Research Square

Preprints are preliminary reports that have not undergone peer review.
They should not be considered conclusive, used to inform clinical practice,
or referenced by the media as validated information.

TraceRIFT: Reference-preserving ensemble selection for discrete fracture networks from incomplete traces

Shuo Wang
Chang'an University
Chen Guo
chenguo@chd.edu.cn
Chang'an University
Bingxin Yang
Chang'an University
Keran Hao
Chang'an University
Yu Zhang
Chang'an University
Fuqiang Zhang
Chang'an University

Method Article

Keywords: discrete fracture network, incomplete observations, hypothesis-set selection, reference preservation, submodular optimization, computational geoscience

Posted Date: November 5th, 2026

DOI: https://doi.org/10.21203/rs.3.rs-11330961/v1

License: This work is licensed under a Creative Commons Attribution 4.0 International License.
Read Full License

Additional Declarations: No competing interests reported.

TraceRIFT: Reference-preserving ensemble selection for discrete fracture networks from incomplete traces

Shuo Wang[1], Chen Guo[1*], Bingxin Yang[1], Keran Hao[1], Yu Zhang[1], Fuqiang Zhang[1]

[1]School of Information Engineering, Chang'an University, Xi'an, 710064, China.

*Corresponding author(s). E-mail(s): chenguo@chd.edu.cn;

Abstract

Incomplete surface traces admit multiple three-dimensional discrete fracture networks with different intensity, size, connectivity and orientation. Reducing these candidates to a small ensemble can discard useful alternatives even when the generated pool contains them. We introduce TraceRIFT, a framework that separates candidate-support loss from retrieval regret and treats ensemble selection as predictive scenario coverage. Candidate-support loss measures the pool's minimum attainable error; retrieval regret measures the additional error introduced by shortlist construction. Set quality is evaluated through the closest retained network, linking ensemble reduction to retention of plausible geological alternatives. Coverage in an eight-attribute space values each network by its contribution beyond a prescribed core of the baseline Reference ranking. TraceRIFT preserves that core while permitting at most two additions beyond the original Reference top ten. Enumerated-greedy optimization of the submodular objective gives a **3/4** approximation guarantee for the finite-scenario proposal before an independent predictive check. On 3,276 held-out simulated parent networks, all selectors retain ten networks from identical query-specific 64-member pools. Under fragmented or missing-surface observations, TraceRIFT reduces mean retrieval regret by 22.68% relative to Reference, from 0.0903 to 0.0698. The paired gain is 0.0205, with a 95% parent-clustered interval of 0.0183–0.0227. Mean regret is also 9.60% lower than rank-based maximal marginal relevance (MMR) and 10.70% lower than predictive MMR under the same constraints. TraceRIFT also lowers regret relative to target-consistency-only and observable-discrepancy References. Workflow deployment and predictive transfer are also assessed using 480 catalogue-derived virtual queries from Wadi Birk.

Keywords: discrete fracture network, incomplete observations, hypothesis-set selection, reference preservation, submodular optimization, computational geoscience

1 Introduction

Fracture traces constrain where discontinuities intersect exposed rock faces, while their continuation through the rock volume controls connections and transport pathways (Bonnet et al. 2001; Lei et al. 2017). Different three-dimensional discrete fracture networks (DFNs) can therefore explain similar surface intensity and orientation while having different connected components. Interpreting finite exposures requires retaining alternative networks consistent with the observed traces.

Digital outcrop models map large or inaccessible exposures while retaining surface position and orientation (Sturzenegger and Stead 2009; Bemis et al. 2014). Automated extraction produces detailed measurements from photogrammetric point clouds (Dong et al. 2026), and DFN simulators support their interpretation through virtual plane and window sampling (Xu and Dowd 2010). Stereological estimators relate trace measurements to volumetric fracture statistics under specified sampling assumptions (Dershowitz and Einstein 1988; Dershowitz and Herda 1992; Mauldon et al. 2001). Connectivity remains partly hidden, and a three-dimensional network may percolate even when its sampled outcrop has no spanning path (Zhu et al. 2022). Figure 1 illustrates this ambiguity with two constructed networks that produce identical window traces but differ in connections beyond the windows.

Stochastic DFN modeling represents this ambiguity through multiple realizations, with increasingly flexible geometries and constraints. Elliptical fractures and learned low-dimensional parameterizations broaden candidate representations (Guo et al. 2024; Teng et al. 2026), while calibrated polygonal growth models reproduce nucleation, abutment and arrest (Kim et al. 2026). Hydraulic and electrical observations support surrogate-assisted trans-dimensional inversion (Sun et al. 2026), while integrated outcrop workflows combine lineaments, exposed planes and geological context (Racolte et al. 2026; Panara et al. 2024a). These approaches improve candidate generation and conditioning, leaving a subsequent decision about which alternatives to retain for inspection or downstream modeling.

Shortlist construction can discard a useful hypothesis even when that hypothesis is already available in the generated pool. Highly ranked networks may cover similar attributes, leaving a lower-ranked candidate as the closest alternative to another plausible configuration. Diversity-aware retrieval and determinantal models formalize the value of complementary items (Carbonell and Goldstein 1998; Kulesza and Taskar 2012). For DFNs, we assess complementarity through the error of the closest retained network in a specified attribute space. Candidate-support loss measures the error floor in the full pool; retrieval regret measures the additional error introduced by shortlist construction.

TraceRIFT addresses this selection problem through predictive scenario coverage around a protected core of a Reference ranking. It retains the first six members and selects a four-member tail from 64 candidates, permitting at most two additions beyond the Reference top ten. The objective values candidates by the eight-attribute scenarios they cover beyond the fixed core. Its monotone-submodular structure (Nemhauser et al. 1978) supports an 11-branch enumerated-greedy solver with a finite-scenario approximation guarantee. A separate set of predictive scenarios checks the proposed tail before adoption.

Table 1 Scientific coordinates used to quantify finite-pool discrepancy. Areas are clipped to domain Ω; graph fractions use retained fractures.

Coordinate	Definition
P_{32}	Total clipped fracture area divided by domain volume.
$\bar{r}$	Mean radius of retained simulator fractures.
$\bar{d}$	Mean intersection-graph degree, $2e/n$.
f_{comp}	Number of connected components divided by fracture count.
f_{LCC}	Fraction of fractures in the largest connected component.
$\lambda^{(O)}_{1,2,3}$	Ordered eigenvalues of $O = (\sum_i A_i n_i n_i^\top)/(\sum_i A_i)$, with clipped areas A_i and unit axial normals n_i.

We test whether complementary selection improves a compact DFN ensemble while preserving its prescribed Reference hypotheses. Selectors receive identical query-specific pools from held-out simulated parents under complete, fragmented and missing-surface observations. Comparisons with constrained relevance–diversity, dispersion and sequential selectors assess the coverage objective and joint search. A separate Wadi Birk experiment uses 480 catalogue-derived virtual queries to assess workflow deployment and predictive transfer.

2 Problem formulation

2.1 Observable traces and candidate networks

Let $G^\star$ denote a parent DFN in a bounded domain Ω. The observation $X = \mathcal{O}_c(G^\star)$ contains finite-window trace graphs, window positions and orientations, and masks identifying requested and observed faces. The condition c specifies complete observation, high trace fragmentation or one missing surface. We extract an observation vector $x = \Psi(X)$ and reproject each candidate G_i through the same operator to obtain observable features. The hidden parent geometry and target vector are retained separately for training and evaluation.

Generation produces a query-specific pool $\mathcal{C}_N(x) = \{G_1, \ldots, G_N\}$ with $N = 64$ explicit three-dimensional networks. Selection returns an ordered subset of ten distinct candidates, the reporting budget throughout this study. Network-attribute vectors define the common comparison space for selection and evaluation.

2.2 Network attributes and discrepancy

We summarize each candidate by

$$y(G) = \big(P_{32}, \bar{r}, \bar{d}, f_{\mathrm{comp}}, f_{\mathrm{LCC}}, \lambda_1^{(O)}, \lambda_2^{(O)}, \lambda_3^{(O)}\big). \tag{1}$$

The coordinates span fracture intensity, size, connectivity and orientation (Table 1). Fractures are vertices in the intersection graph, and their intersections are edges. Mean degree and component fractions describe the organization of this graph.

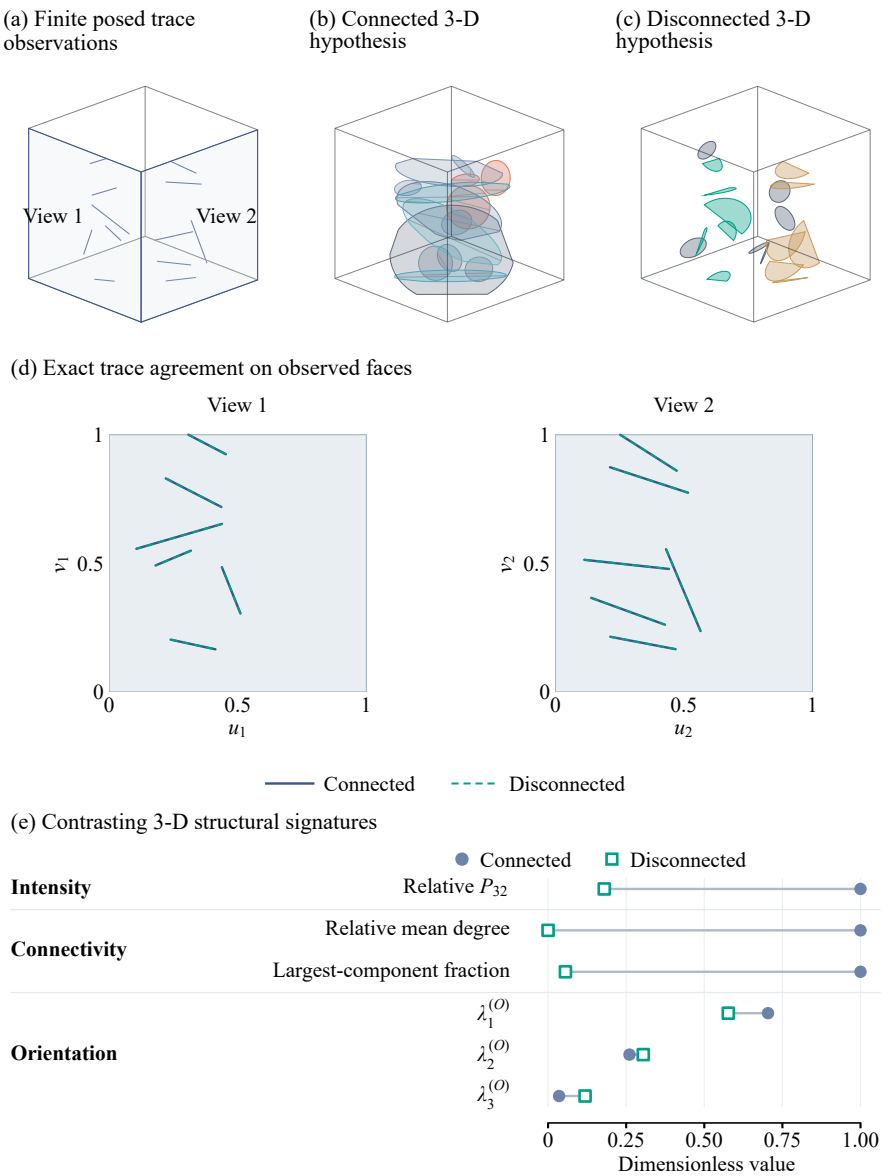


Fig. 1 Nonuniqueness of DFN interpretation from finite traces. (a) Two posed observation windows. (b,c) Constructed connected and disconnected circular-disc networks. (d) Their traces coincide on both windows. (e) These networks have different three-dimensional structural signatures. Window coordinates in this illustrative construction are normalized for display

Training folds determine the scales $s_j > 0$ used to standardize coordinate differences. The four families are intensity, size, connectivity/components and orientation, with respective group sizes $1, 1, 3, 3$. The discrepancy assigns the same total weight to each family:

$$d_s(y, z) = \left\{ \frac{1}{4} \sum_{g=1}^{4} \frac{1}{|I_g|} \sum_{j \in I_g} \left(\frac{y_j - z_j}{s_j} \right)^2 \right\}^{1/2}. \tag{2}$$

2.3 Candidate support and retrieval regret

For a retained index set S, $|S| = 10$, define

$$L_{10}(S; y^\star) = \min_{i \in S} d_s(y_i, y^\star), \tag{3}$$

$$L_{64}^\star(y^\star) = \min_{i \in \mathcal{C}_{64}} d_s(y_i, y^\star), \tag{4}$$

$$\text{Regret @10} = L_{10} - L_{64}^\star \geq 0. \tag{5}$$

The decomposition $L_{10} = L_{64}^\star + \text{Regret @10}$ separates candidate-support error from the additional loss introduced by shortlist selection. The pool-oracle loss $L_{64}^\star$ is the lowest discrepancy available within the query-specific pool. Retrieval regret measures the retained set's excess discrepancy above that floor.

Exact-oracle recall records whether the retained set contains a minimizer of Eq. (4). Near-oracle recall records $L_{10} \leq L_{64}^\star + 0.05$, using an absolute margin in the standardized discrepancy of Eq. (2). The hidden target identifies the best retained member retrospectively; deployment returns all ten candidates for subsequent examination.

3 Reference-preserving hypothesis selection

3.1 Observation-conditioned candidate generation

TraceRIFT first constructs candidates from posed traces, then selects a subset (Fig. 2). The observation vector contains 252 features: 68 geometry-core, 45 axial-orientation, 80 length/intensity and 59 topology features. The first three groups provide a 193-dimensional non-topological context. Window poses and observation masks specify measurement scale, orientation and availability.

The proposal factorizes simulator parameters into spatial process, orientation-family count, radius family, unordered axial family normals and continuous macro parameters. It supports homogeneous Poisson and toroidal Thomas-cluster processes, one to three orientation families, and truncated lognormal or gamma radii. Topological features inform the spatial-process factor; the remaining factors use the non-topological context. Normal signs and permutations of orientation families are treated as equivalent. Conditional neural spline flows model the simulator's continuous parameters (Durkan et al. 2019; Papamakarios et al. 2021). Each simulator attempt draws

$$\theta_a \sim q_\phi^{(-f)}(\cdot \mid x), \qquad \omega_a \sim p_0(\cdot \mid \theta_a), \qquad G_a = \mathcal{S}(\theta_a, \omega_a), \tag{6}$$

where $q_\phi^{(-f)}$ is the observation-conditioned proposal and p_0 governs the unresolved realization variables. Conditioning directs simulation toward the observation while retaining stochastic variability, following simulation-based inference from forward models (Cranmer et al. 2020). All retrieval methods then receive the same query-specific pool of candidates.

The proposal uses five-fold parent-level cross-fitting, with each parent's observations assigned to one fold. Candidate generation for fold f uses models fitted on the other folds, training-fold standardization and an inner early-stopping split. Three independently seeded checkpoints form a uniform mixture across simulation attempts. The six conditional flow states are indexed by spatial process and orientation-family count. Each state uses five transforms, eight spline bins and two 192-unit hidden layers.

Each query receives 80 simulation attempts, followed by validity checks and deterministic deduplication. We retain completed realizations with valid reprojections onto the requested windows. Among realizations sharing a public-reprojection hash, we keep the realization with the smallest content hash. Representatives retain their original attempt order; the first 64 form the candidate pool. Construction fails when fewer than 64 eligible representatives remain after deduplication. Online Resource 1, Section S2, lists the numerical settings and implementation details.

3.2 Predictive scenarios and reference ranking

A separate ensemble of 256 extremely randomized trees predicts the eight-dimensional target vector (Geurts et al. 2006). The predictive covariance combines between-tree variation with residual covariance estimated from training parents. Shrinkage of 0.10 and an eigenvalue floor of 10^{-4} regularize this estimate. The resulting Gaussian risk model, $Y \mid x \sim \mathcal{N}(\mu_x, \Sigma_x)$, supplies target scenarios for the set-selection objective.

Reference fuses full-pool rankings of target consistency and learned public evidence. The first scores regularized consistency with a separately predicted query target. The second uses a learned public-evidence ranker (PyCoRE), scoring discrepancy by the negative root-mean-square of four learned concept energies. If $p_i^{(T)}$ and $p_i^{(P)}$ are the zero-based positions in these rankings, their equal-rank Borda score is

$$s_i^{(R)} = -\frac{p_i^{(T)} + p_i^{(P)}}{2(N-1)}. \tag{7}$$

Sorting this score with content-derived tie breaking yields the ordered reference list $B = (b_1, \ldots, b_{10})$.

3.3 Complementary selection around a fixed core

We partition the Reference into a six-member core $A = \{b_1, \ldots, b_6\}$ and a tail $B_{\text{tail}} = \{b_7, \ldots, b_{10}\}$. For a degraded observation, TraceRIFT retains $A \cup T$, subject to

$$T \subseteq \mathcal{C}_{64} \setminus A, \qquad |T| = 4, \qquad |T \setminus B| \leq K = 2. \tag{8}$$

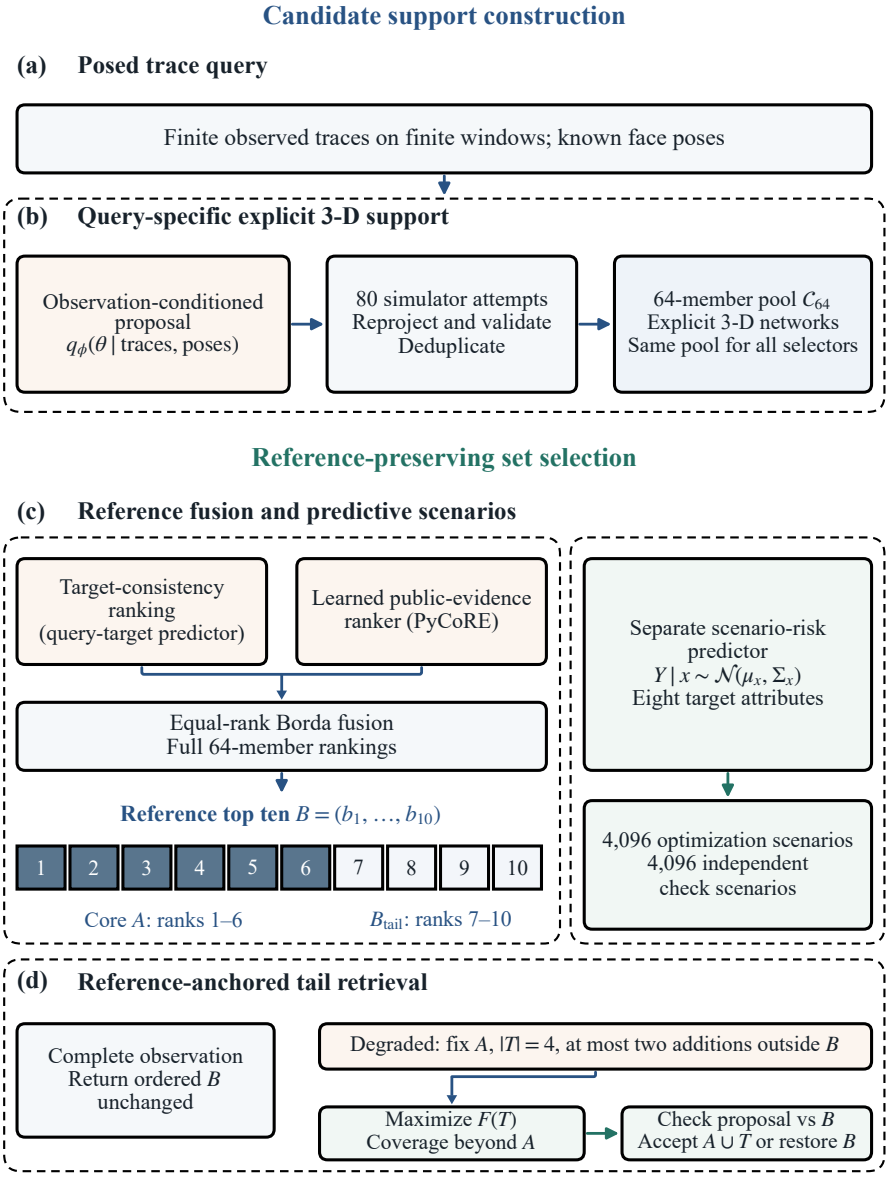


Fig. 2 TraceRIFT workflow. (a,b) Posed finite traces condition 80 simulation attempts; reprojection, validation and deduplication yield a 64-member pool. (c) Target-consistency and learned public-evidence rankings form Reference; a separate eight-attribute risk predictor supplies optimization and independent check scenarios. (d) Complete observations retain the ordered Reference top ten. Under degraded observations, a six-member core is fixed and a four-member tail permits at most two external additions. The independent check accepts the proposal or restores Reference

The first reference candidate remains first, and all six core members are retained. Complete observations return B unchanged, preserving the full ordered Reference list. The ten-member output and six-member core are fixed design choices. We select the replacement budget on the development fold (Section 5.2). At $K = 2$, at least eight retained members belong to the original top ten. This budget defines a discrete trust region around the Reference interpretation.

For scenario $y^{(m)}$, define $d_{mi} = d_s(y^{(m)}, y_i)$ and the core error $a_m = \min_{i \in A} d_{mi}$. Candidate i contributes $g_{mi} = \max(0, a_m - d_{mi})$ beyond the core, and we maximize the tail's mean coverage gain:

$$F(T) = \frac{1}{M} \sum_{m=1}^{M} \max_{i \in T} g_{mi}, \qquad F(\varnothing) = 0. \tag{9}$$

Since $\min(a_m, \min_{i \in T} d_{mi}) = a_m - \max_{i \in T} g_{mi}$, maximizing F minimizes predictive best-member loss. Overlapping scenario coverage reduces a candidate's marginal value as other members enter the set.

Proposition 1 *For fixed candidates, core and predictive scenarios, F is normalized, nonnegative, monotone and submodular. The enumerated-greedy optimization proposal $\widehat{T}$ satisfies $F(\widehat{T}) \geq (1 - 1/e)F(T_K^\star)$, where $T_K^\star$ is the optimal feasible tail in Eq.* (8)*. For the deployed budget $K = 2$, this bound sharpens to $F(\widehat{T}) \geq \frac{3}{4}F(T_2^\star)$. Strictly improving feasible refinement preserves these bounds.*

The scenario-wise maximum has diminishing marginal returns; averaging over scenarios preserves submodularity. Enumeration includes an optimum's retained-reference subset, and greedy external selection attains $1 - (1 - 1/k)^k$ for k additions (Nemhauser et al. 1978). Branches with zero or one external candidate are solved exactly. For branches with two external additions, the factor becomes $1 - (1 - 1/2)^2 = 3/4$, improving the general $1 - 1/e$ bound. Supplementary Section S1 provides the full proof of the optimization guarantee.

The solver enumerates $R \subseteq B_{\text{tail}}$ with $|R| = 4 - k$ for $k = 0, 1, 2$, giving $1+4+6 = 11$ branches (Fig. 3). Each branch greedily adds up to two of the 54 external candidates by marginal coverage. Feasible one-exchange moves refine the tail whenever they strictly improve the objective. Content-based tie breaking makes the result independent of candidate storage order. Optimization uses 4,096 scenarios; an independent set of 4,096 scenarios checks the final proposal. A rejected proposal restores the Reference list returned for that query. The approximation bound applies to the optimization proposal before this independent check.

4 Numerical experiments

4.1 Parent networks and paired observations

The benchmark contains 3,532 independently simulated parent DFNs, divided into development fold 0 (256 parents) and evaluation folds 1–4 (3,276 parents). Each

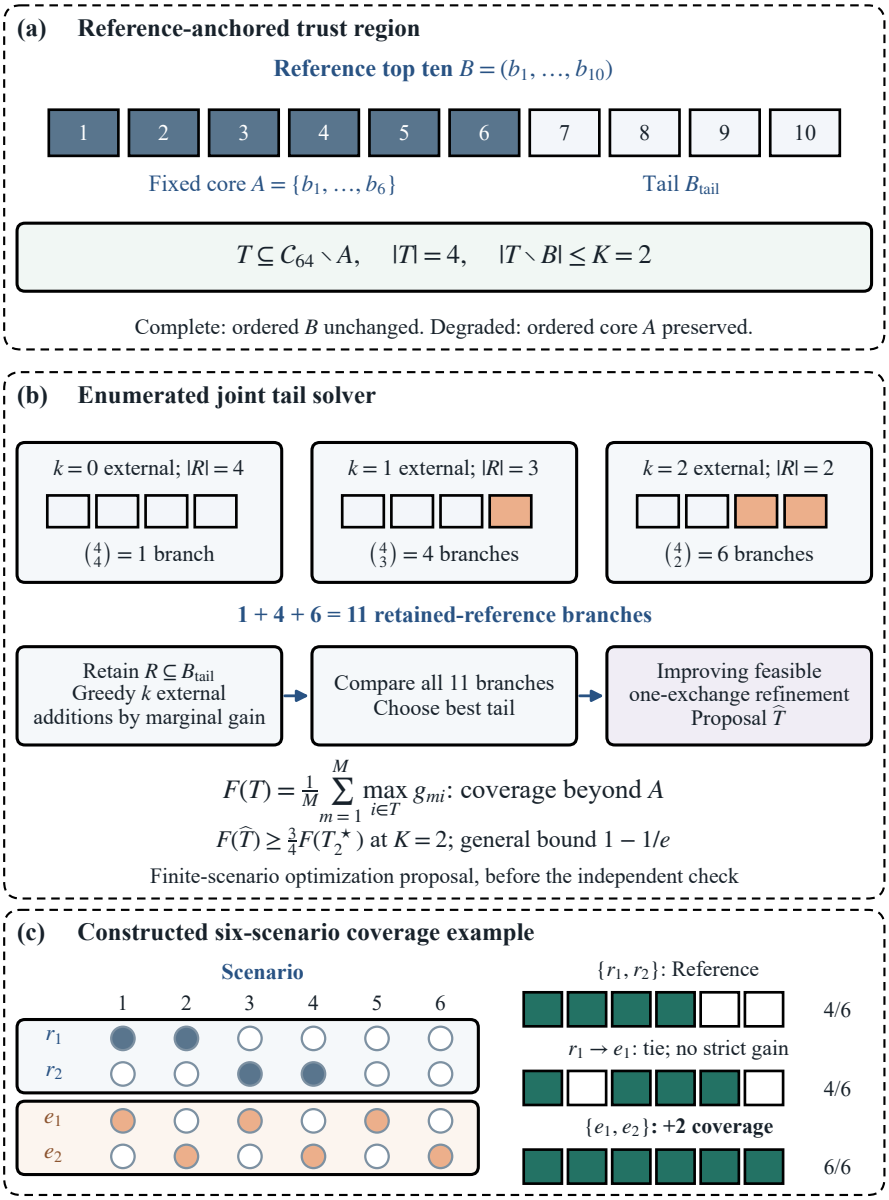


Fig. 3 Reference-preserving tail selection. (a) A fixed six-member core and four tail positions define feasibility. (b) Zero, one or two external additions give 11 retained-reference branches. Each branch is completed greedily; the best branch undergoes improving feasible exchanges. The 3/4 guarantee at $K = 2$ applies to the finite-scenario proposal before the independent check. (c) In the constructed six-scenario example, every single replacement ties Reference coverage, whereas the coordinated pair increases coverage from four to six

parent's proposal and predictive models exclude its outer fold during fitting. Parents and candidates use the simulator family described in Section 3.1. This controlled setting supplies hidden network attributes and a full-pool oracle for separating support loss from retrieval regret. We evaluate transfer to a natural fracture catalogue separately in Section 6.

Each simulated parent occupies a bounded cubic domain measuring 1 m per side. A randomly oriented orthogonal frame defines three finite observation planes clipped to the box. Complete observation retains all requested traces; high fragmentation breaks continuity, and the missing-surface condition removes an entire face. The three conditions share the same parent geometry and observation frame. Feature extraction uses their common physical scale and known poses.

The evaluated cohort contains 9,828 query–condition records, comprising 3,276 complete and 6,552 degraded observations. Pools are constructed from observable inputs, and retrieval policies are finalized on fold 0. The primary endpoint averages the Reference-minus-TraceRIFT regret difference across each parent's two degraded conditions:

$$G_p = \frac{1}{2} \sum_{c \in \mathcal{D}} \left[R_{pc}^{(R)} - R_{pc}^{(T)} \right], \qquad \mathcal{D} = \{\text{fragmented}, \text{missing surface}\}. \tag{10}$$

Positive G_p indicates a reduction in retrieval regret. Complete observations test preservation of the full ordered Reference list.

4.2 Comparators and development settings

TraceRIFT is compared with Reference and two earlier protected-exchange rules. Fixed-donor exchange replaces the rank-ten Reference candidate; Target-guided exchange chooses its donor through target consistency. Both permit one exchange, so these comparisons include differences in selection rules and intervention budgets. The sequential-tail comparator shares TraceRIFT's six-member core, four tail positions and $K = 2$ budget, but uses a strictly improving sequence of single exchanges. This comparison assesses joint search at the same set size and replacement budget.

We additionally compare maximal marginal relevance (MMR; Carbonell and Goldstein 1998) and farthest-first selection (Gonzalez 1985), adapted to the same core and two-addition budget. MMR-R uses Reference ranks for relevance; MMR-P uses singleton expected loss under the same predictive scenarios. Both penalize similarity to retained members. Each external proposal receives the same independent predictive check. The two MMR weights are selected on fold 0, then frozen. These comparisons reuse the archived candidate pools; Supplementary Section S6 gives definitions, parameters and unchecked outputs.

To assess dependence on the Reference ranking, we also use target-consistency-only and observable-discrepancy-only rankings. The latter ranks candidates by increasing root-mean-square discrepancy across geometry, orientation, topology and quality/missingness. Each ranking supplies its own top ten and protected first six. Reference, TraceRIFT and Sequential share that ranking, the original 64-member pool, unchanged predictive scenarios and the two-addition budget. No model or parameter

is refitted or retuned. This retrospective sensitivity analysis uses the existing held-out parents; Supplementary Section S7 defines the rankings and evaluation protocol.

Development ablations test unrestricted joint selection, rank-1-only protection, and removal of the complete-view shield or independent scenario check. We evaluate replacement budgets $K = 0, 1, 2, 3, 4$ using the predefined development criteria. Eligibility includes an exact-oracle recall ratio of at least 0.95 relative to Reference. Parent gain must also have a positive one-sided 95% lower bound. Among eligible budgets, the largest lower bound determines the selected value.

We compare the finite-scenario objective against exhaustive search on 30 fixed degraded queries at each budget from one to four. Both solvers receive identical candidates and predictive scenarios for these comparisons. Supplementary Section S2 reports search counts, development summaries and implementation settings.

4.3 Statistical analysis

We estimate primary effect intervals using 10,000 paired parent-clustered bootstrap replicates. Each resampled parent retains all conditions, with identical resampling indices across methods. We report percentile intervals for paired means and condition- and fold-specific effects. Target-wise and binned summaries are secondary analyses reported with pointwise intervals. Intervals for the additional baseline contrasts are also pointwise.

Candidate support is described by $L_{64}^{\star}$, pairwise target distances and its share of total best-member loss. Target-wise fidelity is the minimum absolute standardized error within each retained set for each coordinate. Gain heterogeneity is summarized by parent-level exceedance curves, condition-wise outcomes and equal-count bins of pool-oracle loss or Reference regret. Nested prefixes evaluate support at increasing candidate counts in retained attempt order. Supplementary Section S3 defines these diagnostics and their aggregation rules.

5 Results

5.1 Separating candidate support from selection loss

The generator produced 64 distinct candidates for every benchmark query across the three observation conditions. Across all 3,532 parents, 847,680 simulation attempts yielded 678,144 retained realizations, including 628,992 candidates in the held-out folds.

Pool-oracle loss and candidate diversity are broadly comparable across observation conditions (Fig. 4). Under degraded observations, mean support loss of approximately 0.272 and TraceRIFT retrieval regret of 0.070 give a total best-member error of about 0.342. Candidate support accounts for 79.6% of this error; the remaining gap measures the loss introduced by shortlist selection.

Across support quintiles, median regret varies less than the upper quantiles, revealing larger missed opportunities in the tail.

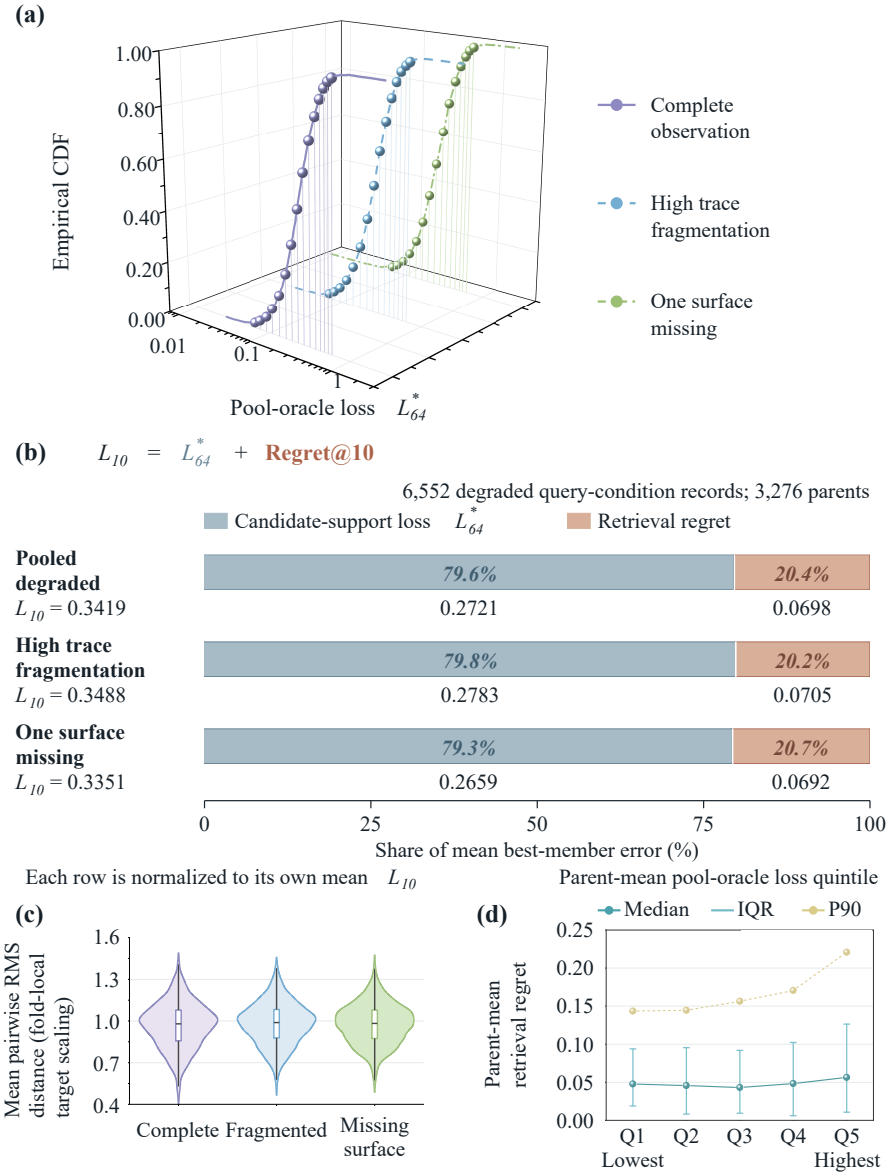


Fig. 4 Candidate support and retrieval loss. (a) Empirical pool-oracle-loss distributions for complete, fragmented and missing-surface observations. (b) Decomposition of mean best-member error over 6,552 degraded queries from 3,276 parents. (c) Pairwise candidate distances in fold-scaled target space. (d) Median, interquartile range and 90th-percentile parent-mean retrieval regret in equal-count support-loss quintiles

5.2 Replacement budget and solver accuracy

The development comparison selects $K = 2$, which gives the larger eligible lower bound on parent gain (Fig. 5). Budgets one and two meet the eligibility criteria; budgets three and four fall below the exact-recall threshold.

The enumerated solver closely matches exhaustive optimization on the checked queries. Across 120 query–budget comparisons, mean relative objective loss is 0.0333% and the maximum is 1.4291%. Agreement is exact in all 30 one-replacement cases, 28 two-replacement cases, 27 three-replacement cases and 28 four-replacement cases.

At $K = 2$, the 8,803 feasible tails are represented by 11 retained-reference branches, within which the solver applies greedy scans and local exchanges. An implementation reusing candidate-column and maximum buffers reproduced every optimization and check record on all 6,552 degraded held-out queries. All 3,276 complete-observation lists also matched exactly. A serial interleaved benchmark on 160 degraded queries, with three repeats per implementation, reduced mean selection time from 19.06 to 5.53 ms (3.44-fold). A separate multi-method run measured 24.71 ms for the original TraceRIFT implementation, 19.51 ms for Sequential, and 1.04, 1.31 and 1.16 ms for MMR-R, MMR-P and farthest-first. Timings include method-specific preprocessing and final checks, excluding shared scenario construction, inference, candidate generation and reprojection. Supplementary Section S6 specifies both runs.

5.3 Reduced regret with the reference core intact

TraceRIFT reduces degraded mean retrieval regret by 22.68% relative to Reference, from 0.090321 to 0.069836 (Table 2; Fig. 6). The paired parent gain is 0.020485, with a 95% interval of [0.018320, 0.022747].

Including shared support loss of approximately 0.2721, total best-member loss decreases from about 0.3624 to 0.3419, a reduction of approximately 5.7%.

Mean regret decreases by 24.04% under fragmentation and 21.25% with a missing surface, supporting the pooled effect across both degraded conditions. Gains across the four held-out folds are 0.022841, 0.022157, 0.019809 and 0.017127. TraceRIFT also reduces mean regret relative to Fixed-donor and Target-guided exchange, with paired gains of 0.013603 [0.011488, 0.015717] and 0.013748 [0.011610, 0.015945], respectively.

TraceRIFT also reduced mean regret relative to MMR-R, MMR-P and farthest-first selection. The paired gains were 0.007420 [0.005541, 0.009351], 0.008366 [0.006602, 0.010181] and 0.019153 [0.017066, 0.021403], respectively, with 95% parent-clustered intervals. These gains compare predictive scenario coverage with relevance–diversity balancing and dispersion in candidate-attribute space.

Relative to Reference, P95 regret falls by 23.37%, from 0.347001 to 0.265895; P90 and P99 fall by 23.26% and 20.59% (Fig. 7). Exact-oracle recall increases from 37.8816% to 39.8199% (+1.9383 percentage points), and near-oracle recall from 53.8462% to 57.6007% (+3.7545 points).

All 3,276 complete observations retain the original ordered top ten, and all 9,828 held-out queries preserve the rank-1 candidate and six-member core. Their rank-1 target and trace errors are consequently identical, with improvements confined to the selected tail.

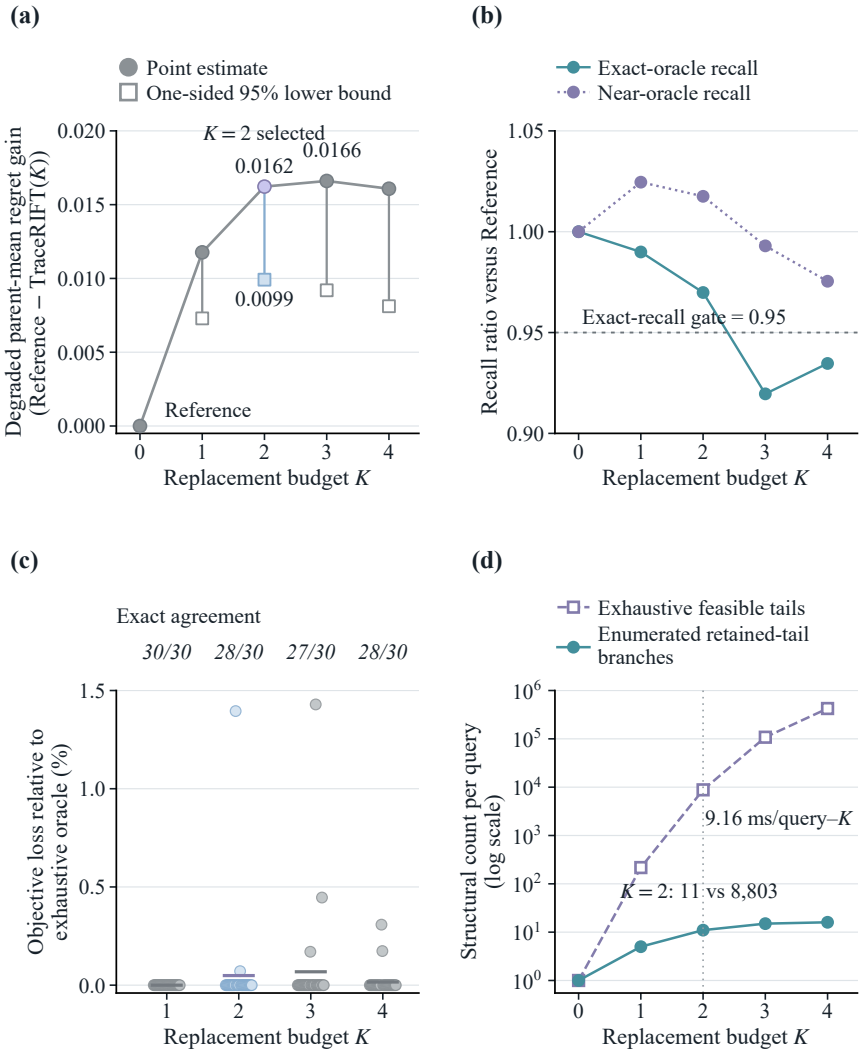


Fig. 5 Replacement-budget selection and solver accuracy on development fold 0 (256 parents). (a) Degraded parent-gain estimates and one-sided 95% lower bounds, connected by vertical segments. (b) Exact- and near-oracle recall ratios, with an exact-recall eligibility threshold of 0.95. (c) Relative objective loss against exhaustive search for 30 queries at each budget; labels give exact-agreement counts. (d) Numbers of feasible tails and retained-reference branches. Budget selection gives $K = 2$

5.4 Attribute coverage and distribution of gains

The retained sets attain lower pooled best-member errors in all eight target coordinates (Fig. 8). Improvements in intensity are smaller than those in orientation, with the relative pattern differing between fragmentation and missing-surface observations.

Table 2 Held-out degraded retrieval results on 6,552 queries from 3,276 parents. Mean and P95 refer to retrieval regret; total loss is best-member attribute error. Recall is a fraction of query records. External baselines use the common predictive check. Fixed-donor and Target-guided allow one exchange; MMR-R, MMR-P, farthest-first, Sequential and TraceRIFT share a six-member core and a two-addition limit.

Method	Mean	P95	Total loss	Exact recall	Near recall
Reference	0.090321	0.347001	0.362420	0.378816	0.538462
Fixed-donor	0.083439	0.317453	0.355538	0.382784	0.549756
Target-guided	0.083584	0.321881	0.355683	0.388126	0.554029
MMR-R	0.077256	0.289792	0.349355	0.376526	0.552656
MMR-P	0.078202	0.299304	0.350301	0.388584	0.563339
Farthest-first	0.088989	0.335183	0.361088	0.372711	0.534341
Sequential	0.069777	0.266553	0.341876	0.397741	0.576007
TraceRIFT	0.069836	0.265895	0.341935	0.398199	0.576007

Under strict-sign classification, approximately 29.3% of parents have positive degraded mean gain and 18.3% have negative gain (Fig. 9). Positive gains extend farther into the magnitude tail, while condition-wise ties account for 71.7% under fragmentation and 71.9% under a missing surface.

Mean gains increase toward the more difficult support bins and are largest in the high-Reference-regret bins. The lower Reference-regret bins include slightly negative effects, consistent with smaller changes where Reference already retains a close candidate.

5.5 Effects of reference protection and joint selection

The development ablations reveal a trade-off between upper-tail improvement and oracle recall (Fig. 10). Unconstrained and rank-1-only selectors lower some upper-tail errors but reduce exact-oracle recall and replace more of the reference list. Six-member-core variants retain more reference members and higher recall while preserving an upper-tail gain. Removing the complete-view shield reduces development mean error but changes the complete-observation list. Removing the independent scenario check leaves the development outputs unchanged because every proposal passes that check.

Joint and same-core sequential selection frequently return different hypotheses with the same best-member error. Their ordered lists differ for 3,684 of 9,828 held-out queries (37.5%), whereas realized regret is identical for 9,682 queries (98.51%). TraceRIFT has lower regret on 75 queries and the sequential selector on 71. For the degraded primary endpoint, the parent-mean contrast, Sequential minus TraceRIFT, is -5.9×10^{-5} with a 95% interval of $[-5.12 \times 10^{-4}, 4.02 \times 10^{-4}]$.

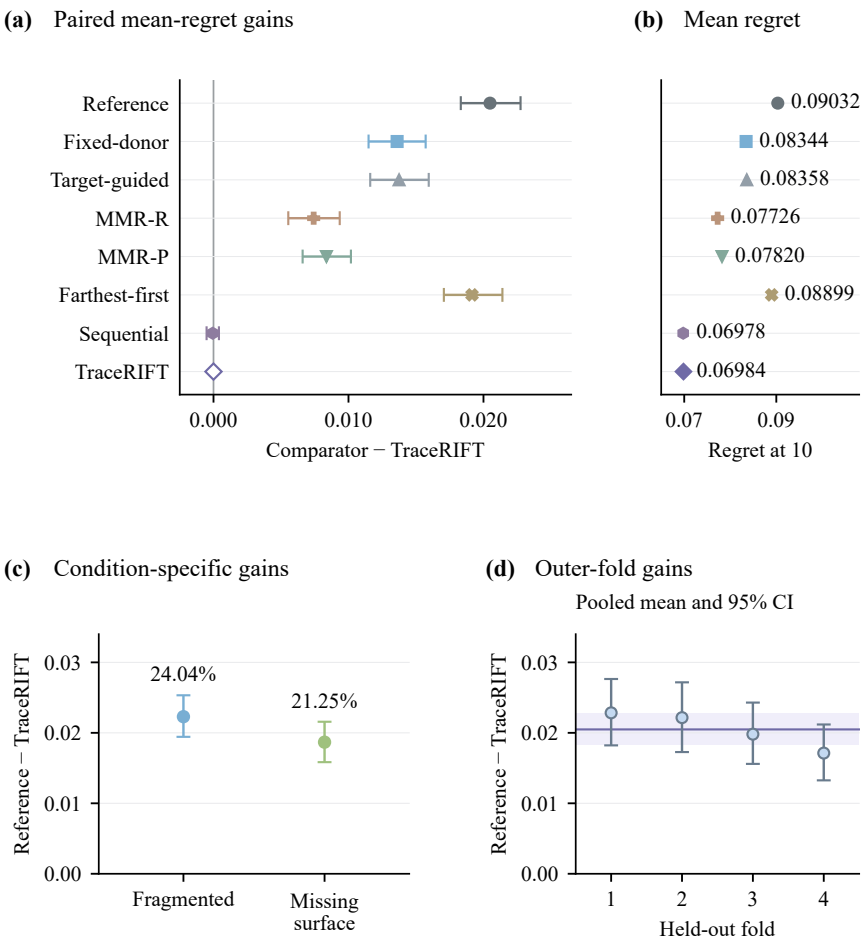


Fig. 6 Held-out mean-regret effects. (a) Comparator-minus-TraceRIFT paired gains for seven comparators, with TraceRIFT shown as a hollow zero-reference marker. Error bars are pointwise two-sided 95% paired parent-bootstrap percentile intervals from 10,000 draws. (b) Absolute degraded mean regret for all eight methods, in the same row order as (a). (c) Reference-minus-TraceRIFT gains under fragmented and missing-surface observations; percentages give relative mean-regret reductions. (d) The same contrast across held-out folds 1–4, with its pooled mean (line) and 95% interval (band). The primary endpoint averages both degraded conditions within each of 3,276 parents, giving 6,552 query records; development fold 0 is excluded. The added MMR and farthest-first comparisons use the common independent predictive check. The Sequential contrast crosses zero

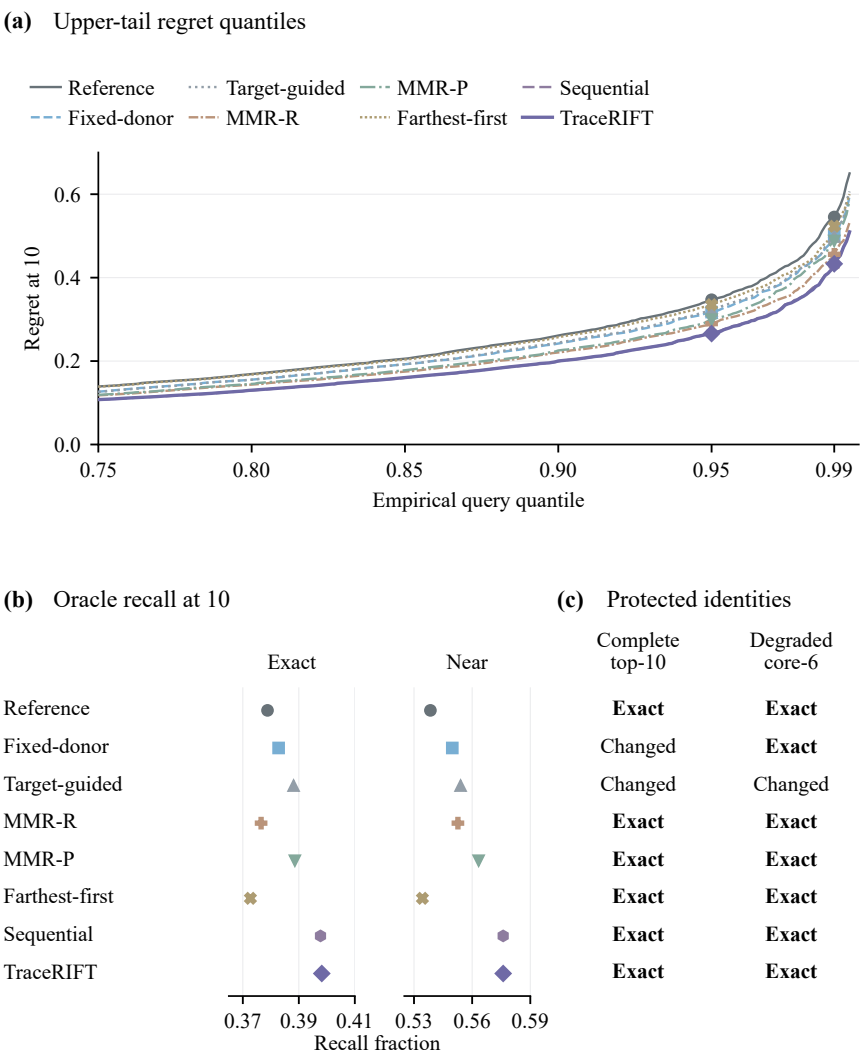
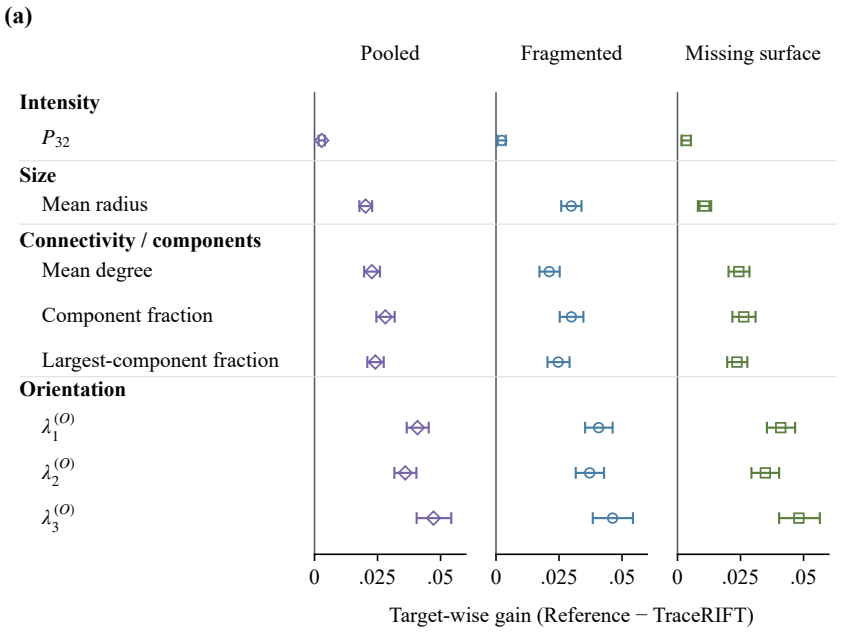


Fig. 7 Upper-tail regret, oracle recall and reference preservation. (a) Empirical upper quantiles of query-level regret for all eight methods; markers identify P95 and P99. Sequential and TraceRIFT nearly coincide. (b) Exact- and near-oracle recall at ten, expressed as fractions of the 6,552 degraded queries; the near-oracle tolerance is 0.05 in standardized discrepancy. The two recall columns use separately labelled scales. (c) Preservation of the ordered complete-observation top ten (3,276 queries) and degraded six-member core (6,552 queries). "Exact" means preservation on every applicable query; "Changed" means at least one differs. Historical one-exchange methods retain their recorded protection outcomes. All methods report ten members from the same query-specific 64-member pools; the added MMR and farthest-first curves use their checked outputs



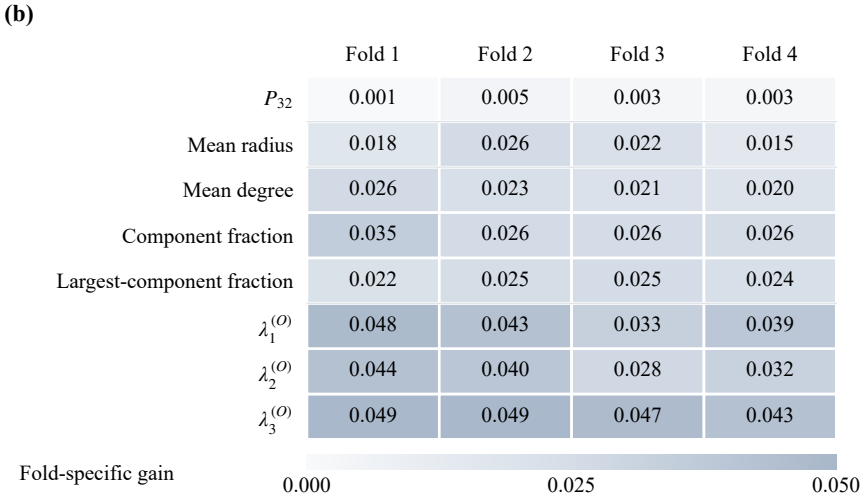


Fig. 8 Target-wise best-member fidelity. (a) Paired reductions in absolute standardized error for all eight attributes, pooled and by degradation condition. (b) Fold-specific mean effects. Positive values favor TraceRIFT. Minima are evaluated separately by coordinate within each retained set, so different attributes may be represented by different candidates. Intervals use parent-level resampling

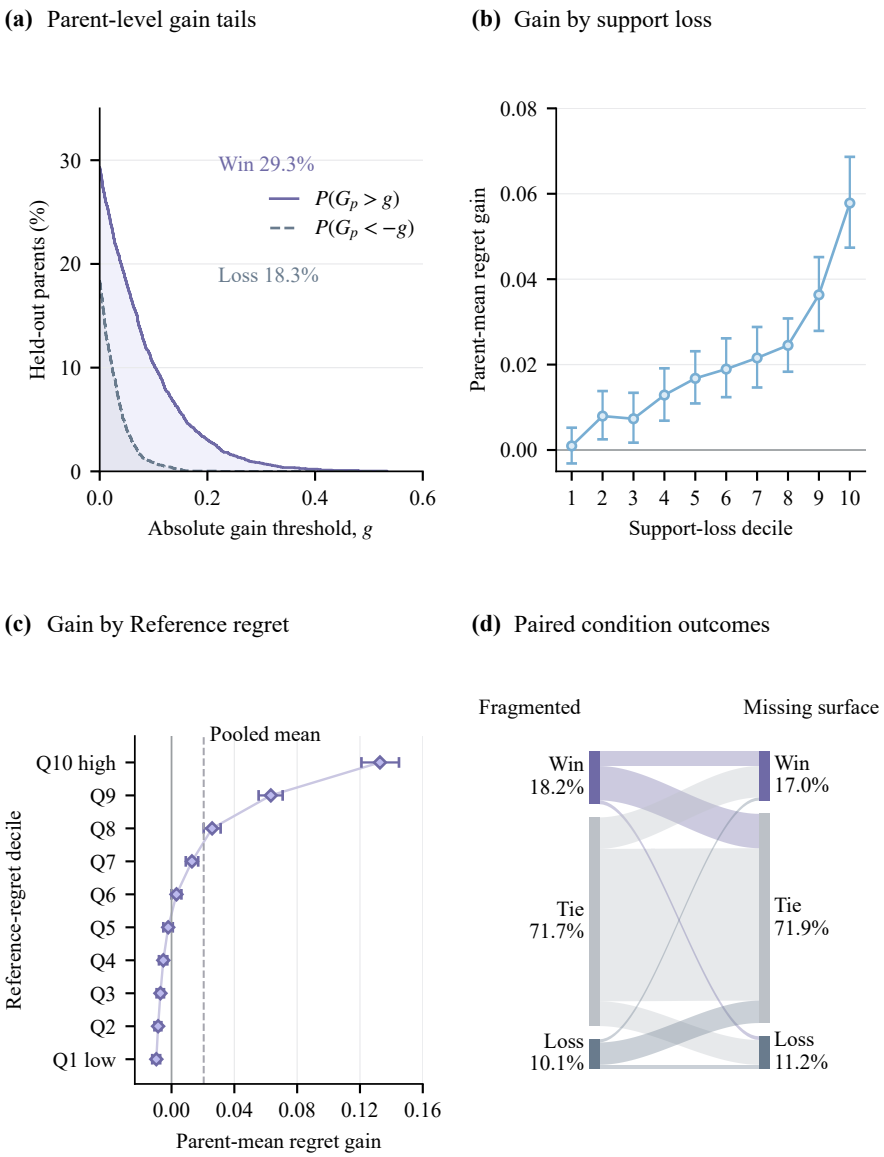


Fig. 9 Distribution of held-out gains. (a) Positive and negative exceedance tails of degraded parent-mean gain, using all held-out parents as a denominator. (b,c) Mean gain in equal-count bins of pool-oracle loss and Reference regret, with parent-based pointwise intervals. (d) Paired condition-wise outcome classes; ribbon widths give parent counts. Panel (a) uses strict-sign pooled gains, whereas (d) classifies each condition with tolerance 10^{-15} (Supplementary Section S3). These bins are descriptive strata defined after evaluation

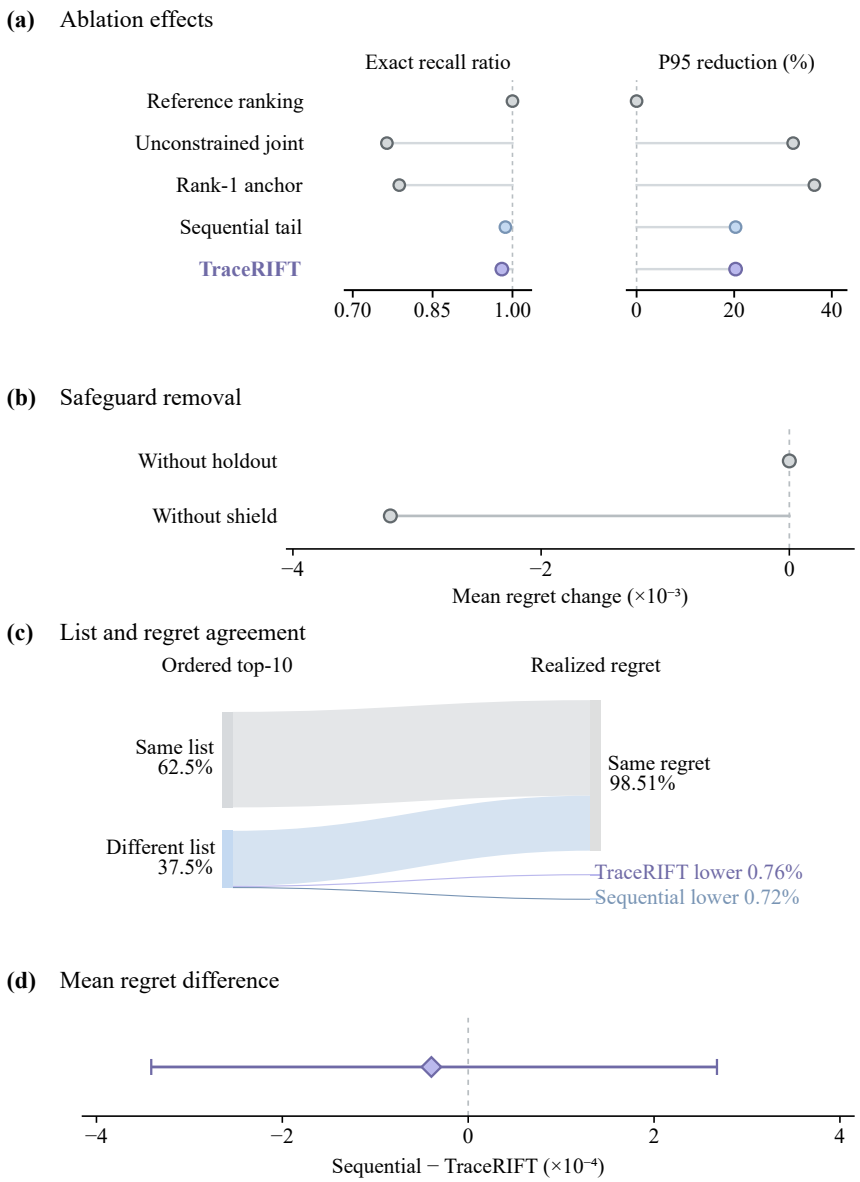


Fig. 10 Reference-protection ablations and solver comparison. (a,b) Development-fold changes in recall, upper-tail regret and mean regret after architectural changes; connectors join point estimates. (c) Joint–sequential ordered-list and realized-regret agreement across 9,828 held-out queries, using one count-to-width scale. (d) All-condition parent-mean regret difference, Sequential minus TraceRIFT, with its 95% parent-clustered interval. This panel averages all three conditions; our primary endpoint averages only fragmented and missing-surface conditions

Table 3 Reference-ranking sensitivity on the same 3,276 held-out parents. Entries are degraded mean regret. Each ranking defines its own top ten and six-member core; TraceRIFT and Sequential allow at most two additions. Gain is Reference minus TraceRIFT, with a pointwise 95% paired parent-bootstrap interval.

Ranking	Reference	TraceRIFT	Sequential	Gain [95% CI]
Borda	0.0903	0.0698	0.0698	0.0205 [0.0183, 0.0227]
Target only	0.1063	0.0750	0.0753	0.0313 [0.0284, 0.0341]
Observable RMS	0.1040	0.0797	0.0800	0.0243 [0.0221, 0.0264]

5.6 Sensitivity to the Reference ranking

Replacing Borda fusion with target-consistency-only or observable-discrepancy-only ranking changes the ordered Reference top ten for every query. TraceRIFT reduces degraded mean regret relative to these References by 29.42% and 23.32%, respectively (Table 3). The positive paired intervals support improvement under both alternative rankings. Absolute TraceRIFT regret remains lowest with the original Borda Reference, indicating that the protected starting interpretation still affects the final ensemble. Within each ranking, TraceRIFT and Sequential yield numerically close mean regrets, and all paired contrast intervals span zero. The gains over the corresponding References are consistent with the predictive-coverage objective shared by both selectors. Supplementary Section S7 reports all nine ranking–method combinations.

5.7 Additional support from larger candidate pools

Increasing the retained prefix from 8 to 64 candidates lowers mean pool-oracle loss (Fig. 11). Early increments produce larger reductions, with further improvement from 48 to 64 candidates at the largest tested pool size.

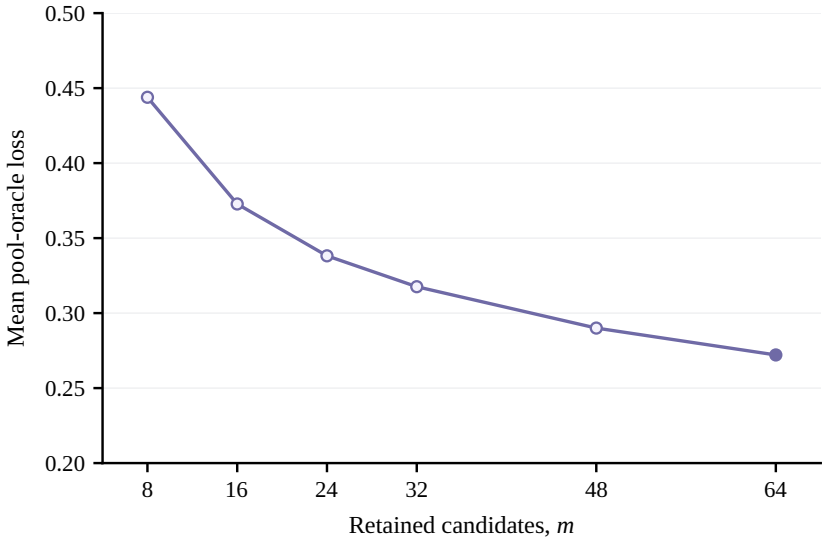


Fig. 11 Candidate-pool size and oracle support. Mean pool-oracle loss for retained prefixes of $m = 8, 16, 24, 32, 48, 64$ candidates in original attempt order. Each point uses nested subsets of common generated pools

6 Transfer to the Wadi Birk catalogue

We applied the unchanged workflow to the public Wadi Birk fracture catalogue derived from a mapped outcrop (Panara et al. 2024a,b). Catalogue centers, normals and radii define disc networks within 24 analysis cells, each measuring 20 m. Four deterministic frames per cell and five observation conditions per frame yield 480 virtual queries. These queries test candidate construction, Reference preservation and predictive transfer under the catalogue's fracture distribution.

The workflow produced 64-member pools for all 480 queries. It preserved the ordered Reference list for all 96 complete observations, rank 1 for all 480 queries and the six-member core for all 384 degraded observations (Fig. 12).

For predictive transfer, field-target regret decreased from 0.133346 to 0.129739; the paired gain was 0.003606 with a 95% interval of $[-0.005004, 0.013503]$. Held-out virtual-trace regret decreased from 0.159305 to 0.147995, with a gain of 0.011310 $[-0.013581, 0.038994]$.

Five catalogue cells were empty, and 185 of the 288 withheld surfaces contained no catalogue traces. The remaining 103 surfaces spanned 17 cells. Mean trace-length coverage on these nonempty surfaces was approximately 3.01% under the specified matching procedure. Their post-hoc parent-mean gain was 0.001162, with a 95% interval of $[-0.002446, 0.005132]$. Supplementary Section S4 reports the separate empty- and nonempty-case analyses.

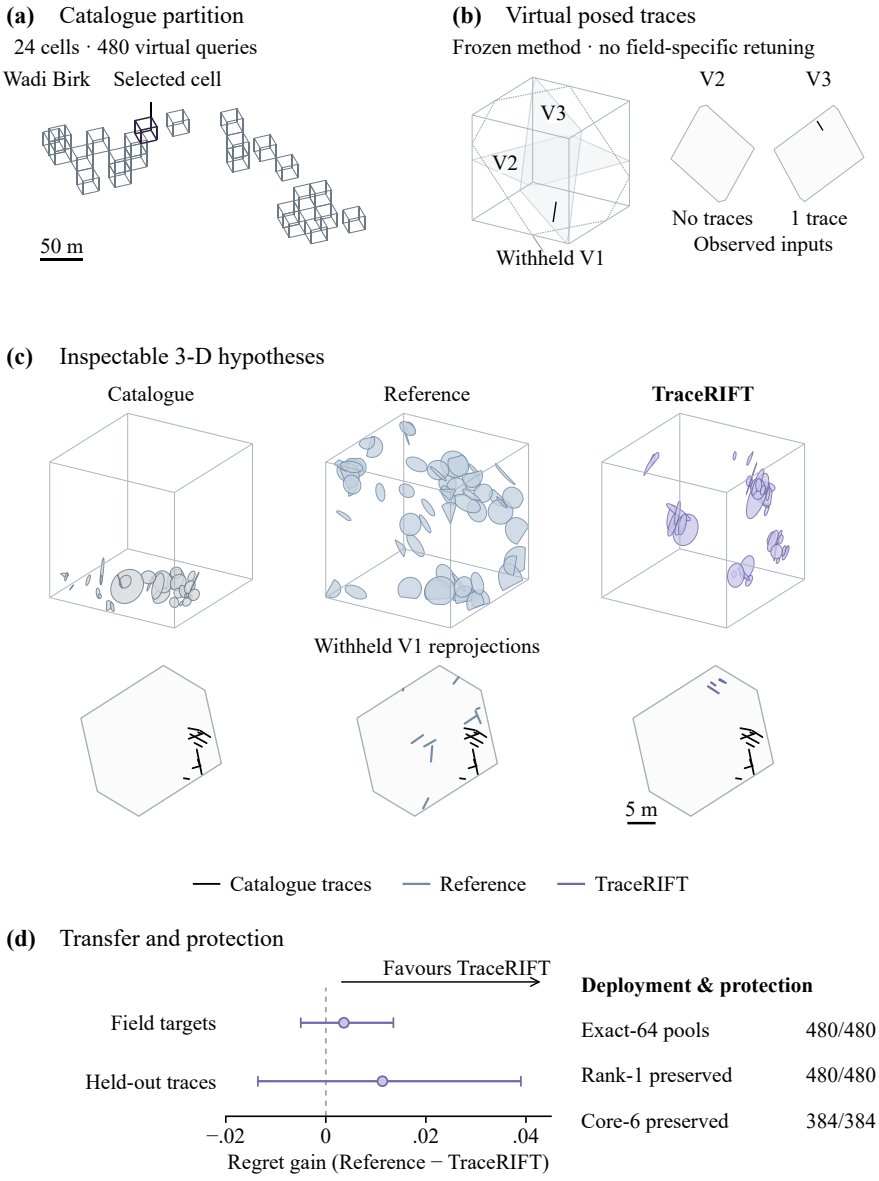


Fig. 12 Transfer to Wadi Birk catalogue data. (a) Spatial layout of the 24 analysis cells used for the Wadi Birk fracture catalogue, with the selected cell highlighted (data from Panara et al. 2024b). (b) Catalogue-derived virtual posed traces. (c) Catalogue and candidate networks with withheld-window reprojections. Reference and TraceRIFT examples are retrospectively identified best members of their retained sets for a fixed display query. (d) Parent-level gain estimates with 95% intervals and reference-preservation counts. This experiment comprises 24 cells, 480 queries, 96 complete observations and 384 degraded observations

7 Discussion

7.1 Choosing complementary fracture-network hypotheses

TraceRIFT makes ensemble reduction an explicit decision within geological interpretation, with candidates valued by their contribution to the retained set. The common-pool comparison isolates shortlist construction from changes in candidate generation.

MMR balances a candidate's individual relevance with its similarity to retained hypotheses (Carbonell and Goldstein 1998). Separation in candidate space, however, need not fill an uncovered region of plausible target attributes. Scenario coverage directly values how much each addition reduces best-member error across the predictive target distribution. MMR-P shares that distribution and its sampled targets, so the comparison supports evaluating coverage of the whole ensemble rather than individual relevance with pairwise redundancy. The similar regret of the coverage-based Sequential selector across all three Reference rankings further supports this interpretation.

The retained geometries allow an interpreter to compare observed-window reprojections and structural attributes while inspecting tail additions alongside the Reference core. Additional traces or borehole observations can screen these alternatives against new evidence, linking shortlist construction to subsequent geological interpretation.

Best-member loss measures whether useful configurations survive ensemble reduction, with the closest member identified retrospectively during evaluation. At deployment, the method returns an unweighted ensemble of ten networks for candidate-specific simulations and conditional response ranges.

7.2 Reference preservation as a bounded update

The protected core and replacement budget specify the permitted change to an existing interpretation while leaving room for complementary alternatives. The two-addition limit preserves at least eight identities from the Reference top ten, while complete observations keep their ordered list. Reference quality therefore affects both the protected hypotheses and the alternatives reachable within the replacement budget.

Enumeration exploits the small replacement budget to express constrained selection as a sequence of cardinality-constrained subproblems. Improvements over target-consistency-only and observable-discrepancy-only References show that the selection gain extends beyond the original Borda fusion. Their different absolute errors also show why the starting Reference remains consequential.

7.3 Separating generation, selection and geological transfer

The support–selection decomposition identifies two routes to better retained networks and provides a separate metric for each. Improved generation lowers the pool-oracle error floor; improved selection closes the gap between that floor and the retained set. Process-informed growth models, richer fracture shapes and additional observations can be evaluated through their effects on candidate support before comparing selection rules.

The target representation determines which geological distinctions the selector rewards through intensity, size, graph structure and orientation concentration. Orientation eigenvalues describe concentration rather than tensor directions, while degree and component fractions summarize connectivity rather than spatially directed pathways. Flow or mechanical applications require response-specific evaluation with aperture, stress dependence, constitutive behavior and boundary conditions specified for the retained geometries.

The controlled benchmark provides exact parent targets, whereas Wadi Birk supplies catalogue-derived virtual observations for testing the unchanged workflow. Independently registered multi-surface or volumetric observations would help distinguish the effects of fracture shape, spatial organization, catalogue truncation and observation geometry. The decomposition could then identify whether added evidence improves candidate support, selection, or both under these observation conditions.

8 Conclusions

TraceRIFT selects complementary fracture networks from a fixed candidate pool while preserving a prescribed Reference core. On 3,276 held-out simulated parent DFNs, it reduced degraded mean retrieval regret by 22.68% and its 95th percentile by 23.37%. Mean regret was also lower than for two constrained MMR variants and farthest-first selection, and reductions relative to Reference persisted under both alternative rankings. Oracle recall increased, and every protected core member was preserved across the evaluated observation conditions. The support–selection decomposition locates these gains within shortlist construction, providing separate criteria for improving candidate generation and ensemble retention. Together with bounded Reference updates and finite-scenario coverage optimization, it makes this selection step measurable within DFN interpretation.

Supplementary information

Online Resource 1: Supplementary methods and results, including the optimization proof, implementation settings, evaluation metrics, Wadi Birk transfer analyses and reproducibility materials.
Online Resource 2: Additional baseline comparison code, an equivalent optimized selector, frozen protocols, development settings, selected policies, source tables and audit records.
Online Resource 3: Reference-ranking sensitivity code, frozen policies, full results and independent audits, together with updated Figure 6 and Figure 7 source data and plotting scripts.

Statements and Declarations

Data and code availability

The accompanying reproducibility package contains the observation and simulator interfaces, trained proposal and predictor assets, Reference ranker, set selector, parent-level evaluation arrays, and figure data and scripts. It provides an end-to-end example and separate procedures for reproducing the reported statistics (Supplementary Section S5). Source code is available at https://github.com/SureWangchd/TraceRIFT-CG; version v0.1.1 and its registered model and data assets are available at https://github.com/SureWangchd/TraceRIFT-CG/releases/tag/v0.1.1. Additional baseline scripts, the equivalent optimized selector, policy arrays and source tables are supplied in Online Resource 2. Reference-ranking sensitivity results and the updated comparison-figure sources are supplied in Online Resource 3. The Wadi Birk source data are available from Panara et al. (2024b) under their original license.

Author contributions

Shuo Wang: Conceptualization, Methodology, Software, Formal analysis, Investigation, Visualization, Writing – original draft.
Chen Guo: Conceptualization, Methodology, Supervision, Project administration, Funding acquisition, Writing – review & editing.
Bingxin Yang: Methodology, Formal analysis, Validation, Writing – review & editing.
Keran Hao: Data curation, Investigation, Visualization, Validation.
Yu Zhang: Investigation, Resources, Data curation, Validation.
Fuqiang Zhang: Validation, Writing – review & editing.

Funding

This work was supported by the National Natural Science Foundation of China (Grant 42374154).

Competing interests

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

References

Bemis SP, Micklethwaite S, Turner D, et al (2014) Ground-based and UAV-based photogrammetry: A multi-scale, high-resolution mapping tool for structural geology and paleoseismology. Journal of Structural Geology 69:163–178. https://doi.org/10.1016/j.jsg.2014.10.007

Bonnet E, Bour O, Odling NE, et al (2001) Scaling of fracture systems in geological media. Reviews of Geophysics 39(3):347–383. https://doi.org/10.1029/1999RG000074

Carbonell J, Goldstein J (1998) The use of MMR, diversity-based reranking for reordering documents and producing summaries. In: Proceedings of the 21st Annual International ACM SIGIR Conference on Research and Development in Information Retrieval, pp 335–336, https://doi.org/10.1145/290941.291025

Cranmer K, Brehmer J, Louppe G (2020) The frontier of simulation-based inference. Proceedings of the National Academy of Sciences 117(48):30055–30062. https://doi.org/10.1073/pnas.1912789117

Dershowitz WS, Einstein HH (1988) Characterizing rock joint geometry with joint system models. Rock Mechanics and Rock Engineering 21(1):21–51. https://doi.org/10.1007/BF01019674

Dershowitz WS, Herda HH (1992) Interpretation of fracture spacing and intensity. In: Proceedings of the 33rd U.S. Symposium on Rock Mechanics, pp 757–766

Dong S, Fu K, Wang L, et al (2026) Intelligent characterization of fractures using 3D point clouds of outcrops generated from 2D photos. Computers & Geosciences 210:106136. https://doi.org/10.1016/j.cageo.2026.106136

Durkan C, Bekasov A, Murray I, et al (2019) Neural spline flows. In: Advances in Neural Information Processing Systems, vol 32. Curran Associates, Inc., URL https://proceedings.neurips.cc/paper_files/paper/2019/hash/7ac71d433f282034e088473244df8c02-Abstract.html

Geurts P, Ernst D, Wehenkel L (2006) Extremely randomized trees. Machine Learning 63(1):3–42. https://doi.org/10.1007/s10994-006-6226-1

Gonzalez TF (1985) Clustering to minimize the maximum intercluster distance. Theoretical Computer Science 38:293–306. https://doi.org/10.1016/0304-3975(85)90224-5

Guo J, Hu J, Zheng J, et al (2024) Estimation of the three-dimensional density of the universal elliptical fractures based on one-dimensional or two-dimensional density. International Journal of Rock Mechanics and Mining Sciences 174:105652. https://doi.org/10.1016/j.ijrmms.2024.105652

Kim J, Choi J, Yoon DH, et al (2026) Genetic discrete fracture network modeling by 3-D polygonal growth-based simulation calibrated with trace map observations. International Journal of Rock Mechanics and Mining Sciences 204:106585. https://doi.org/10.1016/j.ijrmms.2026.106585

Kulesza A, Taskar B (2012) Determinantal point processes for machine learning. Foundations and Trends in Machine Learning 5(2–3):123–286. https://doi.org/10.1561/2200000044

Lei Q, Latham JP, Tsang CF (2017) The use of discrete fracture networks for modelling coupled geomechanical and hydrological behaviour of fractured rocks. Computers and Geotechnics 85:151–176. https://doi.org/10.1016/j.compgeo.2016.12.024

Mauldon M, Dunne WM, Rohrbaugh MBJr. (2001) Circular scanlines and circular windows: New tools for characterizing the geometry of fracture traces. Journal of Structural Geology 23(2–3):247–258. https://doi.org/10.1016/S0191-8141(00)00094-8

Nemhauser GL, Wolsey LA, Fisher ML (1978) An analysis of approximations for maximizing submodular set functions—i. Mathematical Programming 14:265–294. https://doi.org/10.1007/BF01588971

Panara Y, Menegoni N, Finkbeiner T, et al (2024a) High-resolution analysis of 3D fracture networks from digital outcrop models, correlation to plate-tectonic events and calibration of subsurface models (Jurassic, Arabian Plate). Marine and Petroleum Geology 167:106998. https://doi.org/10.1016/j.marpetgeo.2024.106998

Panara Y, Menegoni N, Finkbeiner T, et al (2024b) High-resolution digital outcrop models of late Jurassic outcrop, Wadi Birk, Saudi Arabia. Mendeley Data, Version 1, https://doi.org/10.17632/9bpcypmh2j.1

Papamakarios G, Nalisnick E, Rezende DJ, et al (2021) Normalizing flows for probabilistic modeling and inference. Journal of Machine Learning Research 22(57):1–64

Racolte G, Marques A, Sales V, et al (2026) Towards three-dimensional discrete fracture network modeling using integrated multidimensional outcrop data. Scientific Reports 16:10087. https://doi.org/10.1038/s41598-026-37359-4

Sturzenegger M, Stead D (2009) Close-range terrestrial digital photogrammetry and terrestrial laser scanning for discontinuity characterization on rock cuts. Engineering Geology 106(3–4):163–182. https://doi.org/10.1016/j.enggeo.2009.03.004

Sun Z, Ma L, Li Q, et al (2026) Three-dimensional discrete fracture network identification based on deep learning and reversible jump Markov chain Monte Carlo algorithm. International Journal of Rock Mechanics and Mining Sciences 197:106327. https://doi.org/10.1016/j.ijrmms.2025.106327

Teng Z, Wu H, Zhang K, et al (2026) Projection-based low-dimensional parameterization method for three-dimensional discrete fracture networks through generative adversarial network. Journal of Rock Mechanics and Geotechnical Engineering https://doi.org/10.1016/j.jrmge.2026.02.019

Xu C, Dowd P (2010) A new computer code for discrete fracture network modelling. Computers & Geosciences 36(3):292–301. https://doi.org/10.1016/j.cageo.2009.05.012

Zhu W, Lei G, He X, et al (2022) Can we infer the percolation status of 3-D fractured media from 2-D outcrops? Engineering Geology 302:106648. https://doi.org/10.1016/j.enggeo.2022.106648

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- ESM1.pdf
- ESM2.zip
- ESM3.zip