Resource Allocation and Optimization
Volume 1 · Number 1 · 2026

When Signal-Guided Retention Loses to Index Coverage: A Character-Profile Budget Study

Sean Franklin[1] · Patrick Marshall[2] · Benjamin Lewis[3]

[1] Canada · [2] United States · [3] United Kingdom

Original publication: October 31, 2026 | Article 625

Abstract

A heuristic that directs observations toward changing signals can overlook variation hidden between similar endpoints. We tested such a retention rule on 2,858 released character-velocity profiles, comparing it with repeated splitting of the largest index gap. Both rules retained exactly sixteen or thirty-two observations and used linear reconstruction. Signal-guided retention increased mean profile error by 0.040904 at sixteen points and 0.005159 at thirty-two; character-clustered intervals were 0.015162-0.066646 and 0.000529-0.009790. Doubling the budget lowered pooled error under both rules. These results concern compression of already processed profiles from one writer, and do not evaluate real-time tablet sampling, character readability, or an image-restoration network.

Deterministic retention and reconstruction of public preprocessed pen-velocity profiles from one writer.

When Signal-Guided Retention Loses to Index Coverage: A Character-Profile Budget Study

Deterministic retention and reconstruction of public preprocessed pen-velocity profiles from one writer.

Abstract

A heuristic that directs observations toward changing signals can overlook variation hidden between similar endpoints. We tested such a retention rule on 2,858 released character-velocity profiles, comparing it with repeated splitting of the largest index gap. Both rules retained exactly sixteen or thirty-two observations and used linear reconstruction. Signal-guided retention increased mean profile error by 0.040904 at sixteen points and 0.005159 at thirty-two; character-clustered intervals were 0.015162-0.066646 and 0.000529-0.009790. Doubling the budget lowered pooled error under both rules. These results concern compression of already processed profiles from one writer, and do not evaluate real-time tablet sampling, character readability, or an image-restoration network.

1. Where to spend an observation budget

Keeping more observations generally costs more storage or acquisition effort, but a fixed observation count still leaves a placement decision. One option seeks reasonably even coverage along the sequence index. Another directs its next retained point toward intervals whose currently observed endpoints differ more strongly. The second option uses signal information and can sound more informed. Its effectiveness nevertheless depends on whether endpoint differences reveal the important variation inside an interval.

The research question is whether this particular endpoint-guided allocation rule lowers reconstruction error relative to index-gap allocation at equal point counts. We study sixteen-point and thirty-two-point budgets separately. The directional proposition predicts a negative guided-minus-reference error under each budget. This is a comparison of two explicit deterministic retention procedures, not a theorem about adaptive sampling or a search for the optimal compression algorithm. An unfavorable result for the heuristic would be informative without implying that signal-aware allocation is always inferior.

Character reconstruction work such as Meng et al., 2025 provides context for allocating limited information to restoration. Our experiment operates on normalized velocity vectors from a pen-trajectory collection, rather than pixels from inscription rubbings. It examines where a sparse representation places retained values. There is no execution of the cited image model, no training of a new inpainting network, and no claim that the observed ordering transfers to a historical writing system. The connection is the experimental question of allocating a fixed information budget.

2. Already processed velocity profiles are the object being compressed

The source is Williams, 2006, the UCI Character Trajectories archive. Its MATLAB file contains 2,858 sequences and twenty character labels. All sequences originate from the same writer and involve a single pen-down segment. Their lengths range from 109 to 205 samples, with a median of 174. We inspect array shapes, finite values, and label alignment before applying a retention policy. Every sequence is long enough to support both selected budgets, so no short-profile exclusion changes the evaluation sample.

The source records three channels related to horizontal position, vertical position, and pen-tip force, but the released profiles have been numerically differentiated, smoothed, normalized, and aligned. We analyze only the first two velocity channels. The third, force-related channel is unused. Although the original acquisition used a 200 Hz tablet, the values being reconstructed here are the released processed

velocities. We do not integrate them into physical positions, estimate path length in millimeters, or interpret the numerical loss as displacement of a pen tip.

This differs from Alpaydin and Alimoglu, 1996, whose compact digit attributes represent spatially resampled coordinate points. Equal indices in the present source refer to a processed temporal profile, while the other resource's points have an arc-length construction. We do not combine their observations or compare their numerical error levels. Dataset distinctions also separate these sequences from the oracle-character context of Wang and Deng, 2024 and the character-learning setting in Lake et al., 2019. Shared subject matter does not make their representations interchangeable.

3. A restriction that prevents an oracle allocation

The retention procedures begin with the first and last vector of a profile. At each step they choose an interval between two already retained indices and add the vector at its integer midpoint, rounded downward. Selection then repeats until the required point count is reached. Both rules must decide which interval to split before seeing its midpoint value. They cannot evaluate all candidate midpoint errors and retroactively choose the most helpful one. This information boundary is essential to a non-oracle comparison.

The reference rule selects the eligible interval with greatest index length. Intervals with no unretained integer point are ineligible. Ties go to the interval with the lower left endpoint. Consequently the reference is an index-coverage rule, not an assumption that the underlying velocity is constant. It can discover a hidden excursion by allocating a point to a wide interval even when its observed endpoint vectors are similar. Its eventual spacing depends on integer midpoint splits and need not form an exactly uniform grid.

The guided rule multiplies an interval's index length by the Euclidean distance between its two retained endpoint vectors. It splits the interval with the largest product, using the same tie convention. Endpoint vectors, interval length, and the existing retention set are its only inputs. Character labels, unretained vector values, and eventual reconstruction errors do not enter its priorities. The rule can favor a wide interval with changing endpoints, while giving a low priority to an interval whose endpoints conceal an interior excursion.

4. The comparison is about released-profile retention

An important source restriction precedes these policies. Differentiation, Gaussian smoothing, source normalization, and alignment have already been applied to the complete records. A retained vector can therefore contain information derived from neighboring original observations during source preprocessing. We do not describe the operation as physically deciding when to acquire raw tablet measurements. It is a compression experiment on a fixed public representation, with policy information restricted only during the subsequent retention stage.

Curve reduction in Ramer, 1972 and Douglas and Peucker, 1973 supplies background for the general problem of representing a signal with fewer points. Time-series representation and segmentation work in Keogh et al., 2001b and Keogh et al., 2001a provides additional context for the tradeoff between a compact description and retained shape information. Those algorithms are not benchmarked here. The reference and guided rules have been stated directly so that the reported error cannot be mistaken for the performance of a named published simplification procedure.

The path-graph perspective in Ortega et al., 2017 and Shuman et al., 2012 helps make the remaining information structure explicit: indices order the observations, and selected indices anchor reconstruction

between neighboring retained values. No data-dependent graph, edge-weight learner, or graph neural network is fitted. Labels are used later as covariance groups, without changing which values each policy retains. The resource comparison is exactly the number of retained vector observations, rather than a learned representation's parameter count.

5. A common reconstruction and a common scoring target

After selection, each velocity channel is reconstructed by piecewise linear interpolation in sample index. Endpoints are always present, so every source index lies between retained anchors or is itself retained. We verify that the reconstruction equals the source vectors at all selected indices. The method does not extrapolate, smooth the reconstruction again, or use character-specific templates. Spline foundations in de Boor, 1978 and alternative interpolation ideas in Fritsch and Carlson, 1980 identify possible extensions, but the executed recovery rule remains linear for both allocations.

For every source profile, the score averages squared error across all sample indices and both selected channels. Retained indices contribute zero error because their vectors are reproduced exactly. Evaluating the complete profile gives both policies the same target object, despite their different retained sets. It avoids comparing losses calculated only on two differently composed collections of hidden indices. The resulting value is mean squared error per normalized channel sample, not squared two-dimensional point distance summed across channels.

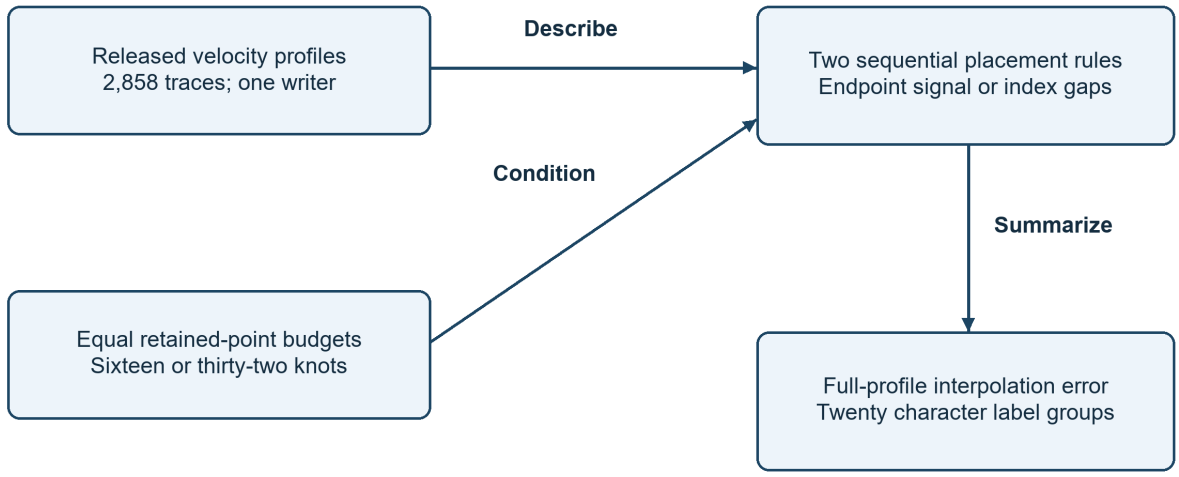


Figure 1. Complete released profiles supply values only at the indices selected by each sequential policy; identical interpolation then reconstructs the same full scoring target.

The sixteen-point budget retains an average of 9.600 percent of each profile's samples; the thirty-two-point budget retains 19.200 percent. These are averages of profile-specific fractions, not the budget divided by the median sequence length. Actual storage would also need indices, coordinate precision, headers, and a representation format. No packed files or runtime measurements are produced, so retained count should not be renamed a measured compression ratio or execution speedup.

6. Paired mean contrasts, grouped by character label

At each budget we subtract index-rule error from guided-rule error for every source profile. An intercept-only ordinary least-squares regression estimates the average of those 2,858 paired differences. A positive coefficient indicates that guided retention is worse on the specified profile loss. The two fits are reported separately rather than treating two budgets applied to the same profile as independent handwriting examples. The score averages profiles equally, so a longer sequence does not automatically receive greater inferential weight.

We cluster covariance by the twenty released character labels, use a finite-sample correction, and apply a t reference with nineteen degrees of freedom. The grouped-dependence framework in Cameron and Miller, 2015 motivates this choice. Heteroskedasticity-consistent estimation in White, 1980 alone would not account for repeated examples within a character group. Nevertheless twenty labels from one writer are not twenty independent writers. The resulting intervals offer a character-grouped summary of this collection, rather than a population guarantee about handwriting styles.

Table 1. Guided-minus-index mean reconstruction error at two exact retained-point budgets.

Model	Term	Estimate (SE)	95% CI	N
M1	Mean outcome/contrast	0.0409 (0.0123)	[0.0152, 0.0666]	2858
M1: fit	R-squared = 0.0000	-	-	2858
M2	Mean outcome/contrast	0.0052 (0.0022)	[0.0005, 0.0098]	2858
M2: fit	R-squared = 0.0000	-	-	2858

Model labels: M1 = Endpoint-signal versus index-gap allocation, 16 knots; M2 = Endpoint-signal versus index-gap allocation, 32 knots.

As these regressions estimate only a mean, their R-squared values are effectively zero. They do not quantify the fraction of the velocity signal reconstructed or the proportion of a character preserved. No classifier outcome, reader response, or latent semantic variable enters the model. Small numerical negative R-squared values arising from floating-point subtraction are treated as zero in the presentation, without changing the estimated mean, covariance, or interval.

7. Signal guidance worsens the pooled error at both budgets

With sixteen retained points, index allocation has average profile error 0.035459490 and guided allocation 0.076363576. Their difference is 0.040904085, with standard error 0.012299056 and a 95% interval from 0.015161865 to 0.066646305. The guided mean is more than twice the reference mean in this setting. Since the budgets and reconstruction rule match, that contrast reflects the particular placement policy interacting with the released velocity profiles, rather than the guided rule being given fewer retained values.

At thirty-two points, average errors are 0.003630634 for index allocation and 0.008789688 for guided allocation. The difference is 0.005159054, with standard error 0.002212375 and interval 0.000528501 to 0.009789607. Doubling the point budget lowers the pooled error under both allocation rules. The remaining guided disadvantage is smaller in absolute units, while its ratio to the reference mean remains substantial. Absolute and relative descriptions therefore emphasize different aspects of the same result.

The proposition favoring this guided rule is not supported by either mean contrast. We do not rescue it by selecting another character subset, changing the tie-breaking rule after seeing outcomes, or choosing a

new signal-priority formula. A stronger heuristic could perform differently. The reported comparison is useful as a negative example of assuming that an endpoint-based signal score automatically identifies the most valuable place to retain additional information.

8. A plausible failure mode is not yet a measured mechanism

The priority formula gives little weight to a wide interval when its retained endpoint vectors are similar. A processed velocity profile can vary within such an interval and return toward its initial vector by the other endpoint. The midpoint rule then receives little encouragement to explore it. This explains a possible weakness of the heuristic from its definition. It does not establish how much of the observed loss difference comes from that specific pattern, since we have not labeled or estimated a mechanism for individual failure intervals.

Influence reasoning associated with Cook, 1977 could motivate inspecting profiles contributing most to the pooled difference. We retain every profile and do not remove high-loss examples. The missing-data perspective in Little and Rubin, 2002 also limits interpretation: deliberate retention from complete processed records differs from naturally unavailable observations. Measurement-error ideas in Carroll et al., 2006 further distinguish the normalized released channels from a noise-free physical velocity. The study does not identify either a natural missingness mechanism or a sensor-error distribution.

Agreement concepts in Bland and Altman, 1986, Lin, 1989, and Koo and Li, 2016 would become relevant if a later study compared reconstruction scores with character-recognition or reader outcomes. We estimate none of those agreement quantities. Two reconstructions with similar profile error could still differ in semantic usefulness, while a large numerical error might occur in a region irrelevant to identifying a character. The current experiment has no evidence with which to establish either relationship.

9. Sparse profile reconstruction and image inpainting remain different tasks

Image-completion methods in Pathak et al., 2016, Liu et al., 2018, Suvorov et al., 2021, and Lugmayr et al., 2022 concern richer reconstruction settings than retaining velocity knots. The image prior discussed in Ulyanov et al., 2017 also uses a different representation and source of inductive structure. Their relevance is to how a limited information set interacts with a reconstruction rule. We do not map a velocity-knot budget into image-mask area, claim equivalent damage mechanisms, or use these results to rank those systems.

Interpretation following Wasserstein and Lazar, 2016 emphasizes the observed mean differences and their restrictions rather than a generic claim that adaptive allocation fails. The reproducibility questions in Ioannidis, 2005 remain pertinent to one writer and one heuristic. Replication could vary the writer, channel representation, budget, or priority score using a declared protocol. Such a program would test whether index coverage is a robust default or merely favorable in this narrow collection. It has not already been carried out by enumerating the profiles available here.

10. Resource guidance needs a verified information boundary

This experiment shows that exploiting observed endpoint values can spend a retention budget less effectively than maintaining index coverage. At both tested budgets, the chosen signal-guided heuristic increases full-profile reconstruction error. Increasing the budget helps both procedures, but it does not reverse their observed ordering. The result is specific enough to reproduce and limited enough to avoid turning a failed heuristic into a universal conclusion about information allocation.

Documentation under Gebru et al., 2021 includes the single writer, processed velocity channels, source sequence lengths, exact retention priorities, midpoint and tie conventions, retained indices, and full-profile loss denominator. Preservation consistent with Wilkinson et al., 2016 keeps these records with the original MATLAB archive and per-profile paired errors. They allow the resource comparison to be checked without inventing a physical sampling experiment. Establishing usefulness for restoration practice still requires task-specific evidence about the representation and outcome that a real application intends to preserve.

References

1. Wasserstein, Ronald L.; Lazar, Nicole A.. (2016). The ASA Statement on p -Values: Context, Process, and Purpose. The American Statistician. https://doi.org/10.1080/00031305.2016.1154108.

2. Suvorov, Roman; Logacheva, Elizaveta; Mashikhin, Anton; Remizova, Anastasia; Ashukha, Arsenii; Silvestrov, Aleksei; Kong, Naejin; Goka, Harshith; Park, Kiwoong; Lempitsky, Victor. (2021). Resolution-robust Large Mask Inpainting with Fourier Convolutions. arXiv:2109.07161. https://arxiv.org/abs/2109.07161.

3. Carroll, Raymond J., David Ruppert, Leonard A. Stefanski, and Ciprian M. Crainiceanu. “Measurement Error in Nonlinear Models.” 2006. https://doi.org/10.1201/9781420010138.

4. Little, Roderick J. A., and Donald B. Rubin. *Statistical Analysis with Missing Data*. Wiley, 2002. https://doi.org/10.1002/9781119013563.

5. Wilkinson, Mark D.; Dumontier, Michel; Aalbersberg, IJsbrand Jan; Appleton, Gabrielle; Axton, Myles; Baak, Arie; Blomberg, Niklas; Boiten, Jan-Willem; da Silva Santos, Luiz Bonino; Bourne, Philip E.; Bouwman, Jildau; Brookes, Anthony J.; Clark, Tim; Crosas, Mercè; Dillo, Ingrid; Dumon, Olivier; Edmunds, Scott; Evelo, Chris T.; Finkers, Richard; Gonzalez-Beltran, Alejandra; Gray, Alasdair J.G.; Groth, Paul; Goble, Carole; Grethe, Jeffrey S.; Heringa, Jaap; ’t Hoen, Peter A.C; Hooft, Rob; Kuhn, Tobias; Kok, Ruben; Kok, Joost; Lusher, Scott J.; Martone, Maryann E.; Mons, Albert; Packer, Abel L.; Persson, Bengt; Rocca-Serra, Philippe; Roos, Marco; van Schaik, Rene; Sansone, Susanna-Assunta; Schultes, Erik; Sengstag, Thierry; Slater, Ted; Strawn, George; Swertz, Morris A.; Thompson, Mark; van der Lei, Johan; van Mulligen, Erik; Velterop, Jan; Waagmeester, Andra; Wittenburg, Peter; Wolstencroft, Katherine; Zhao, Jun; Mons, Barend. (2016). The FAIR Guiding Principles for scientific data management and stewardship. Scientific Data. https://doi.org/10.1038/sdata.2016.18.

6. Gebru, Timnit, et al. "Datasheets for datasets" *Communications of the ACM*, 2021. https://doi.org/10.1145/3458723.

7. Bland, J. Martin, and Douglas G. Altman. "Statistical Methods for Assessing Agreement between Two Methods of Clinical Measurement." *The Lancet*, vol. 327, no. 8476, 1986, pp. 307-310. https://doi.org/10.1016/S0140-6736(86)90837-8.

8. Lin, Lawrence I-Kuei. (1989). A Concordance Correlation Coefficient to Evaluate Reproducibility. Biometrics. https://doi.org/10.2307/2532051.

9. Keogh, Eamonn; Chakrabarti, Kaushik; Pazzani, Michael; Mehrotra, Sharad. (2001b). Dimensionality Reduction for Fast Similarity Search in Large Time Series Databases. Knowledge and Information Systems. https://doi.org/10.1007/pl00011669.

10. de Boor, Carl. (1978). A Practical Guide to Splines. Applied Mathematical Sciences. https://doi.org/10.1007/978-1-4612-6333-3.

11. Meng, Zijie; Zeng, Yuanze; Chang, Xiang; Xu, Tianshuo; Chao, Fei; Cao, Xixin; Shang, Changjing; Shen, Qiang. (2025). Orpaint: a zero-shot inpainting model for oracle bone inscription rubbings with visual mamba block. Science China Information Sciences, 68(8), 189102. https://doi.org/10.1007/s11432-024-4493-4.

12. Keogh, E.; Chu, S.; Hart, D.; Pazzani, M.. (2001a). An online algorithm for segmenting time series. Proceedings 2001 IEEE International Conference on Data Mining. https://doi.org/10.1109/icdm.2001.989531.

13. White, Halbert. “A Heteroskedasticity-Consistent Covariance Matrix Estimator and a Direct Test for Heteroskedasticity.” *Econometrica*, vol. 48, no. 4, 1980, pp. 817-838. https://doi.org/10.2307/1912934.

14. Ramer, Urs. (1972). An iterative procedure for the polygonal approximation of plane curves. Computer Graphics and Image Processing. https://doi.org/10.1016/s0146-664x(72)80017-0.

15. Cameron, A. Colin, and Douglas L. Miller. “A Practitioner’s Guide to Cluster-Robust Inference.” *Journal of Human Resources*, vol. 50, no. 2, 2015, pp. 317-372. https://doi.org/10.3368/jhr.50.2.317.

16. Fritsch, F. N.; Carlson, R. E.. (1980). Monotone Piecewise Cubic Interpolation. SIAM Journal on Numerical Analysis. https://doi.org/10.1137/0717021.

17. Ioannidis, John P. A.. (2005). Why Most Published Research Findings Are False. PLoS Medicine. https://doi.org/10.1371/journal.pmed.0020124.

18. Liu, Guilin; Reda, Fitsum A.; Shih, Kevin J.; Wang, Ting-Chun; Tao, Andrew; Catanzaro, Bryan. (2018). Image Inpainting for Irregular Holes Using Partial Convolutions. arXiv:1804.07723. https://arxiv.org/abs/1804.07723.

19. Shuman, David I; Narang, Sunil K.; Frossard, Pascal; Ortega, Antonio; Vandergheynst, Pierre. (2012). The Emerging Field of Signal Processing on Graphs: Extending High-Dimensional Data Analysis to Networks and Other Irregular Domains. arXiv:1211.0053. https://arxiv.org/abs/1211.0053.

20. Cook, R. Dennis. (1977). Detection of Influential Observation in Linear Regression. Technometrics. https://doi.org/10.1080/00401706.1977.10489493.

21. DOUGLAS, DAVID H; PEUCKER, THOMAS K. (1973). ALGORITHMS FOR THE REDUCTION OF THE NUMBER OF POINTS REQUIRED TO REPRESENT A DIGITIZED LINE OR ITS CARICATURE. Cartographica. https://doi.org/10.3138/fm57-6770-u75u-7727.

22. Ulyanov, Dmitry; Vedaldi, Andrea; Lempitsky, Victor. (2017). Deep Image Prior. arXiv:1711.10925. https://arxiv.org/abs/1711.10925.

23. Lake, Brenden M.; Salakhutdinov, Ruslan; Tenenbaum, Joshua B.. (2019). The Omniglot challenge: a 3-year progress report. https://arxiv.org/abs/1902.03477.

24. Alpaydin, E.; Alimoglu, Fevzi. (1996). Pen-Based Recognition of Handwritten Digits [Dataset]. UCI Machine Learning Repository. https://doi.org/10.24432/C5MG6K.

25. Ortega, Antonio; Frossard, Pascal; Kovačević, Jelena; Moura, José M. F.; Vandergheynst, Pierre. (2017). Graph Signal Processing: Overview, Challenges and Applications. arXiv:1712.00468. https://arxiv.org/abs/1712.00468.

26. Wang, Mei; Deng, Weihong. (2024). A dataset of oracle characters for benchmarking machine learning algorithms. Scientific Data, 11, 87. https://doi.org/10.1038/s41597-024-02933-w.

27. Williams, Ben. (2006). Character Trajectories [Dataset]. UCI Machine Learning Repository. https://doi.org/10.24432/C58G7V.

28. Koo, Terry K., and Mae Y. Li. "A Guideline of Selecting and Reporting Intraclass Correlation Coefficients for Reliability Research." *Journal of Chiropractic Medicine*, vol. 15, no. 2, 2016, pp. 155-163. https://doi.org/10.1016/j.jcm.2016.02.012.

29. Pathak, Deepak; Krahenbuhl, Philipp; Donahue, Jeff; Darrell, Trevor; Efros, Alexei A.. (2016). Context Encoders: Feature Learning by Inpainting. arXiv:1604.07379. https://arxiv.org/abs/1604.07379.

30. Lugmayr, Andreas; Danelljan, Martin; Romero, Andres; Yu, Fisher; Timofte, Radu; Van Gool, Luc. (2022). RePaint: Inpainting using Denoising Diffusion Probabilistic Models. arXiv:2201.09865. https://arxiv.org/abs/2201.09865.