GeoCFM: Positive-Only Conditional Flow Matching for Mineral Occurrence Sampling

Moshe Eliasof[1]* and Eldad Haber[2]

[1] Faculty of Computer and Information Science, Ben-Gurion University of the Negev, Israel
eliasof@bgu.ac.il
[2] Department of Earth, Ocean and Atmospheric Sciences, University of British Columbia, Canada
eldadhaber@gmail.com

Abstract. Critical mineral discovery is a positive-only problem: deposits are observed as sparse locations, while unlabeled regions are not reliable negatives, and similar geophysical signatures can arise from different subsurface states. We therefore model mineral targeting as learning a conditional spatial distribution over occurrence locations, $\pi(p \mid d)$, given geo-images d, rather than predicting a deterministic per-pixel score map. We introduce GeoCFM, a conditional flow-matching model that generates mineral occurrence point sets conditioned on multi-channel geo-images; GeoCFM learns a point-wise transport field in $\mathbb{R}^2$, using UNet features with point-conditioned velocity prediction to bridge dense rasters and sparse supervision without pseudo-negatives. On a synthetic magnetics–geochemistry benchmark with latent activation and on USGS Earth MRI data with a spatially disjoint tile split, GeoCFM improves geometric agreement with observed occurrences over score-map and non-conditional baselines, while representing epistemic uncertainty through conditional sampling.

Keywords: positive-only learning · generative modeling · flow matching

1 Introduction

Mineral exploration is decision-making under extreme uncertainty. Economically viable mineralization is spatially rare, geologically complex, and only indirectly observed through noisy and heterogeneous measurements. In practice, exploration workflows fuse geological, geophysical, and geochemical data layers into mineral discovery maps, also known as mineral prospectivity, that rank targets [3, 4, 33]. Despite advances in sensing, modeling, and machine learning, the global discovery rate of significant mineral deposits has declined over the past decade from a late-2000s peak [25], while drilling remains expensive, irreversible, environmentally damaging and high risk [26].

* Corresponding author.

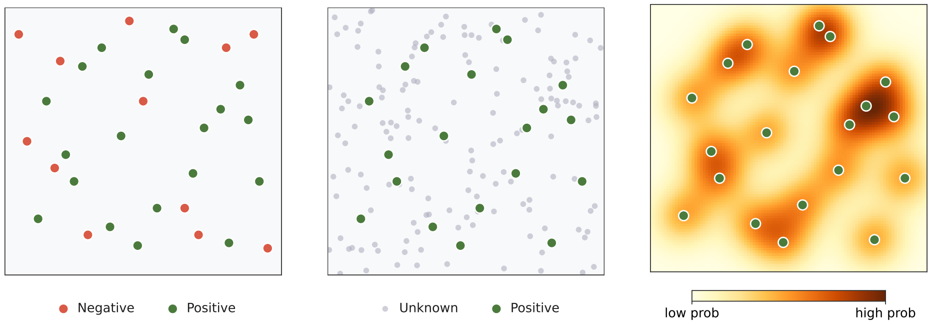

Fig. 1: Mineral prospectivity as a Positive-Only learning problem. Left: Standard supervised learning relies on both known positives and confirmed negatives to learn a decision boundary. Middle: The reality of mineral exploration is positive-only; we observe sparse positive deposits, but the remaining area is unknown/unlabeled (absence of a mine is not evidence of absence). Right: Our GeoCFM approach models a conditional density over mineral occurrences given geo-images, directly yielding a probabilistic mineral discovery map from positive samples alone.

A key difficulty is that mineral exploration is *not* a standard supervised learning problem. We typically observe only a limited set of labeled positives (known mines and mineral anomaly occurrences), but we do not observe trustworthy negatives: an unlabeled location may be untested, undiscovered, or simply uneconomic under current conditions. The supervision is therefore inherently *positive-only* (PO), and in practice *positive and unlabeled* (PU) when the remaining spatial search space is treated as unlabeled data [10, 18]. A common workaround in data-driven prospectivity mapping is to sample pseudo-negatives from unlabeled regions, but many such samples are not true negatives, which can systematically bias the learned decision boundary [34]. This issue is compounded by historical exploration bias: available labels and supporting datasets are concentrated in mature, historically targeted areas, so treating unlabeled regions as negatives amplifies uneven sampling of the search space [15]. We summarize this mismatch between the *true* supervision and the *assumed* supervision in Figure 1.

Beyond missing negatives, mineral exploration is dominated by epistemic uncertainty: the same surface and near-surface measurements can be consistent with *multiple plausible subsurface states*.

Formally, let d denote the observable geoscience measurements (geo-images), e.g., magnetic and gravity grids, geochemical assays, structural geological images, and other sensing modalities, as illustrated in Figure 2. These measurements provide only indirect and partial constraints on the subsurface. We treat the rest of the geological states as a latent variable $g \in \mathcal{G}$, where $\mathcal{G}$ is a high-dimensional geological manifold encoding the lithological structure, fluid pathways, the history of alteration and structural evolution of the Earth that are unobservable. Under this view, mineralization is modeled as the outcome of a mapping

$$p = f(d, g), \qquad g \in \mathcal{G}, \tag{1}$$

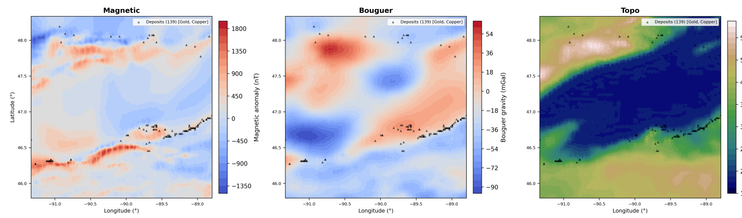

Fig. 2: Magnetics (left), gravity (middle), and topography (right) together with mineral (gold or copper) occurrences, i.e., deposits, in a geo-image over North America.

where p denotes the position of a mineral occurrence. Hence, the inverse mapping $d \mapsto p$ is inherently one-to-many. This non-identifiability induces epistemic uncertainty, and motivates probabilistic predictions rather than a deterministic label map.

Mineral occurrences can be viewed as sparse *keypoints* over geo-image inputs. The central question is not whether an anomaly "contains mineralization," but how to assign a *calibrated probability of mineralization* given the evidence and its uncertainties. Unlike dense prediction (e.g., semantic segmentation), where labels define an (approximately) deterministic pixel-to-class mapping, mineral discovery is one-to-many: similar sensor patterns may correspond to different latent subsurface states, and only a subset yields economic mineralization. The desired output is therefore *probabilistic*, enabling sampling over plausible subsurface outcomes. From a learning perspective, this shifts the objective from discriminative prediction to conditional density sampling. Rather than learning a classifier with *pseudo* negatives, we model a distribution of mineralization outcomes and sample multiple hypotheses consistent with the same evidence. Concretely, we represent deposits by their spatial coordinates p and learn a flow that samples $\pi(p \mid d)$ from positive occurrences alone. Sampling from $\pi(p \mid d)$ exposes epistemic uncertainty, supports uncertainty-aware target ranking, and avoids collapsing ambiguous signals into a single prediction.

In this work, we cast mineral discovery as probabilistic inference under positive-only supervision and materialize it with **GeoCFM**, a conditional flow-matching model that generates mineral occurrence point sets conditioned on geo-images. GeoCFM parameterizes the conditional velocity field with a shared MLP, guided by local features extracted from the geo-image by a UNet. Starting from a 2D Gaussian point cloud, the learned flow transports points into spatially resolved occurrence locations, without requiring any negative samples.

Contributions. Our contributions are:

1. We introduce GeoCFM – a positive-only, image-conditioned point sampler for mineral discovery. Our key contribution is not positive-only learning per se, but casting prospectivity as sampling from a conditional spatial density $\pi(p \mid d)$ that avoids explicit negatives while remaining a sampleable, multi-modal predictive object, instantiated via conditional flow matching with point-conditioned UNet features.

2. We propose an image-to-points architecture that conditions a point-wise velocity field on geo-images via local UNet features sampled at point locations, bridging dense rasters and sparse occurrence supervision without pseudo-negatives.
3. We provide a practical training and evaluation protocol for spatial geoscience data, including a multi-metric, repeated-sampling evaluation (Chamfer distance, Sinkhorn optimal transport, F@5, KDE negative log-likelihood, and top-5% hit rate) and a spatially disjoint remote-sensing split.
4. We show on a controlled synthetic benchmark and on real USGS Earth MRI data that sampling-based mineral discovery captures multi-modality and improves over a broad baseline suite spanning score-map, boosting, one-class, retrieval, and non-conditional methods.

2 Related Work

We briefly review prior work on (i) Mineral discovery/mineral prospectivity mapping, (ii) positive-only and PU learning, and (iii) conditional generative modeling via transport, which motivates our flow-matching formulation.

Mineral Discovery/Prospectivity mapping. Prospectivity mapping traditionally fuses heterogeneous evidence from geology, geophysics, and geochemistry into a spatial favorability or prospectivity score. Classic knowledge-driven workflows use expert rules and evidential weighting, while data-driven variants fit statistical or discriminative models to known occurrences, ranging from logistic and Bayesian-style formulations to random forests and neural networks [3,4,6,12,13,21]. Regardless of modeling choice, most pipelines output a single deterministic score map that is subsequently thresholded or ranked to prioritize follow-up. Since confirmed negatives are rarely available at scale, these methods often introduce pseudo-negatives by sampling from unlabeled regions. When supervision is in fact positive-only, pseudo-negatives can include undiscovered or untested positives and can be spatially biased by historical sampling, which distorts the learned decision rule and undermines calibration [34]. A recurring response is to avoid explicit negatives altogether: one-class formulations such as one-class SVMs estimate the support of the positive distribution [5], while more recent work explores self-supervised and explainability-oriented pipelines that target scalability and uncertainty in exploration [7]. These approaches still produce a per-pixel favorability score; in contrast, we model occurrences as samples from a conditional spatial density, which yields a sampleable, multi-modal predictive object rather than a single deterministic map. We compare against representative score-map, boosting, one-class, and retrieval baselines in Section 5.

Positive-only and PU learning. Learning from positive and unlabeled data (PU learning) is a well-studied weak-supervision setting in which the unlabeled set is a mixture of positives and negatives [1, 10, 18]. Core tools include risk estimators that avoid treating unlabeled samples as negatives, together with class-prior (mixture proportion) estimation to convert positive-conditional models into calibrated posteriors [9, 16, 23, 24]. Much of the PU literature is developed under assumptions such as positives being labeled completely at random, whereas many real domains exhibit selection bias in what becomes labeled [2]. Mineral prospectivity is an extreme case: surveying and drilling are preferentially concentrated in accessible or historically promising regions, so the unlabeled set is spatially biased and far from i.i.d., which can amplify the impact of pseudo-negative construction [8, 15, 22, 31]. Our approach bypasses these difficulties by modeling observed occurrences as samples from a conditional spatial density, avoiding explicit negative sampling and naturally accommodating the fact that similar geo-image evidence can correspond to different outcomes due to unobserved subsurface controls.

Generative models and flow matching. Recent progress in vision has been driven by conditional generative models, including diffusion and score-based models [14, 27]. Flow-based transport models provide an alternative view, generating samples by integrating an ODE defined by a learned vector field. Flow Matching [19] and Conditional Flow Matching (CFM) [28] train such vector fields using a stable regression objective. In contrast to standard CFM where labels are used to guide image generation, our method uses images, d, to guide point-flow: we generate sparse occurrence locations conditioned on images, in the spirit of point-based modules used in vision [17] and flow-based generation of point sets [32]. Flow matching has very recently been applied to geological generation as well: [20] use attention-guided flow matching for sparse 3D geological volume generation. This is related in spirit but distinct in problem setting from our work, as we condition on 2D geo-images to sample occurrence keypoints under positive-only supervision, rather than generating dense 3D geological volumes.

3 Mineral Discovery as a Learning Problem

We now formalize mineral discovery, dubbed prospectivity mapping, as conditional density estimation. Our goal is to turn the conceptual relation in Equation (1) into a tractable learning problem that models the distribution of mineral occurrences conditioned on observed geoscience data, without explicitly reconstructing the latent geological state.

Unobserved geology and non-identifiability. Recall Equation (1), where mineralization depends on both observed geoscientific measurements and latent geology. Let $d \in \mathbb{R}^{C \times H \times W}$ denote geo-images, the observed multi-channel tensor (e.g., magnetics, gravity, geochemistry, structural layers), and let $g \in \mathcal{G}$ represent geological context. Since g is not directly observed and is only weakly constrained by surface measurements, inferring p from d is ill-posed: multiple geological configurations can yield similar observations. We therefore treat g as a nuisance variable and formulate prospectivity probabilistically by marginalizing over latent geology.

Mineral Occurrences as spatial samples. Rather than discretizing mineralization into a dense binary label map, we represent each occurrence by its spatial coordinate. Let $p_k = [p_{k_1}, p_{k_2}] \in \mathbb{R}^2$ denote the position of an occurrence within an image. In practice, in each geo-image there is a set of coordinates $\{p_k\}_{k=1}^N$, where the number of mineral occurrences N varies across samples. This point-based representation matches the available supervision: occurrences are sparse, and unlabeled locations are not reliable negatives, as discussed in Section 1. This representation also decouples positional precision from image resolution, allowing point annotations to be aligned naturally with geo-image inputs. We model occurrences as samples from a *conditional spatial density*, $p \sim \pi(p \mid d)$, and the learning problem is to estimate $\pi(p \mid d)$ from data.

3.1 Implicit marginalization

We note that the latent geology $g \in \mathcal{G}$ is never observed, since it involves inferring about deep earth structures which we have no way to directly measure. Instead, we observe geo-images of size $C \times H \times W$, i.e., $\{d^{(b)}\}_{b=1}^B$, where $d^{(b)} \in \mathbb{R}^{H \times W}$ each paired with a set of occurrence coordinates $p^{(b)}$ where $p^{(b)} \in \mathbb{R}^{N_b \times 2}$. Conceptually, each geo-image is generated under an unknown geological context $g^{(b)}$, so the dataset forms a mixture over geological regimes.

Therefore, a generative description can be formulated as:

$$g \sim \pi(g), \quad d \mid g \sim \pi(d \mid g), \quad p \mid d, g \sim \pi(p \mid d, g), \tag{2}$$

and the conditional density of interest is the marginal

$$\pi(p \mid d) = \int \pi(p \mid d, g)\, \pi(g \mid d)\, dg, \tag{3}$$

obtained by integrating out the latent geology.

Crucially, since we do not have a way to sample g, we do not evaluate Equation (3) explicitly. Instead, we use a parametric model (flow matching model) to sample from $\pi(p \mid d)$. The flow matching model is averaged over all geo-images approximating the expectation over g.

3.2 Synthetic magnetic-geochemistry model problem

To make the role of unobserved geological controls concrete, we introduce a simplified synthetic problem inspired by porphyry-style mineral systems. The goal is conceptual rather than geologically exhaustive: the observable geo-images contain clear, interpretable structure, while occurrence locations depend on hidden variables that are not available during learning.

The observed magnetic anomaly map d over a domain $\Omega \subset \mathbb{R}^2$ (discretized on an $H \times W$ grid) contains multiple ring-like anomalies representing candidate intrusive centers, since magnetic responses often exhibit such structure around intrusions or alteration halos; this provides context but does not uniquely determine mineralization. We additionally generate latent geochemical fields g over the same domain, representing unobserved processes (e.g., fluid transport or enrichment) that determine whether a given structure is mineralized and where

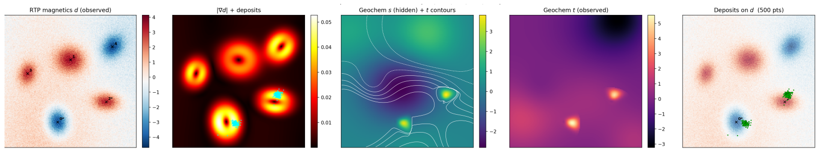

Fig. 3: Synthetic magnetics–geochemistry benchmark. From left to right: (1) observed RTP magnetic geo-image d; (2) contact-zone proxy $\|\nabla d\|$ with sampled deposit locations overlaid (cyan); (3) hidden geochemistry s (background) with contours of the observed proxy t, illustrating that t only partially reflects the latent control; (4) observed geochemistry proxy t; (5) the same deposit set overlaid on d (green, 500 points). Multiple magnetic rings are present, but deposits activate only a subset and localize along arcs, reflecting unobserved geochemical controls and motivating conditional density modeling of occurrence locations.

along it mineralization concentrates; the learner does *not* observe g. Occurrences are then sampled from a spatial density whose support is correlated with magnetic contact zones but whose *activation* and *within-ring localization* are governed by g: magnetic rings indicate where mineralization *could* occur, while favorable latent conditions determine where it *does* occur (the full generator is detailed in Section 5.1).

Consequently, visually similar anomalies can yield different outcomes: some rings host occurrences while others remain barren, and occurrences concentrate along specific arcs. Figure 3 shows an example realization, where occurrences appear only on a subset of the rings and at specific locations, reflecting latent variability not resolvable from d alone. Because one geological state is unobserved, the relevant predictive object is the marginal conditional density $\pi(p \mid d)$ in Equation (3), which can be multi-modal: a single deterministic mapping cannot represent this one-to-many behavior, motivating conditional density estimation.

4 Mineral Discovery using Flow Matching

In Section 3 we cast mineral discovery as learning a conditional spatial density $\pi(p \mid d)$ over occurrence positions $p \in \mathbb{R}^2$ given a multi-channel geoscience geo-image $d \in \mathbb{R}^{C \times H \times W}$. We now describe how to train a conditional velocity model GeoCFM, a geo-image guided conditional flow matching model, to sample from this distribution.

Given a geo-image, d, GeoCFM generates a set of positions $\{p_k\}_{k=1}^N \subset \Omega \subset \mathbb{R}^2$ by sampling N points from the learned conditional density. We work in normalized geo-image coordinates: point locations are scaled to $[0, 1]^2$ for feature sampling and mapped back to pixel coordinates for visualization and evaluation. Since similar observations can correspond to multiple plausible mineralization patterns, GeoCFM is stochastic and produces a distribution over possible occurrence sets rather than a single deterministic prediction.

4.1 Conditional Flow Matching for Mineral Occurrence Locations

For each geo-image d, let $\pi_{\mathrm{data}}(\cdot \mid d)$ denote the empirical distribution of occurrence locations, observed through samples $p \sim \pi_{\mathrm{data}}(\cdot \mid d)$. Conditional flow matching learns a time-dependent velocity field

$$v_\theta(p, t \mid d) : \Omega \times [0, 1] \to \mathbb{R}^2, \tag{4}$$

that transports a simple base distribution p_0 to the target distribution at $t = 1$. Specifically, we define the ODE

$$\frac{dp(t)}{dt} = v_\theta(p, t \mid d), \qquad p(0) \sim \pi_0, \tag{5}$$

and interpret $p(1)$ as a sample from the model $\pi_\theta(\cdot \mid d)$, given by the pushforward of π_0 through the conditional flow. We use $\pi_0 = \mathcal{N}(0, I_2)$ in $\mathbb{R}^2$.

In training, each observed occurrence u_k is treated as a draw from $\pi_{\mathrm{data}}(\cdot \mid d)$, and we train on all points across all geo-images.

4.2 Training Objective

We use the standard conditional flow matching construction with a linear interpolation between a base sample $z \sim \mathcal{N}(0, I_2)$ and a data point u:

$$p_t = t\,p + (1 - t)\,z, \qquad t \sim \mathrm{Unif}[0, 1]. \tag{6}$$

Along this trajectory, the target velocity is constant,

$$v^\star = \dot{p}_t = p - z. \tag{7}$$

We train v_θ by regressing to this velocity:

$$\mathcal{L}(\theta) = \mathbb{E}_d \mathbb{E}_{p \sim \pi_{\mathrm{data}}(\cdot|d)} \mathbb{E}_{z \sim \mathcal{N}(0, I_2)} \mathbb{E}_{t \sim \mathrm{Unif}[0,1]} \left[\| v_\theta(p_t, t \mid d) - (p - z) \|_2^2 \right]. \tag{8}$$

Operationally, for each batch we sample p, z, and t, form p_t via Equation (6), and minimize the mean-squared error between the predicted velocity and $p - z$.

4.3 Conditioning on Image-Like Observations

The main architectural challenge is to condition a point-wise vector field in $\mathbb{R}^2$ on a geo-image d. GeoCFM follows a point-conditioned design: a neural network produces a dense feature map, and local features are sampled at point locations.

UNet feature extractor. We encode each geo-image with a UNet F_ξ, where ξ denotes the UNet parameters, to obtain a dense feature tensor

$$\Phi_\xi = F_\xi(d) \in \mathbb{R}^{B \times D \times H \times W}. \tag{9}$$

The UNet aggregates multi-scale context while preserving spatial resolution, which is important because occurrences are localized.

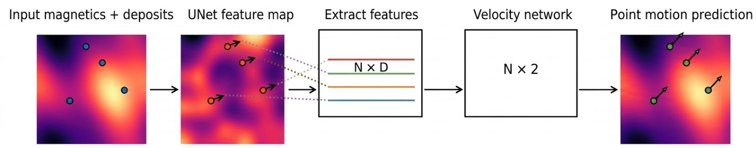

Fig. 4: GeoCFM point-conditioned velocity prediction. The geo-image d is encoded by a UNet into a dense feature map. For the current set of noised points p_t (with N points), we bilinearly sample local features to obtain an $N \times D$ feature matrix, concatenate each feature with the point coordinate (and time embedding), and pass the result through a shared per-point network to predict velocities $v_\theta(p_t, t \mid d) \in \mathbb{R}^{N\times 2}$, which define the point transport in the flow ODE.

Per-point feature sampling. For each point p_t and the computed feature tensor Φ_ξ, we extract a feature vector, ϕ by bilinear sampling Φ. Formally this can be written as

$$\phi_\xi = \mathcal{B}(p_t)\Phi_\xi \tag{10}$$

where $\mathcal{B}(p_t)$ is a bilinear interpolation operator at points p_t obtaining the matrix $\phi_\xi \in \mathbb{R}^D$.

We then parameterize the conditional velocity as

$$v_\theta(p_t, t \mid d) = h_\psi\Big(\phi_\xi,\, p_t,\, t\Big), \tag{11}$$

where h_ψ is an MLP and $\theta = \{\xi, \psi\}$ are the combined network parameters.

Time embedding and velocity head. We use standard sinusoidal time embeddings and a ResNet velocity head h_ψ: a linear projection followed by residual blocks and a final linear map to $\mathbb{R}^2$. The head predicts $\dot{p} = v_\theta(p_t, t \mid d)$ for each point given its local geo-image context and time. Our pipeline is illustrated in Figure 4.

After training, we generate occurrence samples conditioned on a new geo-image d by sampling $p^{(0)} \sim \mathcal{N}(0, I_2)$ and numerically integrating the ODE in Equation (5) from $t = 0$ to $t = 1$. We use an explicit Euler solver with K steps of size $\Delta t = 1/K$. To avoid overloading the occurrence index k in $\{p_k\}_{k=1}^N$, we write Euler iterates with a superscript $p^{(k)}$ and discrete times $t^{(k)}$:

$$p^{(k+1)} = p^{(k)} + \Delta t\, v_\theta\big(p^{(k)}, t^{(k)} \mid d\big), \qquad t^{(k)} = k\Delta t, \qquad k = 0, \ldots, K-1. \tag{12}$$

The final points $p^{(K)}$ are mapped to geo-image pixel coordinates for visualization and evaluation. Drawing N samples yields a generated occurrence set $\{p_k\}_{k=1}^N$ conditioned on the same geo-image. The number of Euler steps K trades inference cost against sample quality; we use $K = 50$ throughout and analyze this trade-off in Section 5.3, where small K is already competitive.

5 Experiments

We evaluate GeoCFM on a controlled synthetic benchmark and on real USGS Earth MRI data. Our experiments are designed to answer:

(Q1) Can GeoCFM learn to *sample* realistic occurrence locations from positive-only supervision, outperforming classical score-map pipelines that rely on pseudo-negatives?
(Q2) Does GeoCFM capture *one-to-many* behavior, where visually similar geophysical signatures activate different subsets of structures due to unobserved controls?
(Q3) Does GeoCFM generalize under a *spatially disjoint* split on real geo-images, where train and test patches come from non-overlapping geographic tiles?

Comparison protocol and baselines. We evaluate GeoCFM as a *stochastic conditional sampler* that learns and draws from a conditional spatial density $\pi(p \mid d)$, rather than as a deterministic score-map predictor. To make every method comparable in this sampling-based setting, we map each baseline into the same point-sampling space: its score or density map is normalized over the admissible pixels into a spatial distribution, from which we draw the same number of points per patch as GeoCFM. We compare against a principled taxonomy of positive-only baselines: *(i) non-conditional spatial samplers* that ignore d – Uniform sampling and a global KDE fit to training occurrences; *(ii) pseudo-negative score-map models* that learn a per-pixel favorability map from positives and sampled pseudo-negatives – Poisson-LR (logistic regression on per-pixel features), Random Forest (RF), gradient-boosted trees (GBDT) [11], and UNet-Seg, a strong dense segmentation network; *(iii) a positive-only one-class model* – a one-class SVM (OCSVM) [5] that estimates the support of the positive distribution; and *(iv) a retrieval-conditioned density estimator* – Retrieval-KDE, which forms a density from retrieved training occurrences. This taxonomy spans the main alternative ways to learn from positive-only spatial labels, so improvements cannot be attributed to a single baseline family.

Evaluation metrics. Because a single nearest-neighbor distance does not capture distribution matching, diversity, or mode coverage, we evaluate with five complementary metrics: the symmetric Chamfer distance (CD, pixels) for geometric agreement; the Sinkhorn optimal-transport divergence for distribution matching; F@5, the F-score of generated vs. observed points matched within a 5-pixel tolerance; the KDE negative log-likelihood (NLL), scoring observed occurrences under a density fit to the generated samples; and the top-5% hit rate (Top5), the fraction of observed occurrences in the top 5% most prospective pixels (arrows in tables indicate the better direction). To assess stability, each metric is averaged over 20 independent draws per input and reported as mean ± standard deviation of the mean across 50 held-out synthetic examples and 200 Earth MRI patches. We also show qualitative overlays and per-pixel prospectivity maps (Figure 5); the supplementary material provides additional qualitative comparisons against all baselines on held-out patches.

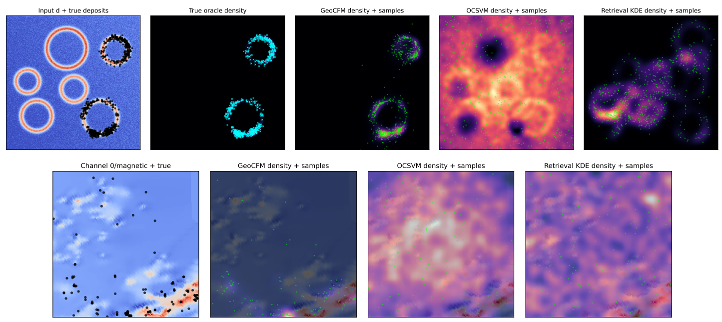

Fig. 5: Uncertainty and prospectivity maps. Top: synthetic input and true deposits, the oracle density, and the conditional densities recovered by GeoCFM, OCSVM, and Retrieval-KDE. Bottom: Earth MRI input and true occurrences with the same three methods. GeoCFM concentrates density on the relevant occurrence structures, exposing epistemic uncertainty through the spread of its conditional density, whereas the one-class and retrieval baselines produce diffuse or misplaced densities.

5.1 Synthetic Magnetics–Geochemistry Benchmark

This benchmark isolates the central modeling challenge: observations contain strong, interpretable geophysical structure, but occurrence locations depend on hidden controls, making the conditional mapping inherently one-to-many. It directly tests whether GeoCFM can learn conditional sampling from positives (Q1) and reproduce one-to-many conditional behavior induced by latent activation (Q2).

Data generation. Each sample is a spatial geo-image over $\Omega \subset \mathbb{R}^2$ ($H \times W$ grid). The generator outputs observed magnetics d, an observed geochemistry proxy t, a latent (unobserved) geochemistry field s, and deposit locations $\{p_k\}_{k=1}^N$. The magnetic field is formed from M non-overlapping elliptical intrusive bodies $\{b_m\}_{m=1}^M$ with random geometry and amplitude over a smooth trend with additive noise, $d = \mathrm{Std}(\sum_m b_m + \mathrm{trend} + \epsilon)$, and a ring-like contact mask $e_m = \mathrm{ContactZone}(\|\nabla d_{\mathrm{clean}}\|, b_m)$ emphasizes intrusion contacts.

Latent activation and deposit sampling. Both s (latent) and t (observed) include smooth backgrounds and localized enrichment along angular arcs of contact rings. Only a subset of intrusions is declared *active* via a stochastic rule based on enrichment statistics (plus noise), so many rings appear but only some host deposits. We define an unnormalized intensity

$$\lambda(p) \propto \Big(\sum_{m=1}^{M} \mathbf{1}\{\mathrm{active}_m\}\, e_m(p) \Big) \exp(\alpha s(p)) \, \exp(\beta t(p)) \, \exp(-\gamma |s(p) - t(p)|), \tag{13}$$

normalize it over Ω to obtain a density $\pi(p \mid d)$, and sample N deposits with sub-pixel jitter. Figure 3 shows a typical realization.

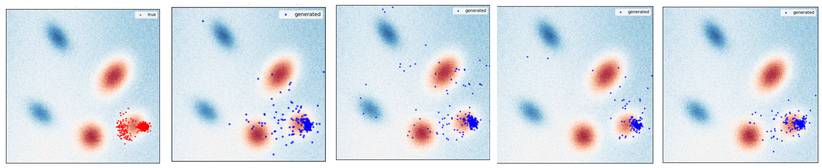

Fig. 6: Synthetic benchmark: true occurrences and four GeoCFM samples conditioned on the same input. Samples align with magnetic contact zones and activate only a subset of anomalies, demonstrating one-to-many conditional behavior.

Task, models, and training. Inputs are the two-channel tensor $[d, t] \in \mathbb{R}^{2\times H\times W}$ and supervision is the deposit set $\{p_k\}_{k=1}^{N}$. GeoCFM is trained with Equation (8) and sampled via Euler integration (Equation (12)), and compared against the full baseline suite above. To remove dataset-size effects we generate samples on-the-fly with $B = 4$, $H = W = 220$, $M = 5$ intrusions, and $N = 500$ deposits per image, training with AdamW (learning rate 3×10^{-4}, weight decay 10^{-4}), cosine annealing, and gradient clipping at norm 1.

Qualitative results. We evaluate on a fixed held-out set of 16 geo-images generated with fixed seeds. Figure 6 visualizes the true occurrence set and four samples generated by GeoCFM for the same input; samples concentrate along contact zones and activate only a subset of visually similar rings, reflecting the latent activation mechanism (Q2).

Quantitative results. Table 1 reports all five metrics on the held-out synthetic set, each averaged over 20 stochastic draws and reported as mean $\pm$ standard deviation of the mean. Among baselines, UNet-Seg is the strongest score-map method (CD 29.46), ahead of GBDT (32.13), RF (37.09), and Poisson-LR (39.93); the one-class and retrieval models (OCSVM, Retrieval-KDE) and the unconditional samplers (Global KDE, Uniform) perform worst. GeoCFM attains the best value on *every* metric, not only Chamfer distance but also Sinkhorn OT, F@5, NLL, and top-5% hit rate, indicating that directly learning a conditional sampler improves geometric agreement *and* distribution matching under positive-only supervision (Q1), and that the gains are not explained by any single baseline family.

5.2 USGS Earth MRI Benchmark

We next evaluate on real continental-scale geo-images from the USGS Earth Mapping Resources Initiative (Earth MRI) [29], paired with mineral occurrence point records (MRDS/MAS–MILS) [30]. This setting features heterogeneous modalities, strong spatial autocorrelation, and positive-only labels. It tests conditional sampling from positives on real data (Q1) and generalization under a spatially disjoint split (Q3).

Table 1: Synthetic benchmark results, averaged over 20 stochastic draws across 50 held-out examples; values are mean ± standard deviation of the mean. We report Chamfer distance (CD, pixels), Sinkhorn OT divergence (Sink.), F@5, KDE negative log-likelihood (NLL), and top-5% hit rate (Top5). Arrows indicate the better direction; the best value in every column is in bold. Baselines are grouped by family (pseudo-negative score-map models; one-class and retrieval-conditioned estimators; non-conditional samplers), and GeoCFM is shown in the last row.

Method	CD ↓	Sink. ↓	F@5 ↑	NLL ↓	Top5 ↑
UNet-Seg	29.46 ± 2.18	0.100 ± 0.006	0.615 ± 0.022	9.143 ± 0.089	0.661 ± 0.031
GBDT	32.13 ± 2.03	0.103 ± 0.006	0.546 ± 0.018	9.288 ± 0.083	0.630 ± 0.027
RF	37.09 ± 2.17	0.116 ± 0.006	0.462 ± 0.018	9.552 ± 0.054	0.586 ± 0.025
Poisson-LR	39.93 ± 2.16	0.122 ± 0.006	0.384 ± 0.016	9.803 ± 0.042	0.500 ± 0.028
OCSVM	45.56 ± 2.20	0.134 ± 0.007	0.246 ± 0.015	10.557 ± 0.020	0.236 ± 0.018
Retrieval-KDE	47.16 ± 2.86	0.132 ± 0.007	0.220 ± 0.016	13.302 ± 0.440	0.058 ± 0.011
Global KDE	41.16 ± 2.04	0.121 ± 0.006	0.255 ± 0.014	10.583 ± 0.030	0.058 ± 0.008
Uniform	49.61 ± 2.28	0.147 ± 0.007	0.206 ± 0.015	10.804 ± 0.006	0.048 ± 0.006
GeoCFM (Ours)	$\mathbf{9.37 \pm 0.62}$	$\mathbf{0.045 \pm 0.003}$	$\mathbf{0.634 \pm 0.012}$	$\mathbf{8.950 \pm 0.110}$	$\mathbf{0.720 \pm 0.020}$

Data, patches, and split. Each sample consists of a multi-channel geo-image tensor $d \in \mathbb{R}^{C\times H\times W}$ and an observed set of occurrences $\{p_k\}_{k=1}^{N} \subset \Omega$, treated as positives with no reliable negatives. We extract windows on a stride lattice, keeping a patch if at least a fraction ρ of its pixels lie within the data footprint; coordinates are normalized to $[0,1]^2$ per patch and geo-images are standardized per channel over non-zero pixels. To avoid spatial leakage from autocorrelation, we partition the region into large rectangular tiles and assign tiles to train vs. test, so patches inherit their tile's split and train/test neighborhoods never overlap (Q3). The supplementary material describes the full Earth MRI preprocessing pipeline (rasterization, gap filling, occurrence alignment, patch extraction, and the tile-based split).

Baselines. We report the full baseline suite described above: the non-conditional Uniform and global KDE samplers; the score-map discriminative models Poisson-LR, RF, GBDT, and UNet-Seg; and the positive-only OCSVM and Retrieval-KDE models. Methods that learn a score map are trained on training tiles and evaluated by sampling points from their predicted maps; Uniform and KDE do not use the input patch for conditioning.

Results. Figure 7 shows a qualitative overlay of sampled occurrences on held-out tiles, and Table 2 reports all five metrics on the test split (mean ± standard deviation of the mean over 20 draws across 200 patches). GeoCFM is best on every metric by a wide margin, supporting conditional sampling from positives on real data (Q1). Because the evaluation uses a spatially disjoint tile split and GeoCFM exhibits only a small train-test Chamfer gap (the supplementary material reports per-method train and test scores on both benchmarks), the im-

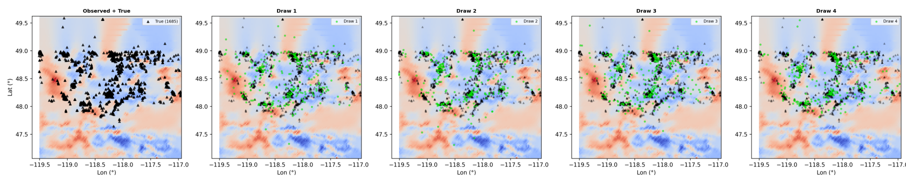

Fig. 7: Earth MRI: Qualitative comparison of conditional mineral occurrence draws. (Left) True mineral occurrence locations (black triangles) overlaid on a held-out geophysics geo-image tile, showing a central cluster. (Draws 1–4) Four unique sampled occurrence locations (green dots) generated by our GeoCFM, all overlaid on the same true occurrences and background geophysics geo-image, demonstrating its ability to capture the complex, non-isotropic distribution of mineral potential and expose non-identifiability. We provide additional examples and comparisons with baselines in the supplementary material.

provement generalizes to new geographic tiles rather than reflecting local spatial leakage (Q3).

5.3 Ablations and Analyses

Sensitivity to the available positives. In a severely imbalanced positive-only setting, which positives are available for training can affect predictions more than the sampler's stochasticity. Retraining GeoCFM with 100%, 75%, and 50% of the positives over three seeds (the supplementary material reports the full positive-subset stability table), GeoCFM is stable: synthetic CD stays within 9.30–9.75 px and Earth MRI CD within 11.76–12.65 px. This separates predictive sampling uncertainty from sensitivity to label availability.

Quality–cost tradeoff via the number of Euler steps. GeoCFM provides a tunable quality–cost tradeoff through the number of Euler steps K (Equation (12)). Small K already preserves strong point-set quality – $K = 5$ gives CD 9.30 on synthetic and 12.00 on Earth MRI, within the standard deviation of the $K = 50$ setting – while the UNet feature map is computed once per patch and each Euler step applies only a bilinear lookup and the lightweight velocity head. The supplementary material provides the full per-patch runtime and Euler-step ablation against the single-pass UNet-Seg reference.

Effect of the sampling budget N. The number of points N drawn per patch is not an estimate of the true deposit count but a Monte Carlo budget for the conditional prospectivity density, fixed here only for a fair point-set comparison. Larger N improves the KDE-NLL and density stability on both benchmarks, with diminishing returns past a few hundred samples (the supplementary material reports the full sampling-budget study); N is thus a free post-hoc knob rather than an a-priori commitment.

Table 2: Earth MRI results with a spatially disjoint (tile-based) train/test split, on the test split, averaged over 20 stochastic draws across 200 patches; values are mean ± standard deviation of the mean. We report Chamfer distance (CD, pixels), Sinkhorn OT divergence (Sink.), F@5, KDE negative log-likelihood (NLL), and top-5% hit rate (Top5). Arrows indicate the better direction; the best value in every column is in bold. Baselines are grouped by family (pseudo-negative score-map models; one-class and retrieval-conditioned estimators; non-conditional samplers), and GeoCFM is shown in the last row.

Method	CD ↓	Sink. ↓	F@5 ↑	NLL ↓	Top5 ↑
UNet-Seg	46.81 ± 2.30	0.205 ± 0.008	0.085 ± 0.006	11.620 ± 0.290	0.130 ± 0.013
GBDT	52.98 ± 2.14	0.215 ± 0.007	0.069 ± 0.004	11.233 ± 0.269	0.102 ± 0.010
RF	54.37 ± 2.23	0.225 ± 0.008	0.071 ± 0.004	11.504 ± 0.297	0.095 ± 0.011
Poisson-LR	64.21 ± 2.26	0.255 ± 0.008	0.055 ± 0.004	11.274 ± 0.239	0.059 ± 0.008
OCSVM	58.38 ± 2.33	0.240 ± 0.008	0.053 ± 0.003	11.946 ± 0.313	0.059 ± 0.008
Retrieval-KDE	61.14 ± 2.34	0.250 ± 0.008	0.053 ± 0.003	11.698 ± 0.298	0.061 ± 0.009
Uniform	64.25 ± 2.34	0.260 ± 0.008	0.052 ± 0.003	11.598 ± 0.288	0.047 ± 0.007
Global KDE	66.35 ± 2.38	0.277 ± 0.008	0.050 ± 0.003	12.053 ± 0.303	0.043 ± 0.007
GeoCFM (Ours)	$\mathbf{12.30 \pm 0.54}$	$\mathbf{0.085 \pm 0.004}$	$\mathbf{0.151 \pm 0.008}$	$\mathbf{10.500 \pm 0.150}$	$\mathbf{0.250 \pm 0.012}$

6 Summary and Outlook

Summary. We reframe mineral prospectivity mapping as learning a *conditional spatial distribution* over occurrence locations, $\pi(p \mid d)$, from positive-only supervision, and instantiate it with GeoCFM, a conditional flow-matching sampler that bridges dense geo-images and sparse deposit supervision via point-conditioned UNet features, without pseudo-negatives. Drawing multiple samples for the same d exposes the epistemic uncertainty induced by unobserved geology. On a synthetic magnetics–geochemistry benchmark and on USGS Earth MRI data with a spatially disjoint split, GeoCFM captures structured one-to-many behavior and outperforms a broad baseline suite across all five evaluation metrics.

Outlook. Supervised prospectivity models often retain a deterministic framing that mismatches positive-only labels and subsurface non-identifiability; the probabilistic generative view adopted here makes ambiguity explicit through sampling and enables uncertainty-aware target ranking. Future extensions can incorporate constrained masks, structural-contact penalties, physics-informed objectives, or geological forward simulators to inject stronger geoscientific priors into the sampler. Coupling conditional sampling with adaptive estimation of the occurrence count, larger multi-modal datasets, and forward simulators further opens the door to calibrated, regional-scale scenario generation.

Acknowledgement. We thank Dr. Bas Peters for valuable discussions and feedback on the conceptual framing of the mineral prospectivity problem, the geoscientific context, and the machine-learning design of the approach.

References

1. Bekker, J., Davis, J.: Learning from positive and unlabeled data: A survey. Machine Learning **109**(4), 719–760 (2020). https://doi.org/10.1007/s10994-020-05877-5
2. Bekker, J., Robberechts, P., Davis, J.: Beyond the selected completely at random assumption for learning from positive and unlabeled data. In: Machine Learning and Knowledge Discovery in Databases (ECML PKDD). pp. 71–85. Springer, Cham (2019). https://doi.org/10.1007/978-3-030-46147-8_5
3. Bonham-Carter, G.F.: Geographic Information Systems for Geoscientists: Modelling with GIS, Computer Methods in the Geosciences, vol. 13. Pergamon (1994). https://doi.org/10.1016/C2013-0-03864-9
4. Carranza, E.J.M. (ed.): Geochemical Anomaly and Mineral Prospectivity Mapping in GIS, Handbook of Exploration and Environmental Geochemistry, vol. 11. Elsevier (2009)
5. Chen, Y., Wu, W., Zhao, Q.: A bat-optimized one-class support vector machine for mineral prospectivity mapping. Minerals **9**(5), 317 (2019)
6. Cracknell, M.J., Reading, A.M., de Caritat, P.: Geological knowledge discovery and minerals targeting from regolith using a machine learning approach. ASEG Extended Abstracts **2015**(1), 1–4 (2015). https://doi.org/10.1071/ASEG2015ab283
7. Daruna, A., Zadorozhnyy, V., Lukoczki, G., Chiu, H.P.: Enabling scalable mineral exploration: Self-supervision and explainability. In: 2024 IEEE International Conference on Big Data (BigData). pp. 2090–2099. IEEE (2024)
8. Diggle, P.J., Menezes, R., Su, T.l.: Geostatistical inference under preferential sampling. Journal of the Royal Statistical Society: Series C (Applied Statistics) **59**(2), 191–232 (2010). https://doi.org/10.1111/j.1467-9876.2009.00701.x
9. Du Plessis, M.C., Niu, G., Sugiyama, M.: Analysis of learning from positive and unlabeled data. Advances in neural information processing systems **27** (2014)
10. Elkan, C., Noto, K.: Learning classifiers from only positive and unlabeled data. In: Li, Y., Liu, B., Sarawagi, S. (eds.) Proceedings of the 14th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, Las Vegas, Nevada, USA, August 24-27, 2008. pp. 213–220. ACM (2008). https://doi.org/10.1145/1401890.1401920, https://doi.org/10.1145/1401890.1401920
11. Friedman, J.H.: Greedy function approximation: a gradient boosting machine. Annals of statistics pp. 1189–1232 (2001)
12. Granek, J., Haber, E.: Data mining for real mining: A robust algorithm for prospectivity mapping with uncertainties. In: Proceedings of the 2015 SIAM International Conference on Data Mining. pp. 145–153. Society for Industrial and Applied Mathematics (SIAM) (2015). https://doi.org/10.1137/1.9781611974010.17
13. Granek, J., Haber, E.: Advanced geoscience targeting via focused machine learning applied to the QUEST project dataset, british columbia. In: Geoscience BC Summary of Activities 2015, pp. 117–126. Geoscience BC (2016), geoscience BC Report 2016-1
14. Ho, J., Jain, A., Abbeel, P.: Denoising diffusion probabilistic models. In: Advances in Neural Information Processing Systems (NeurIPS) (2020). https://doi.org/10.48550/arXiv.2006.11239, https://arxiv.org/abs/2006.11239
15. Hronsky, J.M.A., Kreuzer, O.P.: Applying spatial prospectivity mapping to exploration targeting: Fundamental practical issues and suggested solutions for the future. Ore Geology Reviews **107**, 647–653 (Apr 2019). https://doi.org/10.1016/j.oregeorev.2019.03.016

16. Iyer, A., Nath, S., Sarawagi, S.: Maximum mean discrepancy for class ratio estimation: Convergence bounds and kernel selection. In: International conference on machine learning. pp. 530–538. PMLR (2014)
17. Kirillov, A., Wu, Y., He, K., Girshick, R.: Pointrend: Image segmentation as rendering. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). pp. 9799–9808 (2020). https://doi.org/10.1109/CVPR42600.2020.00982, https://openaccess.thecvf.com/content_CVPR_2020/html/Kirillov_PointRend_Image_Segmentation_As_Rendering_CVPR_2020_paper.html
18. Kiryo, R., Niu, G., du Plessis, M.C., Sugiyama, M.: Positive-unlabeled learning with non-negative risk estimator. In: Guyon, I., von Luxburg, U., Bengio, S., Wallach, H.M., Fergus, R., Vishwanathan, S.V.N., Garnett, R. (eds.) Advances in Neural Information Processing Systems 30: Annual Conference on Neural Information Processing Systems 2017, December 4-9, 2017, Long Beach, CA, USA. pp. 1675–1685 (2017), https://proceedings.neurips.cc/paper/2017/hash/7cce53cf90577442771720a370c3c723-Abstract.html
19. Lipman, Y., Chen, R.T.Q., Ben-Hamu, H., Nickel, M., Le, M.: Flow matching for generative modeling. In: International Conference on Learning Representations (ICLR) (2023). https://doi.org/10.48550/arXiv.2210.02747, https://arxiv.org/abs/2210.02747
20. Lu, Z., Han, M., Guo, P., Bai, T., Su, J., Fang, F., Song, S.: Attention-guided flow-matching for sparse 3d geological generation. arXiv preprint arXiv:2604.09700 (2026)
21. McMillan, M., Haber, E., Peters, B., Fohring, J.: Mineral prospectivity mapping using a vnet convolutional neural network. The Leading Edge **40**(2), 99–105 (02 2021). https://doi.org/10.1190/tle40020099.1, https://doi.org/10.1190/tle40020099.1
22. Phillips, S.J., Dudík, M., Elith, J., Graham, C.H., Lehmann, A., Leathwick, J., Ferrier, S.: Sample selection bias and presence-only distribution models: implications for background and pseudo-absence data. Ecological applications **19**(1), 181–197 (2009)
23. Plessis, M.C., Niu, G., Sugiyama, M.: Class-prior estimation for learning from positive and unlabeled data. Mach. Learn. **106**(4), 463–492 (Apr 2017). https://doi.org/10.1007/s10994-016-5604-6, https://doi.org/10.1007/s10994-016-5604-6
24. Ramaswamy, H., Scott, C., Tewari, A.: Mixture proportion estimation via kernel embeddings of distributions. In: International conference on machine learning. pp. 2052–2060. PMLR (2016)
25. Schodde, R.: Mineral deposit exploration: Discovery trends, 1900-2023. SEG Discovery **142**, 19–35 (Jul 2025). https://doi.org/10.5382/Geo-and-Mining-28
26. Singer, D.A., Kouda, R.: Examining risk in mineral exploration. Natural Resources Research **8**(2), 111–122 (1999). https://doi.org/10.1023/A:1021838618750
27. Song, Y., Sohl-Dickstein, J., Kingma, D.P., Kumar, A., Ermon, S., Poole, B.: Score-based generative modeling through stochastic differential equations. In: International Conference on Learning Representations (ICLR) (2021). https://doi.org/10.48550/arXiv.2011.13456, https://arxiv.org/abs/2011.13456
28. Tong, A., Malkin, N., Huguet, G., Zhang, Y., Rector-Brooks, J., Fatras, K., Wolf, G., Bengio, Y.: Conditional flow matching: Simulation-free dynamic optimal transport. arXiv preprint arXiv:2302.00482 (2023). https://doi.org/10.48550/arXiv.2302.00482, https://arxiv.org/abs/2302.00482

29. U.S. Geological Survey: Earth mapping resources initiative (earth mri). https://www.usgs.gov/special-topics/earth-mri, accessed: 2026-03-04
30. U.S. Geological Survey: Mineral resources data system (mrds) (2026), https://mrdata.usgs.gov/mrds/, accessed: 2026-03-05
31. Warton, D.I., Shepherd, L.C.: Poisson point process models solve the" pseudo-absence problem" for presence-only data in ecology. The Annals of Applied Statistics pp. 1383–1402 (2010)
32. Yang, G., Huang, X., Hao, Z., Liu, M.Y., Belongie, S., Hariharan, B.: Pointflow: 3d point cloud generation with continuous normalizing flows. In: Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV) (2019). https://doi.org/10.1109/ICCV.2019.00464, https://openaccess.thecvf.com/content_ICCV_2019/papers/Yang_PointFlow_3D_Point_Cloud_Generation_With_Continuous_Normalizing_Flows_ICCV_2019_paper.pdf
33. Zuo, R.: Geodata science-based mineral prospectivity mapping: A review. Natural Resources Research **29**, 3415–3424 (2020). https://doi.org/10.1007/s11053-020-09700-9
34. Zuo, R., Wang, Z.: Effects of random negative training samples on mineral prospectivity mapping. Natural Resources Research **29**, 3443–3455 (2020). https://doi.org/10.1007/s11053-020-09668-6