

Supplementary Information

Corpus scale reassembly of fragmented bamboo slips using excavation context and latent slip properties

This file contains Supplementary Figures 1 to 5 and Supplementary Tables 1 to 3. All Methods are given in the main manuscript, as the journal requires.

Contents

Supplementary Figures	2
Supplementary Tables	6

Supplementary Figures

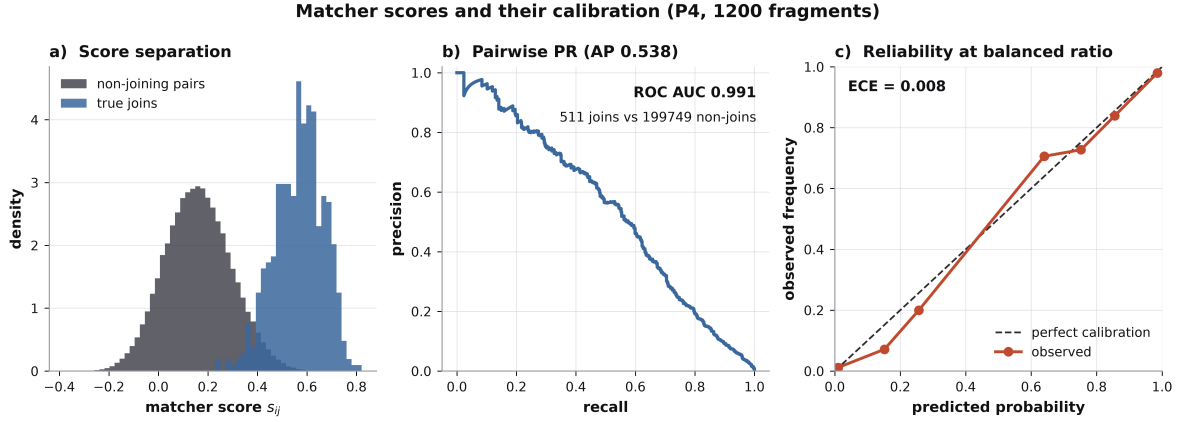


Figure S1: **Matching scores and their calibration at preservation state P4.** a) The score separates joining from non joining pairs. b) Precision against recall at the natural class ratio of about one to four hundred. c) Reliability of the calibration at balanced class ratio, which is the ratio it is fitted at. Platt scaling is fitted on dedicated calibration corpora with known ground truth, disjoint from every evaluation corpus.

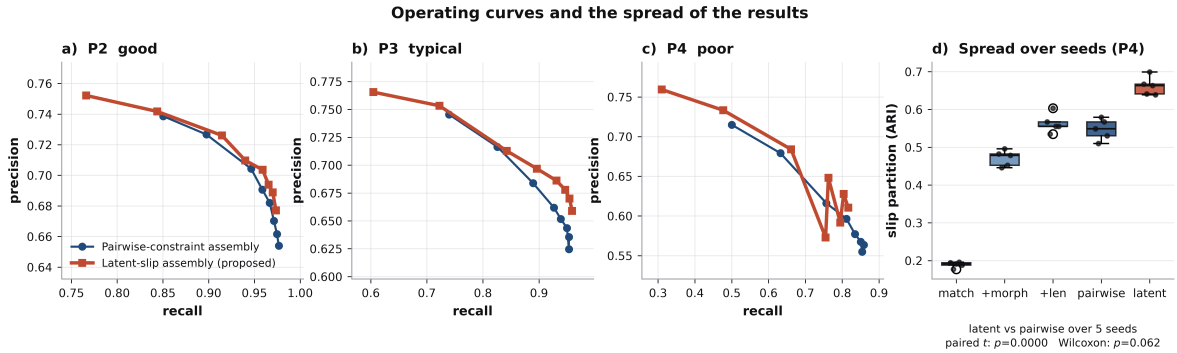


Figure S2: **Operating curves and the spread of the results.** a), b) and c) trace precision against recall as the join threshold varies, at preservation states P2, P3 and P4. d) The distribution over five seeds at P4, with the individual seeds shown, and a paired t test between the latent formulation and the pairwise formulation. The latent formulation wins on all five seeds at P2, P3 and P4, with $p = 0.0004$, $p = 0.0009$ and $p < 0.0001$.

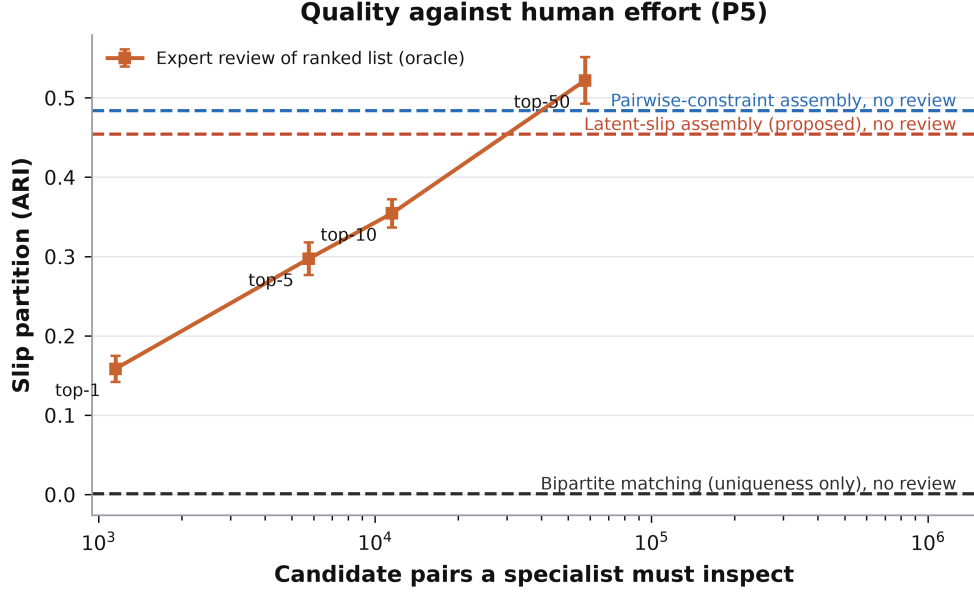


Figure S3: **Reconstruction quality against the specialist time each workflow costs, at preservation state P5.** The orange curve is an oracle specialist working down a ranked list, who finds the correct partner whenever it appears in the top k and never errs, charged the number of candidate pairs that must be inspected. The dashed lines are the optimisation based methods, which return a reconstruction rather than a list and so are charged no review at all. The proposed model reaches a partition index of 0.454 with no review, which the oracle reaches only after inspecting about 39,000 candidate pairs.

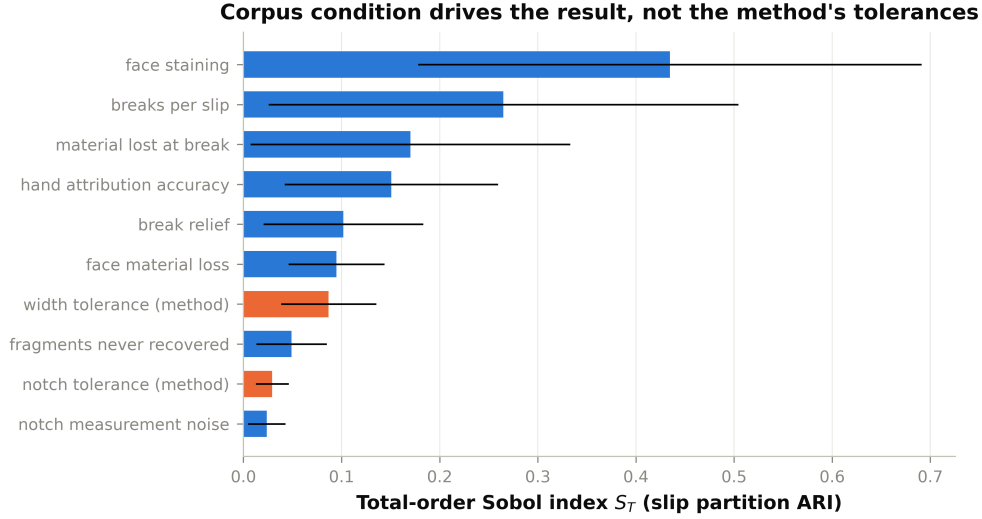


Figure S4: **Total order Sobol indices for the slip partition index.** Orange bars are the two tolerances of the method, the notch window τ_ν and the width gate τ_Ω . Blue bars are properties of the corpus and of the observations. The two tolerances are the least influential inputs, at 0.029 and 0.087, while the independent staining of the two fracture faces dominates at 0.435. Bars are 95% confidence intervals over 384 model evaluations.

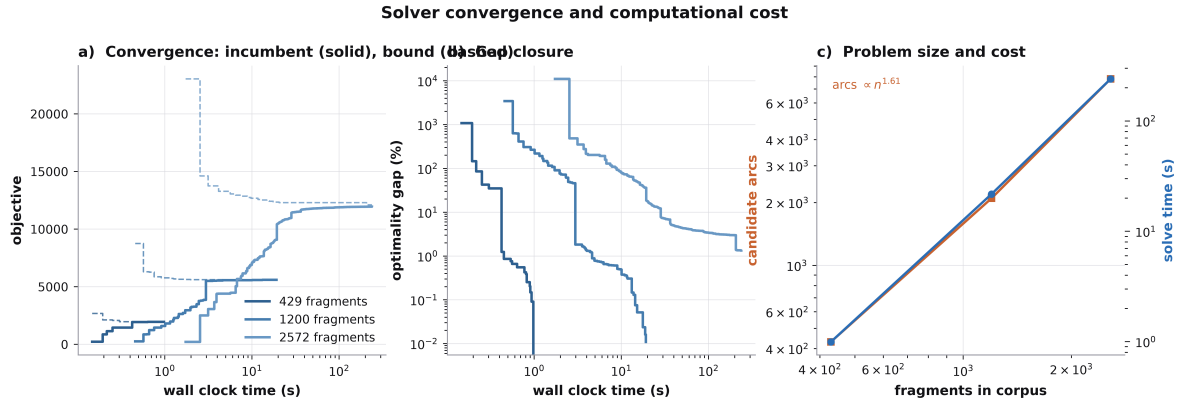


Figure S5: **Convergence and cost of the optimisation.** a) The incumbent solution and the bound during the search. b) The optimality gap closing over time. c) Candidate set size and solve time against corpus size. The model is solved to proven optimality at 429 and 1200 fragments, in 3.2 s and 69.7 s. At 2572 fragments the gap is about 2% when a 240 s budget expires.

Supplementary Tables

Table S1: Hyperparameters. Only θ is tuned, and it is tuned on calibration corpora that are disjoint from every evaluation corpus.

Quantity	Symbol	Value
Encoder embedding dimension	d	128
Descriptor size		$2 \times 32 \times 48$
Training pairs		179952
Training epochs, batch size		40, 512
Learning rate, schedule		3×10^{-4} , one cycle
Candidates kept per break end	k	50
Width gate	τ_{Ω}	0.45 mm
Notch window half width	τ_{ν}	16 mm
Unplaced notch penalty	ρ	1.0
Maximum slip length	$\Lambda_{\max}$	285 mm
Hand attribution accuracy	p	0.80
Number of scribal hands	K	6
Join threshold	θ	tuned per method on calibration corpora
Solver time budget		180 s (900 s for the scaling study)

Table S2: Compared methods. All optimisation based methods are constraint subsets of the same model, solved by the same solver on the same arc set.

Method	Type	What it does	Why it is included
Top-1 candidate	heuristic	take the best candidate for each fragment	the naive use of a ranker
Mutual best	heuristic	accept $i \rightarrow j$ only if j also prefers i	standard practical heuristic
Bipartite matching	optimisation	uniqueness only	strongest method that does not model the slip
+ morphometry, context	optimisation	adds width gate and excavation unit	isolates the cheap gates
+ slip length	optimisation	adds order, acyclicity and length	isolates the chain structure
Pairwise constraints	optimisation	all constraints, pairwise proxies	our first formulation
Latent-slip (proposed)	optimisation	all constraints, latent Λ, Ω, η	the proposed model
Expert review (oracle)	reference	finds the partner if it is in top k	upper bound on ranked-list workflows

Table S3: **Scaling with corpus size** at preservation state P4, mean over three seeds. At 2,540 and 5,060 fragments the proposed model reaches a feasible but not provably optimal solution within a 900 s budget.

Fragments	Method	Join F_1	Slip ARI	Exact slips	Cross-slip	Solve (s)
425	Bipartite matching	0.457	0.305	0.136	0.466	0.0
	Latent-slip (proposed)	0.724	0.732	0.433	0.047	2.5
1151	Bipartite matching	0.353	0.188	0.041	0.617	0.2
	Latent-slip (proposed)	0.676	0.669	0.374	0.117	229.7
2540	Bipartite matching	0.288	0.116	0.016	0.712	1.9
	Latent-slip (proposed)	0.626	0.582	0.349	0.207	905.1
5060	Bipartite matching	0.227	0.065	0.003	0.778	5.2
	Latent-slip (proposed)	0.554	0.491	0.284	0.311	901.6