Uncertainty and Assessment
Volume 1 · Number 1 · 2026

Can a Robust Neighbor Summary Help Reconstruct Manufacturing Signals? A Frozen-Graph SECOM Experiment

Eric Gardner[1] · Owen Brooks[2] · Evan Anderson[3]

[1] Netherlands · [2] Sweden · [3] Finland

Original publication: October 9, 2026 | Article 657

Abstract

A median is an appealing safeguard against an unreliable sensor neighbor, but that appeal does not establish a reconstruction advantage. We sorted SECOM manufacturing records by timestamp, learned a small correlation graph from the first 1,096 records, and withheld selected observed cells in the remaining 471. Among 2,904 cells with enough unmasked donors, signed-neighbor averaging reduced standardized absolute error by 0.116 relative to the training-mean baseline. Replacing that average with a median increased error by 0.014; its two-way clustered interval included zero. In this fixed graph and withholding design, borrowing neighboring signals helped, while the extra robustness heuristic did not clearly help. The result concerns reconstruction of observed measurements, rather than manufacturing yield or behavior under verified sensor faults.

Public-data reconstruction study with deliberately withheld measurements. No production-fault intervention or target-algorithm replication is claimed.

Can a Robust Neighbor Summary Help Reconstruct Manufacturing Signals? A Frozen-Graph SECOM Experiment

Public-data reconstruction study with deliberately withheld measurements. No production-fault intervention or target-algorithm replication is claimed.

Abstract

A median is an appealing safeguard against an unreliable sensor neighbor, but that appeal does not establish a reconstruction advantage. We sorted SECOM manufacturing records by timestamp, learned a small correlation graph from the first 1,096 records, and withheld selected observed cells in the remaining 471. Among 2,904 cells with enough unmasked donors, signed-neighbor averaging reduced standardized absolute error by 0.116 relative to the training-mean baseline. Replacing that average with a median increased error by 0.014; its two-way clustered interval included zero. In this fixed graph and withholding design, borrowing neighboring signals helped, while the extra robustness heuristic did not clearly help. The result concerns reconstruction of observed measurements, rather than manufacturing yield or behavior under verified sensor faults.

1. The choice inside a reconstruction rule

An industrial measurement system can contain many correlated channels even when their physical descriptions are unavailable. Correlation creates an opportunity to reconstruct a missing reading from other channels. It also creates a practical choice: should the reconstruction use an average that retains all donor magnitudes, or a median that limits the influence of an extreme donor? The choice is often discussed through broad claims about robustness. Here it is examined as a paired empirical comparison with a graph and evaluation set fixed before losses are calculated.

The research question is whether median aggregation improves absolute reconstruction error beyond signed-neighbor averaging on later manufacturing records. A second, logically prior question is whether either neighbor rule improves on a constant training-mean prediction. These questions distinguish the value of borrowing cross-channel information from the value of a particular summary of that information. A favorable result for one need not imply a favorable result for the other.

The graph-sign literature, including Peng et al., 2024, motivates studying reliable aggregation over related signals. We do not implement that adaptive algorithm. Our deliberately small reconstruction rule instead isolates an elementary decision that could otherwise be hidden inside a larger method. The empirical result cannot rank adaptive graph estimators, establish convergence, or identify which components of their update equations matter. It can show whether a commonly proposed robust summary helps under one transparent data condition.

2. Connections to graph learning and missing-data analysis

The perspectives in Ortega et al., 2017 and Shuman et al., 2012 explain how a graph can represent relationships among irregularly indexed measurements. In this experiment, channels are graph nodes and training correlations determine a small neighborhood. The graph is an estimated statistical representation, not a recovered factory wiring diagram. Anonymous source columns do not permit a claim that neighboring nodes correspond to physically connected equipment or that an edge identifies a causal transmission pathway.

Learned imputation systems such as Cini et al., 2021, You et al., 2020, Yoon et al., 2018, and Cao et al., 2018 address richer reconstruction problems. Chained-equation and nonparametric approaches in Buuren and Groothuis-Oudshoorn, 2011 and Stekhoven and Bühlmann, 2012 offer other ways to use multivariate information. These works establish relevant alternatives, rather than results obtained in the present experiment. We train none of those systems and do not treat the simple rules as approximations to their published performance. The advantage of the small design is that its donors, signs, exclusions, and errors can be inspected directly.

Robust-estimation arguments in Huber, 1964, Hampel, 1974, and Rousseeuw, 1984 distinguish protection against influential contamination from optimality under every ordinary distribution. A median of three signed donors can be less sensitive to one extreme value, but can discard useful magnitude information when the donors are informative. Thus the directional proposition that a median lowers reconstruction loss is testable rather than assumed. The analysis retains its result even if that proposition receives little support.

3. What the acquired manufacturing file contains

We use the public SECOM release documented by McCann and Johnston, 2008. The numeric file actually contains 1,567 rows and 590 measurement columns, while the repository description refers to 591 features. The analysis records the dimensions read from the file and does not add a fabricated column to reconcile that difference. A separate label file supplies timestamps and pass/fail information. Only the timestamps enter this reconstruction study; pass/fail labels are not an outcome or a predictor.

Rows are sorted stably by their released timestamps. The first 1,096 constitute training data and the last 471 constitute evaluation data. This split asks whether relationships estimated earlier can support reconstruction later. It avoids using evaluation values to estimate the graph, but does not guarantee stability of the manufacturing process across the boundary. The general validation concerns discussed by Roberts et al., 2017 are particularly relevant when nearby observations share operating conditions.

Channel eligibility is determined exclusively in training. A column must have at least 95 percent observed values and a nonzero finite standard deviation. Of 416 eligible columns, the first 64 in source order are selected. This cap makes the experiment bounded and reproducible; it is not a claim that these are the most important process variables. Choosing the columns by their eventual reconstruction performance would introduce a selection opportunity of the kind highlighted by Varma and Simon, 2006.

Each selected channel is centered and scaled by its own training mean and standard deviation. Test data use those frozen parameters. This places anonymous measurements on a comparable standardized scale without inventing physical units. A zero prediction in that scale equals the channel's training mean. It does not equal its median, a physically absent signal, or an engineer's desired operating value. That distinction is essential when interpreting the baseline and the magnitude of its error.

4. A graph frozen before withholding

For each selected channel, we calculate pairwise training correlations and choose the three other channels with largest absolute correlation. Ties follow source-column order. The signs of those correlations orient donor readings toward the target channel. Positively related donors enter with their original standardized sign; negatively related donors enter with the sign reversed. Neighborhood selection and sign assignment are completed without consulting test losses or withheld target values.

This graph is directed because a channel's three preferred neighbors need not select it in return. The procedure uses neither a global graph optimizer nor a learned message-passing layer. It therefore cannot

establish the advantages of graph smoothness regularization or deep graph representations. Matrix-based alternatives such as Candes and Plan, 2010 and Mazumder et al., 2010 show that low-dimensional structure can support incomplete-data estimation through different assumptions. They are relevant comparators for a future study, rather than additional methods executed here.

With random seed 20261107, every originally observed test cell has a ten-percent probability of being withheld. All selected withheld cells are removed from the donor matrix simultaneously. This matters because leaving another evaluation target visible as a donor would give the reconstruction access to information the declared missingness condition removes. The procedure creates 2,979 withheld observed cells. We retain the 2,904 for which at least two of the three frozen neighbors remain available.

The eligibility restriction is shared by both neighbor rules and the constant baseline. Consequently, their errors refer to exactly the same target cells. Cells with insufficient donors are not silently assigned favorable predictions or counted as successful reconstructions. The restriction nevertheless changes the estimand: results concern reconstructable cells under the declared neighbor-availability rule. They do not measure coverage or error for all naturally missing cells in the original file.

5. Predictions, loss differences, and uncertainty

For each retained cell, the baseline predicts zero standardized units. The mean rule takes the arithmetic mean of its available sign-adjusted donors. The median rule takes their median. With two donors the mean and the usual median coincide; with three they can differ. Neither rule is fitted to the withheld targets. Each is evaluated through the absolute difference between its prediction and the actual standardized value saved before withholding.

The first loss contrast is neighbor-mean absolute error minus baseline absolute error. The second is neighbor-median absolute error minus neighbor-mean absolute error. Intercept-only regressions estimate their respective cell-average differences. A negative coefficient favors the first rule named in the contrast. Figure 1 distinguishes graph construction, withholding, and paired evaluation so that the diagram represents the executed design rather than a generic relationship among unspecified constructs.

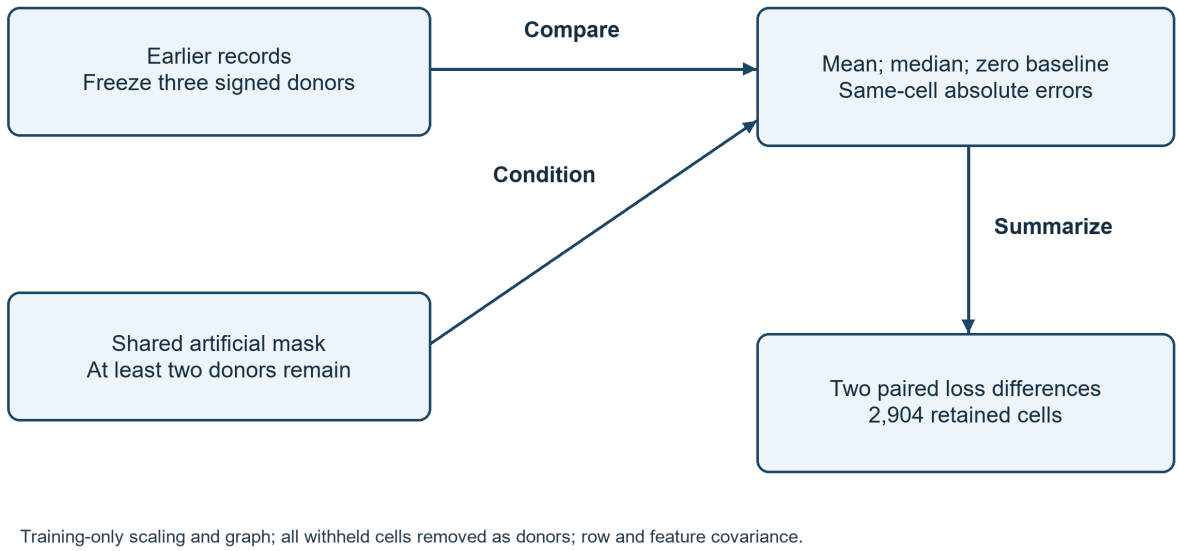

Figure 1. Training-derived donor graph and same-cell comparisons after all simulated missing cells are masked.

Cells share both manufacturing records and measured channels. We therefore use a two-way cluster covariance by record and feature, with 471 record clusters and 64 feature clusters, following the dependence concerns in Cameron and Miller, 2015. Heteroskedasticity correction alone, as introduced by White, 1980, would not represent those shared dimensions. Temporal covariance such as Newey and West, 1987 addresses a different dependence structure and is not additionally fitted here. The reported intervals remain conditional on this chosen clustering scheme.

Table 1. Average paired reconstruction-loss differences in training-standardized units.

Model	Term	Estimate (SE)	95% CI	N
M1	Mean contrast	-0.1157 (0.0330)	[-0.1816, -0.0497]	2904
M1: fit	R-squared = 0.0000	-	-	2904
M2	Mean contrast	0.0136 (0.0070)	[-0.0004, 0.0276]	2904
M2: fit	R-squared = 0.0000	-	-	2904

Model labels: M1 = Signed-neighbor mean minus training-mean MAE; M2 = Signed-neighbor median minus mean MAE.

The intercept models summarize means and have essentially zero explanatory R-squared. Their purpose is to estimate the direction, magnitude, and sampling uncertainty of a declared contrast, not to explain which cells are difficult. The covariance does not include uncertainty from repeatedly choosing new graphs, new training windows, or new withholding seeds. Only one split and one artificial mask were executed. Those limits are more informative than describing the table as a broad robustness validation.

6. Borrowing helped; taking a median did not clearly help

Mean absolute error is 3.7200 for the constant prediction, 3.6044 for signed-neighbor averaging, and 3.6180 for signed-neighbor medians. The mean-minus-baseline coefficient is -0.1157 with standard error 0.0330 and interval -0.1816 to -0.0497. Thus the frozen neighbors supply useful information for the selected later cells under this loss and mask. The improvement is modest relative to the remaining absolute error, rather than evidence of near-complete signal recovery.

The median-minus-mean coefficient is 0.0136 with standard error 0.0070 and interval -0.0004 to 0.0276. Its point estimate favors averaging, while the interval includes no difference. We therefore do not report a confirmed median penalty or a robust-median advantage. A claim that the median must improve reliability would be inconsistent with these observations. The result is narrower: its expected protection was not clearly translated into lower average error in this particular experiment.

Large standardized errors can occur even without injected extreme values because training scales, later process shifts, and unusual readings affect the denominator and target distribution. Influence diagnostics motivated by Cook, 1977 could locate the most consequential cells. They have not been used to remove inconvenient cases or revise the selected graph. A post hoc trimming analysis would answer another question and would need its own declared treatment of genuine process excursions versus measurement corruption.

Agreement measures such as Lin, 1989 do not replace the paired error contrast. High correlation between reconstructed and observed readings could coexist with systematic magnitude errors, while a small average loss gain could coexist with serious errors on rare channels. The loss interpretation issues discussed by Gneiting and Raftery, 2007 and Hyndman and Koehler, 2006 encourage stating the

evaluated quantity explicitly. Here it is absolute standardized reconstruction error, with equal weight per retained cell, rather than probability calibration, pass/fail accuracy, or cost-weighted factory utility.

7. Missingness and deployment remain open

The artificial mask is applied to values that were originally observed. Under Little and Rubin, 2002, this is not equivalent to learning the mechanism that made other source values missing. An actual missing reading may follow a shutdown, saturation, maintenance event, or data-transfer failure and therefore occupy a different part of the process distribution. Since those mechanisms are undocumented here, the experiment cannot establish performance on all operational missingness or on deliberate adversarial sensor attacks.

Measurement-error considerations in Carroll et al., 2006 also matter because the saved observed value is treated as the reconstruction target. It may itself contain noise. A lower error against that value does not demonstrate recovery of an unobserved true physical state. The anonymous columns prevent engineering verification of whether an extreme measurement is wrong, informative, or associated with a legitimate operating regime. We consequently describe the experiment as reconstruction agreement with released readings.

The table contains two planned contrasts rather than a search over many graph sizes and robust estimators. Broader exploration would introduce multiplicity questions addressed by Benjamini and Hochberg, 1995. The interpretation guidance in Wasserstein and Lazar, 2016 supports reporting the confidence interval that overlaps zero, rather than converting it into a decisive label based on a threshold. Neither a favorable mean-rule interval nor an inconclusive median-rule interval certifies the design for production deployment.

Preservation of raw files, the exact selected columns, the mask seed, processed cell losses, and fitted coefficients follows the documentation aims of Gebru et al., 2021 and Wilkinson et al., 2016. A useful extension would vary masks and time windows on a predeclared grid and include naturally missing cases with independently verified targets if such evidence becomes available. For now, the main empirical lesson is specific: information from a frozen correlation neighborhood improved this reconstruction baseline, while replacing its mean by a median supplied no clear additional gain.

The donor-count condition also explains why an apparently simple median comparison has a limited opportunity to differ. With exactly two available donors, the arithmetic mean and the usual sample median give the same prediction. Only retained targets with three donors can contribute a nonzero median-minus-mean difference. The reported coefficient averages across the full eligible set because that is the declared policy population. It should not be mistaken for an estimate conditional on having a potentially influential third donor. Reporting a different conditional comparison after seeing the result would change the question and require a separate statement of eligibility.

Finally, freezing a graph is a substantive restriction, not merely a computational convenience. Later correlations may change as equipment settings or process regimes change. An adaptive method could react to those changes, but could also react to noisy evidence. The current experiment adjudicates neither possibility. It provides a fixed reference with fully recorded training information, which a future adaptive comparison could preserve as a baseline while assessing whether adaptation improves later reconstruction under the same mask and target definitions.

References

1. Varma, Sudhir and Richard Simon. “Bias in error estimation when using cross-validation for model selection.” *BMC Bioinformatics*, vol. 7, no. 1, 2006. https://doi.org/10.1186/1471-2105-7-91.

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

3. 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.

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

5. You, Jiaxuan; Ma, Xiaobai; Ding, Daisy Yi; Kochenderfer, Mykel; Leskovec, Jure. (2020). Handling Missing Data with Graph Representation Learning. arXiv:2010.16418. https://arxiv.org/abs/2010.16418.

6. Cini, Andrea; Marisca, Ivan; Alippi, Cesare. (2021). Filling the G_ap_s: Multivariate Time Series Imputation by Graph Neural Networks. arXiv:2108.00298. https://arxiv.org/abs/2108.00298.

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

8. 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.

9. Roberts, David R., et al. “Cross-validation strategies for data with temporal, spatial, hierarchical, or phylogenetic structure.” *Ecography*, vol. 40, no. 8, 2017, pp. 913-929. https://doi.org/10.1111/ecog.02881.

10. Benjamini, Yoav and Yosef Hochberg. “Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing.” *Journal of the Royal Statistical Society Series B: Statistical Methodology*, vol. 57, no. 1, 1995, pp. 289-300. https://doi.org/10.1111/j.2517-6161.1995.tb02031.x.

11. 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.

12. Huber, Peter J.. (1964). Robust Estimation of a Location Parameter. The Annals of Mathematical Statistics. https://doi.org/10.1214/aoms/1177703732.

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

14. Hyndman, Rob J. and Anne B. Koehler. “Another look at measures of forecast accuracy.” *International Journal of Forecasting*, vol. 22, no. 4, 2006, pp. 679-688. https://doi.org/10.1016/j.ijforecast.2006.03.001.

15. 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.

16. Hampel, Frank R.. (1974). The Influence Curve and its Role in Robust Estimation. Journal of the American Statistical Association. https://doi.org/10.1080/01621459.1974.10482962.

17. Buuren, Stef van; Groothuis-Oudshoorn, Karin. (2011). mice: Multivariate Imputation by Chained Equations in R. Journal of Statistical Software. https://doi.org/10.18637/jss.v045.i03.

18. 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.

19. McCann, Michael; Johnston, Adrian. (2008). SECOM [Dataset]. UCI Machine Learning Repository. https://doi.org/10.24432/C54305.

20. Mazumder, Rahul; Hastie, Trevor; Tibshirani, Robert. (2010). Spectral Regularization Algorithms for Learning Large Incomplete Matrices. Journal of Machine Learning Research 11, 2287-2322. https://jmlr.org/papers/v11/mazumder10a.html.

21. Candes, Emmanuel J; Plan, Yaniv. (2010). Matrix Completion With Noise. Proceedings of the IEEE. https://doi.org/10.1109/jproc.2009.2035722.

22. 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.

23. Newey, Whitney K., and Kenneth D. West. “A Simple, Positive Semi-Definite, Heteroskedasticity and Autocorrelation Consistent Covariance Matrix.” *Econometrica*, vol. 55, no. 3, 1987, pp. 703-708. https://doi.org/10.2307/1913610.

24. Peng, Changran; Yan, Yi; Kuruoglu, Ercan E.. (2024). Adaptive Graph Normalized Sign Algorithm. arXiv:2405.04107. https://arxiv.org/abs/2405.04107.

25. Gneiting, Tilmann; Raftery, Adrian E. (2007). Strictly Proper Scoring Rules, Prediction, and Estimation. Journal of the American Statistical Association. https://doi.org/10.1198/016214506000001437.

26. Yoon, Jinsung; Jordon, James; van der Schaar, Mihaela. (2018). GAIN: Missing Data Imputation using Generative Adversarial Nets. arXiv:1806.02920. https://arxiv.org/abs/1806.02920.

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

28. Cao, Wei; Wang, Dong; Li, Jian; Zhou, Hao; Li, Lei; Li, Yitan. (2018). BRITS: Bidirectional Recurrent Imputation for Time Series. arXiv:1805.10572. https://arxiv.org/abs/1805.10572.

29. Rousseeuw, Peter J.. (1984). Least Median of Squares Regression. Journal of the American Statistical Association. https://doi.org/10.1080/01621459.1984.10477105.

30. Stekhoven, Daniel J.; Bühlmann, Peter. (2012). MissForest-non-parametric missing value imputation for mixed-type data. Bioinformatics. https://doi.org/10.1093/bioinformatics/btr597.