Resource Allocation and Optimization
Volume 1 · Number 1 · 2026

Smaller Marginal Gains from Additional Wind Readings in a Held-Year Forecast

Andrew Franklin[1] · Arthur Allen[2] · Benjamin Lewis[3]

[1] France · [2] Australia · [3] Netherlands

Original publication: October 31, 2026 | Article 626

Abstract

One neighboring wind reading supplied most of the observed gain from using three. After selecting subsets within each input budget using 2014 validation data, we refitted the chosen models and tested 51,384 forecast origins in 2015. The first, second, and third neighboring inputs reduced equal-date mean absolute error by 0.006946, 0.002083, and 0.001216 m/s respectively. The last pointwise interval ended close to zero. Development validation had instead favored the own-only model across budgets, so the later ordering cannot be presented as a prospectively validated budget choice. The experiment measures a small, source-specific information tradeoff; it does not price sensor acquisition or establish an adaptive graph algorithm's performance.

Development-only subset selection followed by a chronological comparison of four neighboring-input budgets.

Smaller Marginal Gains from Additional Wind Readings in a Held-Year Forecast

Development-only subset selection followed by a chronological comparison of four neighboring-input budgets.

Abstract

One neighboring wind reading supplied most of the observed gain from using three. After selecting subsets within each input budget using 2014 validation data, we refitted the chosen models and tested 51,384 forecast origins in 2015. The first, second, and third neighboring inputs reduced equal-date mean absolute error by 0.006946, 0.002083, and 0.001216 m/s respectively. The last pointwise interval ended close to zero. Development validation had instead favored the own-only model across budgets, so the later ordering cannot be presented as a prospectively validated budget choice. The experiment measures a small, source-specific information tradeoff; it does not price sensor acquisition or establish an adaptive graph algorithm's performance.

1. An information budget needs a sequence of decisions

A forecast using several neighboring measurements can improve on a local-only forecast while leaving the value of each added input unclear. A small model with three available neighbors permits a direct comparison of every subset. That comparison still needs a temporal design: choosing a convenient subset after examining the final test year would overstate the evidence for an input budget. Selection, refitting, and final evaluation must therefore have explicit roles.

We ask how the marginal held-year error reduction changes as a linear wind-speed forecast uses zero, one, two, or three neighboring readings. Within each fixed budget, a subset is chosen using only a development validation period. The directional proposition predicts negative later-period loss increments as inputs are added. A separate descriptive question concerns whether the sizes of those increments decline. Ordered point estimates alone do not establish a general law of diminishing information returns.

Wind prediction under the robustness concerns represented by Yan and Kuruoglu, 2023 supplies the application context. Our predictors are ordinary linear regressions with selected contemporary readings. We do not implement adaptive graph-sign diffusion, inject impulsive noise, or compare the target algorithm with new baselines. The experiment instead asks a narrower resource question: what does another observed neighboring wind-speed input buy under this declared source, selection procedure, and later evaluation period?

2. The physical resource and the observed information set

The source is the La Haute Borne SCADA table captured in ENGIE / NREL OpenOA contributors, 2026 within the example material associated with Perr-Sauer et al., 2021. It contains 420,480 wind-speed records for four turbines during 2014 and 2015. Source units are meters per second. We identify ninety-six rows belonging to forty-eight conflicting turbine-time keys and exclude all of those rows. The ambiguous keys are preserved separately rather than resolving simultaneous values by an undocumented row-order choice.

The prediction target is R80711 wind speed ten minutes ahead. Its current reading is available under every budget. The three additional candidates are R80721, R80736, and R80790. A usable origin requires all four contemporary readings, an observed future target, and an exact ten-minute timestamp separation.

Sorting two observations next to one another does not suffice when a gap interrupts the time series. No missing reading is interpolated into this common comparison sample.

The operational perspective in Pinson, 2013 and spatial-resource context in Haslett and Raftery, 1989 motivate examining neighboring information, but fitted weights are not interpreted as causal atmospheric transport. Network forecasting approaches in Li et al., 2017 and Yu et al., 2017 involve additional graph and learning assumptions absent here. Four contemporary predictors do not by themselves establish a learned sensor graph. Likewise a lower absolute wind-speed error is not an electrical-power gain, because no turbine power curve, market consequence, or generation decision is evaluated.

3. Development selection precedes the final year

The development fitting period retains origins whose targets occur before 1 October 2014, giving 39,211 rows. Validation uses origins from October through December with targets still before 2015, giving 13,109 rows over ninety-two dates. The exact target boundary excludes a cross-phase next-interval record from that initial fit/validation division. Once selection is finished, the final model for each budget is refitted on all 52,321 rows whose targets precede 2015.

The final holdout consists of 51,384 origins in 2015. None of their targets enters subset selection or refitting. We do not pick the overall neighbor budget from these later losses. Instead, every predeclared budget is evaluated so that its observed tradeoff can be described. This distinction matters because choosing a single deployed budget and comparing a sequence of available budgets are different activities. The current procedure performs the latter after selecting only within each cardinality.

Temporal-validation considerations in Bergmeir et al., 2018 and Roberts et al., 2017 support keeping this chronology. Model-selection bias discussed in Varma and Simon, 2006 would arise if final-year outcomes were used to choose a preferred subset and then reported as an untouched test of that choice. The design avoids that particular leakage. It still conditions on one development split and one later year, so it does not quantify the variability of subset selection under other validation windows or training periods.

4. Enumerating the small subset space

There are eight subsets of three neighboring candidates. The zero-neighbor budget has one own-only model, the one-neighbor budget has three candidates, the two-neighbor budget has three, and the three-neighbor budget has one. Every forecast contains an intercept and current own speed, plus its selected neighboring terms. Ordinary least squares estimates all development fits, and design matrices are checked for full column rank. No hour, weather, direction, or lagged-error terms are added.

The subset-search traditions in Furnival and Wilson, 1974 and Bertsimas et al., 2016 provide context for comparing available feature combinations. Our space is small enough to enumerate directly; we run neither branch-and-bound nor a mixed-integer optimizer. Validation mean absolute error is averaged first within each date, then equally across dates. The subset with the lowest value within a budget wins, with lexicographic identity ordering available for an exact tie. No candidate is ranked using the later holdout year.

A ridge alternative motivated by Hoerl and Kennard, 1970 could address coefficient stability under correlated readings, but no regularized forecast is fitted. The present result concerns ordinary least squares after cardinality-constrained selection. Retaining one fitting method keeps that choice separate from the information budget. It does not establish that adding neighbors would have the same marginal value with a different model class, a different penalty, or a different temporal validation design.

5. The chosen subsets and their development evidence

The selected one-neighbor model uses R80721. The two-neighbor model uses R80721 and R80790. The final budget includes all three. These choices happen to form a nested sequence in this realization, though the selection rule does not require that property in general. After they are selected, each is estimated again using the full 2014 training set. Refitting changes coefficients, so later performance is not the same fitted-model comparison as the development table.

Table 1. Best validation subset within each neighbor-count budget; values are equal-date validation MAE in m/s.

Neighbor count	Selected additional readings	Development validation MAE
0	None	0.396566
1	R80721	0.399683
2	R80721, R80790	0.401755
3	R80721, R80736, R80790	0.407803

Validation does not favor adding inputs across budgets. If the same criterion had been used to select the overall count, own-only would have won in that development period. We retain this fact even though the later test ordering reverses. The reversal combines a new evaluation period with coefficient refitting on additional 2014 data. It cannot isolate a causal effect of calendar time or establish that the validation procedure correctly anticipated a future gain from purchasing more information.

6. A budget measures input count, with costs left unpriced

Here one additional unit of budget means one neighboring reading at a forecast origin. The own reading is present under every count. We do not measure money, device installation, network traffic, processor time, or delay. Decision-cost reasoning in Elkan, 2001 clarifies why an error gain alone cannot establish that a resource is worth its actual cost. Selective-decision ideas in Geifman and El-Yaniv, 2017 would also require an explicit application objective before translating a forecast comparison into an operational policy.

All budgets are assessed on the same complete four-reading rows. Consequently the own-only policy is not credited with additional dates on which some neighboring observations are missing. This restriction keeps the forecast outcomes comparable, while omitting a potentially valuable availability advantage of a smaller information set. Missing-data principles in Little and Rubin, 2002 make that restriction substantive. We cannot infer which budget would be best under real sensor outages from a sample requiring every candidate sensor to be present.

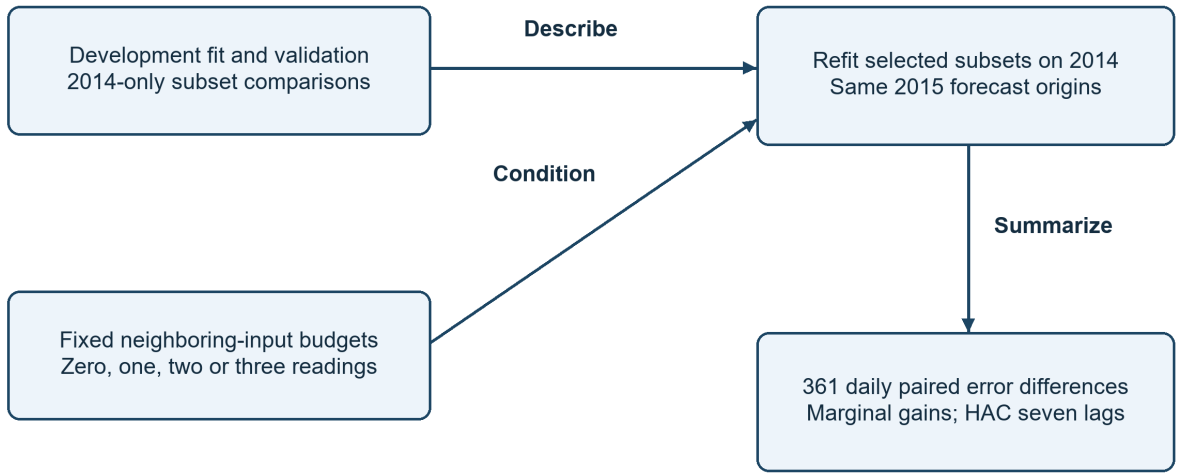


Figure 1. Development-only subset choices are refitted before identical 2015 targets are predicted; daily losses then describe marginal neighbor-count contrasts.

Deployment issues in Sculley et al., 2015 are relevant because the calculation assumes each current reading arrives in time for a ten-minute forecast. Actual transmission or acquisition latency is unknown. Calibration concepts in Guo et al., 2017 address another unmeasured issue: the regressions provide point predictions without a validated uncertainty distribution. The resource diagram therefore ends at observed absolute loss rather than a monetary payoff, a calibrated safety decision, or a measured communication advantage.

7. Daily paired losses and their uncertainty

For every held origin we calculate the absolute difference between prediction and observed target speed. Errors are averaged within each UTC date, including partial dates, leaving 361 available dates. Each date receives equal weight. Accuracy-measure reasoning in Hyndman and Koehler, 2006 helps explain why this target differs from assigning every ten-minute origin equal weight. We do not insert zero loss for dates without eligible observations or estimate the unobserved forecast error on those dates.

The one-minus-zero, two-minus-one, and three-minus-two neighboring-input losses are computed within each date. Separate intercept-only ordinary least-squares fits summarize their means. We use HAC covariance with seven retained-date lags, a finite-sample correction, and a t reference with 360 degrees of freedom, following Newey and West, 1987. Predictive-comparison ideas in Diebold and Mariano, 1995 and West, 1996 motivate keeping the losses paired on the same targets rather than comparing scores from differently available evaluation samples.

Table 2. Held-year marginal loss differences in m/s; negative coefficients favor the larger selected input set.

Model	Term	Estimate (SE)	95% CI	N
M1	Mean outcome/contrast	-0.006946 (0.001128)	[-0.009164, -0.004728]	361
M1: fit	R-squared = 0.0000	-	-	361
M2	Mean outcome/contrast	-0.002083 (0.000548)	[-0.003161, -0.001006]	361
M2: fit	R-squared = 0.0000	-	-	361
M3	Mean outcome/contrast	-0.001216 (0.000606)	[-0.002408, -0.000025]	361
M3: fit	R-squared = 0.0000	-	-	361

Model labels: M1 = 1 minus 0 neighboring inputs; M2 = 2 minus 1 neighboring inputs; M3 = 3 minus 2 neighboring inputs.

The heteroskedasticity-consistent framework in White, 1980 alone would not capture serial dependence among neighboring dates. The dependence discussion in Cameron and Miller, 2015 also cautions that one farm's 361 dates do not amount to independent farm replications. The mean-only fits have zero R-squared, which is unrelated to the explained variance of the original wind regressions. Their intervals condition on the chosen subsets and refitted coefficients, excluding repeated-selection and retraining uncertainty.

8. The first selected neighbor contributes most of the observed gain

The held-year equal-date MAEs are 0.433715403, 0.426769229, 0.424685998, and 0.423469520 m/s as neighbor count rises from zero to three. Thus the first additional reading reduces the mean by 0.006946174 m/s. Its standard error is 0.001127982 and interval -0.009164434 to -0.004727913 for the larger-minus-smaller contrast. This is the largest of the three marginal point reductions, while remaining small in the original physical unit.

The second-neighbor contrast is -0.002083231, with standard error 0.000547891 and interval -0.003160700 to -0.001005763. The third is -0.001216478, with standard error 0.000606027 and interval -0.002408277 to -0.000024679. The third pointwise interval only narrowly stays below zero. We do not interpret that boundary as a strong general guarantee of value from a third reading, particularly when the whole comparison conditions on a single selected subset sequence.

The first step supplies roughly 67.8 percent of the total three-neighbor reduction of 0.010245883 m/s. The marginal point magnitudes are ordered from largest to smallest. We have not fitted a separate test of differences between those marginal effects or a general response curve in sensor count. Their observed ordering supports a descriptive diminishing-gain pattern here, rather than an identified law about how information must accumulate in other wind resources.

9. Validation and test orderings answer different conditional questions

The final test shows gains after full-year refitting, whereas the development validation table shows larger errors as neighbors are added. This is a useful warning against presenting the final curve as a validated purchasing rule. New weather conditions, changed coefficient estimates, and availability selection can all contribute to the difference. None is experimentally isolated. The study reports the complete sequence instead of hiding the earlier unfavorable ordering or claiming that a development-only chooser would have selected the eventual best count.

Measurement concerns in Carroll et al., 2006 also apply to the released wind readings: forecast agreement does not independently certify anemometer accuracy. Influence analysis associated with Cook, 1977 could identify dates driving the small marginal differences, but no high-loss day is removed. Multiplicity principles in Benjamini and Hochberg, 1995 limit a broad inference from several pointwise comparisons. We provide the three declared effects and their intervals rather than searching additional subsets in the final year until a preferred resource curve appears.

Interpretation guided by Wasserstein and Lazar, 2016 keeps physical magnitude and decision scope visible. A statistically distinguishable reduction measured in thousandths of a meter per second may or may not matter to an operation. Its practical value depends on costs and consequences not collected here. Replication on other development windows, farms, or forecast horizons would add evidence about that question. It is not supplied by evaluating many origins from the same four-turbine source.

10. Report the selection history with the resource curve

The completed experiment finds smaller held-year point gains from successive neighboring readings after within-budget subset selection and refitting. The first selected neighbor contributes most of the observed reduction. At the same time, the development criterion would not have selected a larger overall budget, and the final marginal gain remains small. These facts support a careful conditional resource description, without a claim that a deployment should acquire all available readings.

Documentation under Gebru et al., 2021 includes ambiguous-key removal, exact horizon matching, shared availability, target-based date boundaries, every development candidate, and the equal-date loss target. Preservation following Wilkinson et al., 2016 retains these materials with held-year predictions and daily contrasts. They allow another researcher to reconstruct both the selected sequence and the earlier validation evidence. A defensible operational budget still requires independent temporal evidence and actual acquisition costs beyond the numerical comparison reported here.

References

1. Elkan, Charles. “The Foundations of Cost-Sensitive Learning.” *Proceedings of the Seventeenth International Joint Conference on Artificial Intelligence*, 2001, pp. 973-978. https://cseweb.ucsd.edu/~elkan/rescale.pdf.
2. Yan, Yi; Kuruoglu, Ercan E.. (2023). Fast and Robust Wind Speed Prediction Under Impulsive Noise via Adaptive Graph-Sign Diffusion. 2023 IEEE Conference on Artificial Intelligence (CAI), 302-305. https://doi.org/10.1109/CAI54212.2023.00135.
3. 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.
4. 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.
5. Furnival, George M.; Wilson, Robert W.. (1974). Regressions by Leaps and Bounds. Technometrics. https://doi.org/10.1080/00401706.1974.10489231.
6. Gebru, Timnit, et al. "Datasheets for datasets" *Communications of the ACM*, 2021. https://doi.org/10.1145/3458723.
7. Diebold, Francis X.; Mariano, Roberto S.. (1995). Comparing Predictive Accuracy. Journal of Business & Economic Statistics. https://doi.org/10.1080/07350015.1995.10524599.

8. 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.

9. Yu, Bing; Yin, Haoteng; Zhu, Zhanxing. (2017). Spatio-Temporal Graph Convolutional Networks: A Deep Learning Framework for Traffic Forecasting. arXiv:1709.04875. https://arxiv.org/abs/1709.04875.

10. Hoerl, Arthur E.; Kennard, Robert W.. (1970). Ridge Regression: Biased Estimation for Nonorthogonal Problems. Technometrics. https://doi.org/10.1080/00401706.1970.10488634.

11. Bertsimas, Dimitris; King, Angela; Mazumder, Rahul. (2016). Best subset selection via a modern optimization lens. The Annals of Statistics. https://doi.org/10.1214/15-aos1388.

12. Haslett, John; Raftery, Adrian E.. (1989). Space-Time Modelling with Long-Memory Dependence: Assessing Ireland's Wind Power Resource. Applied Statistics. https://doi.org/10.2307/2347679.

13. Geifman, Yonatan, and Ran El-Yaniv. "Selective Classification for Deep Neural Networks." *Advances in Neural Information Processing Systems*, vol. 30, 2017. https://arxiv.org/abs/1705.08500.

14. 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.

15. 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.

16. 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.

17. West, Kenneth D.. (1996). Asymptotic Inference about Predictive Ability. Econometrica. https://doi.org/10.2307/2171956.

18. Perr-Sauer, Jordan; Optis, Mike; Fields, Jason; Bodini, Nicola; Lee, Joseph; Todd, Austin; Simley, Eric; Hammond, Robert; Phillips, Caleb; Lunacek, Monte; Kemper, Travis; Williams, Lindy; Craig, Anna; Agarwal, Nathan; Sheng, Shawn; Meissner, John. (2021). OpenOA: An Open-Source Codebase For Operational Analysis of Wind Farms. Journal of Open Source Software. https://doi.org/10.21105/joss.02171.

19. 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.

20. 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.

21. Sculley, D., et al. "Hidden Technical Debt in Machine Learning Systems." *Advances in Neural Information Processing Systems*, vol. 28, 2015. https://papers.nips.cc/paper/2015/hash/86df7dcfd896fcaf2674f757a2463eba-Abstract.html

22. ENGIE / NREL OpenOA contributors, . (2026). La Haute Borne 2014-2015 SCADA and variable dictionary: captured OpenOA example archive. Local snapshot captured 7 November 2026; observed data years 2014-2015. https://github.com/NREL/OpenOA/blob/main/examples/data/la_haute_borne.zip.

23. Guo, Chuan, Geoff Pleiss, Yu Sun, and Kilian Q. Weinberger. "On Calibration of Modern Neural Networks." *Proceedings of the 34th International Conference on Machine Learning*, PMLR, vol. 70, 2017, pp. 1321-1330. https://proceedings.mlr.press/v70/guo17a.html.

24. 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.

25. Pinson, Pierre. (2013). Wind Energy: Forecasting Challenges for Its Operational Management. Statistical Science. https://doi.org/10.1214/13-sts445.

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

27. Bergmeir, Christoph; Hyndman, Rob J.; Koo, Bonsoo. (2018). A note on the validity of cross-validation for evaluating autoregressive time series prediction. Computational Statistics & Data Analysis. https://doi.org/10.1016/j.csda.2017.11.003.

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

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

30. Li, Yaguang; Yu, Rose; Shahabi, Cyrus; Liu, Yan. (2017). Diffusion Convolutional Recurrent Neural Network: Data-Driven Traffic Forecasting. arXiv:1707.01926. https://arxiv.org/abs/1707.01926.