https://doi.org/10.5194/egusphere-2026-280
Preprint. Discussion started: 12 February 2026

Past the first date: Resolving successive lead-loss episodes in zircon

Lucy M. Mathieson[1] and Christopher L. Kirkland[1]

[1]School of Earth and Planetary Sciences, Timescales of Mineral Systems Group, MinEx Co-operative Research Centre, Curtin University, Perth, WA, Australia

Correspondence: Lucy M. Mathieson (l.mathieson@postgrad.curtin.edu.au)

Abstract. Zircon uranium–lead (U–Pb) geochronology is a cornerstone of Earth science for resolving the timing of deep-time processes, yet the U–Pb system in zircon does not always behave as a closed chronometer. Selective loss of radiogenic Pb can shift isotopic ages and generate discordant analyses that record later alteration, fluid–rock interaction, or heating. Many zircon datasets preserve evidence for more than one Pb-loss episode, yet many Pb-loss modelling approaches return a single "best" loss time. Here, we extend the concordant–discordant comparison (CDC) framework to recover multi-episode Pb-loss histories by using concordant analyses as a reference age distribution and scoring candidate Pb-loss times by how well reconstructed discordant ages reproduce that reference.

The updated workflow can partition discordant analyses into internally coherent sub-arrays, retain reproducible local optima across Monte Carlo realisations rather than collapsing each realisation to a single optimum, and summarise statistically supported candidates as an ensemble catalogue with empirical 95 % uncertainty intervals and support values that quantify run-to-run stability.

Synthetic benchmarks spanning single-stage and two-stage discordance geometries across three scatter tiers show that CDC achieves lower overall median absolute error and higher event-wise coverage than a discordia-likelihood discordance-dating (DD) approach. CDC performs best for single-stage benchmarks and for mixtures in which episodes remain well separated, whereas DD variants are more accurate and attain higher coverage in higher-scatter two-stage cases where likelihood surfaces are broad and competing modes occur. By reporting reproducible local optima rather than a single optimum, the CDC ensemble catalogue enables explicit recovery of multi-episode Pb-loss histories from discordant zircon U–Pb populations. Future work will focus on strongly overlapping multi-episode scenarios that remain difficult to deconvolve when candidate Pb-loss ages are tested one at a time.

1 Introduction

Zircon U–Pb geochronology supports a broad spectrum of Earth science investigations, from constraining the timing of crustal growth to quantifying sediment provenance and metamorphic overprinting (e.g. Belousova et al., 2010; Kuper et al., 2024; Gehrels, 2014). U–Pb age determination relies on concordant growth of the $^{238}\mathrm{U}$–$^{206}\mathrm{Pb}$ and $^{235}\mathrm{U}$–$^{207}\mathrm{Pb}$ decay systems; however, radiation damage and fluid-assisted mobilisation of radiogenic Pb can remove Pb from the crystal lattice, generating discordant analyses (e.g. Nasdala et al., 1998; Palenik et al., 2003; Rizvanova et al., 2000). In the best case, Pb-loss is recent and has no effect on the $^{207}\mathrm{Pb}$–$^{206}\mathrm{Pb}$ age, but in many cases discordance complicates interpretations and, in the worst case,

https://doi.org/10.5194/egusphere-2026-280
Preprint. Discussion started: 12 February 2026

leads to the outright rejection of data. Nonetheless, this discordant component may retain valuable geological information about fluid-alteration histories that would otherwise remain concealed (e.g. Kirkland et al., 2017, 2020; Morris et al., 2015; Mathieson et al., 2025a, b; Reimink et al., 2016, 2025).

Discordance can stem from several, often co-operative, processes. Radiation damage from α-recoil gradually amorphises U-rich zones and once those metamict domains interlink, circulating fluids below $\sim 250\,^{\circ}\mathrm{C}$ can readily strip radiogenic Pb (e.g. Herrmann et al., 2021; Geisler et al., 2002, 2007). Parent-daughter balance may also be perturbed by post-crystallisation U ingress or coupled U-gain/Pb-loss reactions (e.g. Grauert et al., 1974; Seydoux-Guillaume et al., 2015; Garber et al., 2020; Andersen et al., 2022). Radiogenic Pb itself can segregate into nano-sized clusters that are spatially dissociated from their crystallisation isotope and behave as a mobile phase during alteration (Arcuri et al., 2020; Kusiak et al., 2015, 2018, 2023). Further complexity arises when laser-ablation volumes sample mixtures of age-distinct domains within composite analyses (e.g. Mezger and Krogstad, 1997; Schoene, 2014), or when fluid-assisted dissolution–reprecipitation and overgrowth partially reset the isotopic clock but also trap inherited (common) Pb (e.g. Geisler et al., 2007; Hoskin et al., 2003; Rubatto et al., 2008). Shock microstructures and crystal-plastic deformation widen diffusion pathways, leaving zircon susceptible to later pulses of Pb-loss (e.g. Moser et al., 2011; Reddy et al., 2006) or may liberate elements necessary for recrystallisation (e.g. Zhao et al., 2025). Because these mechanisms act at different times and length scales, many operating well below magmatic temperatures, zircon populations commonly record multiple disturbance signatures, producing complex discordance patterns.

Strategies to manage discordance include physical pretreatment methods, such as chemical abrasion thermal ionisation mass spectrometry (CA-TIMS) (e.g. Crowley et al., 2014; Mattinson, 2005), as well as computational approaches designed to recover meaningful geological signals from discordant data (e.g. Kirkland et al., 2017; Mathieson et al., 2025a, b; Morris et al., 2015; Reimink et al., 2016; Sharman and Malkowski, 2023). The former mitigates Pb-loss by selectively dissolving highly damaged zones prior to analysis, but the procedure is perhaps less suited to the high throughput and spatially resolved nature of modern analytical methods such as secondary ion mass spectrometry (SIMS) and laser-ablation inductively coupled plasma mass spectrometry (LA-ICP-MS) (e.g. Crowley et al., 2014; Gehrels, 2014; Mattinson, 2005; Schoene, 2014). Consequently, computational methods have increasingly been adopted to leverage discordant data as a source of geological information, including probabilistic Bayesian frameworks (Morris et al., 2015) developed into the concordant–discordant comparison (CDC) test (Kirkland et al., 2017, 2020; Mathieson et al., 2025a, b), probabilistic discordia modelling (Reimink et al., 2016), and logit-normal statistical modelling (Sharman and Malkowski, 2023). Each of these methods provides valuable insights into discordant zircon data; however, many share a fundamental limitation in assuming or emphasising a single Pb-loss event, thereby struggling to accurately resolve more complex scenarios involving multiple episodes of Pb-loss.

We developed a computational method in Mathieson et al. (2025b) wherein candidate Pb-loss ages are evaluated by how well they reconcile discordant analyses with the known concordant age population (i.e., those within analytical uncertainty of concordia). This method, based on the concordant–discordant comparison (CDC) (Morris et al., 2015; Kirkland et al., 2017), employs the Kolmogorov–Smirnov (K–S) statistic as a quantitative goodness-of-fit measure (Massey, 1951). Specifically, discordant analyses are "reconstructed" to hypothetical pre-loss ages (i.e., the ages they would exhibit if no Pb-loss had occurred since the candidate Pb-loss event) and the resulting distribution is statistically compared against the concordant age distribu-

https://doi.org/10.5194/egusphere-2026-280
Preprint. Discussion started: 12 February 2026

tion. The candidate Pb-loss age that minimises the K–S distance was considered the optimal solution. The algorithm selects only a single "optimal" Pb-loss age (the global minimum of the K–S dissimilarity surface), even if the data contain several plausible minima produced by successive Pb-loss episodes. A kernel density estimate (KDE) of run-level optima can appear multimodal, hinting at secondary events, but local minima that never attain global status within any individual run remain invisible. In addition, discordant analyses from heterogeneous sub-populations are evaluated *en masse*, leaving the method vulnerable to compromise solutions when one population dominates the K–S statistic.

Kirkland et al. (2017) evaluated the CDC test using synthetic zircon populations subjected to various Pb-loss scenarios of increasing complexity, ranging from single-stage, normally distributed Pb-loss to multi-population mixtures experiencing up to three consecutive loss events. Their evaluation demonstrated that the CDC approach faithfully recovers the oldest genuine episode of Pb-loss, provided at least a subset of analyses records that event as single-stage discordance. However, the same modelling campaign revealed two limitations encountered in multistage scenarios: firstly, when all analyses have experienced more than one episode of Pb-loss, the CDC method may produce secondary minima at spurious ages reflecting statistical artefacts from mixed discordia intercepts; secondly, the CDC test alone cannot internally differentiate genuine geological Pb-loss signals from statistical artefacts resulting from complex data mixing. Consequently, the study concluded that the CDC test, in its single-minimum incarnation, cannot deconvolve superimposed Pb-loss events without either demonstrably preserved single-stage domains in a fraction of the sampled population or supplementary geological constraints.

Here we directly address these methodological challenges by implementing three algorithmic improvements to our earlier approach: (i) preserving all significant local optima from each Monte Carlo realisation (local maxima in goodness, equivalently local minima in K–S distance) rather than collapsing the interpretation to a single global optimum, (ii) introducing unsupervised Gaussian mixture clustering to partition discordant analyses into internally coherent sub-populations before the K–S test, preventing a numerically dominant cluster from masking minority signals, and (iii) developing an ensemble-level peak-picking protocol that integrates run-level minima, applies objective statistical criteria (support and height thresholds), and outputs a catalogue of Pb-loss ages with 95 % confidence intervals derived from the 2.5th to 97.5th percentiles of contributing ages. Thus, the new computational pipeline retains the proven ability of the CDC test to identify primary Pb-loss events while improving its capability to detect additional, geologically meaningful episodes when the data warrant that complexity. We validate these advancements using synthetic datasets, demonstrating that our updated approach can resolve multistage Pb-loss complexities that were previously difficult to detect with single-solution methods (Mathieson et al., 2025a, b; Reimink et al., 2016, 2025).

2 Methods

2.1 Baseline CDC workflow

We use the Monte Carlo Concordant–Discordant Comparison (CDC) procedure as coded in Mathieson et al. (2025b) which evaluates candidate Pb-loss ages on a fixed grid and compares reconstructed upper-intercept ages from discordant analyses against the empirical age distribution of concordant analyses. In brief, the method proceeds through six key steps.

https://doi.org/10.5194/egusphere-2026-280
Preprint. Discussion started: 12 February 2026

(i) **Uncertainty propagation.** For each Monte Carlo realisation, each analysis is jittered according to its reported 1σ uncertainties. Independent Gaussian deviates are drawn for the measured ratios $^{238}\mathrm{U}/^{206}\mathrm{Pb}$ and $^{207}\mathrm{Pb}/^{206}\mathrm{Pb}$.

(ii) **Concordant subset.** Analyses are classified as concordant or discordant using one of two user-selected concordance tests: (a) percentage–discordance cutoff, or (b) a concordia–intersection test in which an analysis is concordant if its error ellipse intersects concordia. The concordant subset defines the empirical CDF

$$F_{\mathrm{con}}(t) \;=\; \frac{1}{n_{\mathrm{con}}} \sum_{j=1}^{n_{\mathrm{con}}} \mathbf{1}\Big\{T_j^{(\mathrm{con})} \le t\Big\}. \tag{1}$$

(iii) **Pb-loss grid search.** Within each Monte Carlo run the algorithm steps through a regular Pb-loss age grid $\{t_i\}$ defined by the modelling window (Sect. 2.2). For the synthetic benchmarks we use 1 to 2000 Ma with $N{=}200$ nodes by default, i.e. $\approx 10\,\mathrm{Ma}$ spacing unless otherwise specified.

(iv) **Discordant analysis reconstruction.** For each t_i every discordant analysis is back-projected to concordia using a lower intercept at t_i to obtain a reconstructed upper-intercept age $T_{\mathrm{UI},k}(t_i)$ for analysis k. Let $\mathcal{I}_{\mathrm{val}}(t_i)$ be the indices that yield a valid reconstruction at t_i with size $m(t_i)$. The reconstructed-age CDF used in the two-sample comparison is

$$F_{\mathrm{UI}}(t';t_i) \;=\; \frac{1}{m(t_i)} \sum_{k\in\mathcal{I}_{\mathrm{val}}(t_i)} \mathbf{1}\{T_{\mathrm{UI},k}(t_i) \le t'\}. \tag{2}$$

For a trial lower-intercept age t_i, a discordant analysis admits a *valid* reconstructed upper-intercept age $T_{\mathrm{UI}}(t_i)$ if back-projection to concordia yields a finite intersection consistent with the user's modelling window. Analyses that fail this criterion at t_i contribute to $f_{\mathrm{invalid}}(t_i)$ but are excluded from the reconstructed-age CDF $F_{\mathrm{UI}}(t';t_i)$.

(v) **Goodness of fit and invalid-age penalty.** For each trial Pb-loss age t_i, the two-sample Kolmogorov–Smirnov distance compares the concordant and reconstructed upper-intercept age CDFs,

$$D(t_i) = \sup_{t'\in\mathbb{R}} \big|F_{\mathrm{con}}(t') - F_{\mathrm{UI}}(t';t_i)\big|. \tag{3}$$

Let $m(t_i)$ be the number of normally discordant analyses that yield a valid reconstructed upper-intercept age at t_i, and let n_{disc} be the number of normally discordant analyses. The invalid-reconstruction fraction is

$$f_{\mathrm{invalid}}(t_i) = 1 - \frac{m(t_i)}{n_{\mathrm{disc}}} \in [0,1]. \tag{4}$$

We define the penalised K–S distance

$$D^*(t_i) = D(t_i) + \big[1 - D(t_i)\big]\, f_{\mathrm{invalid}}(t_i), \tag{5}$$

and the corresponding raw and penalised goodness-of-fit measures

$$S_{\mathrm{raw}}(t_i) = 1 - D(t_i), \qquad S_{\mathrm{pen}}(t_i) = 1 - D^*(t_i) = \big[1 - D(t_i)\big]\big[1 - f_{\mathrm{invalid}}(t_i)\big]. \tag{6}$$

https://doi.org/10.5194/egusphere-2026-280
Preprint. Discussion started: 12 February 2026

(vi) **Run-level optimum and summary.** For run n, the optimal age t_n^{opt} maximises $S_{\mathrm{pen},n}(t)$. Across R Monte Carlo realisations we report the median of $\{t_n^{\mathrm{opt}}\}_{n=1}^{R}$ with empirical $95\,\%$ limits given by the 2.5^{th} and 97.5^{th} percentiles.

While this baseline workflow reliably recovers a single dominant Pb-loss age in each Monte Carlo run, it does so at the expense of secondary minima present on the $D(t_i)$ surface. As a result, it may fail to recognise more complex multi-stage Pb-loss histories. In practice, secondary Pb-loss episodes can sometimes emerge indirectly through the KDE of all Monte Carlo best-fit ages, revealing additional peaks that may correspond to meaningful geological events. However, this approach is heuristic and does not consistently capture secondary solutions that appear clearly within individual runs. Section 2.3 addresses this limitation by introducing algorithmic enhancements designed to identify and evaluate all meaningful minima explicitly, allowing multiple geological events to be resolved.

2.2 Prefiltering and modelling window

2.2.1 Reverse-discordant filter

Prior to Pb-loss modelling, we flag analyses as reverse-discordant if they fail the concordance test (Sect. 2.1) and plot geometrically below the Tera–Wasserburg (TW) concordia curve at the same $^{238}\mathrm{U}/^{206}\mathrm{Pb}$ abscissa or to its left at the same $^{207}\mathrm{Pb}/^{206}\mathrm{Pb}$ ordinate. Reverse-discordant points are excluded from the discordant set used for back-projection and K–S comparison because they are not explained by a lower-intercept Pb-loss episode and may reflect spurious U–Pb calibration. Samples are skipped if they lack concordant analyses, lack discordant analyses, or retain fewer than three normally discordant analyses.

2.2.2 Modelling window

For each sample we restrict the Pb-loss age grid to a sample-specific window that is specified by the user, guided by inspection of the data. In practice, we recommend that the upper (older) modelling bound is chosen by the user so that candidate Pb-loss ages do not substantially exceed the youngest concordant $^{238}\mathrm{U}/^{206}\mathrm{Pb}$ age, to avoid fitting Pb-loss episodes older than demonstrably unaffected grains. An automated edge-guard then widens the window once toward any boundary that attracts a large fraction of Monte Carlo run-level optima, and re-runs the sample with approximately constant grid spacing. This allows the grid to expand if the data press against an edge while retaining a conservative default window.

2.3 CDC enhancements

The most recent implementation of the CDC workflow repeated the concordant–discordant comparison hundreds of times under Monte Carlo resampling but, nevertheless, each realisation collapsed to a single Pb-loss age (Mathieson et al., 2025a). The new pipeline introduced here lifts this limitation through three successive enhancements: (i) partitioning discordant analyses into internally coherent clusters to prevent dominance by a single sub-population (Sect. 2.4); (ii) retaining local minima in the per-run dissimilarity surface that pass prominence and width gates, rather than only the global minimum (Sect. 2.5); and (iii) merging these run-level peaks into an ensemble catalogue using a range-agnostic picker that operates in grid units, imposes explicit shoulder-merge and edge-handling rules, and retains only peaks that are reproducible across Monte Carlo runs

https://doi.org/10.5194/egusphere-2026-280
Preprint. Discussion started: 12 February 2026

(Sect. 2.6). Default thresholds are specified as fractions of the grid length or of the trimmed ensemble dynamic range so that results are invariant to the user's age window. The three enhancements are described in detail below.

2.4 Discordant analysis clustering

Discordant zircon populations often comprise several linear sub-arrays, each anchored to a distinct lower-intercept age. Evaluating such mixed populations *en masse* biases the K–S test toward a compromise age that reflects the numerically dominant (or most precise) array. To reduce this bias we optionally partition discordant analyses into clusters of similar concordia-intercept proxies:

1. **Concordia-intercept proxies.** For each discordant analysis we compute a one-dimensional concordia-intercept proxy using a TW lower-intercept search on the user-defined Pb-loss grid. Specifically, we evaluate a grid of candidate concordia ages and, for each trial lower-intercept age t_{L}, construct the chord that joins the concordia point at t_{L} to the measured discordant point. A small geometric residual quantifies how consistently that chord passes through the measured point in both coordinates, and we retain the trial age t_{L} that minimises this residual. The resulting values $T_{\mathrm{LI},k}$ (Ma) are lower-intercept-like concordia-intercept proxies that summarise the position of each grain relative to concordia.

2. **Gaussian mixture clustering and adaptive acceptance.** A one-dimensional Gaussian mixture model is fitted to the set $\{T_{\mathrm{LI},k}\}$ with the number of components $K \le K_{\max} = 4$ chosen by the Bayesian information criterion. The resulting "raw" component labels are then filtered by size and separation gates that adapt to the number of discordant grains n. Each component must contain at least

$$m_{\min}(n) \;=\; \max\bigl(4,\, \lceil 0.12\, n \rceil\bigr) \tag{7}$$

grains and represent at least a fraction $f_{\min}(n)$ of the discordant set, where $f_{\min}(n)$ is tapered from 0.40 for very small samples ($n < 8$), through 0.25 ($8 \le n < 12$) and 0.15 ($12 \le n < 20$), down to 0.10 for $n \ge 20$. Components whose median proxy ages are too close are iteratively merged until every surviving pair satisfies

$$\frac{\bigl|\,\mu_i - \mu_j\,\bigr|}{\max(\hat{\sigma}_i, \hat{\sigma}_j)} \;\ge\; c(n), \tag{8}$$

where μ_i and $\hat{\sigma}_i$ are the median and median absolute deviation (MAD)-based scale of component i, and $c(n)$ lies between 1.05 and 1.20 (stricter for larger n). Components that fail the size or separation thresholds are discarded, and any remaining points are reassigned to the nearest accepted component by median age. If fewer than two components survive, all discordant analyses are treated as a single cluster.

3. **Across-cluster goodness.** For each Monte Carlo run n and grid node t_i we compute the K–S distance and its penalised counterpart for every accepted cluster c, yielding $D_{c,n}(t_i)$ and $D^{*}_{c,n}(t_i)$. We then form the across-cluster minima

$$D_{\min,n}(t_i) = \min_{c} D_{c,n}(t_i), \qquad D^{*}_{\min,n}(t_i) = \min_{c} D^{*}_{c,n}(t_i), \tag{9}$$

https://doi.org/10.5194/egusphere-2026-280
Preprint. Discussion started: 12 February 2026

and the corresponding goodness curves

$$S_{\mathrm{raw},n}(t_i) = 1 - D_{\mathrm{min},n}(t_i), \qquad S_{\mathrm{pen},n}(t_i) = 1 - D^*_{\mathrm{min},n}(t_i). \tag{10}$$

These across-cluster minima are used for both the run-level optima and the ensemble peak catalogue. We take the across-cluster minimum because the goal is to detect whether any internally coherent discordant sub-array is consistent with a candidate loss age, avoiding compromise solutions that can arise when distinct sub-arrays are pooled. Because taking a minimum can inflate apparent goodness if very small or noisy clusters are admitted, we impose explicit size and separation gates and revert to a single-cluster comparison when those conditions are not met.

If too few discordant analyses project successfully to support more than one component or the Bayesian information criterion (BIC) selects a single component, or the size-separation checks above are not met, the algorithm reverts to the original single-cluster CDC behaviour.

2.5 Per–run peak detection

We use R Monte Carlo realisations and a regular age grid of N nodes. After clustering, analytical uncertainties are propagated by Monte Carlo resampling and the K–S dissimilarity is computed on a regular age grid $\{t_i\}_{i=1}^{N}$. For each run n we form the across-cluster minima

$$D_{\mathrm{min},n}(t_i) = \min_c D_{c,n}(t_i), \qquad D^*_{\mathrm{min},n}(t_i) = \min_c D^*_{c,n}(t_i), \tag{11}$$

and define goodness curves

$$S_{\mathrm{raw},n}(t_i) = 1 - D_{\mathrm{min},n}(t_i), \qquad S_{\mathrm{pen},n}(t_i) = 1 - D^*_{\mathrm{min},n}(t_i). \tag{12}$$

Unless stated otherwise, peak detection and ensemble voting are performed on the penalised curve $S_{\mathrm{pen},n}(t)$.

For each run we first compute the robust dynamic range

$$\Delta_n \;=\; P_{95}\big(S_{\mathrm{pen},n}\big) - P_5\big(S_{\mathrm{pen},n}\big), \tag{13}$$

and identify local maxima of $S_{\mathrm{pen},n}$ using the standard peak-finding routines in SciPy (SciPy Developers, 2024a, b). A candidate maximum is retained if it

1. has prominence at least $0.10\,\Delta_n$,
2. is separated from neighbouring candidates by at least five grid nodes, and
3. has full width at half maximum (FWHM) of at least five nodes.

Peak indices are then refined to the nearest local crest (stable on flat-topped plateaus) and the apex age is refined by parabolic interpolation on the grid. A given run may therefore contribute no peaks (if the penalised surface $S_{\mathrm{pen},n}$ is effectively flat),

https://doi.org/10.5194/egusphere-2026-280
Preprint. Discussion started: 12 February 2026

one dominant peak, or several peaks of comparable rank. In contrast to the original CDC implementation, which retained only the single global maximum of $S_{\mathrm{pen},n}(t)$ (i.e., the deepest minimum of $D^{*}_{\mathrm{min},n}(t)$), this procedure keeps all local optima that pass the prominence and width gates and thus preserves evidence for multiple Pb-loss episodes within a single Monte Carlo realisation. To ensure that results are exactly reproducible, we initialise the pseudo-random number generator with a fixed seed for each sample, so that rerunning the algorithm on the same data and settings yields identical Monte Carlo realisations.

2.6 Ensemble peak catalogue

Conceptually, the CDC ensemble peak catalogue is the final list of Pb-loss episodes that emerge as statistically significant peaks when information from all Monte Carlo runs is combined. Each entry corresponds to one such episode and is reported with an age, an empirical 95 % uncertainty interval, and a support value describing the fraction of runs that favour that peak.

Let $S_{\mathrm{pen},n}(t)$ denote the penalised goodness-of-fit curve for Monte Carlo realisation n, evaluated on a regular age grid $\{t_i\}_{i=1}^{N}$, where larger values indicate better agreement between reconstructed and concordant age distributions. We aggregate runs by the pointwise median

$$\tilde{S}(t_i) \;=\; \mathrm{median}_n\, S_{\mathrm{pen},n}(t_i), \tag{14}$$

and apply a light one-dimensional Gaussian smoothing with standard deviation $\sigma = f_{\mathrm{smooth}} N$ grid nodes. We then define the trimmed dynamic range

$$\Delta \;=\; P_{95}(\tilde{S}) - P_{5}(\tilde{S}). \tag{15}$$

If $\Delta < \Delta_{\mathrm{min}}$, the ensemble surface is treated as effectively flat and no CDC ensemble peaks are reported for that sample.

2.6.1 Candidate peaks on the ensemble surface

To prevent broad modes on the ensemble curve from being split into multiple peaks by small shoulders, we adopt a two-scale peak-selection strategy. First, we construct a strongly smoothed version of the ensemble curve, $\tilde{S}_{\mathrm{coarse}}(t)$, by convolving $\tilde{S}(t)$ with a broader Gaussian kernel of width $\sigma_{\mathrm{coarse}} = \max(5, 0.03N)$ grid nodes. Local maxima of $\tilde{S}_{\mathrm{coarse}}(t)$ that exceed a minimum prominence $\pi_{\mathrm{min}} = f_p\,\Delta$ and are separated by at least $0.10N$ nodes define the broad modes of the ensemble surface. If no coarse maxima satisfy this prominence criterion, the prominence threshold is relaxed once to $0.5\,\pi_{\mathrm{min}}$; if still none are found, the global maximum of $\tilde{S}(t)$ is treated as a single broad mode.

We then search for local maxima on the lightly smoothed median curve $\tilde{S}(t)$ itself. Fine-scale candidates are maxima of $\tilde{S}(t)$ that exceed the same prominence threshold π_{min} and satisfy a minimum peak width of $w_{\mathrm{min}} = 3$ grid nodes (width evaluated at relative height 0.5 following standard peak-width definitions). Each broad mode is allowed to contribute at most one ensemble apex: within $\pm 0.05N$ nodes of the corresponding coarse maximum we retain the fine-scale candidate with the largest crest height. If no fine-scale candidate falls within this window, that broad mode does not contribute to the catalogue.

https://doi.org/10.5194/egusphere-2026-280
Preprint. Discussion started: 12 February 2026

2.6.2 Shoulder merge

Fine-scale candidates that lie within $f_d N$ grid nodes of one another are compared pairwise. Let Δ_v be the depth of the intervening valley measured from the lower of the two crest heights. If the valley is shallow relative to the smaller prominence, i.e. if

$$\Delta_v < f_v \min(\pi_i, \pi_j), \tag{16}$$

where π_i and π_j are the prominences of the two candidates, the pair is treated as alternative representations of the same peak and merged, retaining the more prominent (and, for near-ties, higher) apex. Surviving apices are snapped to the local crest of $\tilde{S}(t)$ within a small neighbourhood and the apex age is refined by parabolic interpolation on the grid. Very low-amplitude peaks are then discarded by imposing a global height gate where the crest height must be at least a fraction $h_{\min}$ of the tallest peak on the ensemble curve.

2.6.3 Reproducibility voting and uncertainties

For each ensemble apex at grid index $j^\star$ we construct a local voting window of $\pm f_w N$ nodes. Each Monte Carlo realisation may cast at most one vote in this window. For run n we obtain a set of candidate ages from its per-run peaks (Sect. 2.5) and, if at least one candidate falls inside the window, we take the one with the highest local goodness. If no candidate lies in the window, we fall back to the grid node of maximal goodness within the window.

In both cases we impose a run-specific height gate: a run contributes a vote only if its local height exceeds

$$P_5\big(S_{\mathrm{pen},n}\big) + f_r^{(\mathrm{eff})}\,\big[P_{95}\big(S_{\mathrm{pen},n}\big) - P_5\big(S_{\mathrm{pen},n}\big)\big], \tag{17}$$

where the effective fraction $f_r^{(\mathrm{eff})}$ is allowed to relax for peaks that attract very few candidate per-run peaks. The ensemble support of a peak is defined as the fraction of runs that cast a vote in its window,

$$\mathrm{Support} = \frac{N_{\mathrm{votes}}}{R}. \tag{18}$$

A candidate apex is retained in the catalogue if $\mathrm{Support} \geq \max(r_{\min}/R, f_{\mathrm{support}})$.

For each retained peak we refine the ensemble age $\hat{t}$ by parabolic interpolation on the lightly smoothed median curve and estimate a 95 % confidence interval from the empirical distribution of ages assigned to that peak. When available, we use the subset of run-level optimal ages t_n^{opt} that fall within the voting window; otherwise we use the vote ages themselves. The interval is recentered on the ensemble apex, clipped to the modelling window, and constrained to span at least five grid steps to guard against spuriously tight intervals driven by discretisation.

These steps produce the CDC ensemble peak catalogue: a set of Pb-loss ages that are (i) prominent on the ensemble goodness surface, (ii) reproducibly supported across Monte Carlo realisations, and (iii) equipped with uncertainty intervals that reflect the stability of each peak under Monte Carlo perturbations.

https://doi.org/10.5194/egusphere-2026-280
Preprint. Discussion started: 12 February 2026

2.6.4 Parameterisation and justification of thresholds

Unless otherwise stated, ensemble thresholds are expressed in grid units or as fractions of robust summaries of the ensemble median curve (notably the trimmed dynamic range Δ), so defaults are largely insensitive to the absolute Pb-loss age window used for the grid.

(i) **Dynamic-range gate.** We set $\Delta_{\min} = 0.05$. If $\Delta < \Delta_{\min}$ the ensemble surface is treated as effectively flat and no CDC ensemble peaks are reported.

(ii) **Smoothing scales (fine and coarse).** The lightly smoothed median $\tilde{S}(t)$ uses a Gaussian kernel with $\sigma = f_{\mathrm{smooth}} N$ grid nodes, with $f_{\mathrm{smooth}} = 0.01$ by default (equivalent to $\sigma = 2$ nodes for $N = 200$). The coarse curve $\tilde{S}_{\mathrm{coarse}}$ uses $\sigma_{\mathrm{coarse}} = \max(5, 0.03N)$ nodes.

(iii) **Candidate selection on the ensemble surface (prominence, width, height).** We use a prominence threshold $\pi_{\min} = f_p \Delta$ with $f_p = 0.10$ and identify candidates with standard peak-finding implementations (SciPy Developers, 2024a, b; MathWorks, 2024). Fine-scale candidates are required to satisfy a minimum peak width of $w_{\min} = 3$ grid nodes (width at relative height 0.5). Very low-amplitude peaks are removed by a global height gate requiring the crest height to be at least $h_{\min} = 0.5$ of the tallest ensemble crest.

(iv) **Mode separation and shoulder handling.** Coarse maxima are required to be separated by at least $0.10N$ nodes. Each coarse mode contributes at most one apex, selected as the highest fine-scale candidate within $\pm 0.05N$ nodes of the coarse maximum. Fine-scale candidates within $f_d N$ nodes are considered for shoulder merging (here $f_d = 0.10$); two nearby candidates are merged if the intervening valley satisfies $\Delta_v < f_v \min(\pi_i, \pi_j)$, with $f_v = 0.50$.

(v) **Voting, support, and interval constraints.** Voting windows use a half-width of $\pm f_w N$ nodes (here $f_w = 0.10$). A run casts a vote only if its local height exceeds the run-specific gate based on P_5 and P_{95}; $f_r^{(\mathrm{eff})} = 0.25$ for well-supported peaks and is relaxed to 0.10 for weakly supported peaks as described above, reducing sensitivity to run-to-run fluctuations (Meinshausen and Bühlmann, 2010). A peak is retained only if $\mathrm{Support} \geq \max(r_{\min}/R, f_{\mathrm{support}})$, with $r_{\min} = 5$ and $f_{\mathrm{support}} = 0.10$. Confidence intervals are computed from empirical quantiles of assigned ages, clipped to the modelling window, and constrained to span at least five grid steps. Support values are conditional on the voting-window half-width and the run-level height gate. We therefore interpret them primarily as within-sample stability indicators, not as absolute measures comparable across datasets or parameterisations.

https://doi.org/10.5194/egusphere-2026-280
Preprint. Discussion started: 12 February 2026

3 Synthetic data generation for benchmarking

3.1 Aim and overall design

A controlled library of *in-silico* zircon U–Pb datasets allows us to verify that the new CDC pipeline can (i) recover a single Pb-loss age when only one episode is present; (ii) distinguish superimposed Pb-loss episodes when the data warrant that complexity; and (iii) remains stable to realistic levels of analytical noise and age populations.

Simulations are first generated in Wetherill coordinates,

$$(x, y) \;=\; \left({}^{207}\mathrm{Pb}/{}^{235}\mathrm{U},\, {}^{206}\mathrm{Pb}/{}^{238}\mathrm{U}\right), \tag{19}$$

using the IUGS decay constants $\lambda_{235} = 9.8485 \times 10^{-4}\,\mathrm{Ma}^{-1}$ and $\lambda_{238} = 1.55125 \times 10^{-4}\,\mathrm{Ma}^{-1}$. This choice facilitates direct benchmarking against the probabilistic discordia method of Reimink et al. (2016), which operates natively in Wetherill space.

Simulated data are then converted to TW coordinates

$$(u, v) \equiv \left({}^{238}\mathrm{U}/{}^{206}\mathrm{Pb},\, {}^{207}\mathrm{Pb}/{}^{206}\mathrm{Pb}\right), \tag{20}$$

with the algebraic relations

$$u \;=\; \frac{1}{y}, \qquad v \;=\; \frac{x/y}{R_U}, \qquad R_U \;\equiv\; \frac{{}^{238}\mathrm{U}}{{}^{235}\mathrm{U}} \approx 137.818. \tag{21}$$

Uncertainties are propagated with a fixed off–axis correlation $\rho = +0.85$ in Wetherill space. Let $\sigma_x^2 = \mathrm{Var}(x)$, $\sigma_y^2 = \mathrm{Var}(y)$, and $\mathrm{Cov}(x, y) = \rho\,\sigma_x \sigma_y$. Then

$$\mathrm{Var}\left(\frac{x}{y}\right) = \frac{\sigma_x^2}{y^2} + \frac{x^2 \sigma_y^2}{y^4} - \frac{2\,x}{y^3}\,\mathrm{Cov}(x, y), \tag{22}$$

so the 1σ uncertainty on v is

$$\sigma_v \;=\; \frac{1}{R_U}\sqrt{\mathrm{Var}\left(\frac{x}{y}\right)}. \tag{23}$$

For $u = 1/y$,

$$\mathrm{Var}(u) \;=\; \frac{\sigma_y^2}{y^4}, \qquad \sigma_u \;=\; \frac{\sigma_y}{y^2}. \tag{24}$$

Each panel contains $N = 160$ analyses. All calculations use $R = 200$ Monte Carlo realisations on a 10 Ma Pb-loss age grid with a random seed of 42.

3.2 Dataset scatter tiers

Three dataset scatter tiers (A to C) were simulated, each increasing the geometric scatter about the discordias. For the single-stage suite (Cases 1 to 4) we used perpendicular scatter $\sigma_\perp = \{0.000, 0.005, 0.010\}$ for tiers A, B and C, respectively. For the

https://doi.org/10.5194/egusphere-2026-280
Preprint. Discussion started: 12 February 2026

multi-stage suite (Cases 5 to 7) we used $\sigma_{\perp} = \{0.000, 0.010, 0.020\}$. In both suites, the formal measurement uncertainty on each ratio was drawn from a uniform relative distribution: 0.5 to 1.5 % for Cases 1 to 4 and 1 to 3 % for Cases 5 to 7. Thus, tiers primarily reflect increasing geometric dispersion about the underlying discordia chords rather than changes in analytical precision. A fixed off-axis correlation of $\rho = +0.85$ is used for all tiers.

3.3 Discordance and common-Pb acceptance filter

Before any simulated analysis is accepted it must pass two numerical checks. The first ensures that the intended 25 % concordant population remains unchanged by pruning measurements that are effectively concordant; the second removes points whose signal is dominated by common Pb rather than by genuine Pb mobility (or mixing).

(i) **Minimum fractional discordance.**

$$D_{\mathrm{frac}} = 1 - \frac{\min(t_{76}, t_{68})}{\max(t_{76}, t_{68})} > 0.05, \tag{25}$$

where t_{76} and t_{68} are the individual $^{207}\mathrm{Pb}/^{206}\mathrm{Pb}$ and $^{206}\mathrm{Pb}/^{238}\mathrm{U}$ dates. The 5 % threshold removes analyses whose propagated uncertainties overlap concordia.

(ii) **Upper limit on measured $^{207}\mathrm{Pb}/^{206}\mathrm{Pb}$.**

$$\left(^{207}\mathrm{Pb}/^{206}\mathrm{Pb}\right)_{\mathrm{meas}} < 0.35. \tag{26}$$

On a TW plot the ordinate intercept is controlled by the fraction of common Pb that remains in an analysis. Empirically, ratios above $(^{207}\mathrm{Pb}/^{206}\mathrm{Pb})_{\mathrm{meas}} \sim 0.35$ imply that uncorrected common Pb, rather than radiogenic Pb-loss, dominates the observed discordance. Culling those points keeps the benchmark focused on the processes that the CDC workflow is intended to resolve, rather than on the effects of common Pb.

Analyses that fail either filter are discarded and resampled until the target population size is reached, ensuring that every retained datum is plausible within the modelling assumptions.

3.4 Benchmark cases

Each synthetic panel contains a concordant subset (25 % of analyses in total), with concordant grains drawn around each anchor age t_{up}, and one or more discordant arrays that record either a single Pb-loss event ($t_{\mathrm{up}} \rightarrow t_{\mathrm{low}}$) or two sequential events ($t_{\mathrm{up}} \rightarrow t_{\mathrm{low},1} \rightarrow t_{\mathrm{low},2}$).

3.4.1 Single-stage Pb-loss scenarios (Cases 1 to 4)

Four single-stage Pb-loss benchmarks were devised to isolate the influence of discordia geometry while holding every other simulation parameter constant (Table 1, Fig. 1). Case 1 contains a single chord linking a concordant population at 3000 Ma to a 700 Ma lower intercept. Case 2 uses a single crystallisation age at 3200 Ma but splits the discordant population between

https://doi.org/10.5194/egusphere-2026-280
Preprint. Discussion started: 12 February 2026

Table 1. Single-stage benchmark scenarios. Each trajectory is a single chord from t_{up} to t_{low}.

Case	Tier	Trajectory	Pb-loss path (Ma)	Weight (%)
1	A to C	(i)	3000 → 700	100
2	A to C	(i)	3200 → 1800	50
		(ii)	3200 → 300	50
3	A to C	(i)	3200 → 400	33
		(ii)	3000 → 400	33
		(iii)	2800 → 400	34
4	A to C	(i)	3200 → 500	50
		(ii)	3000 → 1800	25
		(iii)	2800 → 1800	25

two chords that terminate at 1800 Ma and 300 Ma. Case 3 combines three crystallisation ages (3200, 3000, and 2800 Ma) that converge on a common lower intercept at 400 Ma, whereas Case 4 mixes one chord (3200 Ma→500 Ma) with two other chords that both intersect concordia at 1800 Ma (3000 Ma→1800 Ma and 2800 Ma→1800 Ma).

https://doi.org/10.5194/egusphere-2026-280
Preprint. Discussion started: 12 February 2026

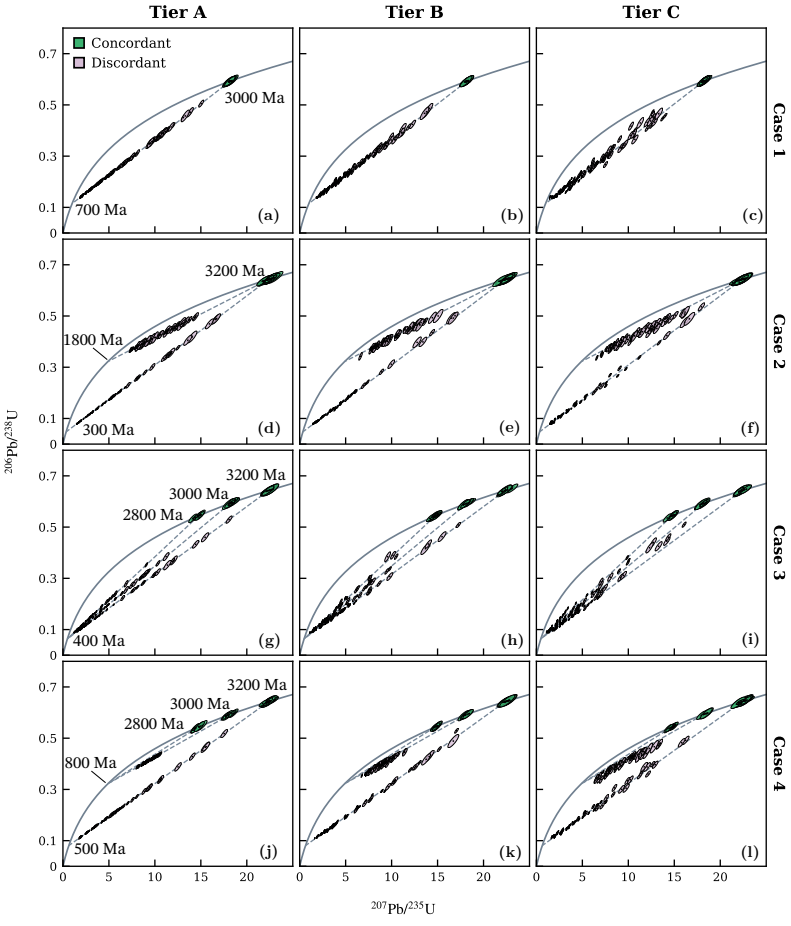

Figure 1. Synthetic zircon U–Pb datasets for single-stage Pb-loss tests. Columns illustrate four Pb-loss histories (Cases 1 to 4; Table 1), while rows step through three dataset scatter tiers (A, B, C) with perpendicular scatter $\sigma_{\perp} = 0$, 0.005 and 0.010, respectively. Each panel contains 160 simulated analyses: 40 concordant (green ellipses) and 120 discordant (purple ellipses), plotted relative to the Wetherill concordia (grey). Increasing scatter from tier A to C broadens the dispersion about each discordia chord, making Pb-loss age recovery progressively more challenging.

https://doi.org/10.5194/egusphere-2026-280
Preprint. Discussion started: 12 February 2026

Table 2. Multi-stage benchmark scenarios. Trajectories showing two sequential Pb-loss steps (i.e. more than one arrow) are splayed by $0.5\,\sigma_{\perp}$.

Case	Tier	Trajectory	Pb-loss path (Ma)	Weight (%)
5	A to C	(i)	3000 → 1500	60
		(ii)	3000 → 1500 → 500	40
6	A to C	(i)	3000 → 1500	45
		(ii)	3000 → 500	33
		(iii)	3000 → 1500 → 500	22
7	A to C	(i)	3200 → 1500	40
		(ii)	3000 → 1500	30
		(iii)	3200 → 1500 → 500	30

3.4.2 Two-stage Pb-loss scenarios (Cases 5 to 7)

The three two-stage benchmarks add an older disturbance at 1500 Ma and a younger one at 500 Ma (Table 2, Fig. 2). In every discordant analysis the segment corresponding to the second episode is displaced by $0.5\,\sigma_{\perp}$ perpendicular to the model chord so that no measurement lies exactly on both chords.

Case 5 uses a single crystallisation age of 3000 Ma. Sixty per cent of the discordant analyses record only the older episode, whereas the remaining 40 per cent record both. Case 6 keeps the same crystallisation age but allocates 45 % of the points to the older episode alone, 33 % to the younger episode alone, and 22 % to the full two-step path. Case 7 superimposes two crystallisation populations at 3200 Ma and 3000 Ma. Forty per cent of the older analyses and 30 % of the younger analyses are disturbed once at 1500 Ma, while an additional 30 % of the older analyses record the full two-event history.

https://doi.org/10.5194/egusphere-2026-280
Preprint. Discussion started: 12 February 2026

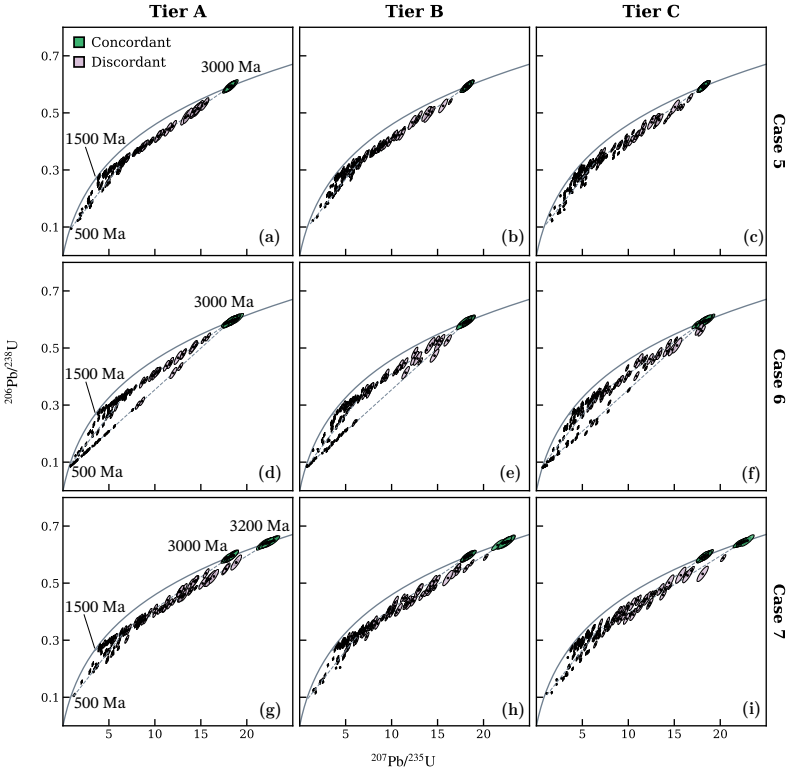

Figure 2. Synthetic zircon U–Pb datasets for multi-stage Pb-loss tests. Columns correspond to dataset scatter tiers A, B and C, having perpendicular scatter $\sigma_\perp = 0$, 0.010 and 0.020, respectively; rows illustrate three increasingly complex Pb-loss histories (Table 2). For analyses that register two Pb-loss events, the younger segment is offset by $0.5\sigma_\perp$ so no point lies exactly on both chords. Each panel contains 160 simulated analyses: 40 concordant analyses (green ellipses) and 120 discordant analyses (purple ellipses), plotted against the concordia curve (grey). Dashed grey chords mark the true Pb-loss trajectories encoded in each case.

3.5 Concordant reference distribution

To examine how the concordant reference distribution affects CDC and the ensemble catalogue, we constructed a concordant–fraction sweep in two synthetic geometries. In both geometries the discordant population is fixed and only the size of the concordant subset is varied.

The first geometry is a single-stage configuration (Case 3, Tier A) with three upper-intercept ages of 1100, 1400, and 1700 Ma that share a common lower intercept at 25 Ma. All discordant chords therefore project toward a single young loss age. The second geometry is a two–loss configuration (Case 2, Tier A) with the same three upper intercepts, but with two

https://doi.org/10.5194/egusphere-2026-280
Preprint. Discussion started: 12 February 2026

distinct lower intercepts at approximately 300 Ma and 1800 Ma. In this case different subsets of chords record loss toward either 300 or 1800 Ma.

For each geometry a discordant set of $n_{\mathrm{disc}} = 120$ analyses was drawn once in TW space as straight lines between the relevant intercept pairs, with Gaussian along– and across–chord scatter. The acceptance filters of Sect. 3.3 were applied (minimum fractional discordance), and for the concordant–fraction sweep we imposed a stricter upper bound on the measured TW ordinate, $(^{207}\mathrm{Pb}/^{206}\mathrm{Pb})_{\mathrm{meas}} < 0.26$, to exclude discordant points that plot well above the concordant reference population in TW space (consistent with added common Pb rather than radiogenic Pb-loss). This discordant array is held fixed throughout the sweep.

Concordant analyses were then added at the appropriate upper intercepts to obtain target concordant fractions $C \in \{5, 10, 20, 30, 40, 50, 60$ where

$$C = \frac{n_{\mathrm{con}}}{n_{\mathrm{con}} + n_{\mathrm{disc}}}, \qquad n_{\mathrm{disc}} = 120. \tag{27}$$

This definition gives $n_{\mathrm{con}} = 6, 13, 30, 51, 80, 120, 180$ and total sample sizes between 126 and 300 analyses. Measurement uncertainties were 0.5 to 1.5 % per ratio. Concordant draws were truncated so that they are not younger than their corresponding t_{up}, and each value of C used a deterministic seed so that only the size of the concordant subset changes across the sweep.

In the single-stage geometry CDC was evaluated on a Pb-loss grid of 200 nodes between 1 and 1000 Ma. In the two–loss geometry the grid extended from 1 to 2000 Ma, again with 200 nodes. In both cases we used $R = 200$ Monte Carlo realisations and recorded both the legacy single-optimum CDC age and the full ensemble catalogue of peaks for each dataset.

3.6 Analysis of synthetic datasets

We benchmarked the new CDC workflow against the probabilistic discordia ("discordance dating", hereafter DD) method of Reimink et al. (2016) on the synthetic suites described in Sect. 3.4 (Cases 1 to 7) and across the three dataset scatter tiers (A to C; Sect. 3.2). Both methods evaluate goodness/likelihood on a 10 Ma grid using $R = 200$ Monte Carlo (CDC) or bootstrap (DD) realisations per dataset.

For CDC, we used the TW implementation and peak–selection strategy described in Sections 2.5–2.6. For each dataset, goodness is evaluated on the grid for each realisation. Candidate ensemble peaks are identified from the lightly smoothed, pointwise median of the penalised surface. The final reported ages correspond to these refined ensemble apices, each with an empirical 95 % interval. A true Pb-loss event is considered covered if it receives an assigned estimate under the truth-window rule and the corresponding 95 % interval overlaps the true event age.

For DD, we followed the published workflow of Reimink et al. (2016) on the same grid. For each bootstrap we record the global maximum of the normalised likelihood surface. The DD point estimate is taken as the median of these maxima, accompanied by equal–tailed 95 % bootstrap confidence intervals.

For both methods we summarise event-level performance by the median absolute error $\mathrm{MAE} = \mathrm{median}\,|\hat{t} - t_{\mathrm{true}}|$ (computed across assigned event estimates only), the maximum absolute error, and empirical coverage.

https://doi.org/10.5194/egusphere-2026-280
Preprint. Discussion started: 12 February 2026

Whenever a dataset contains multiple true Pb-loss events (Cases 2, 4, and 5 to 7), each true event is assigned a truth-centred scoring window. For an event with true age t_{true}, the window half-width is defined as half the nearest-neighbour separation between true ages, clipped to the range $[50, 120]$ Ma. CDC peak ages and DD bootstrap maxima are attributed to an event only if the point estimate $\hat{t}$ falls inside that event's window; estimates outside all windows are left unassigned.

Coverage is reported relative to the full set of true events. An event is counted as covered if it receives an assigned estimate and the corresponding 95 % interval includes t_{true}; unassigned events count as not covered.

3.6.1 DD–PEAKS variant

In addition to the original DD summary, we report a simple peak–aware variant ("DD–PEAKS") that operates on the median likelihood curve. We compute the likelihood on the same grid for each of the bootstraps and form the pointwise median across bootstraps. Local maxima of this median curve are detected using a small prominence threshold (2 % of the tallest median peak) and a minimum width of three grid nodes. For each detected median peak we define a support window equal to its full width at half maximum (FWHM) and assign bootstrap global maxima that fall within this window to the peak. The DD–PEAKS age is the median of assigned maxima with equal–tailed 95 % quantiles. If no median peak meets the prominence/width criteria, no DD–PEAKS estimate is returned for that dataset.

4 Results

4.1 Benchmark accuracy and coverage

4.1.1 Overall behaviour

Pooling all events across Cases 1 to 7 and Tiers A to C, the CDC ensemble catalogue (i.e. the set of Pb-loss ages obtained by applying the CDC ensemble peak picker to the lightly smoothed pointwise median penalised goodness curve $\tilde{S}(t)$) attains smaller typical error and higher empirical coverage than the DD summary based on per-bootstrap global maxima. Considering assigned event–level estimates only, overall MAE is ∼32 Myr for CDC ($n_{\mathrm{assigned}} = 28$) versus 61 Myr for DD ($n_{\mathrm{assigned}} = 17$) (Table 3). A peak–aware variant that summarises the median DD curve by its prominent peaks (DD–PEAKS) reduces MAE to 46 Myr, but does not improve empirical coverage (Table 3).

Performance differs between the two age bands (Table 4). CDC returns assigned estimates for most true events in both bands ($n_{\mathrm{assigned}} = 14$ of 21 for < 1 Ga and $n_{\mathrm{assigned}} = 14$ of 15 for ≥ 1 Ga). In contrast, DD assigns relatively few older events (5 of 15). For < 1 Ga, CDC has both smaller MAE and higher coverage (median bias $\sim +19$ Myr; MAE ∼32 Myr; coverage 33 %) than DD (median bias +71 Myr; MAE 71 Myr; coverage 10 %). For ≥ 1 Ga, DD is more accurate when it yields an estimate (MAE 11 Myr across 5 assigned events), but its low assignment rate reduces coverage to 27 %. CDC assigns 14 of 15 older events and retains higher coverage (53 %), albeit with larger MAE among assigned estimates (∼34 Myr).

https://doi.org/10.5194/egusphere-2026-280
Preprint. Discussion started: 12 February 2026

Table 3. Aggregate performance across all events (Cases 1 to 7; per-event scoring). MAE is the median absolute error. Coverage is the fraction of true events whose 95 % interval includes the true age under the truth-window rule. n_{assigned} is the number of event–level estimates assigned to a truth window (unassigned events have no estimate but still count as not covered).

Method	n_{assigned}	MAE (Myr)	Coverage (%)
CDC	28	32	42
DD	17	61	17
DD–PEAKS	16	46	17

Table 4. Aggregate performance by age band across all tiers and cases (per-event scoring). n_{true} is the number of true events in the age band. n_{assigned} is the number of event–level estimates assigned to a truth window for that band (unassigned events have no estimate). Median bias and MAE are computed across assigned estimates only. Coverage is the fraction of all true events in the band whose 95 % interval includes the true age under the truth–window rule (unassigned events count as not covered).

Age band	Method	n_{true}	n_{assigned}	Median bias (Myr)	MAE (Myr)	Coverage (%)
$< 1\,\mathrm{Ga}$	CDC	21	14	+19	32	33
	DD	21	12	+71	71	10
	DD–PEAKS	21	11	+71	71	10
$\geq 1\,\mathrm{Ga}$	CDC	15	14	-23	34	53
	DD	15	5	+11	11	27
	DD–PEAKS	15	5	+11	11	27

4.1.2 Single-stage benchmarks (Cases 1 to 4)

Across Cases 1 to 4, goodness/likelihood profiles are sharpest in Tier A and broaden with increasing scatter from Tier B to Tier C (Figs. 3, 4). Event–level summaries show that CDC outperforms DD on both MAE and coverage in all three tiers (Table 5). In this subset, the peak–aware DD variant yields identical summary statistics to the baseline global–maximum workflow.

https://doi.org/10.5194/egusphere-2026-280
Preprint. Discussion started: 12 February 2026

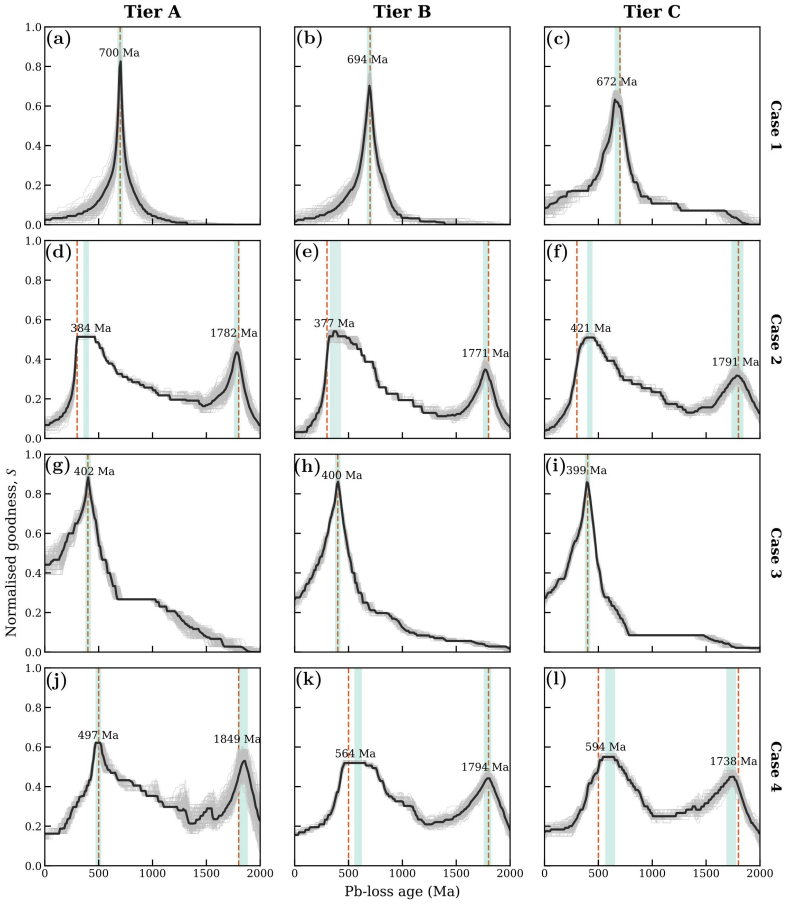

Figure 3. CDC goodness surfaces for the single-stage benchmarks (Cases 1 to 4) across Tiers A to C. Grey curves show individual Monte Carlo realisations, black curves are the lightly smoothed pointwise median penalised goodness $\tilde{S}(t)$, dashed vertical orange lines mark the true Pb-loss ages, and shaded teal bands indicate 95 % confidence intervals from the CDC ensemble catalogue.

https://doi.org/10.5194/egusphere-2026-280
Preprint. Discussion started: 12 February 2026

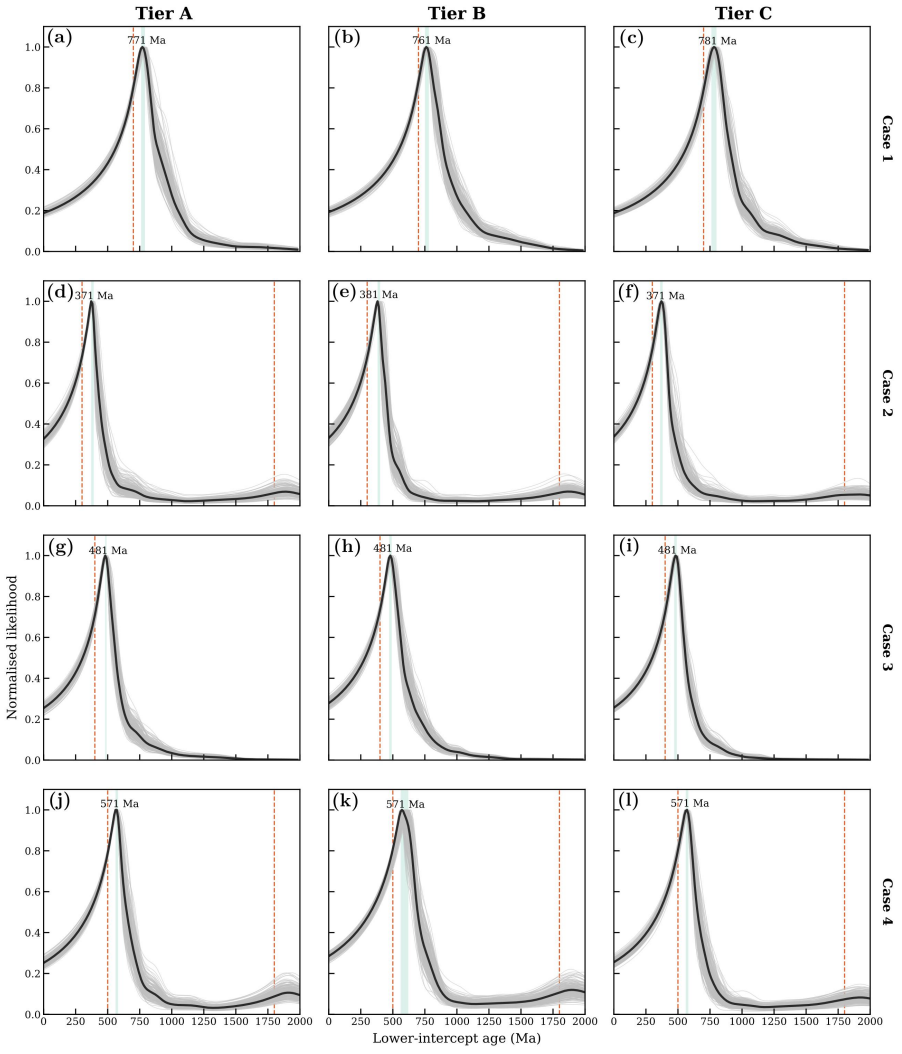

Figure 4. Discordance–dating (DD) likelihood surfaces for the single-stage benchmarks (Cases 1 to 4) across Tiers A to C. Grey curves show individual bootstrap realisations of the normalised summed likelihood, and black curves show the corresponding median profiles. Dashed orange vertical lines mark the true lower–intercept ages. In each panel, the shaded teal band and age label indicate the DD summary lower–intercept age (median of bootstrap maxima) with its 95 % bootstrap confidence interval, as used in the event–level median absolute error (MAE) and coverage summaries.

Across tiers, CDC retains substantially smaller MAE and non-zero coverage, whereas DD frequently fails to yield an assigned truth-window event estimate (reflected in the smaller n) and, when assigned, exhibits large errors and zero coverage in Cases 1 to 4 (Table 5). Even in Tier C, where the goodness/likelihood surfaces are flattest, CDC leads on both metrics (MAE $\sim$28 Myr; coverage 50 %). Here DD–PEAKS coincides with the baseline DD summary because the median summed likelihood curve typically exhibits a single prominent peak.

https://doi.org/10.5194/egusphere-2026-280
Preprint. Discussion started: 12 February 2026

Table 5. Single-stage benchmarks (Cases 1 to 4): event–level accuracy and coverage by tier and method. MAE is the median absolute error (MAE) across events. Coverage is the fraction of true events whose 95 % interval includes the true age under the truth–window rule. n is the number of event–level estimates assigned to a truth window for that method and tier.

Tier	Method	n	MAE (Myr)	Max |bias| (Myr)	Coverage (%)
A	CDC	6	11	84	83
A	DD	2	71	71	0
A	DD–PEAKS	2	71	71	0
B	CDC	6	17	77	67
B	DD	2	76	81	0
B	DD–PEAKS	2	76	81	0
C	CDC	5	28	94	50
C	DD	2	71	71	0
C	DD–PEAKS	2	71	71	0

These counts reflect our truth–window rule at the event level and the choice to score Cases 2 and 4 with an additional Pb-loss event. Consequently, the effective sample size n (number of event–level estimates) can exceed the number of datasets per tier.

4.1.3 Two–stage benchmarks (Cases 5 to 7)

For the two–episode benchmarks (Cases 5 to 7; true ages 500 and 1500 Ma), performance depends strongly on tier and the separability of the two episodes (Figs. 5, 6). Pooled across tiers, DD attains both lower typical error and higher coverage than CDC. Across Cases 5 to 7 (18 true events), CDC achieves an overall MAE of $\sim$42 Myr with 17 % coverage, compared with 21 Myr and 33 % for DD (DD–PEAKS is similar with MAE 20 Myr and 33 % coverage) (Table 6).

https://doi.org/10.5194/egusphere-2026-280
Preprint. Discussion started: 12 February 2026

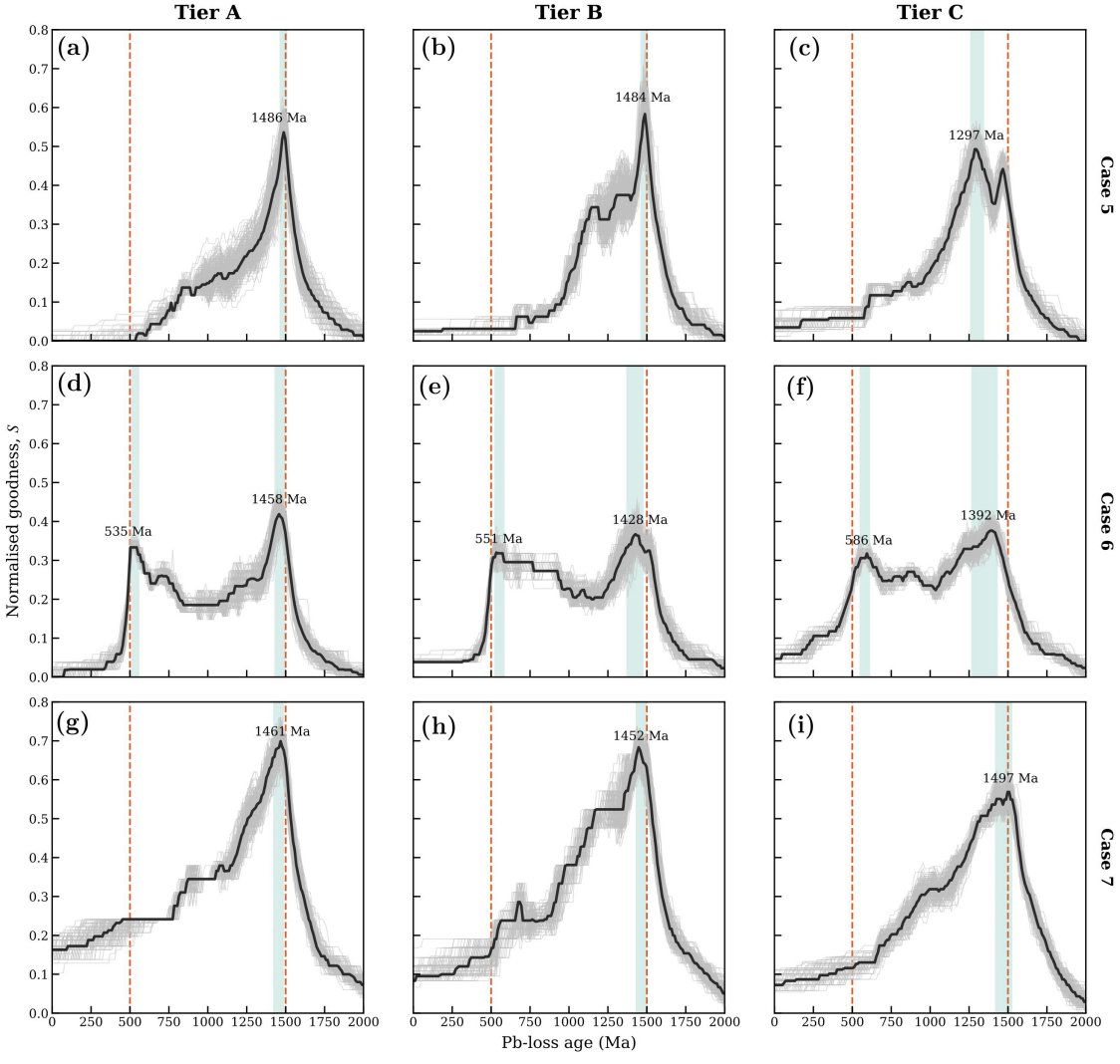

Figure 5. CDC goodness surfaces for the two–stage benchmarks (Cases 5 to 7) across Tiers A to C. Grey curves show individual Monte Carlo realisations, black curves are the lightly smoothed pointwise median penalised goodness $\tilde{S}(t)$, and dashed vertical orange lines mark the true Pb-loss episodes at 500 and 1500 Ma. Shaded teal bands indicate 95 % confidence intervals for the CDC ensemble catalogue peaks that are used in the event–level median absolute error (MAE) and coverage summaries.

https://doi.org/10.5194/egusphere-2026-280
Preprint. Discussion started: 12 February 2026

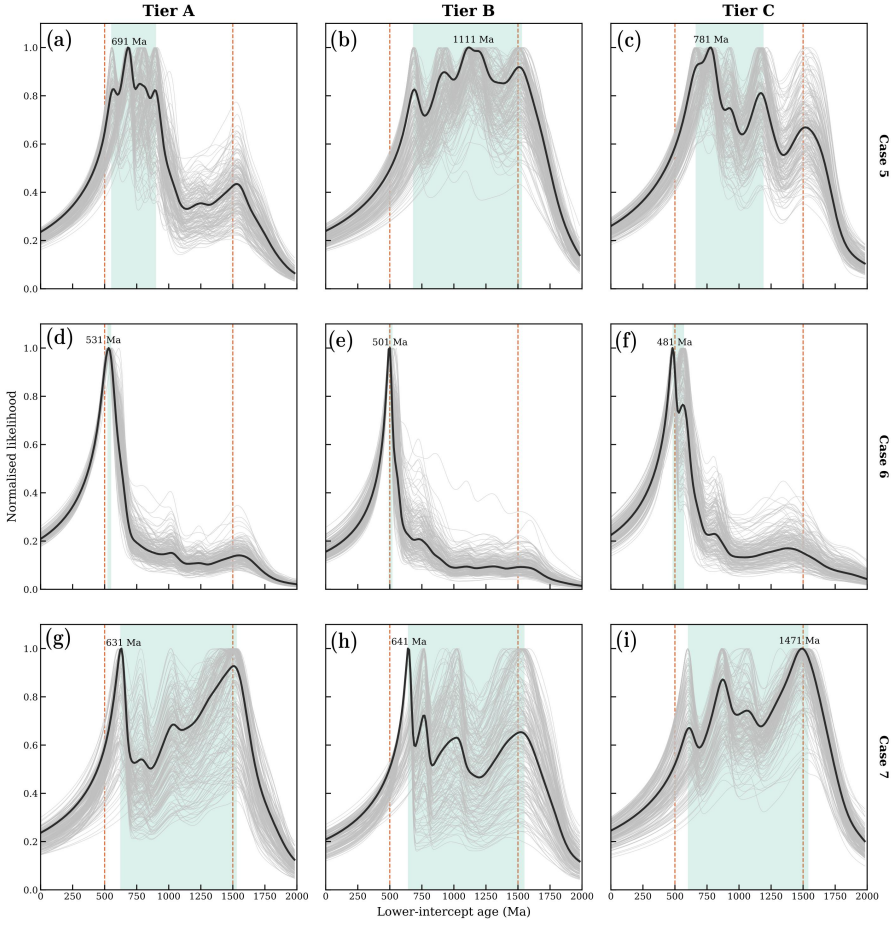

Figure 6. Discordance–dating (DD) likelihood surfaces for the two-stage benchmarks (Cases 5 to 7) across Tiers A to C. Grey curves show individual bootstrap realisations of the normalised summed likelihood, and black curves show the median profiles. Dashed orange vertical lines mark the true lower–intercept ages at 500 and 1500 Ma. In each panel, the shaded teal band and age label indicate the DD summary lower–intercept age (median of bootstrap maxima) with its 95 % bootstrap confidence interval, as used in the event–level median absolute error (MAE) and coverage summaries. In these multi–episode cases, bootstrap maxima can jump between competing modes on the summed likelihood surface, so the intervals may be broad and can span both episodes.

In Tier A, where the two episodes are relatively well separated, CDC and DD–PEAKS are broadly comparable but both exhibit low coverage. CDC yields MAE 37 Myr with 17 % coverage (n=4 assigned event estimates), DD–PEAKS yields 31 Myr with 16.7 % coverage (n=3), and the baseline DD summary is slightly worse in MAE (46 Myr) with the same coverage (17 %).
445 In Tier B, where the episodes overlap more strongly, DD and DD–PEAKS attribute fewer but more accurate event estimates, with markedly smaller MAE (11 Myr) and higher coverage (50 %) than CDC (MAE ∼50 Myr; coverage 17 %). In Tier C,

https://doi.org/10.5194/egusphere-2026-280
Preprint. Discussion started: 12 February 2026

Table 6. Two–stage benchmarks (Cases 5 to 7): event–level accuracy and coverage by tier and method (pooled across cases within each tier). The "All (A to C)" row pools results across tiers (18 true events total). MAE is the median absolute error computed across assigned event estimates; Max |bias| is the maximum absolute error among assigned estimates. Coverage is the fraction of true events whose 95 % interval includes the true age under the truth–window rule; unassigned events count as not covered. n is the number of assigned event–level estimates.

Tier	Method	n	MAE (Myr)	Max |bias| (Myr)	Coverage (%)
A	CDC	4	37	42	17
A	DD	4	46	111	17
A	DD–PEAKS	3	31	61	17
B	CDC	4	50	72	17
B	DD	3	11	21	50
B	DD–PEAKS	3	11	21	50
C	CDC	3	86	109	17
C	DD	4	20	101	33
C	DD–PEAKS	4	20	101	33
All (A to C)	CDC	11	42	109	17
All (A to C)	DD	11	21	111	33
All (A to C)	DD–PEAKS	10	20	101	33

where goodness/likelihood surfaces are flattest and multi-modal structure is most pronounced, CDC is clearly disadvantaged (MAE $\sim$86 Myr; coverage 17 %), while DD and DD–PEAKS attain MAE 20 Myr with 33 % coverage.

These shifts are consistent with the expected trade–off under increasing scatter. The CDC ensemble favours peaks that are reproducible across Monte Carlo realisations, and this is advantageous when the two episodes remain cleanly separated (Tier A) but becomes conservative as shoulders and minor modes proliferate (Tiers B to C). By contrast, the DD workflows reduce each bootstrap to a single global maximum of the summed likelihood curve and summarise those maxima, which can preferentially lock onto steeper shoulders in noisier panels and yield smaller MAE as well as higher coverage under our event-wise rule in the two–stage benchmarks.

4.1.4 Attribution effects

We use a conservative truth–window rule that assigns an estimate to a true event only if it falls within that event's window. Estimates outside all windows are left unassigned. Coverage is therefore reported relative to the full set of true events (18 across Cases 5 to 7), with unassigned events effectively counting as misses. Under this rule, CDC and DD each yield assigned point estimates for 11 events, while DD–PEAKS yields 10. Of the 18 true events, 3 (CDC), 6 (DD), and 6 (DD–PEAKS) have 95 % intervals that include the true age (coverage 17 % for CDC vs. 33 % for DD/DD–PEAKS). These counts explain why

https://doi.org/10.5194/egusphere-2026-280
Preprint. Discussion started: 12 February 2026

Table 7. Dataset-level event recovery by tier across Cases 1 to 7. For each dataset, CDC contributes all catalogue peaks; DD contributes one age (the standard single-age DD summary: the median of bootstrap maxima); DD–PEAKS contributes peak centres from the median DD curve. Matches are assigned using the same truth-window rule as in the event-level analysis. Recall is $n_{\mathrm{matched}}/n_{\mathrm{true}}$, precision is $n_{\mathrm{matched}}/n_{\mathrm{pred}}$, and F_1 is the harmonic mean. MAE (matched) is the median absolute error of matched pairs only.

Tier	Method	n_{datasets}	n_{true}	n_{pred}	n_{matched}	Recall (%)	Precision (%)	F_1 (%)	MAE (Myr; matched)
A	CDC	7	12	10	10	83	100	91	27
A	DD	7	12	7	3	25	43	32	71
A	DD–PEAKS	7	12	16	9	75	56	64	71
B	CDC	7	12	10	10	83	100	91	38
B	DD	7	12	7	3	25	43	32	71
B	DD–PEAKS	7	12	15	7	58	47	52	71
C	CDC	7	12	10	8	67	80	73	45
C	DD	7	12	7	4	33	57	42	50
C	DD–PEAKS	7	12	15	8	67	53	59	71

coverage can remain low even when MAE is modest for the subset of assigned events, and why the effective sample size n varies by method and tier in Table 6.

4.1.5 Dataset–level event recovery

Treating each dataset as a set of reported ages, we also compare event–recovery metrics (recall, precision, and F_1) across tiers (Table 7). Recall is the fraction of true events recovered, precision is the fraction of reported events that are true, and F_1 is their harmonic mean.

In Tiers A and B, CDC recovers 10 of 12 true events (recall 83 %) and precision is 100 % (all reported ages match a true event), yielding $F_1 = 91\%$. In Tier C, CDC recall decreases to 67 % (8 of 12 events) and precision to 80 % ($F_1 = 73\,\%$) as scatter increases.

The baseline DD summary returns one age per dataset and therefore recovers only 3 to 4 of 12 true events per tier (recall 25 to 33%), with precision of 43 to 57 %. DD–PEAKS increases recall relative to DD (Tier A 75 %; Tier B 58 %; Tier C 67 %) but does so by reporting more candidate ages per dataset, reducing precision to 47 to 56 %. Among matched events, CDC also attains smaller median absolute error than DD and DD–PEAKS in all tiers (MAE 27 to 45 Myr for CDC versus 50 to 71 Myr for DD and 71 Myr for DD–PEAKS).

4.1.6 Interval widths and empirical coverage

Median 95 % interval widths (per assigned event estimate) increase with noise for CDC and are generally broader than DD at younger ages. By tier, the CDC medians are $\sim$50 Myr in Tier A, 68 Myr in Tier B, and 90 Myr in Tier C, compared

https://doi.org/10.5194/egusphere-2026-280
Preprint. Discussion started: 12 February 2026

Table 8. Median 95 % interval widths (per assigned event estimate) by tier and method. Width is defined as the difference between the upper and lower bounds of the 95 % interval for each event–level estimate.

Tier	Method	Median width (Myr)
A	CDC	50
A	DD	20
A	DD–PEAKS	20
B	CDC	68
B	DD	50
B	DD–PEAKS	60
C	CDC	90
C	DD	28
C	DD–PEAKS	20

with 20/50/28 Myr for the baseline DD workflow (Table 8). By age band, CDC interval widths are $\sim$50 Myr for $<$ 1 Ga and $\sim$70 Myr for $\geq$ 1 Ga, whereas DD widens substantially at $\geq$ 1 Ga (to $\sim$100 to 130 Myr, baseline vs. DD–PEAKS). Despite broader intervals at younger ages, CDC attains higher overall coverage than DD (Table 3). Conversely, DD's typically tighter intervals under–cover on the single-stage panels but widen at older/high–scatter ages, improving coverage there at the expense of interval length.

4.2 Effect of varying the concordant fraction

4.2.1 Single-stage geometry (25 Ma lower intercept)

With the discordant array held fixed, increasing the concordant fraction C from 5 % to 60 % leaves the legacy single-optimum CDC age at 26 Ma for all datasets. This value corresponds to the grid node nearest the true 25 Ma loss age. The empirical 95 % interval around the optimum narrows as C increases. The intervals are C05: 21 to 46 Ma, C10: 21 to 36 Ma, C20: 21 to 41 Ma, C30: 21 to 36 Ma, and C40 to C60: 21 to 26 Ma. Over the same range the K–S distance between the reconstructed and concordant CDFs decreases from $D = 0.191$ at C05 to $D = 0.131$ at C60.

The CDC ensemble catalogue for this geometry contains a single young peak for every value of C. At C05 the ensemble peak lies at 33 Ma with a 95 % interval of 27 to 58 Ma. For C10 to C60 the peak ages fall between 24 and 25 Ma. The corresponding 95 % intervals are very similar and span roughly 13 to 38 Ma (Table 9). Run support for this peak is 100 % for all concordant fractions. Increasing C therefore tightens both the legacy interval and the ensemble interval and pulls the C05 ensemble peak toward the true 25 Ma loss age, but the overall structure remains that of a single young solution.

https://doi.org/10.5194/egusphere-2026-280
Preprint. Discussion started: 12 February 2026

Table 9. Legacy CDC and ensemble peak ages for the single-stage geometry (25 Ma lower intercept) as a function of the concordant fraction C. Legacy ages are the single-optimum CDC results; ensemble ages are peaks from the CDC ensemble catalogue. Intervals are empirical 95 % ranges over Monte Carlo runs.

C	Legacy age (Ma)	Legacy 95 % CI (Ma)	Ensemble age (Ma)	Ensemble 95 % CI (Ma)
5 %	26	21 to 46	33	27 to 58
10 %	26	21 to 36	25	13 to 38
20 %	26	21 to 41	24	13 to 38
30 %	26	21 to 36	24	13 to 38
40 %	26	21 to 26	24	13 to 38
60 %	26	21 to 26	24	13 to 38

4.2.2 Two–loss geometry (300 Ma and 1800 Ma lower intercepts)

To test whether the same pattern holds when two episodes contribute to the discordant array, we repeated the concordant–fraction sweep for the bimodal Case 2, Tier A geometry (true lower intercepts 300 and 1800 Ma). Here, our legacy CDC single-optimum returns one young Pb-loss age for all concordant fractions. For C05 to C40 the optimum age is 393 Ma. The corresponding 95 % intervals contract from 333–463 Ma to 353–443 Ma at C40. At C50 and C60 the optimum age is 383 Ma, with 95 % intervals of approximately 373–453 Ma (C50) and 363–443 Ma (C60). The K–S effect size remains close to $D \approx 0.47$ for all values of C, showing that the discrepancy between the reconstructed and concordant age distributions is large and does not depend strongly on the number of concordant analyses.

The ensemble catalogue for the same datasets contains two peaks at every concordant fraction. The younger peak lies between about 347 and 362 Ma. Its 95 % ensemble intervals range from roughly 300 to 422 Ma at C05 to ∼330 to 410 Ma at C60. The older peak lies between 1793 and 1795 Ma, with 95 % ensemble intervals of approximately 1760 to 1825 Ma. Both peaks have 100 % run support and their locations and widths vary little across the sweep. Once the discordant array is fixed, increasing the concordant fraction mainly affects the widths of the intervals around each peak and has little influence on the presence, position, or relative support of the two solutions.

In the single-stage case both the legacy CDC age and the ensemble catalogue are dominated by a single young solution whose uncertainty contracts as C increases and whose effect size D decreases. In the two–loss case the effect size remains large for all C, the ensemble catalogue consistently resolves two peaks (a young mode in the 300 to 400 Ma band and an older mode near 1800 Ma), and the legacy statistic reports only a young age around 390 Ma with a slowly narrowing interval.

4.3 Peak reproducibility

Every CDC catalogue entry carries a support value equal to the fraction of Monte Carlo runs that vote for that apex. In the synthetic benchmarks considered here, all reported CDC peaks have high support since low–support modes are excluded by

https://doi.org/10.5194/egusphere-2026-280
Preprint. Discussion started: 12 February 2026

Table 10. CDC runtime by tier (R=200; n_{grid}=200).

Tier	End-to-end median (s)	Min (s)	Max (s)	Per-run median (s)
A	71	61	80	0.35
B	70	58	81	0.35
C	71	52	79	0.35

the ensemble gates and therefore do not enter the catalogue. The support distribution is correspondingly concentrated near 1.0 in all tiers.

Support is not a formal probability of correctness, and given the small and relatively clean benchmark set we do not attempt to calibrate a universal numerical threshold for "acceptable" support. Instead, we regard it as a qualitative indicator of peak stability: high–support modes are reproducible across Monte Carlo realisations, whereas low–support modes (when they occur, e.g. in noisier or more complex datasets) should be interpreted with greater caution and, where possible, checked against independent constraints. In practice, we recommend reporting support alongside age and 95 % interval estimates for CDC peaks, particularly in multi–episode or high–scatter datasets.

4.4 Runtime

4.4.1 CDC

On the benchmark grid ($n_{\text{grid}} = 200$) with R=200 Monte Carlo replicates, the updated CDC workflow (including per-run across–cluster minima and the ensemble catalogue) completed typical Case 1 to 4 datasets in a median of 71 s end-to-end (overall range 52 to 81 s across tiers). By tier, end-to-end medians were: Tier A 71 s (61 to 80 s), Tier B 70 s (58 to 81 s), and Tier C 71 s (52 to 79 s) (Table 10).

The Monte Carlo stage dominated wall time, with a per-run median of 0.35 s across tiers. Non–Monte Carlo overhead (concordance classification, per-run peak detection, ensemble catalogue, I/O) contributed only ∼2 s per job. Consequently, CDC runtime scales nearly linearly with R at fixed grid resolution:

$$T_{\text{CDC}} \approx R \times 0.35\,\text{s} + \text{overhead}, \quad \text{for } n_{\text{grid}} = 200. \tag{28}$$

4.4.2 DD

Across tiers, the per-bootstrap median was 18 to 19 s, yielding end-to-end medians of 3565 s (Tier A; ∼59 min), 3981 s (Tier B; ∼66 min), and 4604 s (Tier C; ∼77 min) for n_{boot}=200. These values are based on complete runs with end-to-end time between 1000 and 20000 s, excluding clearly incomplete or anomalously slow runs.

https://doi.org/10.5194/egusphere-2026-280
Preprint. Discussion started: 12 February 2026

Table 11. DD runtime by tier (n_{boot}=200).

Tier	End-to-end median (s)	Min (s)	Max (s)	Per-bootstrap median (s)
A	3565	3185	3917	18
B	3981	3677	5004	19
C	4604	2861	6680	18

Table 12. Runtime comparison between CDC and DD at R=200 and n_{grid}=200. Speedup is defined as DD/CDC. CDC per-run is the per-replicate median; DD per-boot is the per-bootstrap median.

Tier	CDC E2E median (s)	DD E2E median (s)	CDC per-run (s)	DD per-boot (s)	Speedup (×)
A	71	3565	0.35	18	50
B	70	3981	0.35	19	57
C	71	4604	0.35	18	65
Overall[†]	71	4050	0.35	18	57

[†]Mean across tier medians (for context only).

4.4.3 Runtime performance of CDC vs. DD

At identical replicate counts and grid spacing, the updated CDC implementation is consistently faster across tiers (Tables 10, 11). Table 12 summarises tier-wise end-to-end medians, per-replicate costs, and the implied speedups, defined as DD/CDC.

CDC is faster by 50 to 65 times end-to-end depending on tier, with an overall ∼57 times speedup using the mean of tier medians. Per–replicate cost is ∼0.35 s for CDC versus ∼18 to 19 s for DD (roughly ∼50 times faster per replicate). At fixed grid resolution (n_{grid}=200), CDC scales nearly linearly with replicates, while DD scales approximately as

$$T_{\mathrm{DD}} \approx n_{\mathrm{boot}} \times (18 \text{ to } 19)\,\mathrm{s} + \text{overhead}. \tag{29}$$

Non–Monte Carlo overheads in CDC (classification, peak detection, catalogue, I/O) are small (∼2 s per job) relative to the Monte Carlo stage.

5 Discussion

5.1 Why a single-minimum K–S grid search fails for multiple Pb-loss episodes

The original CDC implementation evaluates a grid of candidate Pb-loss ages t. For each t, discordant analyses are back-projected to reconstructed upper-intercept (UI) ages under the assumption that the lower intercept is fixed at t. The empirical cumulative distribution function (CDF) of concordant ages is then compared with the CDF of reconstructed UI ages using the

https://doi.org/10.5194/egusphere-2026-280
Preprint. Discussion started: 12 February 2026

K–S distance $D(t)$ (Eq. (3)). When the invalid-age penalty is enabled, we use the penalised distance $D^*(t)$ (Eq. (5)) and define the corresponding goodness as $S_{\mathrm{pen}}(t) = 1 - D^*(t)$, with the unpenalised form $S_{\mathrm{raw}}(t) = 1 - D(t)$.

Multi-episode Pb-loss and heterogeneous resetting are expected in complex crustal samples where fluid–rock interaction is episodic and zircon domains differ in susceptibility, as illustrated by modelling and application to the Acasta Gneiss Complex (Kirkland et al., 2020). When discordant analyses record more than one genuine Pb-loss episode, the reconstructed UI ages form a mixture distribution. In this setting the one-dimensional K–S surface can be weakly informative with respect to t. Rather than producing a distinct optimum at either true disturbance age, $D^*(t)$ may develop a broad trough, or equivalently $S_{\mathrm{pen}}(t)$ may develop a broad plateau, across the range of candidate ages where the reconstructed-age mixture changes only weakly. A single-minimum grid search must nevertheless return one value, so the inferred Pb-loss age can drift within this plateau and need not correspond to a discrete disturbance. This behaviour is amplified by discretisation of the age grid.

In the legacy implementation, flat minima are resolved deterministically. If a contiguous set of adjacent grid nodes share the same minimum misfit (to numerical precision), the reported optimum is taken as the midpoint of that minimum plateau (or as the nearest boundary if the plateau reaches the edge of the modelling window). This produces a reproducible single age, but it does not restore uniqueness, since any age within the plateau yields near-indistinguishable K–S misfit.

Figure 7 illustrates these effects for a synthetic mixture (Case 2, Tier A) containing two Pb-loss episodes at 300 Ma and 1800 Ma. With the penalty enabled, $S_{\mathrm{pen}}(t)$ exhibits a broad young plateau and the distribution of Monte Carlo run-level optima spans 373 to 403 Ma with a median of 383 Ma. In contrast, the unpenalised curve is dominated by the older episode and yields a narrow run-optima interval of 1759 to 1799 Ma with a median of 1779 Ma.

The difference between the raw and penalised behaviour reflects suppression of the older mode by infeasible reconstructions. At candidate ages near the older episode, many discordant analyses back-project to geologically impossible reconstructed UI ages (or fail to intersect concordia) and are counted as invalid under the penalty rule. For Case 2, Tier A, the penalised calculation at $t = 1779$ Ma yields 48 invalid reconstructed ages, which increases $D^*(t)$ and reduces $S_{\mathrm{pen}}(t)$ near 1800 Ma relative to the young plateau. As a result, the penalised single-optimum summary concentrates on the young plateau, while the raw summary preferentially selects the older disturbance. Neither single-age summary captures both Pb-loss episodes simultaneously, highlighting a fundamental limitation of a single-minimum K–S grid search for mixed-discordance datasets.

We return to this example in Sect. 5.2, where we show how the ensemble workflow preserves and reports multi-modality rather than collapsing the result to a single grid node.

5.2 Advances in the CDC workflow

Figure 7 shows that, in mixed-discordance cases, the K–S distance can be effectively non-unique over a broad plateau, so any one-number summary (whether a single surface optimum or a run-optima median) can drift within that interval without uniquely identifying a geological episode. The updated CDC workflow therefore aims to preserve and report multi-modality in the goodness surface rather than collapsing the result to a single grid node.

In earlier applications of the CDC framework, we often summarised the Monte Carlo output by plotting a histogram or kernel–density estimate (KDE) of the R run–level optima and visually selecting one or more peaks from that density (e.g.

https://doi.org/10.5194/egusphere-2026-280
Preprint. Discussion started: 12 February 2026

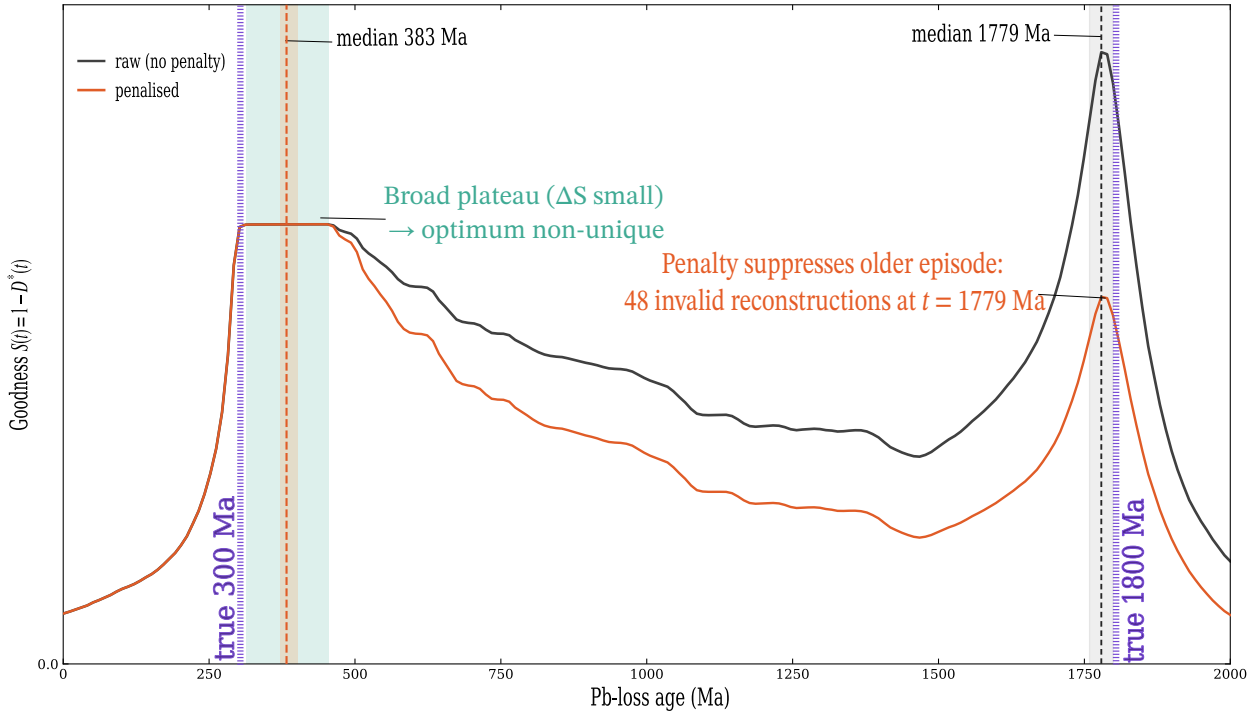

Figure 7. K–S goodness as a function of candidate Pb-loss age for synthetic Case 2, Tier A with two Pb-loss episodes (true ages shown as dotted purple vertical lines at 300 Ma and 1800 Ma). The raw goodness curve (no invalid-age penalty) is shown in black $\left(S_{\mathrm{raw}}(t) = 1 - D(t)\right)$ and the penalised curve is shown in orange $\left(S_{\mathrm{pen}}(t) = 1 - D^*(t)\right)$. Shaded bands show the empirical 95 % interval of Monte Carlo run-level optima; dashed vertical lines mark the medians (penalised 383 Ma, 373 to 403 Ma; raw 1779 Ma, 1759 to 1799 Ma). The penalised curve contains a broad young plateau (teal shaded band), so the optimum is effectively non-unique on the evaluated grid; when several adjacent nodes share the same minimum misfit (to numerical precision), our legacy implementation reports the midpoint of that tied interval. Near the older episode the penalty reduces $S_{\mathrm{pen}}(t)$ because many reconstructions are infeasible (48 of the 120 discordant analyses yield invalid reconstructions at $t = 1779$ Ma), suppressing the older mode relative to the young plateau even though the older mode dominates the unpenalised curve.

https://doi.org/10.5194/egusphere-2026-280
Preprint. Discussion started: 12 February 2026

pooled–sample analyses in Kirkland et al., 2017, 2020; Mathieson et al., 2025a, b). This diagnostic can reveal multi–modality when distinct episodes take turns as the global optimum across runs, but it has two limitations. First, it discards non–global local minima within each run, so minority episodes that never win globally are invisible. Second, it depends on a user-chosen bandwidth, which sets an implicit smoothing scale.

The new workflow differs in three ways. First, candidate modes are detected on the robust median ensemble curve $\tilde{S}(t)$ using range-normalised thresholds, reducing sensitivity to the modelling window. Second, each catalogue entry carries a support value (the fraction of Monte Carlo runs voting for that apex), providing an internal reproducibility measure alongside ages and intervals. Third, when discordant clustering is accepted, misfit is evaluated separately for each discordant sub-array and the across-cluster minimum is used at each trial age (Sect. 2.4), reducing dominance by a numerically large array.

5.3 Summary of benchmark behaviour

Across the synthetic benchmarks, the new CDC workflow and the discordance dating (DD) method of Reimink et al. (2016) show complementary strengths. In the least scattered setting (Tier A), goodness and likelihood profiles as functions of candidate Pb-loss age are relatively sharp. In panels with a single true episode this usually produces one narrow maximum. In panels that contain two distinct disturbance ages at the dataset level (i.e. mixtures of two single-stage discordance arrays), such as Cases 2 and 4 in Tier A, a second mode is often resolved and is commonly well separated. Here we use *single-stage* to describe the discordance geometry of individual analyses (a single chord to a single lower intercept), not the number of disturbance ages present within a dataset; Cases 2 and 4 are therefore single-stage per analysis but multi-episode at the dataset level. As scatter increases in Tiers B and C, the same episodes produce broader features with stronger shoulders and additional competing maxima.

Across the single-stage benchmarks (Cases 1 to 4), CDC consistently attains smaller median absolute error and higher empirical coverage than DD in all tiers (Table 5). In this subset, DD–PEAKS typically coincides with the baseline DD summary because the median summed likelihood curve usually contains only one prominent peak. In the two-stage benchmarks (Cases 5 to 7), performance depends more strongly on separability. Under low scatter (Tier A), CDC and the peak-aware DD variant (DD–PEAKS) give broadly similar results and coverage remains modest for both approaches (Table 6). As scatter increases and the two episodes overlap more strongly (Tiers B and C), the DD variants yield lower median absolute error and higher coverage than CDC (Table 6). At the dataset level, CDC can recover multiple disturbance ages per panel and therefore shows higher recall of true events in Tiers A and B, whereas the baseline DD summary returns a single age per dataset and has low recall in multi-episode cases (Table 7).

The DD–PEAKS variant partially closes this gap by allowing multiple reported ages per dataset, which increases recall relative to the single-age DD summary, but at the cost of reduced precision because more candidate ages are returned per dataset. This recall–precision trade-off is expected whenever a broad or multi-modal likelihood surface is summarised by a discrete list of peaks rather than by a single optimum.

These tier-wise patterns reflect differences in how CDC and DD summarise the same Pb-loss age profiles. The roles of the concordant reference, clustering, the invalid-age penalty, and peak-selection criteria are discussed in Sect. 5.6.

https://doi.org/10.5194/egusphere-2026-280
Preprint. Discussion started: 12 February 2026

Interpretation of empirical coverage requires attention to the event-attribution rule used in scoring. We use a conservative truth-window assignment: an estimated age contributes to a true event only if it falls within that event's window, while estimates outside all windows are left unassigned and therefore count as not covered. Consequently, coverage reflects both interval calibration and the propensity of a workflow to return an attributable estimate in the first place. This also explains why the effective sample size of assigned event estimates can vary by method and tier.

5.4 Geological scenarios represented by Cases 2–7

The synthetic discordia geometries were chosen to mimic patterns that can naturally arise in zircon U–Pb datasets (Table 13). A mixed upper-intercept (UI) age structure may reflect inheritance and mixed provenance, whereas more than one lower-intercept (LI) age may arise where a sample undergoes polyphase tectonothermal and/or hydrothermal reworking. Resetting is rarely uniform across a zircon population because radiogenic Pb mobility depends on access of reactive fluids to damaged or microstructurally compromised zircon domains (e.g. Nasdala et al., 2001), and can be inhibited by annealing (e.g. Zhang et al., 2000) or by zircon overgrowths that shield older domains (Kirkland et al., 2020). These behaviours are documented for zircon from the Acasta Gneiss Complex, where concordant–discordant comparison modelling and zircon domain observations support episodic radiogenic Pb-loss linked to fluid–rock interaction and overgrowth development. Synthetic overprinting experiments in the same study further show that mixtures of single- and two-stage discordance can generate secondary similarity maxima whose position depends on discordia geometry and the degree of overprint, and that recovering more than one disturbance episode requires that at least some analyses retain a single-stage record of the older episode (i.e. are not fully overprinted by the youngest Pb-loss event) (Kirkland et al., 2020).

Cases 2–4 represent mixtures of single-stage discordia arrays. Case 2 (one UI, two LIs) is consistent with polyphase disturbance of a single crystallisation population, where different fractions of grains or domains were affected during different episodes. Such partitioning can arise when fluid access and strain are localised (e.g. along fracture networks or shear zones) and susceptibility varies between grains (e.g. differences in damage state or shielding by overgrowths) (e.g. Kirkland et al., 2020; Nasdala et al., 2001). Case 3 (multiple UIs, one LI) represents inherited or detrital zircon populations spanning multiple growth ages that experience a single later metamorphic or hydrothermal episode, so distinct UI components converge on a shared LI. Case 4 (multiple UIs and multiple LIs) corresponds to samples that contain a composite zircon population (e.g. inherited cores plus younger zircon growth and/or mixed detrital components) and that have experienced more than one disturbance episode, such that different zircon components preferentially record different parts of the disturbance history.

Cases 5–7 introduce sequential Pb-loss to represent partial overprinting. Case 5 captures incomplete overprinting, in which a younger event permits renewed Pb mobility in only some domains that were already disturbed during an older episode, leaving a mixed population of single-stage (old-only) and two-stage (old+young) trajectories. Case 6 allows old-only, young-only, and two-stage trajectories to coexist, which can occur when permeability pathways and zircon susceptibility evolve through time and a younger event variably erases evidence of an older disturbance in some grains. Case 7 combines mixed UI populations with differential overprinting, as expected when the probability of repeated Pb mobility differs between zircon components.

https://doi.org/10.5194/egusphere-2026-280
Preprint. Discussion started: 12 February 2026

Table 13. Examples of geological situations that can generate benchmark discordia mixtures in Cases 2–7.

Case	Synthetic feature	Representative natural scenario(s)
2	One UI, two LIs	A single crystallisation population affected by two disturbance episodes, with different fractions reset during each episode due to localised fluid/strain access and variable zircon susceptibility.
3	Multiple UIs, one LI	Inherited or detrital mixtures spanning multiple growth ages overprinted during one later metamorphic or hydrothermal episode, producing a shared LI across UI components.
4	Multiple UIs, multiple LIs	A sample containing a composite zircon population (e.g. inherited cores plus younger zircon growth and/or mixed detrital components) that experienced more than one disturbance episode, such that different components preferentially record different episodes.
5	Two-stage path in part of the population	Incomplete overprinting: a younger event induces renewed Pb mobility in only some domains that recorded an older episode, leaving a mix of old-only and old+young trajectories.
6	Mixture of old-only, young-only, and two-stage paths	Evolving permeability pathways and time-dependent susceptibility, so some grains record only the older event, some only the younger event, and some both. Older records may be variably erased.
7	Mixed UIs with differential overprinting	Differential susceptibility between UI components (e.g. damage-dependent susceptibility and/or microstructural partitioning of fluid/strain access between components).

UI: upper intercept; LI: lower intercept.

One contributing factor can be differences in accumulated radiation damage (for a given U and Th content, older domains have had longer to accrue α-damage) (e.g. Murakami et al., 1991; Weber, 1990).

5.5 Illustrative natural example: multi-modal Pb-loss signal in a Gawler Craton gneiss, South Australia

To illustrate behaviour on natural data, we applied our updated CDC workflow to zircon U–Pb analyses from a gneiss sample in the Gawler Craton, South Australia (Geoscience Australia, 2026a). In the interpreted results for this sample, the dominant zircon population is attributed to metamorphism at $2427 \pm 4\,\mathrm{Ma}$ (weighted mean $^{207}\mathrm{Pb}/^{206}\mathrm{Pb}$; 95 % confidence), and a discordia regression yields an upper-intercept age of $2823 \pm 37\,\mathrm{Ma}$ with a lower-intercept age of $2161 \pm 250\,\mathrm{Ma}$ (95 % confidence; MSWD = 1.8) (Geoscience Australia, 2026b).

For this example, we model candidate Pb-loss ages over a 1–1500 Ma window using a uniform age grid (approximately 10 Ma spacing) and $R = 100$ Monte Carlo runs (Fig. 8). The concordant subset is dominated by Archean–Palaeoproterozoic

https://doi.org/10.5194/egusphere-2026-280
Preprint. Discussion started: 12 February 2026

ages (median concordant age 2754 Ma; p_{16}–p_{84} 2386–2806 Ma; $n = 34$). Within the modelling window, the penalised ensemble catalogue contains two supported Pb-loss maxima at 815 Ma (95 % CI 780–874 Ma) and 269 Ma (95 % CI 231–380 Ma).

These two younger modelled ages are best interpreted as Pb-loss ages recording episodes of zircon-domain disturbance rather than zircon growth, given their large offset from the main zircon growth and metamorphic ages in this sample. The ∼815 Ma Pb-loss age is consistent with Rodinia breakup–related extension along the Gawler–Adelaide margin and coeval ∼820–830 Ma mafic dyke emplacement, including the Gairdner Dyke Swarm at ∼827 Ma (Wingate et al., 1998; Preiss, 2000; Huang et al., 2015). Mafic intrusion and associated extensional faulting can promote radiogenic Pb-loss in zircon by generating transient thermal perturbations and hydrothermal fluid pathways that preferentially mobilise Pb from radiation-damaged domains (e.g. Cherniak and Watson, 2003; Pidgeon et al., 1966; Kirkland et al., 2017). The younger ∼269 Ma Pb-loss age instead points to a late Palaeozoic–Permian intraplate reactivation/thermal–hydrothermal pulse, where modest reheating and/or fluid flow along inherited structures can selectively reset metamict zircon without requiring magmatism or high-grade metamorphism (e.g. Hall et al., 2018; Glorie et al., 2019; Weisheit et al., 2014; Geisler et al., 2007). This example demonstrates that the updated CDC workflow can preserve and report more than one disturbance-age in a natural dataset. The resulting geometry is consistent with benchmark end-members in which one dominant concordant component is affected by more than one disturbance episode (Cases 2 and 4).

5.6 CDC vs. DD: sources of bias, coverage, and robustness

5.6.1 Clustering acceptance and fallback

Discordant clustering is accepted only when the mixture model yields at least two components that pass size and separation checks that scale with the number of discordant analyses. In our implementation each component must contain at least $m_{\min}(n) = \max\big(4, \lceil 0.12\, n \rceil\big)$ analyses and must exceed a minimum fractional size that tapers to 0.10 for larger n. Components whose median ages are too close relative to a MAD based within cluster scale are merged until their separation exceeds an n dependent factor $c(n)$ (about 1.05 to 1.20). If fewer than two components survive these checks, or the BIC selects a single component, CDC reverts to the single cluster comparison.

5.6.2 Penalty structure and shoulder competition

The invalid–age penalty stabilises CDC against geologically impossible reconstructions. It can also reduce the apparent amplitude of a genuine episode when many analyses from a competing episode reconstruct to impossible ages at the same trial time. In the highest scatter two–stage panels this effect can reduce the prominence of the older episode represented by shorter (more reset) chords so that the ensemble surface favours a single broader maximum. This outcome is controlled by two thresholds: a per-run relative-height threshold that determines whether a run casts a vote near a candidate, and an ensemble-level prominence threshold that determines whether a mode on the median curve is treated as a candidate at all.

https://doi.org/10.5194/egusphere-2026-280
Preprint. Discussion started: 12 February 2026

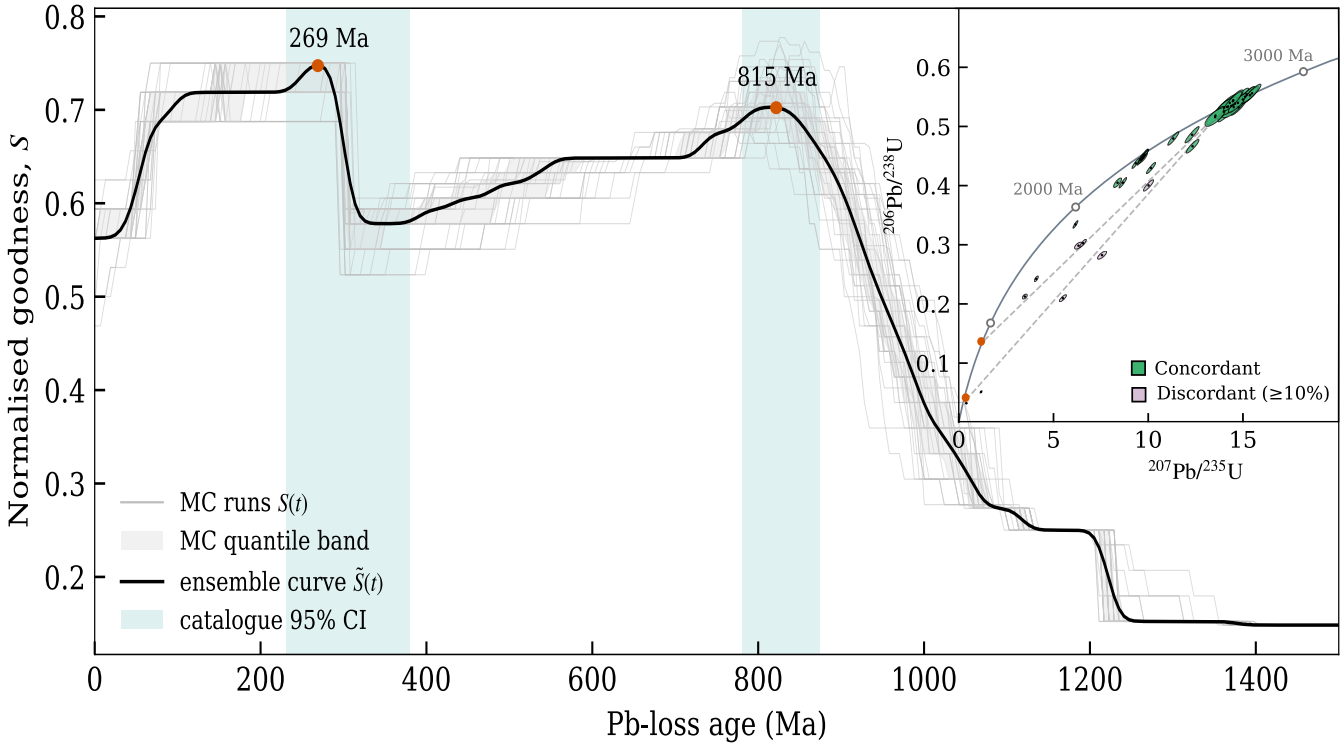

Figure 8. Illustrative natural example (Gawler Craton, South Australia). Penalised CDC goodness curves $S_{\mathrm{pen}}(t)$ from Monte Carlo resampling (grey), their interquartile (25th–75th percentile) envelope (light grey), and the ensemble median curve $\tilde{S}_{\mathrm{pen}}(t)$ (black). Shaded blue bands show the 95 % catalogue confidence intervals (CI) for two supported Pb-loss maxima (markers) at 269 and 815 Ma. Inset: Wetherill concordia. Data ellipses are coloured by a 10 % discordance threshold (green: concordant; purple: discordant). Dashed chords are illustrative connections between concordia at the median concordant age (upper-intercept) and concordia at the catalogue Pb-loss ages (lower-intercepts).

5.6.3 Grid resolution and peak criteria

In the main benchmark suite, candidate Pb-loss ages are evaluated on an approximately 10 Ma grid. CDC uses $N = 200$ nodes spanning 1 to 2000 Ma and DD uses a 10 Ma grid over the same range. CDC then refines each reported apex by local quadratic interpolation and applies run-level and ensemble-level gates on prominence, width, separation, and reproducibility. These rules reduce edge artefacts and discourage shallow shoulders from entering the catalogue. In the flattest multi-episode panels they also make CDC conservative, so some true episodes remain unassigned under the truth-window rule. DD behaves differently because each bootstrap contributes a single global maximum, and the bootstrap spread can widen when the likelihood profile is flat or multi-modal.

https://doi.org/10.5194/egusphere-2026-280
Preprint. Discussion started: 12 February 2026

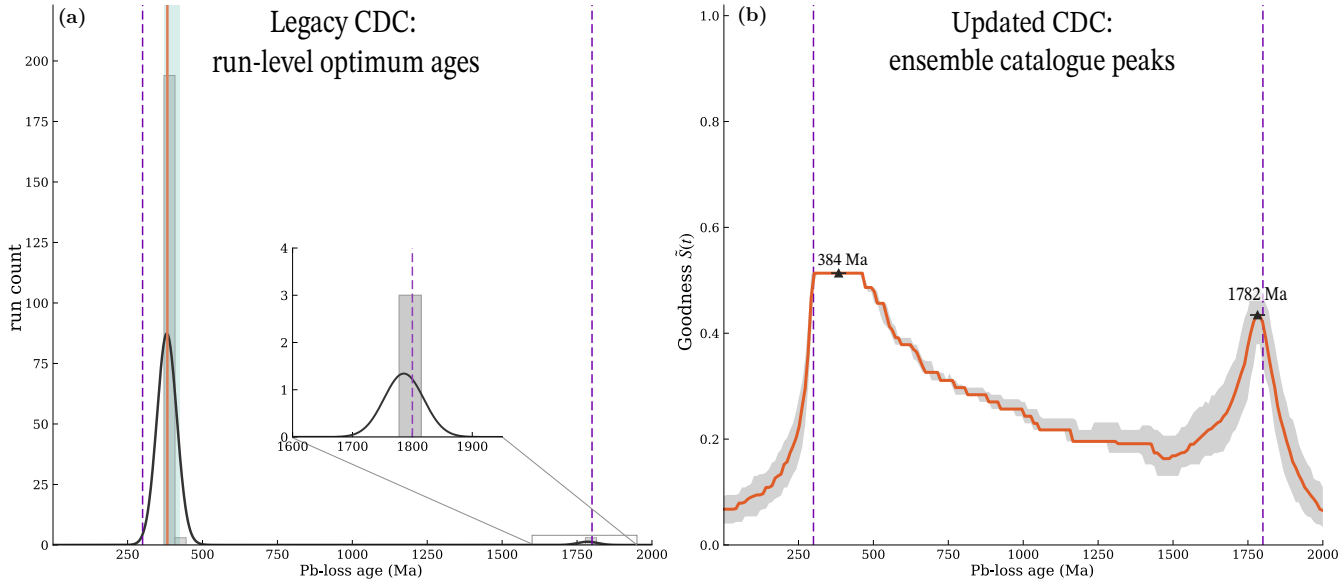

Figure 9. CDC upgrade: legacy per-run optimum versus penalised ensemble catalogue. (A) Legacy diagnostic summary based on the global optimum from each Monte Carlo realisation (histogram of per-run optimum ages; black curve shows a KDE for visualisation). The inset zoom highlights a low-frequency older optimum near ∼1800 Ma that occurs as the global optimum in only a few runs, despite being a real disturbance episode in this synthetic case. **(B)** Updated CDC summary using the penalised ensemble surface: the orange curve is the robust median goodness $\tilde{S}(t)$ across runs and the teal band is the 95 % run-surface envelope. Triangles mark peaks reported in the penalised ensemble catalogue; peak intervals are shown by horizontal bars. Purple vertical dashed lines mark the imposed disturbance episodes at 300 and 1800 Ma.

5.7 Interpretation and reporting

The benchmarks show that the central challenge is not locating a single optimum on a one-dimensional profile, but deciding how to summarise a Pb-loss age signal when the underlying objective is broad or multi-modal. CDC and discordance dating (DD) reduce this structure in different ways, so results should be reported in a form that preserves the information content of the full profile.

For CDC, we recommend reporting the penalised ensemble catalogue together with the median goodness curve $\tilde{S}(t)$ (and, where possible, an uncertainty envelope derived from run-to-run variability). Catalogue support values are particularly important because they quantify the reproducibility of a mode across Monte Carlo realisations and help distinguish robust peaks from shoulders. Here, support is defined as the fraction of Monte Carlo runs that cast a vote for the peak within its voting window (after applying the run-level height gate), rather than the fraction of runs for which the peak is the single global optimum. Figure 9 illustrates why a global-optimum histogram can understate minority episodes, whereas the ensemble catalogue can retain them through vote-based support.

https://doi.org/10.5194/egusphere-2026-280
Preprint. Discussion started: 12 February 2026

The concordant–fraction sweeps provide a useful calibration. In a single-stage geometry, the ensemble catalogue contains one dominant peak whose apex age and interval remain stable as the concordant fraction increases, and a single CDC age is a reasonable one-number summary. In a two-loss geometry with disturbance episodes near $\sim$300 and $\sim$1800 Ma, the ensemble catalogue consistently identifies two well-separated modes with high support, even when a legacy single-age summary remains pinned to the younger plateau. In such cases a single CDC age would hide real structure and the ensemble catalogue should be reported explicitly.

When CDC reports no catalogue peaks within the modelling window, the underlying surface is commonly monotonic or dominated by a boundary optimum. We interpret such outcomes as evidence that the disturbance age is poorly resolved by the available data; any boundary age should be treated as a limit rather than a well-resolved event.

In the present synthetic benchmarks, the reported catalogue peaks have support close to 1.0 because low-support modes are excluded by the ensemble gates; in noisier real datasets, support should be more discriminating and is best interpreted as a stability indicator rather than a probability of correctness.

For DD, it is useful to show both the summed likelihood curve and the distribution of bootstrap maxima. DD makes direct use of the likelihood surface within each bootstrap realisation, but it is more computationally intensive than CDC and, in its standard form, often returns a single summary age per dataset. When the likelihood surface is flat or multi-modal, bootstrap maxima can jump between competing shoulders. This behaviour can be geologically informative, but it also implies that a single point estimate should be interpreted as a summary of a broader, weakly constrained profile.

Agreement between CDC and DD supports a stable interpretation. Persistent disagreement can reflect assumptions intrinsic to each approach. CDC uses the concordant subset as the reference population, so performance is best when concordant and discordant fractions sample the same underlying age distribution and when disturbance does not preferentially target one age component. If discordance is age-dependent (e.g. older grains are more readily damaged and become over-represented in the discordant fraction), then the comparison can be driven by differences in relative peak heights as much as by Pb-loss timing (Reimink et al., 2025). Conversely, DD does not require a concordant reference, but it can be strongly influenced by the most reset analyses; inferred lower-intercept ages can be pulled toward the portion of the population that exerts the greatest likelihood leverage (often the most-reset analyses), which in some geometries yields lower-intercept estimates that are older than the youngest disturbance episode. In such cases the inferred lower intercept may be more naturally interpreted as a maximum-age constraint on the timing of the youngest disturbance rather than as a uniquely resolved event age. DD was introduced and first demonstrated in detrital-zircon-focussed studies (Reimink et al., 2016, 2025), whereas CDC has been applied in both detrital and crystalline (igneous/metamorphic) zircon datasets, provided a representative concordant reference distribution is available.

Finally, computational cost may influence method choice in large studies. At matched replicate counts and grid spacing, CDC is $\sim$50 to 65$\times$ faster end-to-end than DD in our benchmark implementation. This makes CDC well suited to high-throughput screening when a representative concordant reference population is available. DD remains valuable where concordant reference data are sparse or absent, or as a likelihood-based cross-check on key samples, but its bootstrap cost encourages more targeted deployment.

https://doi.org/10.5194/egusphere-2026-280
Preprint. Discussion started: 12 February 2026

5.8 Limitations

CDC behaviour depends on the peak-selection gates, the invalid-age penalty, the age-grid resolution, and how well the concordant subset represents the relevant upper-intercept population. We fixed thresholds a priori using range-normalised criteria tied to the grid and to robust summaries of the median surface. This choice supports reproducibility and limits false positives, but it can make CDC conservative in the most scattered multi-episode panels under the truth-window scoring.

Exact grid endpoints are excluded from peak selection, so profiles that are effectively monotonic within the modelling window can yield no reported peaks. We interpret such outcomes as boundary optima rather than as geologically meaningful events.

Interval width is an important complement to point accuracy. In the benchmarks, CDC 95 % intervals broaden with increasing scatter, while DD intervals are often tighter at younger ages but can under-cover in the single-stage suite and then widen substantially in older or high-scatter settings as the likelihood surface flattens. These patterns highlight that differences in empirical coverage arise from both peak selection/assignment and from how each method's uncertainty summary responds to multi-modality.

5.8.1 Splayed-chord geometry

A failure mode arises when two Pb-loss episodes share a common upper intercept, but the younger episode splays off the older chord. At one candidate age, many analyses from the other episode reconstruct to impossible ages and incur the invalid-age penalty. This can collapse the ensemble surface toward a single intermediate maximum even when both episodes are real. Discordant clustering helps only when the clustering view separates the sub-arrays. We therefore regard the single-upper-intercept splay as a structural limitation of the present CDC method. Addressing this case likely requires a model that allows different discordant sub-arrays to be evaluated against different candidate episodes simultaneously, rather than penalising feasibility under a single shared trial lower intercept. This behaviour is consistent with prior synthetic overprinting experiments showing that multiple Pb-loss events cannot be deconvolved when no analyses retain a single-stage record of the older episode (Kirkland et al., 2020).

6 Conclusions

We present a concordant–discordant comparison (CDC) workflow that extends a single-optimum K–S grid search to a multi-peak, ensemble summary. The workflow optionally partitions discordant analyses into coherent sub-arrays, preserves run-level peaks that pass explicit width and prominence criteria, and merges these candidates into an ensemble catalogue using reproducibility voting and range-normalised peak metrics. Each reported peak is accompanied by an empirical 95 % interval and a support value that summarises peak stability under Monte Carlo resampling.

On synthetic benchmarks spanning Cases 1 to 7 and Tiers A to C, CDC and discordance dating (DD) show complementary behaviour under our truth-window scoring and the configurations tested here. Pooled across all events, CDC attains lower

https://doi.org/10.5194/egusphere-2026-280
Preprint. Discussion started: 12 February 2026

overall median absolute error and higher empirical coverage than the single-age DD summary. In the single-stage suite (Cases 1 to 4), CDC achieves lower event-wise error and higher coverage across tiers. In the two-stage suite (Cases 5 to 7), DD variants achieve lower error and higher coverage than CDC in the higher-scatter tiers (B to C), while Tier A is broadly comparable. Computational cost is also a practical consideration for batch applications: at R=200 and n_{grid}=200, CDC completes in a median of ∼71 s, whereas DD requires ∼59 to 77 min depending on tier.

These benchmarks demonstrate that discordant U–Pb data can, in principle, be mined for geologically meaningful disturbance-age information. By explicitly reporting multi-modality when it is supported, the CDC ensemble catalogue allows more complex Pb-loss histories to be reconstructed from zircon populations than is possible with single-age summaries, provided that CDC outputs are interpreted in conjunction with independent geological constraints and that concordant and discordant subsets plausibly sample the same underlying zircon age structure (i.e., the concordant subset is a valid reference population).

Several limitations remain. CDC depends on how well the concordant subset represents the relevant upper-intercept population, and the invalid-age penalty and peak-selection gates can make CDC conservative when profiles are flat or when episodes overlap strongly. In these settings, disagreement between CDC and DD should be treated as a warning that the Pb-loss age signal is weakly constrained and that any one-number summary will depend on how a broad or multi-modal profile is reduced.

Code and data availability. The LeadLoss software and the data required to reproduce all figures and tables in this manuscript are archived on Zenodo as the release `paper-2025-peak-picking-v1.2.2` (Mathieson, 2026) (version DOI: 10.5281/zenodo.18576518; concept DOI (all versions): 10.5281/zenodo.14039112). The development repository is hosted on GitHub at https://github.com/Curtin-Timescales-of-Mineral-Systems/LeadLoss.
The source SHRIMP analytical data and interpreted age results used for the illustrative Gawler Craton example (Fig. 8) are publicly available from the Geoscience Australia Portal dataset records Geoscience Australia (2026a) and Geoscience Australia (2026b) (last access: 6 February 2026). The portal records provide dataset descriptions and metadata, and the data were retrieved via the OGC Web Feature Service (WFS) endpoints listed in those records. The sample subset used for Fig. 8 is included in the Zenodo archive under `papers/2025-peak-picking/data/inputs/ga_gawler_fig08/`.

Author contributions. Lucy M. Mathieson: Conceptualisation, Methodology, Software, Validation, Formal analysis, Investigation, Data curation, Visualisation, Writing – original draft, Writing – review & editing. Christopher L. Kirkland: Funding acquisition, Methodology, Validation, Supervision, Writing – review & editing.

Competing interests. The authors declare that they have no competing interests.

https://doi.org/10.5194/egusphere-2026-280
Preprint. Discussion started: 12 February 2026

Acknowledgements. This work was supported by the Mineral Exploration Cooperative Research Centre (MinEx CRC), whose activities are funded by the Australian Government's Cooperative Research Centre Programme. This is MinEx CRC document 2026/XX. Portions of the LeadLoss software development workflow were supported by ChatGPT (OpenAI) as a programming assistant (e.g., debugging, refactoring, and drafting code documentation) (OpenAI, 2026). All AI-suggested code and documentation changes were reviewed and validated by the authors.

https://doi.org/10.5194/egusphere-2026-280
Preprint. Discussion started: 12 February 2026

References

Andersen, T., van Niekerk, H., and Elburg, M. A.: Detrital zircon in an active sedimentary recycling system: Challenging the 'source-to-sink' approach to zircon-based provenance analysis, Sedimentology, 69, 2436–2462, https://doi.org/10.1111/sed.12996, 2022.

Arcuri, G., Moser, D., Reinhard, D., Langelier, B., and Larson, D.: Impact-triggered nanoscale Pb clustering and Pb loss domains in Archean zircon, Contributions to Mineralogy and Petrology, 175, 59, https://doi.org/10.1002/9781119227250.ch13, 2020.

Belousova, E., Kostitsyn, Y., Griffin, W., Begg, G., O'reilly, S., and Pearson, N.: The growth of the continental crust: constraints from zircon Hf-isotope data, Lithos, 119, 457–466, https://doi.org/10.1016/j.lithos.2010.07.024, 2010.

Cherniak, D. J. and Watson, E. B. a.: Diffusion in Zircon, Reviews in mineralogy and geochemistry, 53, 113–143, https://doi.org/10.2113/0530113, 2003.

Crowley, Q. G., Heron, K., Riggs, N., Kamber, B., Chew, D., McConnell, B., and Benn, K.: Chemical abrasion applied to LA-ICP-MS U–Pb zircon geochronology, Minerals, 4, 503–518, https://doi.org/10.3390/min4020503, 2014.

Garber, J. M., Smye, A. J., Feineman, M. D., Kylander-Clark, A. R., and Matthews, S.: Decoupling of zircon U–Pb and trace-element systematics driven by U diffusion in eclogite-facies zircon (Monviso meta-ophiolite, W. Alps), Contributions to Mineralogy and Petrology, 175, 55, https://doi.org/10.1007/s00410-020-01692-2, 2020.

Gehrels, G.: Detrital zircon U–Pb geochronology applied to tectonics, Annual Review of Earth and Planetary Sciences, 42, 127–149, https://doi.org/10.1146/annurev-earth-050212-124012, 2014.

Geisler, T., Pidgeon, R. T., Van Bronswijk, W., and Kurtz, R.: Transport of uranium, thorium, and lead in metamict zircon under low-temperature hydrothermal conditions, Chemical Geology, 191, 141–154, https://doi.org/10.1016/S0009-2541(02)00153-5, 2002.

Geisler, T., Schaltegger, U., and Tomaschek, F.: Re-equilibration of Zircon in Aqueous Fluids and Melts, Elements, 3, 43–50, https://doi.org/10.2113/GSELEMENTS.3.1.43, 2007.

Geoscience Australia: Sensitive High Resolution Ion Microprobe (SHRIMP) — Analyses, Geoscience Australia Portal [data set], https://portal.ga.gov.au/metadata/33d9d8e7-f310-43d5-b50a-7fbdd03cab43, Last access: 6 February 2026; retrieved via OGC WFS (CSV): https://services.ga.gov.au/gis/geochronology-isotopes/wfs?request=GetFeature&service=WFS&version=1.1.0&typeName=isotope:shrimp_analyses&outputFormat=csv, 2026a.

Geoscience Australia: SHRIMP — Sample group ages (interpreted results), Geoscience Australia Portal [data set], https://portal.ga.gov.au/metadata/ccbaa8eb-60b0-432d-b2c5-b13ce5eda30c, Last access: 6 February 2026; retrieved via OGC WFS (CSV): https://services.ga.gov.au/gis/geochronology-isotopes/wfs?request=GetFeature&service=WFS&version=1.1.0&typeName=isotope:shrimp_sample_group_ages&outputFormat=csv, 2026b.

Glorie, S., Hall, J. W., Nixon, A., Collins, A. S., and Reid, A.: Carboniferous fault reactivation at the northern margin of the metal-rich Gawler Craton (South Australia): Implications for ore deposit exhumation and preservation, Ore Geology Reviews, 115, 103 193, https://doi.org/https://doi.org/10.1016/j.oregeorev.2019.103193, 2019.

Grauert, B., Seitz, M., and Soptrajanova, G.: Uranium and lead gain of detrital zircon studied by isotopic analyses and fission-track mapping, Earth and Planetary Science Letters, 21, 389–399, https://doi.org/10.1016/0012-821X(74)90178-2, 1974.

Hall, J. W., Glorie, S., Reid, A. J., Boone, S. C., Collins, A. S., and Gleadow, A.: An apatite U–Pb thermal history map for the northern Gawler Craton, South Australia, Geoscience Frontiers, 9, 1293–1308, https://doi.org/10.1016/j.gsf.2017.12.010, 2018.

https://doi.org/10.5194/egusphere-2026-280
Preprint. Discussion started: 12 February 2026

Herrmann, M., Söderlund, U., Scherstén, A., Næraa, T., Holm-Alwmark, S., and Alwmark, C.: The effect of low-temperature annealing on discordance of U–Pb zircon ages, Scientific Reports, 11, 1–11, https://doi.org/10.1038/s41598-021-86449-y, 2021.

Hoskin, P. W., Schaltegger, U., et al.: The composition of zircon and igneous and metamorphic petrogenesis, Reviews in mineralogy and geochemistry, 53, 27–62, https://doi.org/10.2113/0530027, 2003.

Huang, Q., Kamenetsky, V. S., McPhie, J., Ehrig, K., Meffre, S., Maas, R., Thompson, J., Kamenetsky, M., Chambefort, I., Apukhtina, O., et al.: Neoproterozoic (ca. 820–830 Ma) mafic dykes at Olympic Dam, South Australia: links with the Gairdner large igneous province, Precambrian Research, 271, 160–172, https://doi.org/10.1016/j.precamres.2015.10.001, 2015.

Kirkland, C. L., Abello, F., Danišík, M., Gardiner, N. J., and Spencer, C.: Mapping temporal and spatial patterns of zircon U–Pb disturbance: A Yilgarn Craton case study, Gondwana Research, 52, 39–47, https://doi.org/10.1016/j.gr.2017.08.004, 2017.

Kirkland, C. L., Johnson, T. E., Kinny, P. D., and Kapitany, T.: Modelling U-Pb discordance in the Acasta Gneiss: Implications for fluid–rock interaction in Earth's oldest dated crust, Gondwana Research, 77, 223–237, https://doi.org/10.1016/j.gr.2019.07.017, 2020.

Kuper, K. M., Armstrong, R., Kirkland, C. L., Olierook, H. K., Clark, C., and Evans, K.: Implications of high-grade metamorphism on detrital zircon data sets: A case study from the Fraser Zone, Western Australia, Chemical Geology, 647, 121 918, https://doi.org/10.1016/j.chemgeo.2023.121918, 2024.

Kusiak, M. A., Dunkley, D. J., Wirth, R., Whitehouse, M. J., Wilde, S. A., and Marquardt, K.: Metallic lead nanospheres discovered in ancient zircons, Proceedings of the National Academy of Sciences, 112, 4958–4963, https://doi.org/10.1073/pnas.1415264112, 2015.

Kusiak, M. A., Wilde, S. A., Wirth, R., Whitehouse, M. J., Dunkley, D. J., Lyon, I., Reddy, S. M., Berry, A., and De Jonge, M.: Detecting Micro-and Nanoscale Variations in Element Mobility in High-Grade Metamorphic Rocks: Implication for Precise U-Pb Dating of Zircon, Microstructural Geochronology: Planetary Records Down to Atom Scale, pp. 277–291, https://doi.org/10.1002/9781119227250.ch13, 2018.

Kusiak, M. A., Wirth, R., Wilde, S. A., and Pidgeon, R. T.: Metallic lead (Pb) nanospheres discovered in Hadean and Eoarchean zircon crystals at Jack Hills, Scientific Reports, 13, 895, https://doi.org/10.1038/s41598-023-27843-6, 2023.

Massey, F. J.: The Kolmogorov-Smirnov Test for Goodness of Fit, Journal of the American Statistical Association, 46, 68–78, https://doi.org/10.2307/2280095, 1951.

Mathieson, L.: LeadLoss: Pb-loss peak picking tools and manuscript reproduction workflow (paper-2025-peak-picking-v1.2.2), Zenodo [software], https://doi.org/10.5281/zenodo.18576518, concept DOI (all versions): 10.5281/zenodo.14039112, 2026.

Mathieson, L. M., Kirkland, C., Bodorkos, S., and Daggitt, M.: From discordance to discovery: extracting fluid–rock interaction timescales through zircon U–Pb analyses in the Arunta region, Central Australia, Journal of the Geological Society, 182, https://doi.org/10.1144/jgs2024-111, 2025a.

Mathieson, L. M., Kirkland, C., and Daggitt, M.: Turning trash into treasure: Extracting meaning from discordant data via a dedicated application, Geochemistry, Geophysics, Geosystems, 26, https://doi.org/10.1029/2024GC012066, 2025b.

MathWorks: Prominence (Signal Processing Toolbox), https://www.mathworks.com/help/signal/ug/prominence.html, accessed 2025-11-13, 2024.

Mattinson, J. M.: Zircon U–Pb chemical abrasion ("CA-TIMS") method: combined annealing and multi-step partial dissolution analysis for improved precision and accuracy of zircon ages, Chemical Geology, 220, 47–66, https://doi.org/10.1016/j.chemgeo.2005.03.011, 2005.

Meinshausen, N. and Bühlmann, P.: Stability Selection, Journal of the Royal Statistical Society: Series B, 72, 417–473, https://doi.org/10.1111/j.1467-9868.2010.00740.x, 2010.

https://doi.org/10.5194/egusphere-2026-280
Preprint. Discussion started: 12 February 2026

Mezger, K. and Krogstad, E. J.: Interpretation of discordant U-Pb zircon ages: An evaluation, Journal of Metamorphic Geology, 15, 127–140, https://doi.org/10.1111/J.1525-1314.1997.00008.X, 1997.

Morris, G. A., Kirkland, C. L., and Pease, V.: Orogenic paleofluid flow recorded by discordant detrital zircons in the Caledonian foreland basin of northern Greenland, Lithosphere, 7, 138–143, https://doi.org/10.1130/L420.1, 2015.

Moser, D. E., Cupelli, C., Barker, I., Flowers, R., Bowman, J., Wooden, J., and Hart, J.: New zircon shock phenomena and their use for dating and reconstruction of large impact structures revealed by electron nanobeam (EBSD, CL, EDS) and isotopic U–Pb and (U–Th)/He analysis of the Vredefort dome, Canadian Journal of Earth Sciences, 48, 117–139, https://doi.org/10.1139/E11-011, 2011.

Murakami, T., Chakoumakos, B. C., Ewing, R. C., Lumpkin, G. R., and Weber, W. J.: Alpha-decay event damage in zircon, American Mineralogist, 76, 1510–1532, 1991.

Nasdala, L., Pidgeon, R., Wolf, D., and Irmer, G.: Metamictization and U-Pb isotopic discordance in single zircons: a combined Raman microprobe and SHRIMP ion probe study, Mineralogy and Petrology, 62, 1–27, https://doi.org/10.1007/BF01173760, 1998.

Nasdala, L., Wenzel, M., Vavra, G., Irmer, G., Wenzel, T., and Kober, B.: Metamictisation of natural zircon: accumulation versus thermal annealing of radioactivity-induced damage, Contributions to Mineralogy and Petrology, 141, 125–144, https://doi.org/10.1007/s004100000235, 2001.

OpenAI: ChatGPT, Large language model, https://chat.openai.com/, accessed: 2026-01-16, 2026.

Palenik, C. S., Nasdala, L., and Ewing, R. C.: Radiation damage in zircon, American Mineralogist, 88, 770–781, https://doi.org/10.2138/am-2003-5-606, 2003.

Pidgeon, R. T., O'Neil, J. R., and Silver, L. T.: Uranium and Lead Isotopic Stability in a Metamict Zircon under Experimental Hydrothermal Conditions, Science, 154, 1538–1540, https://doi.org/10.1126/SCIENCE.154.3756.1538, 1966.

Preiss, W.: The Adelaide Geosyncline of South Australia and its significance in Neoproterozoic continental reconstruction, Precambrian Research, 100, 21–63, https://doi.org/10.1016/S0301-9268(99)00068-6, 2000.

Reddy, S. M., Timms, N. E., Trimby, P., Kinny, P. D., Buchan, C., and Blake, K.: Crystal-plastic deformation of zircon: A defect in the assumption of chemical robustness, Geology, 34, 257–260, https://doi.org/10.1130/G22110.1, 2006.

Reimink, J. R., Davies, J. H., Waldron, J. W., and Rojas, X.: Dealing with discordance: a novel approach for analysing U–Pb detrital zircon datasets, Journal of the Geological Society, 173, 577–585, https://doi.org/10.1144/jgs2015-114, 2016.

Reimink, J. R., Beckman, R., Schoonover, E., Lloyd, M., Garber, J., Davies, J. H. F. L., Cerminaro, A., Perrot, M. G., and Smye, A.: Discordance dating: A new approach for dating alteration events, Geochronology, 7, 369–385, https://doi.org/10.5194/gchron-7-369-2025, 2025.

Rizvanova, N., Levchenkov, O., Belous, A., Bezmen, N., Maslenikov, A., Komarov, A., Makeev, A., and Levskii, L.: Zircon reaction and stability of the U-Pb isotope system during interaction with carbonate fluid: Experimental hydrothermal study, Contributions to Mineralogy and Petrology, 139, 101–114, https://doi.org/10.1007/s004100050576, 2000.

Rubatto, D., Müntener, O., Barnhoorn, A., and Gregory, C.: Dissolution-reprecipitation of zircon at low-temperature, high-pressure conditions (Lanzo Massif, Italy), American Mineralogist, 93, 1519–1529, https://doi.org/10.2138/am.2008.2874, 2008.

Schoene, B.: U-Th-Pb Geochronology, Treatise on Geochemistry, 4, 341–378, https://doi.org/10.1016/B978-0-08-095975-7.00310-7, 2014.

SciPy Developers: scipy.signal.find_peaks, https://docs.scipy.org/doc/scipy/reference/generated/scipy.signal.find_peaks.html, accessed 2025-11-13, 2024a.

SciPy Developers: scipy.signal.peak_prominences, https://docs.scipy.org/doc/scipy/reference/generated/scipy.signal.peak_prominences.html, accessed 2025-11-13, 2024b.

https://doi.org/10.5194/egusphere-2026-280
Preprint. Discussion started: 12 February 2026

Seydoux-Guillaume, A.-M., Bingen, B., Paquette, J.-L., and Bosse, V.: Nanoscale evidence for uranium mobility in zircon and the discordance of U–Pb chronometers, Earth and Planetary Science Letters, 409, 43–48, https://doi.org/10.1016/j.epsl.2014.10.044, 2015.

Sharman, G. R. and Malkowski, M. A.: Modeling apparent Pb loss in zircon U-Pb geochronology, Geochronology Discussions, 2023, 1–25, https://doi.org/10.5194/gchron-6-37-2024, 2023.

Weber, W.: Radiation-induced defects and amorphization in zircon, Journal of Materials Research, 5, 2687–2697, https://doi.org/10.1557/JMR.1990.2687, 1990.

Weisheit, A., Bons, P. D., Danišík, M., and Elburg, M. A.: Crustal-scale folding: Palaeozoic deformation of the Mt Painter Inlier, South Australia, Geological Society, London, Special Publications, 394, 53–77, https://doi.org/10.1144/SP394.9, 2014.

Wingate, M. T., Campbell, I. H., Compston, W., and Gibson, G. M.: Ion microprobe U–Pb ages for Neoproterozoic basaltic magmatism in south-central Australia and implications for the breakup of Rodinia, Precambrian Research, 87, 135–159, https://doi.org/10.1016/S0301-9268(97)00072-7, 1998.

Zhang, M., Salje, E. K., Capitani, G. C., Leroux, H., Clark, A. M., Schlüter, J., and Ewing, R. C.: Annealing of-decay damage in zircon: a Raman spectroscopic study, Journal of Physics: Condensed Matter, 12, 3131, https://doi.org/10.1088/0953-8984/12/13/321, 2000.

Zhao, J., Xiao, L., Xiao, Z., Wu, X., He, Q., Hao, J., Li, R., and Lin, Y.: Element Redistribution and Age Resetting in Shock-Deformed Zircon from the Chicxulub Impact Structure, Geochimica et Cosmochimica Acta, 393, 219–237, https://doi.org/https://doi.org/10.1016/j.gca.2025.01.021, 2025.