Masked Mineral Modeling: Continent-Scale Mineral Prospecting via Geospatial Infilling

Sujay Nair[1,2*], **Evan Austen Coleman**[1*], **Sherrie Wang**[1], **Elsa Olivetti**[1]

[1]Massachusetts Institute of Technology,
[2]Georgia Institute of Technology
{sujaynr, ecol, sherwang, elsao}@mit.edu

Abstract

Minerals play a critical role in the advanced energy technologies necessary for decarbonization, but characterizing mineral deposits hidden underground remains costly and challenging. Inspired by recent progress in generative modeling, we develop a learning method which infers the locations of minerals by masking and infilling geospatial maps of resource availability. We demonstrate this technique using mineral data for the conterminous United States, and train performant models, with the best achieving Dice coefficients of 0.31 ± 0.01 and recalls of 0.22 ± 0.02 on test data at 1×1 mi^2 spatial resolution. One major advantage of our approach is that it can easily incorporate auxiliary data sources for prediction which may be more abundant than mineral data. We highlight the capabilities of our model by adding input layers derived from geophysical sources, along with a nation-wide ground survey of soils originally intended for agronomic purposes. We find that employing such auxiliary features can improve inference performance, while also enabling model evaluation in regions with no recorded minerals.

1 Introduction

Addressing climate change by transitioning the global economy away from its dependence on fossil energy sources is anticipated to require the mining of unprecedented quantities of raw material (Daehn, Allanore, and Olivetti 2024; The White House 2021; Franks, Keenan, and Hailu 2023). Many of these deposits of metals have only recently been characterized, with notable discoveries having even occurred within developed countries where thorough searches (also known as mineral prospecting) have already been performed (Wang et al. 2024b). The McDermitt Caldera on the border of Nevada and Oregon is a salient case: evidence suggests it may be the largest deposit of lithium-bearing rock in the world, but its potential magnitude was detailed in 2023 (Benson, Coble, and Dilles 2023). Motivated by this picture and the present need to identify and characterize mineral deposits around the globe, we seek to discover scalable approaches for mineral prospecting and subsurface mapping which are robust to the reality that existing records can be incomplete.

Meanwhile, powerful data-driven techniques to infill missing information have recently been developed and applied successfully in other fields. These include the use of masked autoencoders (MAE) for restoring patches of otherwise complete images (He et al. 2021), and applications of bidirectional encoder representations from transformers (BERT) to natural language generation (Devlin 2018). Motivated by these developments in generative modeling, we explore whether a similar paradigm can enable more powerful inference for mineral prospecting over large survey areas.

In this work, we train such models and find that infilling techniques can indeed recover missing or ablated mineral data. Notably, we observe that incumbent transformer-based architectures for infilling tasks (e.g. ViT) behave poorly out of the box in the subsurface infilling domain due to the highly sparse and spatially clustered nature of mineral deposits. To mitigate this, we propose a ResNet U-Net-based approach, which makes use of the strong spatial bias in mineral co-occurrence and achieves signficiantly higher performance than both ViT and classical geostatistical baselines.

As a bonus, the structure of our proposed model makes it easy to include auxiliary data. In fact, there are many datasets potentially worth applying to mineral inference which are far less sparse than mineralogical data. Over the past two decades, sizable aerial and ground-based efforts have developed substantial open-source datasets of terrestrial features (NRCS 2023; Orgiazzi et al. 2018; Kokaly et al. 2017; Qian 2021; Hajaj et al. 2024). Due to their different intended applications, these datasets can have drastically different spatial and depth coverages than those applied to traditional mineral prospecting. An example is hyperspectral imaging of field samples: geologists have recently used these advanced cameras to perform field-scale prospecting (Wan, Fan, and Jin 2021). However continent-scale datasets of these same measurements have already been collected over the US and Europe to model soil and ecological health (NRCS 2023; Orgiazzi et al. 2018). The growing number of individual such large-scale measurement campaigns presents an opportunity to fuse these diverse data sources and exploit variations in their coverage to accelerate the development, refinement, and testing of hypotheses regarding the subsurface resources accessible on Earth (Qian 2021).

In this work, we present 4 contributions. ***(1)*** We introduce **M**asked **M**ineral **M**odeling **(M3)**, a novel method and practical baseline for continent-scale mineral prospecting via geospatial infilling. ***(2)*** We perform a series of experiments

*These authors contributed equally.

on the proposed model, illustrating improved performance over both learning-based and classical prior work, and favorable scaling properties. ***(3)*** We explore conditioning the model on auxiliary data, finding significant performance improvements and extension of model capabilities, and lastly ***(4)*** We analyze the capabilites of the learned model and the implicitly learned relationships between different minerals.[1]

2 Related Work

Subsurface mapping. The geospatial infilling procedure studied in this work is a form of prospectivity mapping (Yousefi et al. 2021), also known as mineral abundance inversion (Chen et al. 2024), or as fossicking when done by hobbyists (Franks, Keenan, and Hailu 2023). This is a task of interest within the fields of geology, lithology, and mineralogy. It falls within the broader category of subsurface mapping and characterization, the modeling of below-ground resources based on sparse subsurface data combined with surface-accessible features (Misra, Li, and He 2019; Vohra et al. 2024). Remote sensing and data fusion have been applied with varying degrees of success to such challenges, for mineral resource analysis but also other climate-relevant applications, including geological CO_2 reservoir modeling (Liu and Grana 2020; Narayan et al. 2024; Wang et al. 2024a) and drilling viability for the extraction of geothermal energy (Wang et al. 2025). Many metrics are used to compare the performance of feature reconstruction, depending on the structure of the outputs. For supervised multi-class classification of lithological phenotypes, accuracy (overall and per-class) and Cohen's κ are common, e.g. (Dong et al. 2024). For regression of elemental content, metrics are standard to spectroscopy: mean squared error (MSE), bias, ratio of performance to deviation (RPD), and the R^2 of a linear fit between expected and predicted abundances, e.g. (Jiang et al. 2024). Focusing on ML-based mineral resource estimation, most studies prior to 2022 used MLPs or SVMs (60%), analyzed one mineral (84%), and relied on field-level data collection (88%) (Dumakor-Dupey and Arya 2021).

Data fusion for improving geospatial inference. Many geospatial features are frequently analyzed together in subsurface mapping, including ground-based magnetometry, ground-penetrating radar, and seismometry data, as well as satellite-based aeromagnetic and imaging spectroscopy data (Vohra et al. 2024). Due to its growing availability, applications of hyperspectral remote sensing and ground surveys to subsurface and mineral resource mapping have drawn recent interest within the literature (Hajaj et al. 2024; Wan, Fan, and Jin 2021). Hyperspectral satellites such as AVIRIS-NG provide more granular color information than traditional (multispectral) satellites like Landsat, measuring $\sim$100$\times$ more spectral bands ($\sim$1000 colors instead of $\sim$10) with spectral resolutions $\sim$10$\times$ finer (Qian 2021). While using such data requires processing a large number of input channels, it can support more performant models. Most recently, using feature reduction techniques and fusion of hyperspectral and multispectral remote sensing data, the authors of (Dong et al. 2024) showed that continent-scale lithological mapping (roughly, a precursor of the data we use as ground truth for M3) could be achieved remotely across the Tibetan plain with 97% accuracy and only 1% of the available training data, using ViT with dynamic graph convolutional layers.

Masked geospatial infilling of satellite imagery. There has been some recent interest within the literature regarding geospatial applications of MAE. Techniques and foundation models such as SatMAE and ScaleMAE have been developed which can perform spatial, spectral, and temporal infilling of satellite imagery (Cong et al. 2022), impute continuous surface-accessible features such as land use/land cover (LULC), and extend geospatial correlations across distance scales (Reed et al. 2023; Tang et al. 2023). Due to the continuous nature of remotely sensed imagery, infilling tasks on satellite data fall well-within the domain of vanilla MAE designed for RGB images, and the performance of ViT architectures is strong in these settings. However, in this work we focus on sparse subsurface mineral readings, rather than surface observations and satellite imagery, and find empirically that ViT-based approaches struggle over our ResNet U-Net approach, as it leverages a spatial inductive bias which is well-suited to the spatial clusters characteristic of mineral prospecting data.

This work: contributions and differences. Multiple aspects of the present study are unaddressed by existing literature. The application of subsurface mapping techniques to mineral availability on the scale of the conterminous United States (which makes up roughly 33% of the North American continent's surface) represents a significant increase in effective survey area, and an initial step away from site-level analysis toward Earth-scale prediction. We employ a hyperspectral ground survey dataset which was developed by the US Department of Agriculture for agronomic purposes, and which has not been used for any subsurface mapping tasks. Our problem specification is also distinct, since we perform multiple binary classifications instead of a single multi-class classification, to flag the locations of mineralogical resources over a masked region. This infilling technique is a generative approach similar to MAE, which has not been studied in this context (He et al. 2021). In addition, relative to previous works which infer lithological labels from geospatial features, the ground truth data we infill is more indicative of mining viability, i.e. it flags the locations of ore of some measurable grade (concentration) which could in principle be separated from waste rock and refined by a metallurgical process.[2] This change in interpretation has consequences for the choice of performance metrics appropriate to the method. For example, because the classes of interest are not mutually exclusive (two resource types can coexist), the κ score is not appropriate for model comparison. Overall accuracy is also problematic, as the mineralogical data we employ as ground-truth are an incomplete collection of true positives (Schweitzer 2019). Instead, we consider metrics appropriate to image segmentation and object detection which de-emphasize true negative detection. We use recall to track reproduction of true positives, and

[1]Appendices at github:sujaynr/M3 and arXiv:2511.09722.

[2]In this initial work, we do not differentiate resources by the economics of separation, although resource grades are logged.

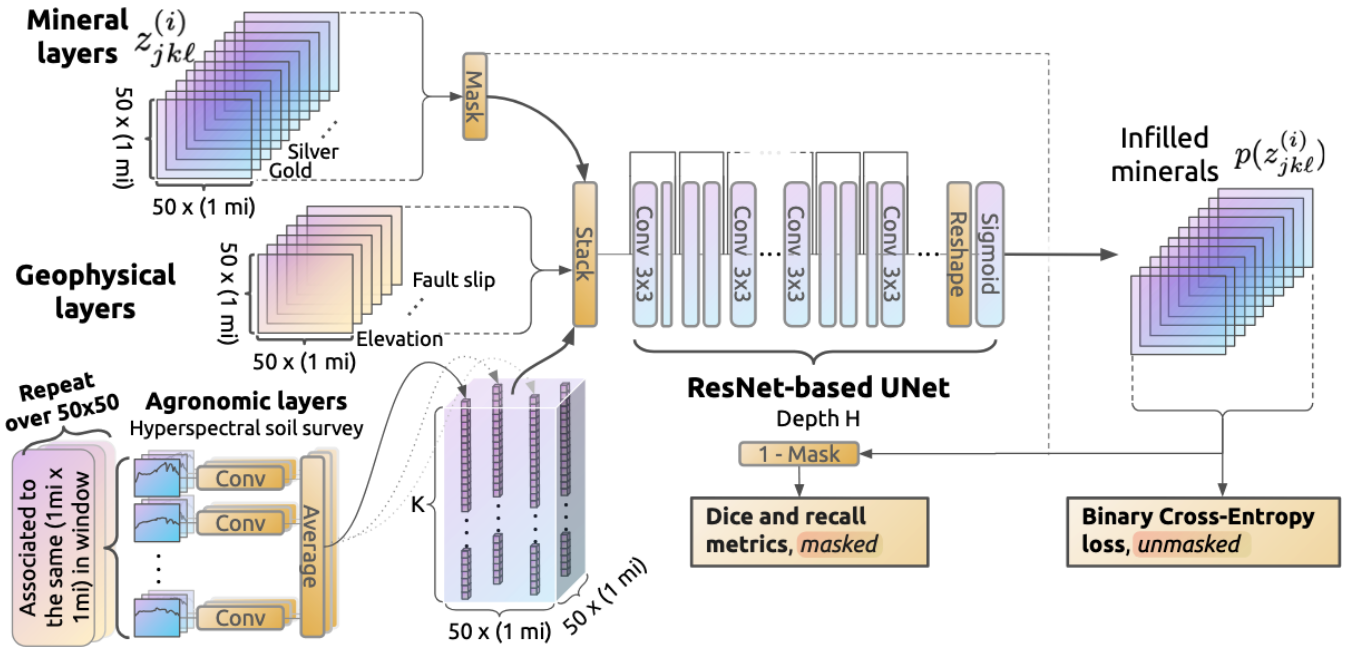

Figure 1: **General model architecture and data pipeline.** For a location (x_j, y_k) in the i^{th} (50 mi)×(50 mi) context window, the availability of resource ℓ is marked with a per-pixel binary flag $z^{(i)}_{jk\ell} \in \{0, 1\}$. The mineral data layers are masked, stacked with geophysical and agronomic inputs, and jointly passed through either a ResNet of depth H (shown) or a ViT. Masked minerals are infilled by the network, and the output is a per-pixel probability $p(z^{(i)}_{jk\ell})$ of the presence of resource ℓ. We train to maximize Dice coefficients computed over the masked region. Due to their combination of spatial sparsity and feature richness, the agronomic data for a given pixel pass through a feature reduction network f which maps them to vectors of length K at that location.

Dice coefficients to indicate rates of new resource prediction. Both metrics fall between 0 and 1. Larger values indicate greater agreement between predictions and ground-truth.

3 Data

We aggregate multiple geospatial data layers spanning the conterminous United States, grouped into three categories: mineral, geophysical, and agronomic.

Mineral layers. Mineral presence layers were collected from the United States Geological Survey (USGS) Mineral Resource Data System (MRDS) (McFaul et al. 2000). We extracted resource geolocations for the 10 most data-rich metallic ores; these metals are listed in Figure 2. We excluded all records labeled as processing plants (as opposed to occurrences, prospects, or producers), due to inconsistent documentation on the geographic origin of processed ore.

Geophysical layers. For this initial study, we included a minimal suite of 6 geophysical layers from a mix of USGS-based sources. Elevation data was pulled from the USGS National Map Elevation Point Query Service (Arundel et al. 2018). We used fault presence and slip rate maps from the USGS Quaternary Fault and Fold Database (USGS et al. 2025). Rock types and geological age ranges were derived from the USGS Geologic Map of North America (GMNA) (Garrity et al. 2009). The rock type classifications considered in the GMNA are derived from the mode of formation: sedimentary, metamorphic, igneous volcanic, igneous plutonic, and ice. We purposefully excluded the lithological classification layers of the GMNA to prevent indirectly passing mineral labels to the model and to maintain the simplicity of the geophysical inputs in our analysis, as these data are highly stratified but cannot be arranged along a single axis by a unique physical observable (e.g. geological age).

Agronomic layers. We use soil survey data from the USDA Rapid Carbon Assessment Project (RaCA) (NRCS 2023), a campaign which covered the conterminous United States and collected single-pixel hyperspectral images (scans) of over 120K soil samples using an ASD LabSpec spectrophotometer at 1 nm resolution within the Visible-Near Infrared (Vis-NIR) spectral window of 350-2500 nm. Prior to scanning, each sample was air-dried, ground, and sieved to particle sizes below 2 mm. A representative fraction of the samples was sent to a laboratory for elemental analysis (via dry combustion). We identified 23,591 spectra with complete records. We removed one outlier with 48% water content by mass, as well as 40 spectra having unlogged reflectances for one or more wavelengths. We removed wavelengths below 365 nm from consideration, due to their higher incidence of pixel failures. After applying all selection cuts, we obtain 23,550 soil spectra within a spectral window of 365-2500 nm, alongside corresponding geolocation data.

Context window datasets. To produce training and validation data for our analysis, we selected 100K square context windows of side length 50 mi at random from the conter-

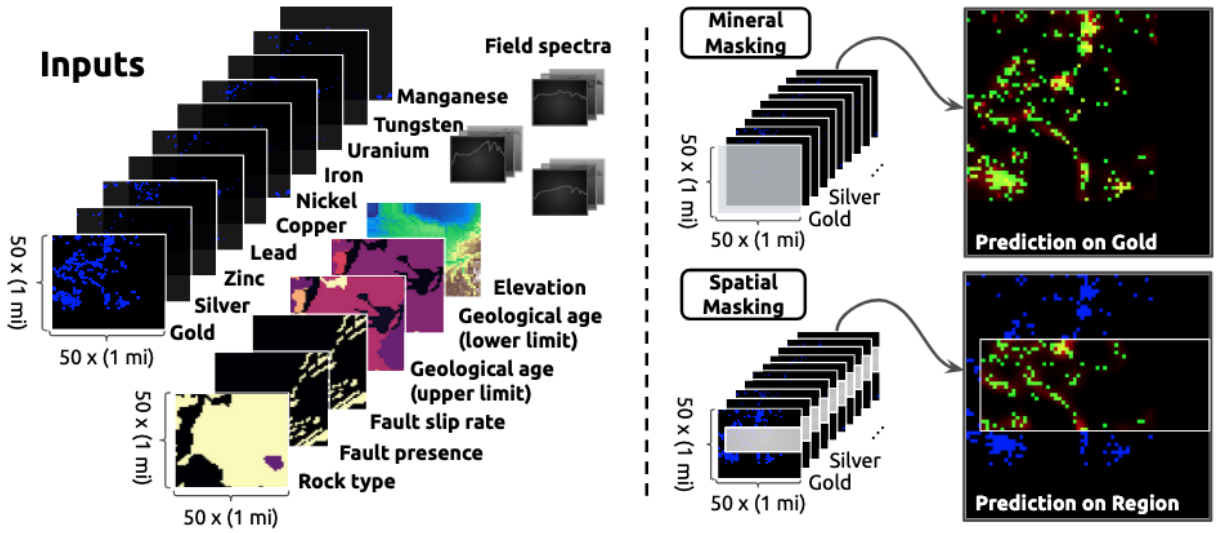

Figure 2: **Example input features and model outputs.** The displayed predictions are generated by the model trained in Experiment 5.3 with $N = 10$K, $H = 152$, and $K = 64$. The left panel shows all possible forms of input data for this context window. The first 10 layers consist of the availability of 10 minerals [Gold, Silver, Zinc, Lead, Copper, Nickel, Iron, Uranium, Tungsten, Manganese] on (50 mi)×(50 mi) regions. The next 6 layers consist of geophysical inputs [Rock type, Fault presence, Fault slip rate, Geological age (upper/lower limit), Elevation]. Finally, the agronomic data is embedded and included. The panel on the right shows the two possible masking strategies, mineral and spatial, and example predictions. Mineral masking removes an entire mineral layer, while spatial masking removes a random rectangular region. The blue points are ground truth outside of the masked region, the green points are ground truth within the masked region, and the red is prediction.

minous US, and all data appropriate to that region were aggregated. Test data were prepared similarly under both in-distribution (ID) and out-of-distribution (OOD) configurations. In the in-distribution case, we pulled a single fixed set of 5K windows at random from the US. In the OOD case, we pulled 447 windows from a 300 mi square region centered on the McDermitt Caldera, and used as a validation set 417 windows from within the 50 mi square annulus surrounding it. To allow for the development of a sensible model baseline which used only mineral layers as inputs, it was required that at least one resource be present in each sampled context window. The scale of the spatial context in pixels is chosen as $2\times$ the optimal value determined in (Dong et al. 2024), so that the reduction of features due to masking still leaves sufficient information for inference. The spatial resolution was 1 mi, set by the minimum viable resolution of our choice of digital elevation model (Arundel et al. 2018). We created a separate dataset for visualization, which included context windows spanning the conterminous US scanning in 50 mi increments along lines of constant latitude and longitude.

4 Method

The models trained in this work all use the full set of mineral layers $\mathcal{M}$ as a subset of their inputs, and perform an infilling task over these layers. For a given longitude and latitude (x_j, y_k) in the i^{th} context window $(\mathcal{X}_i, \mathcal{Y}_i)$, the availability of resource ℓ is marked with a per-pixel, per-mineral binary class $z^{(i)}_{jk\ell} \in \{0, 1\}$. The model output is a (50 mi)×(50 mi) per-pixel probability estimate $p(z^{(i)}_{jk\ell})$ of the presence or absence of a given mineral resource. Before prediction, mineral layers are masked in both mineral space and location. Masked values of z are represented as -1 in the inputs. The mask for mineral ℓ at location (x_j, y_k) in the i^{th} context window is denoted by $M^{(i)}_{jk\ell} \in \{0, 1\}$, with 1 indicating removal from the inputs. The mask emulates two behaviors that might realistically cause a mineral record to be missing: incomplete recordkeeping due to lack of interest in specific mineral species, and regional inaccessibility due to geographic obstacles. To account for these possibilities, we apply 2 kinds of masking with equal probability: mineral-only (i.e. mask is 1 for all ℓ in a subset of $\mathcal{M}$) and spatial-only (mask is 1 across all ℓ for all (x_j, y_k) in a rectangle). An aggressiveness hyperparameter A determines the maximum fraction of the mineral layers which can be masked (see Appendix F for its impact on performance). We sought a high, fixed $A < 1$ of 0.8 for all main results to induce a nontrivial inference task while avoiding training scenarios which represent hallucination in mineral-only inference, as doing so is expected to exaggerate the impacts of auxiliary features.

In our experiments, the geophysical and agronomic layers represent two optional sets of inputs. If selected, geophysical input layers are simply stacked with the masked mineral layers. The agronomic layers require a more sophisticated approach, as multiple hyperspectral scans can be associated to a single pixel in the context window (up to ∼30 vectors of length 2,135), and only up to 15 such pixels in a given 2500-pixel window have associated ground survey data. To avoid memory constraints, we apply a feature reduction f: each scan $R^{(i)}$ within a given longitude and latitude

(x_j, y_k) is passed through a 1D CNN with average pooling, ReLU activations, and a fully connected layer mapping it into $f(R^{(i)}(x_j, y_k))$, a vector of length K (see Appendix B). All outputs for a given site are averaged, and injected into a $50\times50\times K$ block at the associated geolocation (x_j, y_k). That block is then stacked with the other input data layers.

The combined input channels are passed through a ResNet architecture with varying depth H, passed through a sigmoid, and output as an array with the same shape as the mineral layers. Predictions $\hat{z}^{(i)}_{jk\ell}$ are set to $p(z^{(i)}_{jk\ell}) > 0.5$. We use the per-pixel, per-mineral binary cross-entropy as the loss. As an evaluation metric, we consider the Dice coefficient for all masked pixels having positive flags, i.e. $\{\hat{z}^{(i)}_{jk\ell} M^{(i)}_{jk\ell} = 1\}$ and $\{z^{(i)}_{jk\ell} M^{(i)}_{jk\ell} = 1\}$, see Figure 1.

5 Experiments

In our experiments, we seek to understand how well M3 compares to prior work and scalably leverages large data sources. In **Section 5.1**, we evaluate the approach against learning-based and classical prior work, trained only on mineral data. Then, in **Section 5.2** we study how inference performance scales as a function of the dataset and model sizes. In **Section 5.3**, we extract performance improvements by introducing auxiliary input layers. Lastly, in **Section 5.4** we probe the capabilities of M3, specifically its performance in scenarios where mineral records are nonexistent, and the dependencies of its predictions on the different mineral classes. We report seed-level standard errors of measurement alongside results.

5.1 Comparisons to Prior Work

In the first experiment, we compare to baselines for mineral-to-mineral inference and understand how model generalization varies based on size and inductive bias. We train a ViT architecture as a learning-based baseline (see Appendix C), a Gaussian Process (GP) as a geostatistical baseline (also known as *kriging*; Appendix D), and finally our proposed M3 ResNet architecture in Figure 1. We perform an initial grid sweep in model sizes and optimizer hyperparameters for a training dataset of size $N = 10$K instances (details in Appendix E). The models use masked mineral layers as inputs. We train each ML model for 100K gradient steps, on either a 90:10 train-validation split (for ID) or the OOD split described in Section 3. We repeat on 3 fixed random seeds in each configuration, and 7 seeds in the OOD ViT and M3 runs. We set masking aggressiveness to $A = 0.8$. We save checkpoints every 10K steps, and evaluate performance when the validation Dice coefficient asymptotes.

Figure 3 summarizes the performance statistics, demonstrating equivalent or superior generalization performance of the M3 architecture over both baselines. The GP outperforms M3 in training set performance, but sharply drops during validation and testing. This behavior is expected as kriging is the best linear unbiased predictor (BLUP) to interpolate features in the training distribution without straying from ground-truth, yet we explicitly added nonlinear effects through masking and geographic scale which violate kriging's assumptions of homoskedasticity, stationarity, and isotropy (Kleijnen 2017). However, any superior generalization performance of our approach over ViT is surprising, given that the attention mechanisms in transformers lend themselves to extraction of long-range correlations. Coarse-graining the inputs $10\times$ for the best hyperparameter combinations and retraining destroys this advantage. These findings suggest that in this setting the mineralogical features are a unique combination of locally clustered and sparsely sampled, so that the inductive bias of convolutional downsampling is more enabling than the versatility of geospatial attention. This argument is similarly supported by the poor performance of the GP, because that baseline hinges upon successful regression of a faithful model of the correlations between sampled points as a function of their relative distances. We propose that this phenomenon might be more widespread within subsurface mapping, where true positives for resources such as critical minerals, drilling locations for geothermal energy harvesting, and safe CO_2 injection sites may be rare or sparsely-sampled but locally-clustered due to a larger-scale feature situated deep within the Earth's crust (e.g. magmatic veins, viable caprock).

5.2 How Does Performance Scale with Data Quantity and Model Size?

Next we seek to understand how the M3 model scales over dataset and model sizes. In Figure 4 we sweep over ID training sets of size $N \in [1\text{K}, 10\text{K}, 100\text{K}]$ and model sizes from ResNet50 backbone to ResNet152. We see clearly favorable scaling properties and significant gains in test performance as the training dataset size grows. We also see benefit from increased model size, but interestingly only at the ResNet152 scale, where test performance is considerably higher than both ResNet50 and 101. Overall, these results suggest that as a method M3 has potential to continue improving in performance as more comprehensive datasets are assimilated and resources become available to train larger models.

5.3 Do Auxiliary Inputs Improve Performance?

To understand the potential benefits of including the geophysical and agronomic layers, in this experiment we train models ID from scratch for 30K gradient steps, progressively adding the auxiliary feature sets as inputs alongside the mineral data. Based on the results of Experiment 5.2 with $N = 10$K, we choose $H = 152$ and $K = 64$, and repeat on 3 fixed random seeds. Figure 5 quantifies the performance differences between these 2 approaches and the corresponding mineral-only baseline. We find significant training and generalization improvements from including geophysical data. The agronomic layers contribute a marginal boost in performance, likely due to their spatial sparsity, but did generally lead to better performance throughout the full training process.

5.4 Model Capabilities and Learned Features

In this section, we probe deeper into model capabilities. First, we study how the model predicts with no mineral input, only geophysical and agronomic data. Then we explore how mineral prediction depends on number and type of input minerals.

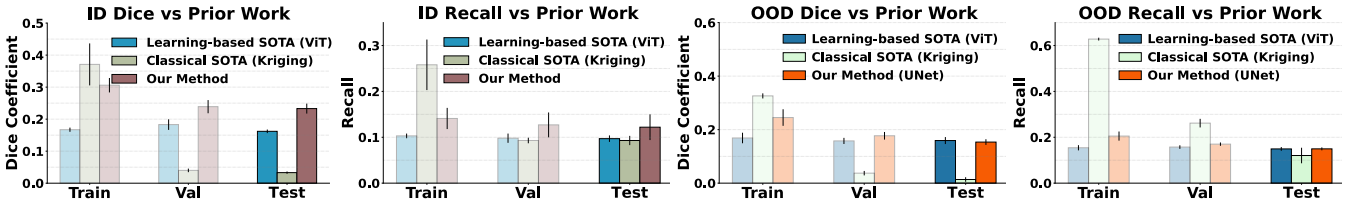

Figure 3: **Comparing M3 to prior work.** Our proposed method, M3, generally outperforms prior work from both learning-based approaches (ViT), and classical geostatistical approaches (Kriging). We observe that ViT struggles to fit the data as effectively due to a lack of spatial inductive bias, and that a ResNet-based approach is better-suited to inference over sparse clustered mineral deposits. Kriging fits well to training data due to its use of a Gaussian Process, but completely fails to generalize to unseen samples. These conclusions hold especially for ID (left) generalization. For OOD generalization (right), ViT required $5\times$ as many gradient steps to converge to equivalent validation and test performance, with significantly lower training performance. The OOD region is a 90K mi^2 square centered at the McDermitt Caldera, and the validation region is the surrounding 50-mi annulus.

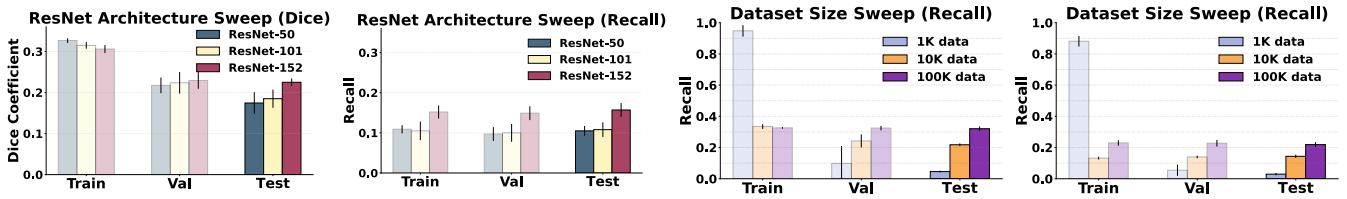

Figure 4: **Scaling properties of M3.** We study how the M3 model performance scales as a function of architecture depth H (**Left**) and dataset size N (**Right**). Dice coefficients (Top) over the masked regions, averaged over each mineral layer, and the recall on test data (**Bottom**) over the masked regions in each considered scenario. We observe favorable scaling, with improving performance with bigger models and more data.

Evaluation on empty mineral records A noteworthy consequence of including auxiliary features is that they provide more complete coverage. For models which only consider mineral layers as inputs, inference over regions with zero prospects is effectively hallucination, i.e. infilling a blank image. However, with auxiliary features, we can evaluate the model in regions not containing any minerals and transfer learned copresence features to new localities using these additional inputs. Figure 5 demonstrates this capability over the conterminous US, and plots locations where any resources are predicted when a model from Experiment 5.3 is evaluated 150 times on the visualization dataset (i.e. it infills 150 masks per context window in the dataset). The model is conservative in that it reproduces ground-truth without rampant hallucination in previously unseen regions. In 150 evaluations, it flags only 3,775 mi^2 of land in the 819,050 mi^2 region having no training data (0.46%). These sites include 1 iron and 3 zinc clusters on the mainland, 3 matching locations to existing sand, stone, and clay mines. (See Appendix H for details.)

Evaluation on systematically-removed mineral records In this experiment, we compute Dice coefficients one mineral population at a time, to understand whether minerals with more prospecting records have outsized impact on performance metrics. Using the best model in Experiment 5.1, and for each mineral layer, we evaluate on test data after masking all but 1 mineral layer (1:1 prediction). We then compute evaluation metrics for positive predictions of only that mineral layer, and repeat with 1 extra mineral unmasked. We continue this process, unmasking 1 layer at a time until 9 of the 10 mineral layers are implemented. Finally, we unmask the predicted layer, and evaluate one last time. With the exception of the predicted layer, which is always unmasked last, the unmasking sequence is in ascending order of the minerals' commonality in the training dataset, so that an increasing quantity of mineral data is added in with each unmasking step. Figure 6 shows the performance results on test data for all 10 runs. The vertical axis is ordered by the abundance of mineral data, with the top (gold) being the most common and the bottom (nickel) being the first unmasked layer.

We observe a general trend of increasing performance as we progressively unmask, with some minerals requiring fewer unmaskings to predict effectively. Interestingly, iron and uranium exhibited lower performance despite being the 4$^{\text{th}}$ and 6$^{\text{th}}$ most abundant layers. Though relative inference power suffers for these minerals, their addition as an input feature can seemingly boost inference for other species. An example is iron for silver: iron ties for second-to-last in scale-up, with a Dice coefficient of 0.22 after unblinding all other minerals, yet after unblinding it in the silver test it leads to a Dice coefficient boost of 0.10, taking third only to gold (0.11) and tungsten (0.13). In addition, the substantial inference power obtained after 9 of 10 layers are unmasked is surprising in the context of the literature, because machine learning-based prospectivity mapping has tended to analyze data from only one or two minerals at a time (48/51 or 94% before 2022, maximum of 4) (Dumakor-Dupey and Arya 2021; Leite and de Souza Filho 2009). Our result demonstrates that scaling the number of mineral inputs in M3 bolsters inference, and can be pivotal to achieving predictive power for minerals

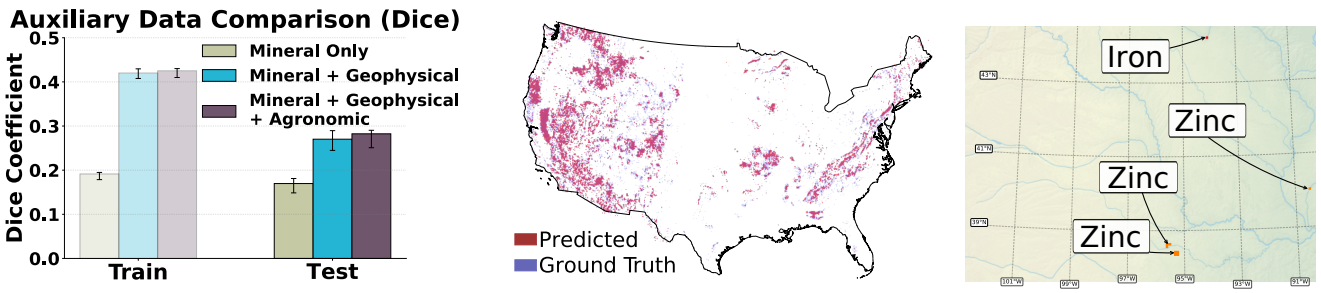

Figure 5: **Adding auxiliary data to M3 (Left).** Dice coefficients on training and test data as a function of input layer combinations ($N = 10$K, $A = 0.8$, $H = 152$, $K = 64$, 30K gradient steps). The inclusion of auxiliary data leads to faster training and significant generalization improvements, driven by geophysical data. **Visualizing M3 predictions over the continental US (Center).** An M3-generated map of resources in the conterminous United States (EPSG:5070). The model including agronomic and geophysical input layers in Experiment 5.3 was evaluated 150 times over a visualization dataset whose (50 mi)×(50 mi) tiles covered the United States. Model outputs from the masked regions (red) were summed over mineral layers, with ground truth overlaid in blue. **The model predicts 4 mainland clusters (Right)** of iron and zinc resources where no prospecting records on the top 10 minerals exist; 3 match existing sand, silica, stone, and clay mines. (See Appendix H for more detailed maps.)

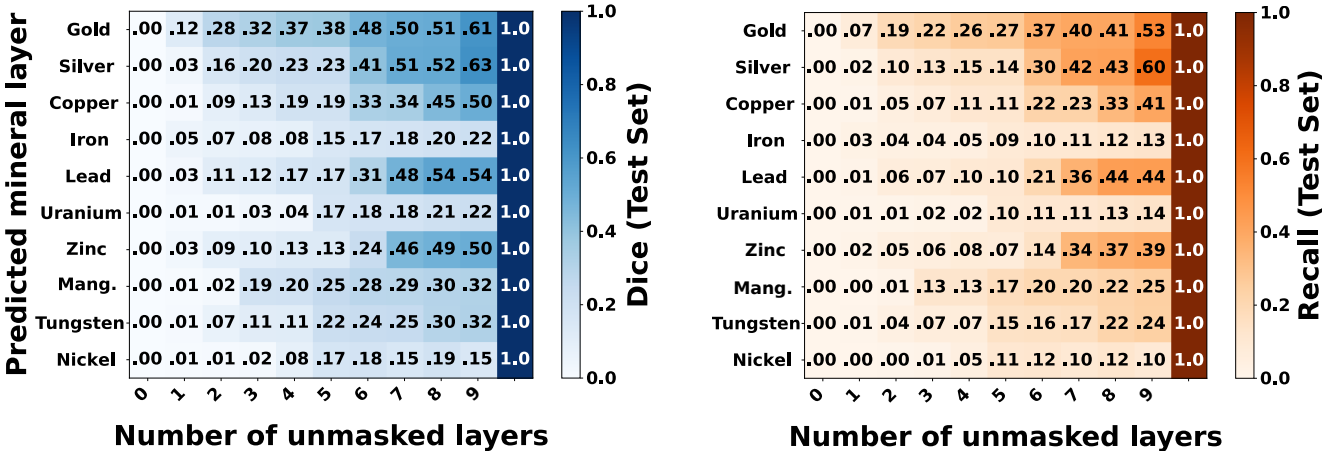

Figure 6: **Performance as a function of number of masked input minerals.** Matrix of single-mineral Dice coefficients and recalls. In each of 10 trials, we start from 1:1 mineral-only inference. We predict the mineral on the vertical axis from only one unmasked mineral layer, excluding the output layer's data until the final step. Unmasking occurs in order of increasing statistics, depicted as the vertical axis labels read from bottom to top. The model was trained on only mineral layers as inputs, with $H = 152$, $N = 100$K, and $A = 0.8$. We observe a general trend of increasing performance as we unmask, with some minerals requiring fewer unmasked minerals to predict effectively.

that have fewer prospecting records, as is the case for many minerals critical to addressing climate change (e.g. lithium and the platinum group metals). (See Appendix A for details.)

6 Future Work

We introduced Masked Mineral Modeling (M3), a method to infer locations of mineral resources by infilling masked geospatial maps of prospecting records. Using presence data for the 10 most abundant minerals in the conterminous US, we trained ResNet-backed infilling models, benchmarked their performance against both ML and geostatistical baselines, and demonstrated their superiority in this setting. Many avenues for investigation remain. Further scaling of both the number of input features and parameters is likely a straightforward way to improve generalization. It is also worth evaluating whether graph-based architectures extract more predictive power from sparse hyperspectral data in the agronomic layers. Applying satellite imagery and pretrained models such as SatMAE (Cong et al. 2022) to this problem are tantalizing possibilies. However, as mining activity is easily identified by satellite (Fonseca, Marshall, and Salama 2024), some effort is required to understand how these data might bias M3. Lastly, M3 can act as the main building block for a more sophisticated technique to reverse-engineer a masking procedure end-to-end, and determine when resources have been systematically ablated from records. (See Appendix G.)

Ethical Statement

The potential consequences of the presented approach are multi-fold, and span the full range of positive to negative impacts. As established above, addressing climate change will require unprecedented quantities of minerals, and the supply of high-grade resources is small relative to the anticipated demand (Olivetti and Cullen 2018; Daehn, Allanore, and Olivetti 2024). Inability to meet this demand is expected to exacerbate global inequality and inhibit developing regions and disadvantaged communities from accessing the materials, technology, and infrastructure required to survive, adapt to, and reverse the effects of climate change (Wang et al. 2024b). Meanwhile, mining as an industrial practice can have adverse impacts on communities (e.g. noise and dust from detonation, groundwater contamination, energy consumption, and more). Rapid acceleration of mining generally accompanies deregulation and lax permitting, which can exacerbate these negative impacts.

A particularly pressing sequence of events within this sector is unfolding in the present day: the implementation of nation-wide bans on the export of critical minerals could require affected nations to develop redundant stockpiles of those minerals, even though they are already rare in the absence of supply shock (Nassar et al. 2024). Inability to extract sufficient value from known sources has led to some speculation surrounding the viability of mining in new biomes, e.g. long-studied deposits on or under the sea floor (United States Department of the Interior 2025). Meanwhile, some evidence suggests that sizable deposits of relevant minerals have been missed by imperfect record-taking or incomplete prospecting, possibly because many of these metals have only gained economic value alongside the recent onset of climate change (Benson, Coble, and Dilles 2023). In addition, newer approaches to mining such as *in situ* recovery can leverage resources in locations and at depths not generally accessible via traditional methods. Such modern techniques generally reduce the environmental footprint of resource extraction by minimizing their intrusion into the Earth's surface. This collection of circumstances reflects multiple important reasons to re-evaluate existing prospecting records.

In this setting, M3 offers continent-scale transparency regarding the locations of mineral deposits. In turn, this knowledge can provide a clearer picture of the quantities of different resources which are readily available for extraction, can support the planning of a just energy transition (Wang et al. 2024b), and can potentially reduce the need to mine in pristine or nontraditional environments to meet the challenge of the global energy transition. Our method can also in principle be applied to subsurface mapping of sparse features more generally, e.g. for the identification of potential geothermal energy wells (Wang et al. 2025) and geological reservoir injection sites for CO_2 sequestration (Wang et al. 2024a).

A major technical challenge with M3 is hallucination. Unlike LLM hallucinations, which are generally straightforward to reject on an evidence basis, mineralogical hallucinations may be more difficult to distinguish from a genuine discovery without intensive ground-based data collection efforts. Consistent prediction of a resource excess across multiple models might reasonably motivate a direct sampling campaign, however another danger is the ignition of geopolitical tensions when new mine sites are flagged in regions governed by fragile states. Any unexpected outputs to M3 should be treated with significant skepticism until ground-based data collection has occurred.

References

Arundel, S.; Bulen, A.; Adkins, K.; Brown, R.; Lowe, A.; Mantey, K.; and Phillips, L. 2018. Assimilation of the National Elevation Dataset and launch of the 3D elevation program through the USGS spatial data infrastructure. *International Journal of Cartography*, 4(2): 129–150.

Benson, T. R.; Coble, M. A.; and Dilles, J. H. 2023. Hydrothermal enrichment of lithium in intracaldera illite-bearing claystones. *Science Advances*, 9(35): eadh8183.

Chen, W.; Li, X.; Qin, X.; and Wang, L. 2024. Hyperspectral Remote Sensing Inversion of Mineral Abundance Based on Sparse Unmixing Method. In *Remote Sensing Intelligent Interpretation for Geology: From Perspective of Geological Exploration*, 211–229. Springer.

Cong, Y.; Khanna, S.; Meng, C.; Liu, P.; Rozi, E.; He, Y.; Burke, M.; Lobell, D.; and Ermon, S. 2022. Satmae: Pre-training transformers for temporal and multi-spectral satellite imagery. *Advances in Neural Information Processing Systems*, 35: 197–211.

Cressie, N. 1990. The origins of kriging. *Mathematical geology*, 22: 239–252.

Daehn, K. E.; Allanore, A.; and Olivetti, E. A. 2024. A key feedback loop: building electricity infrastructure and electrifying metals production. *Philosophical Transactions A*, 382(2284): 20230234.

De Luca, G. B.; and Silverstein, E. 2022. Born-Infeld (BI) for AI: Energy-Conserving Descent (ECD) for Optimization. *PMLR*, 162: 4918.

Devlin, J. 2018. Bert: Pre-training of deep bidirectional transformers for language understanding. *arXiv preprint arXiv:1810.04805*.

Dong, Y.; Yang, Z.; Liu, Q.; Zuo, R.; and Wang, Z. 2024. Fusion of GaoFen-5 and Sentinel-2B data for lithological mapping using vision transformer dynamic graph convolutional network. *International Journal of Applied Earth Observation and Geoinformation*, 129: 103780.

Dosovitskiy, A. 2020. An image is worth 16x16 words: Transformers for image recognition at scale. *arXiv preprint arXiv:2010.11929*.

Dumakor-Dupey, N. K.; and Arya, S. 2021. Machine learning—a review of applications in mineral resource estimation. *Energies*, 14(14): 4079.

Fonseca, A.; Marshall, M. T.; and Salama, S. 2024. Enhanced detection of artisanal small-scale mining with spectral and textural segmentation of Landsat time series. *Remote Sensing*, 16(10): 1749.

Franks, D. M.; Keenan, J.; and Hailu, D. 2023. Mineral security essential to achieving the Sustainable Development Goals. *Nature Sustainability*, 6(1): 21–27.

Garrity, C. P.; et al. 2009. Database of the geologic map of North America-adapted from the map by JC reed, Jr. and others (2005).

Hajaj, S.; El Harti, A.; Pour, A. B.; Jellouli, A.; Adiri, Z.; and Hashim, M. 2024. A review on hyperspectral imagery application for lithological mapping and mineral prospecting: Machine learning techniques and future prospects. *Remote Sensing Applications: Society and Environment*, 101218.

He, K.; Chen, X.; Xie, S.; Li, Y.; Dollár, P.; and Girshick, R. 2021. Mae: Masked autoencoders are scalable vision learners. *arXiv preprint arXiv:2111.06377*.

Iakubovskii, P. 2019. Segmentation Models. https://github.com/qubvel/segmentation_models.

Jiang, G.; Chen, X.; Zhou, K.; Wang, J.; Zhou, S.; and Bai, Y. 2024. Estimation of lithium content in rock debris based on spectral feature coefficients. *Ore Geology Reviews*, 106167.

Kleijnen, J. P. 2017. Kriging: methods and applications.

Kokaly, R.; Clark, R.; Swayze, G.; Livo, K.; Hoefen, T.; Pearson, N.; Wise, R.; Benzel, W.; Lowers, H.; Driscoll, R.; et al. 2017. USGS Spectral Library version 7 data: US Geological Survey data release. *United States Geological Survey (USGS): Reston, VA, USA*, 61.

Leite, E. P.; and de Souza Filho, C. R. 2009. Artificial neural networks applied to mineral potential mapping for copper-gold mineralizations in the Carajás Mineral Province, Brazil. *Geophysical Prospecting*, 57(6): 1049–1065.

Liu, M.; and Grana, D. 2020. Petrophysical characterization of deep saline aquifers for CO2 storage using ensemble smoother and deep convolutional autoencoder. *Advances in Water Resources*, 142: 103634.

McFaul, E.; Mason, G.; Ferguson, W.; and Lipin, B. 2000. US Geological Survey mineral databases; MRDS and MAS/MILS. Technical report, US Geological Survey,.

Misra, S.; Li, H.; and He, J. 2019. *Machine learning for subsurface characterization*. Gulf Professional Publishing.

Narayan, S.; Kumar, V.; Mukherjee, B.; Sahoo, S.; and Pal, S. 2024. Machine learning assisted reservoir characterization for CO2 sequestration: A case study from the Penobscot field, Canada offshore. *Marine and Petroleum Geology*, 169: 107054.

Nassar, N. T.; Shojaeddini, E.; Alonso, E.; Jaskula, B.; and Tolcin, A. 2024. Quantifying potential effects of China's gallium and germanium export restrictions on the US economy. Technical report, US Geological Survey.

NRCS, U. 2023. Rapid Carbon Assessment (RaCA).

Oliver, M. A.; Webster, R.; et al. 2015. *Basic steps in geostatistics: the variogram and kriging*, volume 106. Springer.

Olivetti, E. A.; and Cullen, J. M. 2018. Toward a sustainable materials system. *Science*, 360(6396): 1396–1398.

Orgiazzi, A.; Ballabio, C.; Panagos, P.; Jones, A.; and Fernández-Ugalde, O. 2018. LUCAS Soil, the largest expandable soil dataset for Europe: a review. *European Journal of Soil Science*, 69(1): 140–153.

Poniewierski, J. 2019. Block model knowledge for mining engineers–an introduction. *Available at: Available at*, 1–33.

Qian, S.-E. 2021. Hyperspectral satellites, evolution, and development history. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 14: 7032–7056.

Reed, C. J.; Gupta, R.; Li, S.; Brockman, S.; Funk, C.; Clipp, B.; Keutzer, K.; Candido, S.; Uyttendaele, M.; and Darrell, T. 2023. Scale-mae: A scale-aware masked autoencoder for multiscale geospatial representation learning. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, 4088–4099.

Schweitzer, P. 2019. Record quality tables for the mineral resources data system. *US Geological Survey data release*.

Shepard, D. 1968. A two-dimensional interpolation function for irregularly-spaced data. In *Proceedings of the 1968 23rd ACM national conference*, 517–524.

Tang, M.; Cozma, A.; Georgiou, K.; and Qi, H. 2023. Cross-scale mae: A tale of multiscale exploitation in remote sensing. *Advances in Neural Information Processing Systems*, 36: 20054–20066.

The White House. 2021. Building Resilient Supply Chains, Revitalizing American Manufacturing, and Fostering Broad-based Growth. Accessed from The White House official website.

United States Department of the Interior. 2025. Interior Highlights Critical Mineral Prospects on the Seafloor. Accessed from DOI.gov official website.

USGS et al. 2025. Quaternary fault and fold database for the United States. Accessed January 15, 2025.

Vohra, M.; Gupta, A.; Umair, M. M.; Shukla, A.; Karunamurthy, J. V.; and Gupta, A. 2024. Automated Underground Mapping of Buried Utilities: A Review of Robotic Solutions and Sensor Technologies. In *2024 9th International Conference on Control and Robotics Engineering (ICCRE)*, 168–172. IEEE.

Wan, Y.-q.; Fan, Y.-h.; and Jin, M.-s. 2021. Application of hyperspectral remote sensing for supplementary investigation of polymetallic deposits in Huaniushan ore region, northwestern China. *Scientific Reports*, 11(1): 440.

Wang, H.; Williams-Stroud, S.; Crandall, D.; and Chen, C. 2024a. Machine learning and deep learning for mineralogy interpretation and CO2 saturation estimation in geological carbon Storage: A case study in the Illinois Basin. *Fuel*, 361: 130586.

Wang, P.; Yang, Y.-Y.; Heidrich, O.; Chen, L.-Y.; Chen, L.-H.; Fishman, T.; and Chen, W.-Q. 2024b. Regional rare-earth element supply and demand balanced with circular economy strategies. *Nature Geoscience*, 17(1): 94–102.

Wang, Y.; Zhang, X.; Qian, J.; Li, X.; Liu, Y.; Wu, W.; Lu, Z.; and Xie, B. 2025. Machine and deep learning-based prediction of potential geothermal areas in Hangjiahu Plain by integrating remote sensing data and GIS. *Energy*, 134370.

Wolfram Research. 2024. Mathematica, Version 14.2. Champaign, IL, 2024.

Yousefi, M.; Carranza, E. J. M.; Kreuzer, O. P.; Nykänen, V.; Hronsky, J. M.; and Mihalasky, M. J. 2021. Data analysis methods for prospectivity modelling as applied to mineral

exploration targeting: State-of-the-art and outlook. *Journal of Geochemical Exploration*, 229: 106839.