d) supports the automated classification: the pit on the left corresponds to coal mining, while the one on the right clearly exhibits characteristics of hard rock metal extraction. This example highlights a common pitfall in expert-based labelling, the risk of commission errors when nearby but unrelated mining infrastructures are aggregated under a single label. Such errors can lead to significant overestimation of the land footprint of specific commodities.

This finding emphasises a crucial point: both manual expert labelling and automated data-driven clustering must take into account a complete set of mines in the region of interest. Focusing exclusively on a single mine without considering the surrounding mining context and all co-located features can easily lead to commission errors and incorrect labelling, compromising the accuracy of the final land-use attribution. An accurate association between polygons and inventory sites requires completeness of data covering all mining activities in a region. Even if a spatial assessment targets only a subset of sites or commodities, omitting contextual information can compromise precision. This reinforces the value of scalable methods that integrate broader regional context in the delineation and classification of mine land use.

A limitation of this study is the temporal misalignment of the integrated datasets. The integrated data sources span several years. Polygons, in particular, have been derived from a 2019 image mosaic Maus et al. (2022b), and from images acquired in multiple years in the case of Tang and Werner (2023a) and contributors (2017). This integration assumes that the land-use polygons and inventory points are roughly contemporaneous, which may introduce errors, for example, by linking a new mine (point) to an older land-use polygon, or vice-versa. However, it is reasonable to assume for a global-scale static footprint analysis that large-scale mining infrastructure persists over many years. Future work could leverage land-use and operational time-series data to provide a more precise assessment of the global mineral-specific land footprint.

Another important consideration is that small and isolated features tend to remain unassigned, as the optimisation process prioritises minimising multi-commodity ambiguity over forced grouping. This trade-off results in a higher proportion of unassigned areas (A_u), which are primarily composed of small, fragmented polygons. Crucially, these unassigned features are typically located far from any recorded inventory point; consequently, assigning them would require a substantial increase in the distance threshold, which would inevitably lead to unrealistic associations and false positives. The Pareto front can visually confirm this in Fig. 4, which illustrates that even considerable threshold distances (e.g., 10 km) would still keep nearly 20% of the area unassigned. The prevalence of unassigned small polygons, therefore, stems primarily from the incompleteness of the inventory data rather than a limitation of the clustering threshold.

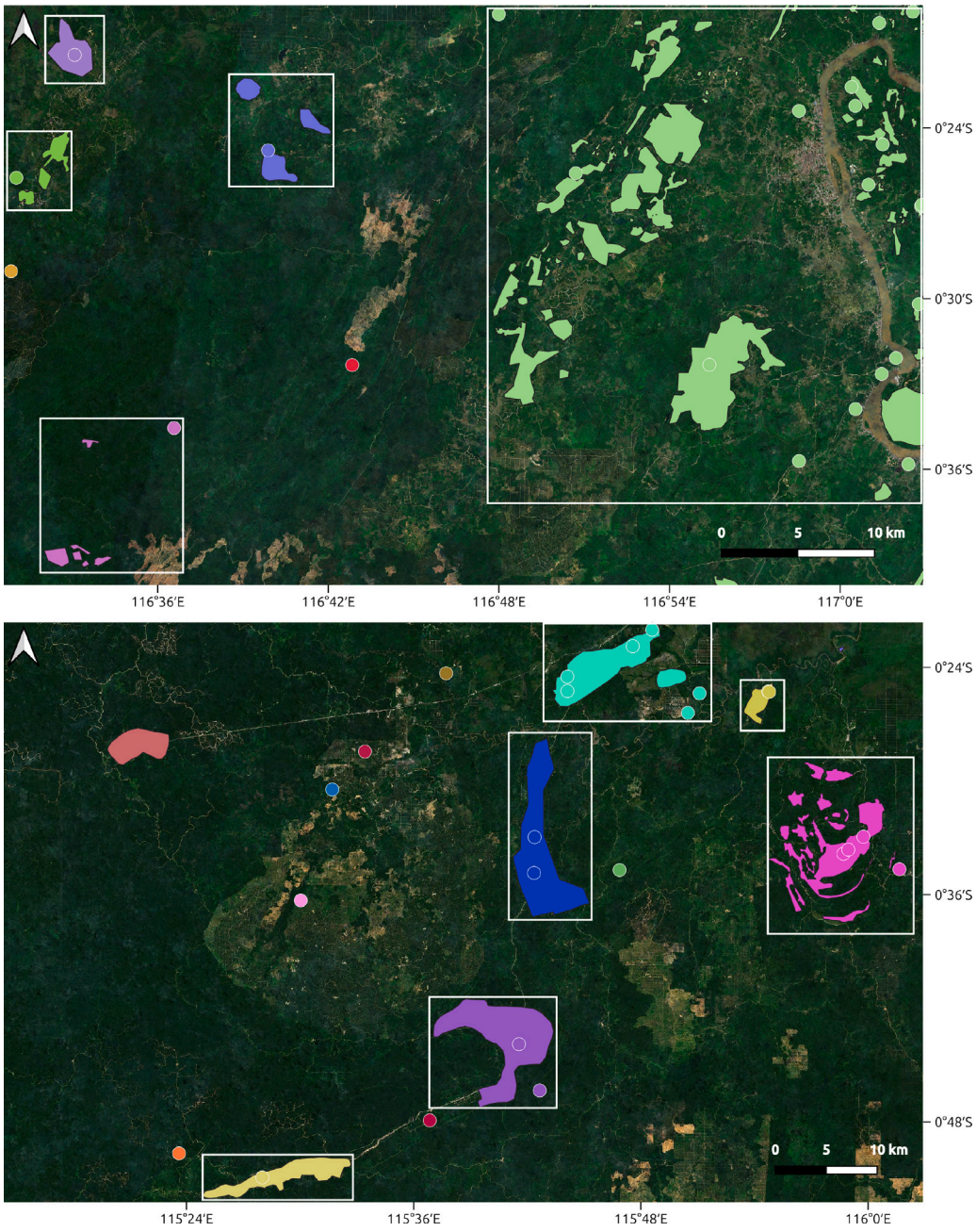


Fig. 7. Illustration of mining clusters in East Kalimantan, Indonesia. The map on the top shows larger coal mining clusters in the Kutai Kartanegara Regency, and the map on the bottom shows smaller clusters with coal and gold extraction in the West Kutai Regency. Colours indicate different clusters. Image: Sentinel-2 cloudless https://s2maps.eu by EOX IT Services GmbH (Contains modified Copernicus Sentinel data 2019). (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

A further limitation involves the level of data aggregation inherent in the clustering process. By grouping polygons and inventory points into unified spatial units, the reported operational values are aggregated to the cluster level, leaving the specific relationships between individual land-use polygons (e.g., specific pits or waste facilities) and distinct inventory entries unresolved. Resolving these precise one-to-one associations is impractical with currently available global datasets and would require extensive field knowledge to disentangle complex infrastructure. However, since the optimisation algorithm is explicitly designed to minimise the mixing of different primary commodities within a group, the thematic consistency of the clusters is preserved. While the internal granular relationships remain generalised, the environmental impacts associated with the total area of a cluster can be confidently attributed to the single primary commodity assigned to that spatial unit.

3.4. Mine land footprint overview

Fig. 9 presents an overview of mine land use disaggregated by primary commodity, country, and biome. The plot on the left shows the distribution of mine area by primary commodity type. Coal and gold each account for over 22.5% and 21.1% of the global mining footprint, respectively, while copper, iron, lithium, and phosphate represent smaller shares, ranging from about 6.6% to less than 2.0%. The spatial distribution of predominant commodities worldwide is illustrated in Fig. 10. Notably, 26.8% of the mining area was not assigned

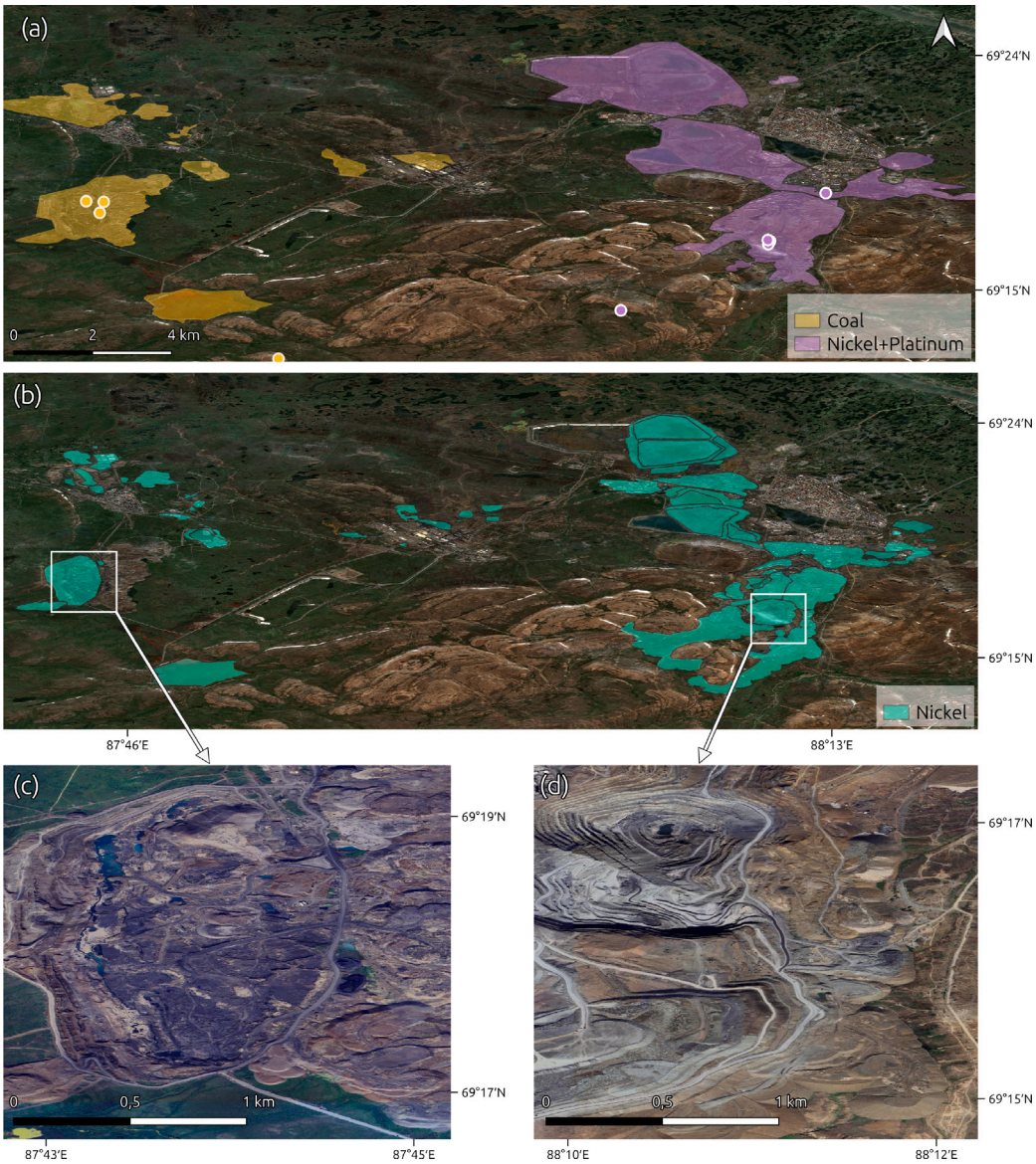


Fig. 8. Example of polygons with mismatching commodities. (a) Show two clusters with their respective assigned commodities, (b) polygons and respective primary commodity reported by Werner et al. (2020b), (c) a closer view of a coal pit, and (d) a closer view of a hard rock pit. Images: Sentinel-2 cloudless https://s2maps.eu by EOX IT Services GmbH (Contains modified Copernicus Sentinel data 2019).

to a specific commodity (dark gray areas in the map), emphasising the existing gaps in global mining data (Maus and Werner, 2024). The unassigned area is distributed over more than 131,000 relatively small polygons (see Table 2), which can be mainly attributed to small-scale artisanal mining, as well as the large number of quarrying sites identified in the OpenStreetMap database.

The top right panel in Fig. 9 shows the total mine area by country. The "Rest of World" group encompasses 24.6% of the total mining land footprint, followed by major mining economies such as Russia (12.1%), China (10.7%), and the United States (10.6%). Australia, Indonesia, Brazil, and Canada each contribute between 4.5% and 8.2%, while several countries in South America, Africa, and Asia account for smaller but regionally significant shares. Mining land use in the top-listed countries and the rest of the world is mainly associated with coal and gold. Mine land use in Chile, however, is primarily associated with Copper extraction. These results illustrate the geographic concentration of mining activity in a small number of countries.

The bottom right panel in Fig. 9 disaggregates mine land by biome, based on the Ecoregions 2017 classification (Dinerstein et al., 2017). Mining activities are most prevalent in tropical and subtropical moist broadleaf forests (22.7%), temperate broadleaf and mixed forests (16.5%), and deserts and Xeric shrublands (15.8%). Together, these three biomes account for over half of the global mining area. Mining is also widespread in temperate grassland and savanna ecosystems (10.4%), as well as boreal forests (9,9%). The primary commodity impacting tropical and subtropical moist broadleaf forests is gold, while temperate broadleaf and mixed forests are primarily affected by coal. Deserts and Xeric shrublands, in contrast, have a more heterogeneous distribution of mined land per commodity. These findings highlight that mine-driven land-use changes in each biome are associated with a distinct mineral commodity.

4. Conclusion

This study presents a scalable and replicable methodology for integrating heterogeneous mining datasets, addressing the challenge of spatial mismatches between mine inventories and land-use data. By

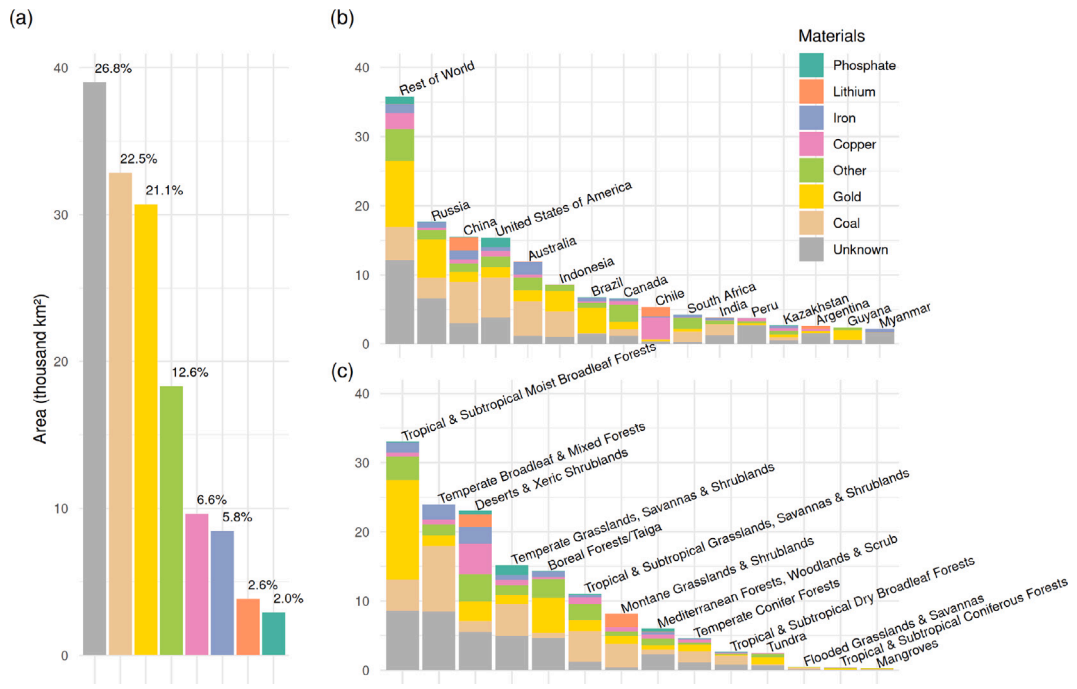


Fig. 9. Global mine land use profiles associated with primary commodities in different countries and biomes. (a) Shows the distribution of land associated with different commodities, (b) the distribution across the country, and (c) the distribution across biomes. The colours indicate the same commodity across all plots. Data available as a supplementary file (details in subsection Appendix A.3). (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

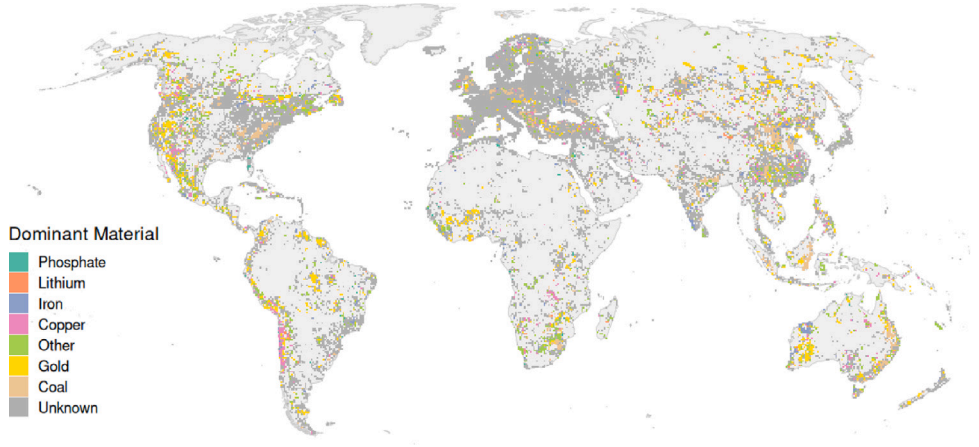


Fig. 10. Spatial distribution of material-specific mining land use. The colours indicate the material with the largest area within each 50 × 50 km grid cell. The coordinate reference system is Robinson. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

automatically constructing coherent *mine clusters*, the approach demonstrated accurate assignment of commodity information to georeferenced mine infrastructures. This is demonstrated by high validation agreement, with over 95% commodity overlap with expert-labelled data. While the clustering method is robust, its accuracy is inherently dependent on the quality and completeness of the input inventory and land-use datasets. The primary approach's significance lies in removing expert subjectivity from the data integration process, offering a consistent, data-driven approach for global analysis. Furthermore, the method also provides a flexible way to fine-tune the clustering to specific needs, as it remains stable for a wide range of maximum distance thresholds.

The application of this methodology produced the most comprehensive commodity-specific map of global mining land use to date, covering 145,738.1 km^2. The results reveal the dominant footprint of coal and gold mining and the concentration of impacts in a few countries and biodiversity-rich biomes, such as tropical forests. These findings underscore the need to integrate spatial and thematic data to enable more detailed environmental assessments of mineral extraction and advance understanding of mining impacts.

The significance of this data integration approach extends beyond commodity attribution. Production value, ownership, and other financial information of specific properties linked to polygons via clustering can enhance transparency of environmental impacts in the mining sector. While the inventory points used in this study provide site-specific commodity data, other economic indicators are often reported at the corporate or company level. Future attempts to integrate such aggregated variables must carefully account for these granularity mismatches. However, other operational information (e.g., production volumes, ore grades) is often reported at the site level, opening new avenues for analysis, such as land-use intensity relative to commodity

production. These will enable future applications to develop more precise, spatially explicit Life Cycle Inventory (LCI) intensity factors, thereby enabling spatially disaggregated Life Cycle Assessments (LCA) and supporting the development of science-based targets for nature by precisely attributing ecological pressures to specific mineral supply chains.

The resulting commodity-specific mine land use offers immediate utility for environmental governance. For instance, the finding that gold mining is the primary driver of land use in tropical moist forests highlights a critical intervention point for deforestation-free supply chain policies and biodiversity offsets. Furthermore, the substantial land footprint associated with coal extraction suggests that energy transition strategies could yield measurable co-benefits for land rehabilitation and ecosystem restoration. This further highlights that this data integration provides a critical foundation for more granular assessments of mining expansion, supply chain pressures, and sustainability assessments.

CRediT authorship contribution statement

Victor Maus: Writing – review & editing, Writing – original draft, Visualization, Validation, Software, Project administration, Methodology, Investigation, Funding acquisition, Formal analysis, Data curation, Conceptualization.

Declaration of Generative AI and AI-assisted technologies in the writing process

During the preparation of this work the author used ChatGPT and Grammarly in order to enhance readability and ensure coherent structure of the text. After using this tool/service, the author reviewed and edited the content as needed and takes full responsibility for the content of the publication.

Declaration of competing interest

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

Acknowledgements

Funded by the European Union. This work was supported by the European Union's Horizon Europe project RAWCLIC (grant agreement no. 101183654 https://doi.org/10.3030/101183654) and European Research Council (ERC) project MINE-THE-GAP (grant agreement no. 101170578 https://doi.org/10.3030/101170578).Views and opinions expressed are however those of the author(s) only and do not necessarily reflect those of the European Union or the European Research Council Executive Agency. Neither the European Union nor the granting authority can be held responsible for them.

Appendix A. Supplementary data

Supplementary material related to this article can be found online at https://doi.org/10.1016/j.jclepro.2025.147437.

Data availability

All input data sources can be obtained from open data repositories or purchased from proprietary data providers. The data inputs used in this work are subject to licenses that are incompatible with each other, which hinders the distribution of the mixed dataset produced in this study under a unified open license. Additional data collected and results not available elsewhere are provided as supplementary materials: Additional Polygons (*'s01-additional_polygons.gpkg'*) and Mine Land Use Accounting (*'s02-mine_area_accounting.csv'*). A global map of the mining land use by material is available for visualisation from https://maps.minethegap.eu. Other datasets will be made available on request.

The code developed in this study is openly available under the GPL-3.0 License from Maus (2025). The repository includes comprehensive documentation to support the reproducibility of results and the reuse of the method. A development version of the code is maintained on the GitHub repository https://github.com/vwmaus/mining-spatial-data-integration.

References

Bebbington, A., Williams, M., 2008. Water and mining conflicts in peru. MRD 28, 190–195. http://dx.doi.org/10.1659/mrd.1039.

Berthet, E., Lavalley, J., Anquetil-Deck, C., Ballesteros, F., Stadler, K., Soytas, U., Hauschild, M., Laurent, A., 2024. Assessing the social and environmental impacts of critical mineral supply chains for the energy transition in europe. Glob. Env. Chang. 86, 102841. http://dx.doi.org/10.1016/j.gloenvcha.2024.102841.

Cabernard, L., Pfister, S., 2022. Hotspots of mining-related biodiversity loss in global supply chains and the potential for reduction through renewable electricity. Environ. Sci. Technol. 56, 16357–16368. http://dx.doi.org/10.1021/acs.est.2c04003.

Conde, M., 2017. Resistance to mining. A review. Ecol. Econ. 132, 80–90. http://dx.doi.org/10.1016/j.ecolecon.2016.08.025.

contributors, O., 2017. Planet dump. Retrieved from {Data}filefrom{November}25, 2021, https://www.openstreetmap.org.

Crona, B., Parlato, G., Lade, S., Fetzer, I., Maus, V., 2023. Going beyond carbon: An "Earth system impact" score to better capture corporate and investment impacts on the earth system. J. Clean. Prod. 429, 139523. http://dx.doi.org/10.1016/j.jclepro.2023.139523.

Dinerstein, E., Olson, D., Joshi, A., Vynne, C., Burgess, N.D., Wikramanayake, E., Hahn, N., Palminteri, S., Hedao, P., Noss, R., Hansen, M., Locke, H., Ellis, E.C., Jones, B., Barber, C.V., Hayes, R., Kormos, C., Martin, V., Crist, E., Sechrest, W., Price, L., Baillie, J.E.M., Weeden, D., Suckling, K., Davis, C., Sizer, N., Moore, R., Thau, D., Birch, T., Potapov, P., Turubanova, S., Tyukavina, A., de Souza, N., Pintea, L., Brito, J.C., Llewellyn, O.A., Miller, A.G., Patzelt, A., Ghazanfar, S.A., Timberlake, J., Klöser, H., Shennan-Farpón, Y., Kindt, R., Lillesø, J.P.B., van Breugel, P., Graudal, L., Voge, M., Al-Shammari, K.F., Saleem, M., 2017. An ecoregion-based approach to protecting half the terrestrial realm. Biosci. 67, 534–545. http://dx.doi.org/10.1093/biosci/bix014.

Fonseca, A., McAllister, M.L., Fitzpatrick, P., 2014. Sustainability reporting among mining corporations: a constructive critique of the GRI approach. J. Clean. Prod. 84, 70–83. http://dx.doi.org/10.1016/j.jclepro.2012.11.050.

Giljum, S., Maus, V., Kuschnig, N., Luckeneder, S., Tost, M., Sonter, L.J., Bebbington, A.J., 2022. A pantropical assessment of deforestation caused by industrial mining. PNAS 119, http://dx.doi.org/10.1073/PNAS.2118273119.

Giljum, S., Maus, V., Sonter, L., Luckeneder, S., Werner, T., Lutter, S., Gershenzon, J., Cole, M.J., Siqueira-Gay, J., Bebbington, A., 2025. Metal mining is a global driver of environmental change. Nat. Rev. Earth Env. http://dx.doi.org/10.1038/s43017-025-00683-w.

Global Energy Monitor, 2023. Global coal mine tracker. https://globalenergymonitor.org/projects/global-coal-mine-tracker/.

Gunantara, N., 2018. A review of multi-objective optimization: Methods and its applications. Cogent Eng. 5, 1502242. http://dx.doi.org/10.1080/23311916.2018.1502242.

Guo, J., He, T., Xiao, W., Lei, K., 2024. Time series procession for monitoring land disturbance caused by surface coal mining in China. J. Clean. Prod. 448, 141585. http://dx.doi.org/10.1016/j.jclepro.2024.141585.

Jasansky, S., Lieber, M., Giljum, S., Maus, V., 2022. Open database on global coal and metal mine production. http://dx.doi.org/10.5281/zenodo.7369478, https://zenodo.org/records/7369478.

Jasansky, S., Lieber, M., Giljum, S., Maus, V., 2023. An open database on global coal and metal mine production. Sci. Data 10, 52. http://dx.doi.org/10.1038/s41597-023-01965-y.

Kobayashi, H., Watando, H., Kakimoto, M., 2014. A global extent site-level analysis of land cover and protected area overlap with mining activities as an indicator of biodiversity pressure. J. Clean. Prod. 84, 459–468. http://dx.doi.org/10.1016/J.JCLEPRO.2014.04.049.

Lagos, G., Peters, D., Videla, A., Jara, J.J., 2018. The effect of mine aging on the evolution of environmental footprint indicators in the Chilean copper mining industry 2001–2015. J. Clean. Prod. 174, 389–400. http://dx.doi.org/10.1016/j.jclepro.2017.10.290.

Lèbre, E., Sharma, V., Corzo-Remigio, A., 2024. Extracting minerals for the energy transition – local data for global decision making. J. Clean. Prod. 143563. http://dx.doi.org/10.1016/j.jclepro.2024.143563.

Luckeneder, S., Giljum, S., Schaffartzik, A., Maus, V., Tost, M., 2021. Surge in global metal mining threatens vulnerable ecosystems. Glob. Env. Chang. 69, 102303. http://dx.doi.org/10.1016/j.gloenvcha.2021.102303.

Maus, V., 2025. Spatial data integration for mine land use analysis (version v2.0). http://dx.doi.org/10.5281/zenodo.15611392.

Maus, V., da Silva, D.M., Gutschlhofer, J., da Rosa, R., Giljum, S., Gass, S.L.B., Luckeneder, S., Lieber, M., McCallum, I., 2022b. Global-scale mining polygons (version 2). http://dx.doi.org/10.1594/PANGAEA.942325.

Maus, V., Giljum, S., da Silva, D.M., Gutschlhofer, J., da Rosa, R.P., Luckeneder, S., Gass, S.L.B., Lieber, M., McCallum, I., 2022a. An update on global mining land use. Sci. Data 9, 433. http://dx.doi.org/10.1038/s41597-022-01547-4.

Maus, V., Giljum, S., Gutschlhofer, J., da Silva, D.M., Probst, M., Gass, S.L., Luckeneder, S., Lieber, M., McCallum, I., 2020. A global-scale data set of mining areas. Sci. Data 7, 1–13. http://dx.doi.org/10.1038/s41597-020-00624-w.

Maus, V., Werner, T.T., 2024. Impacts for half of the world's mining areas are undocumented. Nat. 625, 26–29. http://dx.doi.org/10.1038/d41586-023-04090-3.

Müllner, D., 2013. Fastcluster: Fast hierarchical, agglomerative clustering routines for R and python. J. Stat. Softw. 53, 1–18. http://dx.doi.org/10.18637/jss.v053.i09.

Murguía, D.I., Bringezu, S., Schaldach, R., 2016. Global direct pressures on biodiversity by large-scale metal mining: Spatial distribution and implications for conservation. J. Env. Manag. 180, 409–420. http://dx.doi.org/10.1016/j.jenvman.2016.05.040.

OECD, 2019. Global Material Resources Outlook to 2060: Economic Drivers and Environmental Consequences. Organisation for Economic Co-operation and Development, Paris, https://www.oecd-ilibrary.org/environment/global-material-resources-outlook-to-2060_9789264307452-en.

OpenStreetMap contributors, 2017. Planet dump. Retrieved from {Data}filefrom{November}25,2021, https://www.openstreetmap.org.

Owen, J.R., Kemp, D., Lechner, A.M., Harris, J., Zhang, R., Lebre, E., 2022. Energy transition minerals and their intersection with land-connected peoples. Nat. Sustain. 6, 203–211. http://dx.doi.org/10.1038/s41893-022-00994-6.

R Core Team, 2024. R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna, Austria, https://www.R-project.org/.

Siqueira-Gay, J., Metzger, J.P., Sánchez, L.E., Sonter, L.J., 2022. Strategic planning to mitigate mining impacts on protected areas in the Brazilian Amazon. Nat. Sustain. 5, 853–860. http://dx.doi.org/10.1038/s41893-022-00921-9.

Siqueira-Gay, J., Sonter, L.J., Sánchez, L.E., 2020. Exploring potential impacts of mining on forest loss and fragmentation within a biodiverse region of Brazil's northeastern Amazon. Resour. Policy 67, 101662. http://dx.doi.org/10.1016/j.resourpol.2020.101662.

Sonter, L.J., Ali, S.H., Watson, J.E.M., 2018. Mining and biodiversity: key issues and research needs in conservation science. Proc. R. Soc. B 285, 20181926. http://dx.doi.org/10.1098/rspb.2018.1926.

Sonter, L.J., Moran, C.J., Barrett, D.J., Soares-Filho, B.S., 2014. Processes of land use change in mining regions. J. Clean. Prod. 84, 494–501. http://dx.doi.org/10.1016/j.jclepro.2014.03.084.

S&P, 2024. SNL metals and mining database capital IQ pro. https://www.spglobal.com/marketintelligence/en/solutions/sp-capital-iq-pro.

Sun, X., Giljum, S., Maus, V., Schomberg, A., Zhang, S., You, F., 2025. Robust assessments of lithium mining impacts embodied in global supply chain require spatially explicit analyses. Environ. Sci. Technol. 59, 7081–7094. http://dx.doi.org/10.1021/acs.est.4c12749.

Tang, L., Werner, T.T., 2023a. Global mining footprint mapped from high-resolution satellite imagery. Commun. Earth Env. 4, 134. http://dx.doi.org/10.1038/s43247-023-00805-6.

Tang, L., Werner, T.T., 2023b. Global mining footprint mapped from high-resolution satellite imagery (version 2). http://dx.doi.org/10.5281/ZENODO.7894216.

Thebault-Spieker, J., Hecht, B., Terveen, L., 2018. Geographic biases are 'Born, not made': Exploring contributors' spatiotemporal behavior in OpenStreetMap. Proceedings of the 2018 ACM International Conference on Supporting Group Work. ACM, New York, NY, USA, pp. 71–82. http://dx.doi.org/10.1145/3148330.3148350.

Werner, T.T., Bach, P.M., Yellishetty, M., Amirpoorsaeed, F., Walsh, S., Miller, A., Roach, M., Schnapp, A., Solly, P., Tan, Y., Lewis, C., Hudson, E., Heberling, K., Richards, T., Chia, H.C., Truong, M., Gupta, T., Wu, X., 2020a. A geospatial database for effective mine rehabilitation in Australia. Min. 10, 745. http://dx.doi.org/10.3390/min10090745.

Werner, T.T., Mudd, G.M., Schipper, A.M., Huijbregts, M.A., Taneja, L., Northey, S.A., 2020b. Global-scale remote sensing of mine areas and analysis of factors explaining their extent. Glob. Env. Chang. 60, 102007. http://dx.doi.org/10.1016/j.gloenvcha.2019.102007.