Supplementary Information

From sensor rankings to restoration decisions in traffic forecasting

Bailin Xiong*, Kai Du†, Jinhui Tang†, Renxiang Wang†

Chongqing Jiaotong University, No. 66 Xuefu Road, Nan'an District, Chongqing, 400074, China.

*Correspondence: Bailin Xiong. E-mail: 622250991028@mails.cqjtu.edu.cn.

†Kai Du, Jinhui Tang, and Renxiang Wang contributed equally and share second authorship.

Co-author e-mails: Kai Du: 622250991054@mails.cqjtu.edu.cn; Jinhui Tang: 622250991019@mails.cqjtu.edu.cn; Renxiang Wang: 622250991037@mails.cqjtu.edu.cn.

This Supplementary Information contains 51 tables and ten figures. Sections S1–S18 document the original GWN retrospective studies; S19 reports GWN PEMS03 confirmation and rolling PEMS08 evaluation; S20 reports their DCRNN transfer; S21 covers the initial METR-LA DCRNN transfer; S22 retains detailed original diagnostic tables; S23 reports additional METR-LA DCRNN seeds and LSTM. S24 supplies original absolute restoration outcomes and the exact METR42 gain and interval provenance; S25 reports post hoc correlation-based mRMR-style selection; S26 audits PEMS03 training records, performance differences and reversal strata; S27 reports the added blockwise policy-label exchange. Supplementary Data 1 contains all 180 correlation-policy mask outcomes. Supplementary Data 2 contains all fault-anchor, mask, week and returned-count strata. The unified Zenodo archive, version 1.7.1 (https://doi.org/10.5281/zenodo.23614390), provides the code, source sets, outputs and audit records; its README and manifests map these files to the analyses.

S1. Intervention and scoring details

Table S1 defines the original ten inexpensive ranking controls. The reference deletion intervention changes both standardized values and availability masks. A value-gradient direction changes only values while holding availability and the position of the last valid history fixed. The gradient controls and complete-history deletion therefore evaluate distinct interventions.

Table S1. Original inexpensive ranking controls and their information requirements.

Score	Definition	Information used
Weighted outdegree	Source-column sum after prior self-loop removal and receiver-row normalization	Prior graph
Betweenness	Unweighted betweenness on directed source-to-receiver support	Prior graph
PageRank	Weighted directed PageRank, damping 0.85	Prior graph
Core number	Core number on undirected prior support	Prior graph
Fit variance	Population variance over valid parameter-fitting readings	Fitting readings and validity
Fit missing rate	One minus valid fraction in parameter-fitting period	Fitting validity
Fit mean strength	Source-column sum multiplied by mean valid fitting reading	Prior graph and fitting readings
Persistence damage	Error change when deletion makes own-source persistence fall back to fitting mean	History and future targets, offline
Signed Linear value gradient	Sum of minus input value times absolute-loss gradient in the value channel	Fixed proxy, history and future targets, offline
Magnitude Linear value gradient	Sum of absolute values of the preceding directional terms	Fixed proxy, history and future targets, offline

Auxiliary GWN value-gradient, random and constant scores are retained in complete figures/tables. Their presence does not change the finite ten-score family used for the proxy-versus-best-control contrast. A constant score produces an undefined Spearman correlation. This occurs for the PEMS08 fitting missing-rate score; no zero is substituted. All eligible signed effects are retained. Positive-loss coverage and top-set overlap describe rankings, not the joint recovery outcome.

For secondary partial correlations, score and reference columns are ranked first. Static control columns are rank-transformed, constant directions are removed, and an intercept is included in an SVD-based projection onto the control column space. Correlation is then calculated between residual ranks. The fitted projection design remains fixed at its point-estimate value throughout date resampling. Near-constant residuals give undefined correlations. The one-control and seven-control adjustments quantify descriptive residual-rank associations.

S2. Statistical units, complete source tables and reference sensitivity

The original target date is the encoded date of the first future target in a window. Five mask numerators are pooled within a date. Repeated dates in bootstrap draws repeat both the error numerator and its valid-target denominator; policies within a contrast share draw indices. Percentile intervals use 2,000 draws. Circular two-date blocks preserve two adjacent encoded dates, including circular wrap-around, as a

sensitivity to short temporal dependence. Longer dependence scales are assessed separately where the temporal coverage permits.

The manuscript retains the original METR Linear-versus-variance interval [-0.000517, 0.002368] mph and separately reports the later reference-sensitivity reaggregation [-0.000495356967, 0.002407313127]. Both have the same point gain to saved numerical precision but differ in the date-draw seed. The latter was generated jointly with the reference-43 contrast. It is reported as a separate reaggregation of the same observed data.

Table S2. Primary contrasts retained in the main manuscript. Positive gain is lower error under the candidate. Intervals are conditional paired-date percentile intervals, not training-seed confidence intervals.

Dataset / comparison	Gain	95% date interval	Unit	Interpretation
METR42 Linear vs variance, original draws	0.000868415	[-0.000517, 0.002368]	mph	Increment uncertain
METR42 Linear vs variance, common-date sensitivity	0.000868415070	[-0.000495356967, 0.002407313127]	mph	Same point, different draws
METR43 Linear vs variance	0.001098920220	[-0.000218330264, 0.002437488888]	mph	Increment uncertain
METR reference-43 minus reference-42 gain	0.000230505151	[-0.000635280755, 0.000902035258]	mph	Change in contrast uncertain
METR43 Linear vs Bounded	-0.001187389662	[-0.002296219535, 0.000180682669]	mph	No established Linear advantage
PEMS08 reference 42 Linear vs variance	0.283076669	[0.256921714, 0.310155305]	vehicles/5 min	Positive conditional contrast
METR HJ vs original best family	0.002560616848	[0.000571064705, 0.005033009318]	mph	Below original 0.003 point criterion

Complete absolute errors, paired contrasts, random-policy instances, source scores and denominator definitions are supplied with the reproducibility bundle. The technical record index maps each scientific table and figure to these data without adding independent observations or changing the original estimates.

S3. PEMS08 grouped-source refitting and original hypothesis outcomes

The PEMS08 grouped-source refitting analysis uses 2,040 ridge heads, 4,085 design columns and 170 grouped deletions across five scoring masks. The model is refit after deleting a source's design columns; this is distinct from deleting a history from a fixed deep network. Persistent Bernoulli availability with seed 2026101403 defines training, whereas scoring uses the original exact-51-source masks. A normalized ridge objective is used. Alpha is selected from 0.01, 0.1 and 1 by nominal calibration MAE, with ties favoring larger alpha; the selected value is 0.01.

Scoring uses the 1,774-window validation cohort and compares its rankings against the later 3,561-window GWN reference. The inexpensive dynamic scores and Linear proxy are recomputed on that scoring cohort. Consequently, the resulting Linear correlation of 0.754437 is not the same quantity as the historical-cohort correlation 0.841022. The nominal errors refer to different cohorts and are reported with their respective evaluation periods. Seven circular-block refitting draws have zero scoring-cohort denominator and remain undefined, leaving 1,993 valid draws rather than substituting zeros. Three source groups, three prediction heads and 32 validation inputs were also checked by direct column removal and re-solving against saved fast-refit predictions; differences were below 0.0000001 standardized units. This checks numerical agreement with direct refitting.

Table S3. Original PEMS08 pattern criteria and their outcomes. “Met” refers only to the retained single-reference branch. The criteria and their timing are retained from the original replication plan.

Criterion	Original expectation	Observed outcome	Status
R1	Four topology correlations have magnitude at most 0.15	All magnitudes below 0.087	Met
R2	Variance correlation at least 0.50	0.207131	Not met
R3	Linear correlation at least 0.65 and above best finite inexpensive control	0.841022; margin 0.609156	Met
R4	Agreement across two reference initializations	Only one PEMS reference in this branch	Not evaluable
R5	Linear Other correlation below Own	Ordering retained	Met
R6	Refitted Global below best inexpensive control and Own above Global	Global 0.458953 versus control 0.316094; Own 0.729230	Conjunction not met
R7	Both Linear and variance outperform random-policy mean in primary condition	Linear does; variance worse by 0.027666169	Conjunction not met

S4. Joint greedy selection and original continuation rule

The HJ primary Global gain compares its MAE against the minimum MAE of Linear ranking, fit variance and Bounded ranking. The best member is reselected within each paired-date draw. This is distinct from always fixing Bounded as comparator, although Bounded attains the best primary point estimate. Twenty random policies, the auxiliary gradient and coverage controls remain in the archived primary table but do not replace the original three-policy family for its stopping decision.

The criterion combined a minimum point gain of 0.003 mph and a strictly positive date-interval lower endpoint. HJ satisfied the latter but not the former. The subsequent conditions, including positive gains in at least three fault/budget cells and a positive reference-43 estimate, were conditional on passing the first step and were not executed. They are not treated as failures with zero effect. Supplementary Figure S1 retains the original gain, threshold and descriptive restoration-effort comparison; Figure S2 retains the broader METR joint-restoration matrix. The HJ selection procedure uses earlier targets offline and does not adapt separately to each later case.

S5. Related archived experiments outside the principal estimand

Earlier risk-objective, interaction-student and query experiments involved different models, fitting procedures, windows and decisions. Their existence is disclosed here to prevent a misleading account of uniformly successful development; they are not pooled into the current reference count or interpreted as independent repetitions of its policy.

In the earlier METR historical-holdout risk-objective comparison, the six-condition Global means were 3.956438 mph for bounded worst and 3.880295 for its matched bounded empirical-risk control; all three model seeds lacked an improvement. A separate validation illustration selected after inspection showed a slightly smaller ratio-defined relative degradation alongside larger absolute error in one random-group condition. The illustration describes that selected condition; the reported policy comparisons use absolute error, with the training objective identified separately.

The C1 pilot trained one student per type and evaluated two saved GWN references over six exposed windows and three dates. Its primary Global gains against the single student were -0.000253229 and -0.002004694 mph. Local or static-control positives do not replace those primary results. A later one-shot family-regret query diagnosis had a small positive primary gain, but 18 of 20 scenarios produced identical inputs and removing one window reversed the gain. It was neither a closed-loop deployment study nor a set of 20 independent observations. These branches support caution in interpreting intermediate labels and development positives; their small, incompatible designs do not constitute a systematic benchmark of every method family.

Later codec allocation, low-bit Moment decoding and M2 experiments address different interventions and actual-byte rules. They are documented separately from the fixed source-count restoration experiments reported here. The original PEMS08 branch retains its single-reference role; the later rolling GWN and DCRNN evaluations are specified in S19 and S20.

S6. Data and code organization

The unified computational archive supplies numerical inputs to the tables and figures, source code, fitted checkpoints, policy sets and reproduction commands. Its README and manifests map analyses to package-relative files. Raw data access is given in the main Data availability statement: METR-LA through DCRNN, PEMS08 through ASTGNN, and the exact PEMS03 NPZ/CSV/TXT files at https://zenodo.org/records/7816008. All three PEMS03 checksums match the study inputs. The arithmetic reaggregation and same-environment inference replays are documented separately in S16 and S19–S23; detailed provenance remains in the archive.

S7. Exact dates, denominators and additional descriptive checks

Table S4. Primary-condition denominator audit. Each occurrence is one input mask/window pair; windows overlap and masks share the same fitted predictor. All policies share these denominators.

Dataset	Global valid targets	Failed-source valid targets	Tail window occurrences	Encoded target dates
METR-LA	35,845,775	3,622,771	15,210	2012-06-16 through 2012-06-27
PEMS08	20,451,000	2,045,100	10,025	2016-08-25 through 2016-08-31

PEMS08 timeline upper boundaries are fit 11,249, calibration 12,499, validation 14,284, total 17,856, all exclusive. Window starts follow the archived DevelopmentData construction: fitting starts [0,fit_end-23), calibration [fit_end-12,train_end-23), validation [train_end-12,val_end-23); historical reference uses all 12 targets in [val_end,total). The later PEMS restoration period begins at first-target date 2016-08-25. Exact per-date denominators and effects are in DATE_WISE_PRIMARY_GAINS.csv, with epoch-day integers so no unverified timezone conversion is implied.

Additional saved-statistics summaries describe variation across evaluation dates. Linear-versus-variance daily Global gains are positive on seven of 12 METR42 dates, eight of 12 METR43 dates and all seven PEMS08 dates. Full date-level numerators are supplied in METR_PRIMARY_DATE_INPUTS.npz and PEMS_PRIMARY_DATE_INPUTS.npz.

The original circular-date algorithm draws ceil(D/2) start indices uniformly from [0,D), expands each into that date and its circular successor, concatenates and truncates to D sampled positions. The circular algorithm joins the end and start of the observed date sequence. The refitting-ranking calculation uses the

sorted union of scoring and reference cohort dates; the same union-date multiplicities are restricted separately to each cohort's dates before recomputing its loss scores. A draw can put zero multiplicity on every scoring-cohort date. Undefined statistics are retained as nonfinite, and intervals use finite values for that statistic. The original refitting circular calculation has 1,993 valid draws. This shared-union-calendar calculation reweights the two observed periods and is reported as a conditional sensitivity.

Table S5. Post hoc descriptive per-case restoration. Cases are window/mask pairs. Overlap refers to the policy set intersecting the fault set before accounting for native/base missingness. The complete summary includes all available listed ranking policies in PER_CASE_RESTORATION_SUMMARY.json.

Dataset / policy	Cases	Restored reads: min / median / max	Cases with zero restored reads	Median policy-fault source overlap
METR Linear	15,210	0 / 12 / 96	6,046	1
METR variance	15,210	0 / 12 / 120	6,676	1
METR Bounded	15,210	0 / 12 / 96	5,753	2
PEMS Linear	10,025	0 / 12 / 72	3,505	1
PEMS variance	10,025	0 / 12 / 60	3,534	2
PEMS GWN gradient	10,025	0 / 12 / 72	3,575	2

Only saved restored_read_counts, policy_fault_intersection and METR restored_active_sources fields were read for these distributions. No raw traffic or sensor-error tensor was decoded for this descriptive addition. The summaries compare realized returned readings under equal eligible-source counts. Zero restored reads indicates that the policy changes no available historical input in that case.

S8. Analysis timing, exposure and implementation details

The original METR-LA and PEMS08 studies were developed and examined before the present manuscript was assembled. Their datewise summaries, restoration-count checks and additional controls are retrospective analyses of exposed data. The HJ continuation criterion and its unmet point threshold remain in S4. Repeated masks, saved checkpoints and continuation of one fitting do not create additional unused datasets or reference architectures. PEMS03 confirmation and rolling PEMS08 evaluation were subsequently specified as separate experiments (S19).

For the original evaluation, the graph-fault anchor is drawn from NumPy default_rng(fault_seed + window_start*1009 + mask_seed*1000003), with fault_seed 20261103 for evaluation and 20261211 for HJ selection. The ordered graph traversal and tail calculation use the fixed shared rules. Sources with no valid future targets are excluded from the within-window tail ranking; the highest ceil(5% of remaining sources) source MAEs are averaged before pooling window–mask occurrences. Completely target-empty historical windows are excluded before constructing this cohort. These definitions are unchanged by the editorial revision.

S9. Early-ranking correspondence and realized restoration

These post hoc analyses use the exact frozen selection/evaluation starts and exclude the 11 crossing-target windows. Each reconstructed Linear, variance and Bounded eligibility set matched the archived policy at both budgets. Complete source vectors, temporal metrics and selected-set comparisons are supplied in the bundle and mapped in its record index. Later reference top sets are used solely to describe selection correspondence.

The high full-period correlations are therefore not identical to the early-policy diagnostics. METR early Linear/later reference rho is 0.657442/0.655677, with primary overlap 9/21 and 7/21. PEMS is 0.748243

with overlap 14/17. Early Linear and variance sets overlap 16/21 for METR and 2/17 for PEMS; the resulting Jaccards are 0.615385 and 0.0625. These patterns leave temporal ranking mismatch and set similarity as alternatives to a specific interaction explanation. An early GWN exact-loss policy is not the already evaluated GWN value-gradient policy. All exact-reference and Linear magnitude-gradient sets differ from the original policies. Their subsequent complete primary restoration evaluation is reported in S14–S16; the 5% diagnostic candidates were not automatically evaluated.

Table S6. Realized primary restoration by source, not by nominal eligible-set count. Window–mask cases are dependent. Percentages are rounded; full distributions for all policies and fault/budget cells are distributed.

Dataset / policy	Zero sources (%)	One source (%)	At least two (%)	Mean active sources	Nominal eligible sources
METR Linear	39.75	22.59	37.66	1.487	21
METR variance	43.89	19.32	36.79	1.498	21
METR Bounded	37.82	23.79	38.38	1.434	21
PEMS Linear	34.96	38.43	26.60	1.076	17
PEMS variance	35.25	30.61	34.13	1.131	17

The supplied source-level distributions report fault frequencies, policy–fault overlap, eligibility fractions, active-source counts and returned-reading counts for every retained condition. Eligibility coverage is not identical to information actually returned under the base mask. The archive provides joint restoration outcomes for fixed sets and aggregate information-access fractions. Component-specific restoration under a common fault background would require a separate comparison.

S10. Identical-sequence group sensitivity

The complete published value trajectories give 145 singleton groups plus groups of sizes 8, 8, 7 and 2; the fitting prefix gives 142 singleton groups plus groups of sizes 8, 8, 7, 3 and 2. Full maps, sequence hashes and column-to-group membership are in PEMS_DUPLICATE_GROUP_MAP.csv. These groups describe equality of the published sequences. The full-period map is retained as a post hoc diagnostic; evaluated group-aware policies use fitting-period sequences.

Table S7. PEMS primary target-weight sensitivity under unchanged model inputs/graph/predictions. Group weighting first averages the target error contributions within a sequence group, then gives each group equal weight. All 24 original policies and four conditions remain in PEMS_GROUP_WEIGHT_SENSITIVITY.csv. Entries are reported as point summaries.

Weighting	Groups	Linear MAE	Variance MAE	Variance minus Linear	Linear distinct groups selected	Variance distinct groups selected
Original columns	170	18.194477	18.477554	0.283077	17	10
Full-period identical groups	149	19.626756	19.953786	0.327030	17	10
Fitting-only identical groups	147	19.764265	20.095820	0.331555	17	10

Each column's weight is inverse group size. Counts and per-day weighted differences are in PEMS_GROUP_WEIGHT_SENSITIVITY.csv and PEMS_GROUP_WEIGHT_DAILY.csv. Positive pooled point differences survive both weight schemes. The operation reweights target errors while preserving predictor inputs, graph structure and predictions. The group-aware restoration policies are

evaluated separately. All selected Linear sources are singleton groups under both definitions; variance includes redundant selected columns. The original policies retain their selected groups; equal group coverage is evaluated through the separate group-aware rules.

S11. Fitting trajectories and checkpoint stability

Available histories contain per-epoch validation MAE and learning rates for the six relevant historical fitted models. The available logs contain validation MAE and learning rates for GWN, and additionally contain training objectives for the proxy fits where recorded. Figure S3 shows validation and learning-rate trajectories. The lowest logged validation epoch is descriptive and does not replace the original checkpoint-update tolerance. Exact training histories and checkpoint identities are retained in the reproducibility bundle.

The capped fits retain selected epochs 92/96 for METR Linear/Bounded and 92/93 for PEMS Linear/GWN. The best/last comparison directly evaluates ranking and restoration sensitivity to the saved checkpoint. The best-reference results were reaggregated from saved original batches, and the last-reference results were evaluated by new CPU forwards, without training or new data. A CPU replay of the identical primary Linear set provided a numerical check; the other best-reference outcomes use their retained original inference batches. The frozen Linear proxy reproduced its saved earlier scores to 4.7e-8 (METR-LA) and 6.3e-7 (PEMS08) and reproduced the top 11, 21 and 42 METR-LA sets exactly and the top 17 PEMS08 set exactly; at the PEMS08 5% and 20% budgets one source of nine and one of 34 differ. Across the complete later cohort and all five base masks, absolute Global MAE moved by 0.035%, 0.183% and 0.535% for references 42, 43 and PEMS08, in opposite directions across datasets. Relative gains over no restoration moved by at most 0.03 percentage points for reference 42 and PEMS08 but by 0.14 points for reference 43, and the ordering of policies was not preserved (Spearman 0.93, 0.41 and 0.99). The primary 10% Linear sets were identical and all source sets used for the later reference comparison were held fixed, so those changes concern the reference weights. Outside the primary budget, the PEMS08 5% and 20% diagnostic Linear sets each change by one source. Table S8 lists every policy and both files. For PEMS08 the group-aware reduction of the Linear-versus-variance gap was 31.4% with the best file and 32.3% with the last file, and it was positive on all seven observed dates under both. Excluding any single date leaves that reduction between 29.6% and 32.4% with the best file and between 30.5% and 33.3% with the last file; the seven dates remain dependent descriptions of one exposed cohort.

What was frozen and what was recomputed. The earlier and later window lists, the eleven excluded boundary-crossing windows, the five base masks, the graph-block fault sets (seed 20261103), target eligibility, the fitting-period variance score, the 147 identical-sequence groups and every policy's source set were carried over unchanged. The exact-reference set therefore remains the one selected from each reference's best-checkpoint earlier deletion scores; it was not reselected under the last file. Only the weight file used for the later-period forward passes changed. The Linear proxy's earlier deletion scores were recomputed for both saved files to test selection stability, and the resulting top sets were identical at every budget except the PEMS08 5% and 20% budgets, where a single source differs. Recomputing a reference's deletion ranking under the last file would require about 675,000 single-source forward passes per reference and was outside the scope of this diagnostic. Same-background individual-restoration comparisons remain open.

Table S8. Best versus last saved checkpoint under the unchanged primary condition (10% graph-block fault, 10% restoration budget, five base masks, complete later cohort). best rows re-aggregate the original per-window batches; last rows are CPU forward passes of the saved last file. The 16 published cells reproduce to a maximum absolute difference of 2.1e-14 and identical restored-entry counts. Global MAE is in mph for METR-LA and in vehicles per five minutes for PEMS08. The paired differences condition on the saved checkpoints, shared masks and overlapping windows.

Reference	Policy	best	last	last - best	Relative (%)	Gain vs none, best to last (%)
METR42	No restoration	3.234207	3.233085	-0.001122	-0.035	0.000 to 0.000
METR42	Linear (best set)	3.203683	3.203009	-0.000674	-0.021	0.944 to 0.930
METR42	Linear (last set)	3.203683	3.203009	-0.000674	-0.021	0.944 to 0.930
METR42	Variance	3.204551	3.203917	-0.000634	-0.020	0.917 to 0.902
METR42	Bounded	3.203185	3.202600	-0.000585	-0.018	0.959 to 0.943
METR42	Exact-reference	3.204506	3.204323	-0.000183	-0.006	0.918 to 0.890
METR43	No restoration	3.270726	3.270434	-0.000292	-0.009	0.000 to 0.000
METR43	Linear (best set)	3.244299	3.239320	-0.004978	-0.153	0.808 to 0.951
METR43	Linear (last set)	3.244299	3.239320	-0.004978	-0.153	0.808 to 0.951
METR43	Variance	3.245398	3.239452	-0.005946	-0.183	0.774 to 0.947
METR43	Bounded	3.243111	3.239059	-0.004052	-0.125	0.844 to 0.959
METR43	Exact-reference	3.241603	3.239874	-0.001729	-0.053	0.890 to 0.934
PEMS08 reference 42	No restoration	18.518179	18.617305	+0.099125	+0.535	0.000 to 0.000
PEMS08 reference 42	Linear (best set)	18.194477	18.286825	+0.092347	+0.508	1.748 to 1.775
PEMS08 reference 42	Linear (last set)	18.194478	18.286825	+0.092347	+0.508	1.748 to 1.775
PEMS08 reference 42	Variance	18.477554	18.575815	+0.098261	+0.532	0.219 to 0.223
PEMS08 reference 42	Group-aware variance	18.388647	18.482422	+0.093775	+0.510	0.699 to 0.724
PEMS08 reference 42	Exact-reference	18.208726	18.300708	+0.091982	+0.505	1.671 to 1.701
PEMS08 reference 42	GWN-gradient	18.357873	18.449458	+0.091586	+0.499	0.866 to 0.902

The CAAF primary author-preprint abstract and methods informed the comparison with correlation-assisted placement. The seven static PEMS rank controls contain six nonconstant directions; standardized-rank SVD has rank seven including the intercept, with singular values and tolerance supplied in the computational records. The unified Zenodo archive, version 1.7.1 (https://doi.org/10.5281/zenodo.23614390), contains the original release, twelve original checkpoint files, subsequent reference fits, correlation-selection results, saved-output audit and blockwise label-exchange materials. Original files retain their historical layout under original_release_1.6.1; additions and block_exchange contain the later analyses. Sections S19–S23 document the same-environment inference replays and their coverage.

S12. Complete primary observed-date contrasts

Table S9. All observed dates and their shared Global valid-target counts, pooled across five masks. Gains are variance minus Linear in the dataset unit; a dash means that reference is absent. The date-level summaries retain dependence from overlapping windows and shared masks.

Dataset	Encoded date	Valid targets	Gain, ref. 42	Gain, ref. 43
METR-LA	2012-06-16	2,196,515	0.007682936	0.006988220
METR-LA	2012-06-17	899,340	0.001130467	-0.000520169
METR-LA	2012-06-18	1,940,850	-0.000397718	0.000499867
METR-LA	2012-06-19	3,512,530	0.001237395	0.002487848
METR-LA	2012-06-20	3,517,990	-0.000570202	0.000391153
METR-LA	2012-06-21	3,480,100	-0.001930031	-0.000047453
METR-LA	2012-06-22	3,475,365	0.004588821	0.001716773
METR-LA	2012-06-23	3,388,560	-0.002259092	-0.002572582
METR-LA	2012-06-24	3,413,595	-0.001508792	-0.002153867
METR-LA	2012-06-25	3,354,320	0.001269334	0.002304362
METR-LA	2012-06-26	3,400,810	0.000522043	0.001041794
METR-LA	2012-06-27	3,265,800	0.002819942	0.003803308
PEMS08	2016-08-25	2,937,600	0.307708231	—
PEMS08	2016-08-26	2,937,600	0.320675280	—
PEMS08	2016-08-27	2,937,600	0.260419609	—
PEMS08	2016-08-28	2,937,600	0.236525656	—
PEMS08	2016-08-29	2,937,600	0.273825862	—
PEMS08	2016-08-30	2,937,600	0.247413490	—
PEMS08	2016-08-31	2,825,400	0.337029244	—

S13. Model configurations and selected checkpoints

All three proxy fits use OutageResidual in development_runtime_v05.py: a two-channel GRU with hidden width 32; node, hour and weekday embeddings of dimensions 4, 4 and 3; local head Linear(43,32), ReLU, Linear(32,12); graph head Linear(36,12) without bias; top-k prior width eight; and cache_empty=False in these runs. METR Linear/Bounded differ by additive versus tau*tanh(d/tau), tau=2; both have 6,637 parameters. PEMS Linear inherits the additive construction with 170 rather than 207 node embeddings and has 6,489 parameters. Fitting scalers and priors are dataset-specific.

The two METR references and the PEMS reference call the same gwn_mask_adapter_v09.make constructor. It uses four input channels, 12 outputs, two directed prior supports and learned adaptive adjacency. graph_wavenet_compat_v09.gwnet defaults are residual/dilation widths 32/32, skip/end widths 256/512, kernel width two, four blocks with two layers each, and dropout 0.3. The METR references each have 309,464 parameters; no PEMS parameter count was inferred by merely copying that number. Its node-dependent adaptive embeddings use the 170-node constructor. All six use Adam with initial learning rate 0.001, weight decay 0.0001 and gradient clipping at five; the plateau scheduler halves the rate after five plateau epochs down to 0.00001. Their distinct epoch histories and stopping outcomes remain in S11.

Historical selected epochs are 92/96 for METR Linear/Bounded, 54/52 for METR GWN42/GWN43 and 92/93 for PEMS Linear/GWN42. The bundle contains exact configurations, available training histories and the twelve best/last checkpoint files, with their source mapping and hashes in the technical record

index. S16 specifies which branches have saved-output verification and which have model inference replays.

S14. Exact-reference, gradient-magnitude and group-aware controls

The added primary-condition experiment evaluates METR references 42/43 and PEMS reference 42 under every original later start and five masks. Earlier policy scores, identifiers, sets and starts are distributed in analysis_R3/FROZEN_CONTROLS.npz and FROZEN_POLICY_SOURCES.csv. The group constraint uses the fitting-only 147-group map, not the full-period 149-group sensitivity map. It retains 17 different groups by scanning the original descending score order (source identifier resolves ties) and skipping previously selected groups. Group-aware Linear is exactly the original eligible set. Its predicted source errors, counts and dates are bitwise equal after mapping the shared input prediction, so the group-aware label is recorded as an alias of Linear.

Table S10. Complete added Global primary contrasts. Gain is control MAE minus candidate MAE. Date and circular-two-date intervals use the same fitted model, source sets and observed dates; they do not quantify independent training or future-traffic uncertainty. The symmetric group-Linear row duplicates Linear exactly and is labeled as such. Intervals are descriptive, unadjusted post hoc sensitivities.

Reference	Control → candidate	Gain	Date 95%	Circular 2-date 95%
METR42	Linear → Exact GWN	-0.000823143	[-0.002028831, 0.000341590]	[-0.001922832, 0.000528998]
METR42	Linear → Linear magnitude	-0.001464955	[-0.003852307, 0.000718254]	[-0.003823347, 0.000311268]
METR42	Variance → Exact GWN	0.000045272	[-0.002163412, 0.002496614]	[-0.002237461, 0.002653295]
METR42	Variance → Linear magnitude	-0.000596540	[-0.001956696, 0.000745744]	[-0.001947090, 0.000640073]
METR43	Linear → Exact GWN	0.002696101	[0.000784567, 0.004424564]	[0.001162783, 0.004240033]
METR43	Linear → Linear magnitude	-0.001397561	[-0.003862979, 0.000635553]	[-0.004069148, 0.000525756]
METR43	Variance → Exact GWN	0.003795021	[0.001319153, 0.005769153]	[0.001442685, 0.006143770]
METR43	Variance → Linear magnitude	-0.000298641	[-0.002031311, 0.001108295]	[-0.002162581, 0.001200270]
PEMS08 reference 42	Linear → Exact GWN	-0.014248509	[-0.017793558, -0.011092147]	[-0.017625898, -0.011805139]
PEMS08 reference 42	Linear → Linear magnitude	-0.135404512	[-0.150526197, -0.121221583]	[-0.149180481, -0.123042962]
PEMS08 reference 42	Variance → Exact GWN	0.268828161	[0.241952012, 0.296420546]	[0.239722615, 0.298224366]
PEMS08 reference 42	Variance → Linear magnitude	0.147672157	[0.128878183, 0.167860617]	[0.127067134, 0.170363065]
PEMS08 reference 42	Variance → Group variance	0.088906626	[0.072939285, 0.106862751]	[0.072869078, 0.109768097]
PEMS08 reference 42	Group variance → Linear	0.194170043	[0.179134115, 0.209185755]	[0.180749936, 0.205787154]
PEMS08 reference 42	Group variance → Group Linear (alias)	0.194170043	[0.179134115, 0.209185755]	[0.180749936, 0.205787154]

ABSOLUTE_PRIMARY.csv retains all 20 policy rows including original controls, three MAE scopes, denominator units, restored occurrences and actual zero/one/multiple-source counts. ALL_OBSERVED_DATES.csv gives every policy/date numerator and denominator. There are 12 METR

dates per reference and 7 PEMS dates; no later-label policy selection or new dates were added. Source-count equality does not match restored entries. Under group-aware variance, PEMS restores 153,804 occurrences, compared with 136,032 for original variance and 129,432 for Linear. The 31.4% reduction describes the error-gap change between these specified policies, which also return different readings.

S15. Numerical consistency of the added controls

The initial PEMS comparison of computational batch size one and batch size 32 gave a maximum valid-source MAE discrepancy of 0.000178485359, exceeding the unchanged 0.0001-unit tolerance. Shape-dependent float32 differences were reproduced on CPU. The final evaluator uses a fixed computational shape of 32 for direct, packed and incomplete-tail calls. Repeated padding outputs are discarded before calculating numerators, counts and dates; BatchNorm statistics and dropout remain fixed in evaluation mode.

The shared error calculation was checked by 26 CPU forwards across 20 policy cases, including changed inputs and incomplete tails, with zero single/packed discrepancy and unchanged model state. Three-reference CUDA checks also gave zero discrepancy. On the complete evaluated cohort, original-policy Global MAEs differed from saved values by at most 0.000000000381 dataset units, and all twelve baseline restoration counts matched exactly. A separate NumPy reaggregation checked 619 arithmetic, date and resampling quantities. These checks support numerical comparability of the specified computation; they do not constitute independent model fitting or scientific generalization. The original failed check and detailed verification logs remain in the technical archive.

S16. Reproduction scope of the historical added controls

The bundle supplies the earlier selected sets, per-date error aggregates, source tables and replay code for the original added-control analyses. Their table and interval arithmetic can be recomputed without the original private machine paths. The original evaluators used archived benchmark files and saved weights; they did not freshly refit every predictor or repeat every earlier score computation. The best/last CPU comparison in S11 remains a separate analysis, and exact-reference policies were not reselected under the last reference file.

Historical ridge-refit, joint-greedy and negative-development branches are supplied as retained code and saved-output records, with inference-replay coverage specified below. A complete fresh historical prediction run also requires matching preprocessing, node order, scalers, target validity, calendars, graph construction, masks, source sets and dependencies. Some archived evaluators depend on the original project layout. The historical evaluators and their environment requirements are identified in the archive. S19 describes the separately completed same-environment CUDA replays for the seven new validation cases. Default saved-output verification is distinct from inference and from optional document regeneration.

S17. Exact temporal boundaries and archived model settings

We used archived versions of METR-LA speed and PEMS08 flow data. The benchmark sources are identified in the DCRNN and Graph WaveNet materials and the ASTGCN data repository [2,1,33]; the exact local input versions and provenance records are listed in Supplementary Information. Dataset-specific units are retained and no speed/flow effects are pooled. A window contains 12 historical steps and 12 future targets at five-minute resolution, with stride one. Timeline partitions are applied to original indices

before window construction: all future targets must lie within their assigned role, while history may precede that role's first target.

For METR-LA, the exclusive upper boundaries are 21,591 for parameter fitting, 23,990 for training-tail calibration and 27,417 for development validation. The corresponding counts are 21,568, 2,388 and 3,416 windows. Calibration is not used to fit neural model parameters, and validation selects checkpoints. Of 6,844 original historical evaluation windows, 531 have no valid targets and are excluded, leaving 6,313. The fitting-only mean and standard deviation are 58.62550693096417 and 12.801036685068539. METR readings are valid when finite and nonzero. This convention defines native history availability and valid targets; it does not diagnose device health.

PEMS08 contains 17,856 time steps and 170 published columns. Its archived cohorts contain 11,226 parameter-fit windows, 1,239 calibration windows, 1,774 validation windows and 3,561 historical reference windows. The exclusive parameter-fit, calibration, validation and historical upper boundaries are 11,249, 12,499, 14,284 and 17,856. Its fitting-only mean and standard deviation are 229.7117835310851 and 145.4811702933454. Flow zero is valid. The original graph, repeated columns and released interpolated values are retained. Neither finite entries nor the speed dataset's zero rule are used to infer the frequency of real flow-sensor failures.

Historical evaluation is divided using the date of the first future target. For METR-LA, the earlier 12 encoded dates yield 3,260 selection windows after excluding 11 target-crossing windows; the later 12 dates yield 3,042 evaluation windows. For PEMS08 the corresponding counts are 1,545, 11 and 2,005, with seven later dates. These are within-archive roles. All periods had prior development exposure.

The proxy has node-independent two-channel value/mask GRU encoders with hidden dimension 32, node and calendar embeddings, a local prediction head and a graph-output head. Its local prediction includes the last available value, or an empty-history state if none is available. The additive graph correction uses receiver-normalized prior weights after removing self-loops, retains at most eight sources per receiver, and is not renormalized after truncation or deletion. A coefficient is multiplied by the source's fraction of available history. Linear sums coefficient-weighted source/receiver output differences; Bounded replaces each difference by the bounded transform $g(d) = \tau \tanh(d/\tau)$, with $\tau = 2$, applied before multiplying by its graph coefficient. Here d is the source-minus-receiver graph-head output in standardized prediction units. Both METR proxies have 6,637 parameters.

Node-independent encoding and a fixed one-hop additive correction allow exact reuse of proxy states when complete source histories are removed. Deleted sources' own predictions and receiver contributions must both change. We recompute absolute errors from the resulting joint predictions; we do not sum individual loss increments to obtain a set loss. The proxy architecture supports this computational reuse; Graph WaveNet deletion uses explicit masked-input inference.

The METR-LA Graph WaveNet references use 309,464 parameters with four input channels: masked standardized values, an availability mask, the final historical hour divided by 23 and weekday divided by 6, broadcast over histories and sources. The adapter supplies receiver-normalized prior and reverse transition supports and enables adaptive adjacency [1]. Values are standardized before unavailable entries are zeroed; the explicit mask distinguishes standardized zero from absence. Calendar channels remain available under the interventions.

The archived METR references use Adam (initial learning rate 0.001, weight decay 0.0001), gradient norm clipping at 5, validation-based learning-rate halving with patience 5, and a minimum rate of 0.00001. A checkpoint requires improvement exceeding 0.00001 mph at fixed 30% missingness and mask 1001.

After at least 30 epochs, 15 non-improving epochs stop training; the cap is 96. Reference 42 completed 69 epochs with best epoch 54 and reference 43 completed 67 with best 52. Both reached patience. Their training order seed is 42 and mask base seed is 1043; initialization and dropout randomness differ. Four persistent mask views (10%, 20%, 30%, 50%) expand a batch of 16 windows to 64 views. The masking epoch argument is fixed at zero rather than independently resampling views every epoch.

The METR Linear and Bounded fits completed 96 epochs, selecting epochs 92 and 96 without satisfying patience. PEMS08 Linear completed 96 epochs, with checkpoint epoch 92 selected. The original PEMS08 GWN run stopped after 82 complete epochs for resource reasons; continuation restored the saved model, optimizer, scheduler and random states, and completed epochs 83–96, selecting epoch 93 with patience count 3/15. Neither PEMS08 fit is described as converged. Only that original PEMS08 reference branch is included in S1–S18; four newly fitted rolling branches are reported in S19. PEMS uses the same four-channel GWN adapter and the same Linear GRU/graph-head construction, with 170 node embeddings and its own fitting scaler/prior; the Linear parameter count is 6,489. Supplementary section S13 and MODEL_CONFIG_MAP.csv map all six fitted predictors to their exact constructor, archived configuration/log copies and checkpoint hashes. Separately named development branches are outside this manuscript's evidence cohort.

S18. Ranking summaries retained from the previous main text

Full-period single-source Linear correlations with METR-LA references 42 and 43 were 0.748974 and 0.738737, exceeding the strongest of ten inexpensive controls by 0.115639 and 0.070950. The strongest control changed from variance to Linear value-gradient magnitude. In PEMS08, the Linear correlation was 0.841022, with a conditional date interval of 0.804102–0.855392, while the best inexpensive control reached 0.231866. These diagnostics use the full-period cohort, not the earlier source-selection period.

Table S11. Single-source Global ranking associations. The best inexpensive score is chosen from the original ten-score family, not from the auxiliary references. Spearman correlations are unitless.

Dataset / reference	Linear proxy	Best inexpensive score	Control correlation	Proxy minus control
METR-LA / 42	0.748974	Fit variance	0.633335	0.115639
METR-LA / 43	0.738737	Linear value-gradient magnitude	0.667787	0.070950
PEMS08 / 42	0.841022	Linear value-gradient magnitude	0.231866	0.609156

Table S12. Decision-aligned ranking, selected-set correspondence and original primary utility. Reference denotes a fixed GWN initialization. Later reference top sets are diagnostic only. Complete directly evaluated policies appear in Supplementary Table S43.

Dataset / reference	Early Linear vs later reference rho	Early Linear vs later reference top-set overlap	Early Linear vs variance set overlap	Linear gain vs variance	Unit
METR / 42	0.657	9 / 21	16 / 21	0.000868	mph
METR / 43	0.656	7 / 21	16 / 21	0.001099	mph
PEMS / 42	0.748	14 / 17	2 / 17	0.283077	vehicles / 5 min

PEMS partial rank correlations are 0.869110 after variance adjustment and 0.869873 after the nominal seven-static-score adjustment (six nonconstant directions plus intercept). These are descriptive residual-rank associations. The two METR reference full-period rankings correlate at 0.845414 and share 16 of their top 21 sources; both quantities summarize the two observed fitted references.

Main-text citation numbers apply to this supplement; its cited benchmark and model sources are listed in the main bibliography. The scientific results above are unchanged by editorial relocation.

S19. GWN confirmation and rolling evaluation

This section reports the separately specified PEMS03 confirmation with three full fitting repetitions and four rolling PEMS08 folds, distinct from the historical experiments in S1–S18. Fitting, source selection and final evaluation use the disjoint roles in Table S13. The original specified contrasts remain separate from the post hoc additions in Tables S18–S20.

PEMS03 independence was assessed by a local text, structured-record and artifact-name audit plus the authors' declaration of no known relevant use outside this project. The audit and author declaration record the known exposure status within the available project history. The raw file was hashed without initially decoding its final suffix. Training decoded only through D54 and selection only through D62; all three sets of fitted weights and selected policies were fixed before evaluation decoded D64–D91. PEMS08 was already exposed during development and is described only as retrospective rolling validation.

Table S13. Inclusive encoded day-block roles in the additional experiment. All 12 history and 12 target positions must lie inside their role. Calibration is reserved and unused. PEMS03 D55/D63 are gaps. Each PEMS08 fold has a one-day gap immediately before evaluation; fitting prefixes expand. Nominal calendars do not verify physical event timestamps.

Case (seed)	Fit	Calibration	Validation	Selection	Evaluation
PEMS03 (42/43/44)	1–42	43–47	48–54	56–62	64–91
PEMS08 fold 1 (42)	1–24	25–26	27–30	31–33	35–41
PEMS08 fold 2 (42)	1–31	32–33	34–37	38–40	42–48
PEMS08 fold 3 (42)	1–38	39–40	41–44	45–47	49–55
PEMS08 fold 4 (42)	1–45	46–47	48–51	52–54	56–62

PEMS03 supplies 12,073 fitting windows, 1,417 calibration windows, 1,993 validation windows, 167 hourly selection windows and 8,041 evaluation windows per complete seed. Each PEMS08 fold supplies 71 hourly selection windows and 1,993 evaluation windows; the four evaluation periods are disjoint and sum to 7,972 windows over 28 day blocks. Each PEMS03 policy has 172,720,680 valid target occurrences per seed, and each PEMS08 policy has 20,328,600 per fold. Five masks are pooled, not treated as independent training repetitions. The new role containment and hourly selection differ from the original historical cohort, so its seven-day PEMS08 values remain unchanged and are not merged with these results.

Table S14. All named policy families in the added experiment. Global, failed-source and tail MAE are in vehicles per five minutes. PEMS03 is the equal-weight mean of three complete fitted pipelines; restored occurrences are the mean per seed. PEMS08 is target-pooled over four disjoint evaluation folds; restored occurrences are summed across folds. Random mean is the equal mean of 20 retained fixed policies. Complete individual policies, every seed/fold and target counts are distributed in ALL_POLICIES.csv. Aliases are retained as identities, not repeated evidence.

Dataset	Policy	Global MAE	Failed MAE	Tail MAE	Restored occurrences
PEMS03	none	17.318014	20.787069	65.508372	0.0
PEMS03	Linear	17.187704	19.533052	64.043923	1,021,520.0
PEMS03	Bounded	17.183516	19.516297	64.028067	1,067,704.0
PEMS03	variance	17.239775	20.135388	64.956717	1,136,508.0
PEMS03	Linear gradient	17.233250	20.081487	64.873786	1,171,348.0
PEMS03	exact reference	17.158306	19.312699	63.873843	1,096,012.0
PEMS03	group variance	17.239775	20.135388	64.956717	1,136,508.0
PEMS03	group-aware Linear	17.187704	19.533052	64.043923	1,021,520.0
PEMS03	random mean	17.240757	20.148184	65.082549	1,191,747.0
PEMS08	none	18.864327	23.746328	82.830080	0.0
PEMS08	Linear	18.603757	21.142382	79.911971	441,600.0
PEMS08	Bounded	18.603684	21.138484	79.915077	431,172.0
PEMS08	variance	18.813285	23.285706	82.566760	555,480.0
PEMS08	Linear gradient	18.741983	22.542773	81.407725	523,332.0
PEMS08	exact reference	18.600240	21.102037	79.870618	428,484.0
PEMS08	group variance	18.740237	22.563563	81.553872	619,680.0
PEMS08	group-aware Linear	18.603757	21.142382	79.911971	441,600.0
PEMS08	random mean	18.799056	23.132977	82.340697	595,650.6

Table S15. Prespecified GWN contrasts and block-length sensitivities. Gain is control minus candidate. Intervals are unadjusted and conditional on fitted weights and source sets. Noncircular blocks weight edge days differently, so a percentile interval need not be centered on the full-cohort point estimate. Rolling draws may cross fold boundaries and reweight the realized sequence of four fitted configurations. Post hoc boundary and edge-weight sensitivities are reported separately in Tables S18–S20; the later DCRNN label-exchange analysis is in S27.

Dataset	Control − candidate	Gain	7-day interval	3-day interval	14-day interval
PEMS03	variance − Linear	0.052071	0.035786 to 0.068018	0.038395 to 0.064428	0.040759 to 0.067702
PEMS03	exact reference − Linear	-0.029397	-0.035095 to -0.025459	-0.035875 to -0.024182	-0.034061 to -0.026344
PEMS03	variance − group variance	0.000000	0.000000 to 0.000000	0.000000 to 0.000000	0.000000 to 0.000000
PEMS08	variance − Linear	0.209529	0.194260 to 0.250113	0.181891 to 0.248656	0.213965 to 0.245479
PEMS08	exact reference − Linear	-0.003517	-0.010451 to 0.004527	-0.011266 to 0.003710	-0.006103 to 0.003705
PEMS08	variance − group variance	0.073049	0.050160 to 0.108543	0.046958 to 0.106614	0.064767 to 0.108665

Table S16. All 21 fitting endpoints, with the recorded stopping rule and selected checkpoint. Four fits reached the 192-epoch cap. Checkpoint selection used the raw validation improvement threshold. Best and last weights and available histories are retained.

Case	Model	Epochs	Best epoch	Bad epochs	Stop
PEMS03_s42	GWN	129	114	15	patience
PEMS03_s42	Linear	192	192	0	cap
PEMS03_s42	Bounded	192	192	0	cap
PEMS03_s43	GWN	108	93	15	patience
PEMS03_s43	Linear	192	182	10	cap
PEMS03_s43	Bounded	131	116	15	patience
PEMS03_s44	GWN	120	105	15	patience
PEMS03_s44	Linear	179	164	15	patience
PEMS03_s44	Bounded	192	191	1	cap
PEMS08_F1_s42	GWN	143	128	15	patience
PEMS08_F1_s42	Linear	151	136	15	patience
PEMS08_F1_s42	Bounded	151	136	15	patience
PEMS08_F2_s42	GWN	174	159	15	patience
PEMS08_F2_s42	Linear	130	115	15	patience
PEMS08_F2_s42	Bounded	130	115	15	patience
PEMS08_F3_s42	GWN	140	125	15	patience
PEMS08_F3_s42	Linear	152	137	15	patience
PEMS08_F3_s42	Bounded	152	137	15	patience
PEMS08_F4_s42	GWN	121	106	15	patience
PEMS08_F4_s42	Linear	127	112	15	patience
PEMS08_F4_s42	Bounded	124	109	15	patience

Table S17. Weekly point contrasts and leave-one-week-out ranges. PEMS03 weekly values average the three fitted pipelines on each week; PEMS08 weeks coincide with folds. The folds use one seed and overlapping earlier fitting prefixes.

Dataset	Control − candidate	Week 1	Week 2	Week 3	Week 4	Leave-one-week-out range
PEMS03	variance − Linear	0.065321	0.066158	0.043906	0.032873	0.047357 to 0.058448
PEMS03	exact reference − Linear	-0.018359	-0.031550	-0.034704	-0.032929	-0.033061 to -0.027622
PEMS03	variance − group variance	0.000000	0.000000	0.000000	0.000000	0.000000 to 0.000000
PEMS08	variance − Linear	0.187525	0.233663	0.254787	0.162140	0.194443 to 0.225325
PEMS08	exact reference − Linear	0.000000	-0.001424	0.010355	-0.022999	-0.008141 to 0.002977
PEMS08	variance − group variance	0.083083	0.100571	0.103974	0.004567	0.062740 to 0.095876

Post hoc direct comparator and fold-boundary sensitivities

The following analyses were added on 19 October 2026 after inspection of the completed experiments and the review request. The frozen list specified variance minus Linear, exact-reference minus Linear and variance minus group-aware variance; it did not specify the direct group-aware-variance-minus-Linear contrast or these additional resampling schemes. Their post hoc analysis record was written after the

completed results were inspected and before these additional summaries were computed. No model, selected set, raw input, original result or tolerance changed.

Table S18. Newly added direct group-aware-variance-minus-Linear comparison on rolling PEMS08. MAEs and gains are in vehicles per five minutes. Relative gain divides by group-aware variance MAE. Fold rows describe different fitted models and selected sets with overlapping training prefixes. The pooled estimate uses common valid-target counts (20,328,600 per fold; 81,314,400 total).

Evaluation	Group-aware variance MAE	Linear MAE	Gain	Relative gain (%)
Pooled four folds	18.740237	18.603757	0.136480	0.728
Fold 1	18.878169	18.773726	0.104442	0.553
Fold 2	19.352503	19.219411	0.133092	0.688
Fold 3	18.881294	18.730481	0.150813	0.799
Fold 4	17.848982	17.691409	0.157573	0.883

For the new direct contrast, the original noncircular shared indices give a seven-day conditional percentile interval of 0.122096–0.151043, a three-day interval of 0.120589–0.152730 and a fourteen-day interval of 0.125432–0.152570. Each draw recomputes the difference of paired sampled absolute-error numerators divided by their shared sampled target count. Neither endpoint subtraction nor subtraction of independently sampled intervals was used. The same 10,000 indices are shared across all three variance/group/Linear contrasts at each block length; draw-wise contrast additivity was checked. These are unadjusted post hoc conditional percentile summaries.

The rolling estimand is the difference between target-pooled errors over the four realized evaluation segments using their actual fold-specific fitted predictor and selected set. A cross-boundary sampled block traverses a change in this configuration; the block therefore spans multiple fitted model and selection configurations. The original fourteen-day noncircular draws give expected day-slot totals of 3.733333, 10.266667, 10.266667 and 3.733333 for folds 1–4, respectively, compared with seven observed day blocks each. Individual expected day multiplicities range from 0.133333 to 1.866667. This upweights the middle folds, where the variance-minus-Linear gains are larger. The ratio of expected reweighted error-difference numerator to expected target count is 0.231210, distinct from both the unweighted point 0.209529 and the expectation of a ratio. These inclusion weights explain why the original conditional percentile interval can exclude the full-cohort point estimate. Complete day/fold weights are supplied in EXPECTED_DAY_INCLUSION.csv and EXPECTED_FOLD_INCLUSION.csv.

Table S19. Post hoc comparison of resampling objects on the same rolling outputs. V denotes variance; GV denotes group-aware variance; L denotes Linear. Intervals are 2.5th–97.5th percentiles of 10,000 paired draws in vehicles per five minutes. Original noncircular V−L rows are retained unchanged; only GV−L and the circular/fold sensitivities are new. Circular concatenation balances marginal day inclusion but adds an artificial end-to-start join and still crosses model/selection boundaries. Independently resampling circular blocks inside each fold keeps its model and selected set fixed and prevents cross-fold blocks; it does not account for dependence from shared training prefixes. The schemes provide alternative conditional reweightings of the same observed outputs.

Contrast	Sampling scheme	3-day interval	7-day interval	14-day interval
V − L	Original noncircular path	0.181891 to 0.248656	0.194260 to 0.250113	0.213965 to 0.245479
V − L	Circular path	0.177997 to 0.244361	0.173684 to 0.243984	0.176041 to 0.242787
V − L	Within-fold circular	0.188349 to 0.231638	Degenerate; no interval	Inadmissible
GV − L	Original noncircular path	0.120589 to 0.152730	0.122096 to 0.151043	0.125432 to 0.152570
GV − L	Circular path	0.121051 to 0.152048	0.120753 to 0.150042	0.123053 to 0.149327
GV − L	Within-fold circular	0.126447 to 0.146878	Degenerate; no interval	Inadmissible

The within-fold seven-day diagnostic samples every day exactly once per fold, so its estimate is numerically unchanged in all draws. This deterministic result is reported as a diagnostic rather than an uncertainty interval. The seven-day fold duration restricts admissible within-fold block lengths to seven days or fewer. Three-day circular within-fold intervals describe only within-fold day reweighting; folds themselves, training records and selected sets are never redrawn. There are only four observed evaluation weeks and one fitted seed per fold, with shared earlier fitting observations.

Table S20. Every leave-one-fold-out pooled gain, retaining the original fitted models and sets in the other folds. Omitting a fold does not retrain them. Both contrasts favor Linear on each separate fold (Tables 2 and S18), and the new GV−L leave-one-out gains remain positive. This checks concentration of the observed contrast across evaluation weeks.

Contrast	Omit fold 1	Omit fold 2	Omit fold 3	Omit fold 4
Variance − Linear	0.216863	0.201484	0.194443	0.225325
Group-aware variance − Linear	0.147159	0.137609	0.131702	0.129449

The supplied post hoc analysis code and common draw indices allow all direct contrasts and sensitivity intervals to be recomputed from the saved daily numerators and counts. A separate day-multiplicity-matrix calculation reproduced those intervals. This verifies the saved arithmetic for the recorded analyses. The original specified contrast results remain unchanged.

Policy identities and numerical reproduction boundaries

Linear and variance share 6/36, 9/36 and 7/36 selected sources in the three PEMS03 fits. Group-aware variance is exactly variance, and group-aware Linear is exactly Linear, for all three. In rolling PEMS08 their Linear–variance intersections are 2/17, 3/17, 1/17 and 3/17; each original variance set represents ten fitting-period exact-sequence groups and each group-aware set represents 17. Exact-reference and Linear coincide in fold 1; Bounded and Linear coincide in folds 3 and 4. Every alias was checked against the per-window outputs. These identities preclude treating the equal outputs as independent corroboration. The 28 policies, including all random instances, were retained in every case.

Saved-output verification covers all 28 policies in each of the seven cases, including shared window/mask identities, target denominators, restored counts, selected-set sizes and aliases. The six originally specified

dataset-by-contrast estimates and their three-, seven- and fourteen-day intervals were recomputed without changing the reported values. Detailed execution receipts and accounting are retained in the technical archive and are not treated as evidence of an effect.

Each of seven raw-file inference evaluations was actually repeated in separate directories and fresh processes using copied inference sources, best weights, selected policies and explicit input configurations. The repeated inference used the same raw input files and hashes. The same installed Python/Torch/CUDA environment was used. Policy/mask/start/day identity, valid-target/failed-target/restoration counts matched exactly, and maximum per-window Global/failed MAE and tail differences were zero. Package files preserve both original and fresh outputs/status records. This seven-case replay repeats inference in the recorded environment using the same implementation and saved weights. Historical branch coverage and clean-environment requirements are documented separately. A CPU output audit only recomputes saved arithmetic and is labeled accordingly.

An additional packaged-reader CPU regression evaluated the first 32 PEMS03 seed-42 windows under no restoration and Linear, with mask 1001 (64 policy-window rows). Identities and counts matched, but maximum Global/failed-source MAE and tail differences from the saved GPU outputs were 0.000026755, 0.000101664 and 0.000137846. The failed-source and tail differences exceed the unchanged 0.0001-unit gate. CPU numerical portability therefore remains unvalidated. The failed result was retained and the threshold was not relaxed. It does not negate the exact seven-case CUDA replays or supply new independent scientific evidence.

The historical version 1.2.0 bundle was initially supplied locally. Its supporting records are now included within the original-release component of the unified public archive identified in S6; correspondence is available at 622250991028@mails.cqjtu.edu.cn. It contains package-relative data/code mappings, input identities, checkpoints, shared indices and numerical verification logs. The reader guide gives dependency requirements and separate commands for saved-output verification and raw-file inference. The guide distinguishes saved-output verification from fresh inference and describes the environment needed for each.

S20. DCRNN transfer of fixed source sets

This section reports a separately frozen extension on the already known GWN splits. It does not relabel PEMS03 as a new blind confirmation, replace historical GWN numbers, or treat a validation usability gate as a restoration-policy benefit. The original GWN-specific observations in S1–S19 retain their historical scope.

The DCRNN protocol was frozen before its outputs were inspected and after the GWN results were known. Three PEMS03 seeds (42–44) and four PEMS08 rolling folds (seed 42) inherited the exact raw inputs, temporal roles, 12-step history/horizon, mask seeds 1001–1005, graph-block fault seed, 30% background missingness, 10% fault and 10% source-eligibility budget. Only the reference predictor was refitted; the original proxy and inexpensive policy sets were transferred unchanged. DCRNN-specific exact single-source deletion scores used only the same earlier selection windows, with the old exact-GWN set retained as an auxiliary transfer policy, yielding 29 labels including aliases.

The local DCRNN adaptation uses two 64-unit diffusion-GRU layers, diffusion order two, curriculum decay 2000 and 374,273 registered parameters. Its four channels provide masked values, availability, last-history hour and weekday, with no future labels available at inference. Pinned TensorFlow and PyTorch

source commits and licenses are supplied; the local missing-input adaptation and its source versions are documented in Supplementary section S20.

Adam, weight decay, clipping, plateau scheduling, raw-unit improvement threshold, minimum 30 epochs, maximum 192 epochs and patience 15 match the frozen plan. Each optimizer update aggregates 16 windows and four missingness views using valid-target-weighted microbatches of four windows. Computation uses float32 with TF32 disabled. Full model, optimizer, scheduler and random states are retained at complete-epoch continuations. Mask 1001 alone selects the validation checkpoint; five-mask diagnostics never repick it. Each reference must outperform same-mask last-observed persistence before final evaluation.

After all seven references and source sets were jointly frozen, all 29 policy labels were evaluated on five common masks, followed by all seven cases in a fresh directory and fresh CUDA processes in the same installed environment. Every decompressed per-window replay output matched its original exactly; the replay checks numerical consistency within that environment. Cross-architecture comparisons matched mask/start/day identities and valid/failed target denominators.

MAE pools raw error numerators over valid targets, then equally averages PEMS03 fits; rolling PEMS08 pools four realized fold segments. All policies, fitted seeds and architectures share 10,000 day-index draws.

The primary PEMS03 interval uses circular seven-day blocks over 28 days (seed 2026101910); primary PEMS08 uses circular three-day blocks independently within each seven-day fold (seed 2026101911), never crossing a fitted-model boundary. Circular 3/7/14-day concatenated sensitivities, within-fold seven-day degeneracy, and every fold/leave-one-fold-out point are retained. Fourteen-day within-fold blocks are inadmissible. No interval was selected by its favorable direction.

These intervals condition on the fitted references, source sets and observed periods, retaining the shared training-prefix structure. Historical GWN intervals remain unchanged; GWN reaggregation with new shared indices is labeled separately. An independent day-multiplicity calculation reproduced the new intervals within 1e-10 units. The subsequently added blockwise label-exchange analysis is reported separately in S27.

Table S21. All seven new DCRNN reference endpoints. Validation uses the selected mask-1001 checkpoint and is distinct from final fault/restoration evaluation. Every model stopped by patience without reaching the 192-epoch cap. Five-mask validation usability checks passed for all cases.

Case	Completed epochs	Best epoch	Validation MAE	Persistence MAE	Stop
PEMS03 / 42	147	132	19.021818	54.115436	patience
PEMS03 / 43	134	119	18.940684	54.115436	patience
PEMS03 / 44	161	146	18.972835	54.115436	patience
PEMS08 F1 / 42	149	134	18.129558	54.929597	patience
PEMS08 F2 / 42	144	129	17.240655	54.662037	patience
PEMS08 F3 / 42	176	161	16.871523	54.702168	patience
PEMS08 F4 / 42	118	103	19.445044	54.830230	patience

Table S22. Every DCRNN case under final fault/restoration evaluation. MAEs are in vehicles per five minutes. PEMS03 rows share evaluation days; PEMS08 folds use different dates and one reference seed each. The reference-specific exact policy is DCRNN exact selection.

Case	None	Variance	Group variance	Linear	DCRNN exact	Variance minus Linear
PEMS03 / 42	25.324685	24.573522	24.573522	24.919139	23.892199	-0.345617
PEMS03 / 43	25.190700	24.450910	24.450910	24.635957	23.772782	-0.185047
PEMS03 / 44	25.263705	24.503130	24.503130	24.962307	23.846166	-0.459177
PEMS08 F1 / 42	19.238393	19.163447	19.093967	19.026815	19.023302	0.136632
PEMS08 F2 / 42	19.539818	19.462079	19.421937	19.278422	19.282147	0.183657
PEMS08 F3 / 42	18.994927	18.919263	18.867479	18.748657	18.734585	0.170605
PEMS08 F4 / 42	17.830095	17.731995	17.738527	17.603194	17.559704	0.128801

Table S23. Pooled DCRNN contrasts and primary conditional intervals. All contrasts are control minus candidate; positive favors the candidate. Negative exact-reference minus Linear favors DCRNN exact selection. The extra exact-reference minus exact-GWN-transfer comparison is explicitly auxiliary. Group variance is an alias of variance in PEMS03.

Dataset	Control minus candidate	Gain	Conditional 95% interval	Sampling
PEMS03	variance minus Linear	-0.329947	-0.350773 to -0.313547	7-day circular blocks
PEMS03	Group-aware variance minus Linear	-0.329947	-0.350773 to -0.313547	7-day circular blocks
PEMS03	DCRNN exact minus Linear	-1.002085	-1.053646 to -0.933719	7-day circular blocks
PEMS03	DCRNN exact minus Transferred GWN exact	-0.965348	-1.018483 to -0.898199	7-day circular blocks
PEMS08	variance minus Linear	0.154924	0.132292 to 0.178756	3-day circular blocks within folds
PEMS08	Group-aware variance minus Linear	0.116205	0.105992 to 0.126633	3-day circular blocks within folds
PEMS08	DCRNN exact minus Linear	-0.014338	-0.018936 to -0.010091	3-day circular blocks within folds
PEMS08	DCRNN exact minus Transferred GWN exact	-0.013372	-0.017984 to -0.009027	3-day circular blocks within folds

Table S24. Every weekly/fold contrast and corresponding leave-one-block-out effect. Lists follow blocks 1–4; PEMS03 weekly effects equally average three references, whereas each PEMS08 fold has a different reference. Omitting a block does not refit models or reselect sources. The negative PEMS03 variance-minus-Linear contrasts persist after every omission. Both PEMS08 variance comparators favor Linear in every fold and omission. DCRNN exact versus Linear varies in fold 2, although its pooled effect remains favorable.

Dataset	Control minus candidate	Week/fold gains 1–4	Omit block gains 1–4
PEMS03	variance minus Linear	-0.366649, -0.311243, -0.318440, -0.323640	-0.317764, -0.336205, -0.333797, -0.332042
PEMS03	Group-aware variance minus Linear	-0.366649, -0.311243, -0.318440, -0.323640	-0.317764, -0.336205, -0.333797, -0.332042
PEMS03	DCRNN exact minus Linear	-1.070845, -1.033016, -0.902252, -1.002641	-0.979261, -0.991736, -1.035490, -1.001901
PEMS03	DCRNN exact minus Transferred GWN exact	-1.040708, -0.996889, -0.870381, -0.953801	-0.940333, -0.954795, -0.997125, -0.969184
PEMS08	variance minus Linear	0.136632, 0.183657, 0.170605, 0.128801	0.161021, 0.145346, 0.149696, 0.163631
PEMS08	Group-aware variance minus Linear	0.067152, 0.143515, 0.118822, 0.135333	0.132557, 0.107102, 0.115333, 0.109830
PEMS08	DCRNN exact minus Linear	-0.003513, 0.003725, -0.014072, -0.043491	-0.017946, -0.020359, -0.014426, -0.004620
PEMS08	DCRNN exact minus Transferred GWN exact	-0.003513, 0.005075, -0.023522, -0.031527	-0.016658, -0.019521, -0.009988, -0.007320

Table S25. Paired DCRNN-minus-GWN changes in the indicated gain. The GWN values here use the new DCRNN-plan shared indices, separately from every unchanged historical interval. Effects are computed draw by draw before taking percentiles; interval endpoints were not subtracted.

Dataset	Gain contrast	DCRNN minus GWN	Conditional 95% interval
PEMS03	variance minus Linear	-0.382018	-0.410082 to -0.357129
PEMS03	Group-aware variance minus Linear	-0.382018	-0.410082 to -0.357129
PEMS08	variance minus Linear	-0.054605	-0.063781 to -0.045460
PEMS08	Group-aware variance minus Linear	-0.020275	-0.024059 to -0.016405

Table S26. Actual restored input occurrences for the fixed source-count policies. These counts are repeated source-by-history entries across overlapping windows and masks. They quantify returned information and differ between policies; physical costs were unspecified.

Case	Variance	Group variance	Linear	DCRNN exact	Transferred GWN exact
PEMS03 / 42	1136508	1136508	1045644	1494408	1129968
PEMS03 / 43	1136508	1136508	1092132	1485312	1094076
PEMS03 / 44	1136508	1136508	926784	1475136	1063992
PEMS08 F1 / 42	136596	157740	101712	101904	101712
PEMS08 F2 / 42	136392	152856	117228	101556	98676
PEMS08 F3 / 42	136140	151836	106464	111264	102420
PEMS08 F4 / 42	146352	157248	116196	110700	125676

The inherited sets were checked against the original GWN policies without modification. DCRNN exact and transferred GWN exact intersect in 1, 2 and 2 of 36 sources on PEMS03, and 16, 14, 15 and 13 of 17 sources on PEMS08. The intersections describe how source-set composition changes across fitted predictors. The PEMS03 variance-minus-Linear effects are negative in all 15 fitted-seed-by-mask summaries, while all 20 PEMS08 fold-by-mask summaries are positive. The counts summarize seed/mask and fold/mask coverage of the fixed evaluation; no post-result tuning or source reselection was applied.

The complete machine-readable data include all 29 absolute policy rows per case, per-mask metrics, Global/failed/tail errors, daily numerators and denominators, restored counts, every circular sensitivity and the exact 10,000 shared index arrays. Seven-case fresh-directory CUDA inference covered 35 case-mask combinations. All per-window identity/count fields and decompressed output contents matched exactly. This establishes recorded computational consistency in the same installed environment. Clean-environment installation, CPU numerical portability, new data acquisition and third-party full-pipeline replication remain unverified. Neither that audit nor the earlier timing probe is treated as scientific effect evidence.

S21. Separate METR-LA reference-42 DCRNN transfer

This extension follows the completed seven-case PEMS comparison in S20. It adds one seed-42 DCRNN fit on the previously exposed original METR-LA cohort, without retraining the proxies or altering any of the 27 inherited policy sets. A new DCRNN-specific exact-deletion set gives 28 distinct policy labels and sets. The original GWN fits and historical intervals, all seven PEMS DCRNN fits and the previously released figures remain unchanged. This section reports the initial retrospective single-fit transfer; S23 adds the subsequent reference seeds.

The exclusive fitting, calibration and validation boundaries are 21,591, 23,990 and 27,417. Window counts are 21,568, 2,388 and 3,416. The training-only finite-nonzero-speed scaler uses mean 58.62550693096417 mph and standard deviation 12.801036685068539 mph. Parameter fitting, calibration and checkpoint validation remain distinct. Each input and target sequence contains twelve five-minute steps. Earlier selection uses all 3,260 original dense windows; later evaluation uses 3,042 windows over twelve encoded first-target dates, retaining the original eleven excluded boundary windows. No PEMS stride-12 selection rule is imported into this case.

The DCRNN constructor and optimizer settings match S20: two 64-unit diffusion-GRU layers, order two, curriculum decay 2000, 374,273 parameters, four masked-value/availability/calendar channels, valid-target-weighted microbatches of four windows accumulated into sixteen-window updates, and four training missingness views. Adam, plateau scheduling, minimum thirty epochs, maximum 192 epochs, improvement threshold 0.00001 and patience 15 remain fixed. Training completed 89 epochs and selected epoch 74 by mask 1001. Validation MAE at that checkpoint was 2.951929 mph, compared with 5.494108 mph for same-mask persistence. This is a validation usability check, distinct from final restoration effects. Five-mask diagnostics did not change the checkpoint.

Masks 1001–1005 use 30% background missingness; the graph-block fault rate and source-eligibility budget are 10%, giving 21 of 207 sources. Eligibility sets for none, Linear, variance, Bounded, Linear gradient, GWN gradient, inherited GWN exact and twenty fixed random orders are unchanged. DCRNN exact deletion uses only earlier labels and all five scoring masks. The new model and all source sets were jointly frozen before decoding the later DCRNN evaluation. Final denominators are 35,845,775 valid targets and 3,622,771 failed-source targets per policy, across 15,210 window–mask occurrences. Both error scopes are in mph. The sensor-tail metric averages within-window source tails and is not target-pooled MAE.

Table S27. All 28 absolute policy rows from the separate METR-LA DCRNN evaluation. Every policy uses the same target denominators. Restored entries count repeated returned input occurrences. Individual random policies are fixed source orders, not new fitted-model replicates. Exact-reference denotes the new DCRNN-specific set; exact-GWN-transfer denotes the inherited original reference-42 set. No group-aware policy was added.

Saved policy label	Global MAE	Failed MAE	Tail metric	Restored entries
none	3.196830	3.703198	15.769678	0
Linear	3.169708	3.473992	15.492025	264633
variance	3.170085	3.478618	15.490492	266610
GWN_gradient_transfer	3.173129	3.508282	15.508392	256471
Bounded	3.169461	3.471483	15.487056	255881
random_00	3.181521	3.579227	15.659772	252675
random_01	3.182199	3.592773	15.672860	271555
random_02	3.179893	3.563727	15.645654	266273
random_03	3.183554	3.589025	15.655775	237877
random_04	3.176797	3.535858	15.594074	277989
random_05	3.180966	3.578583	15.657237	237602
random_06	3.182477	3.586688	15.664442	271138
random_07	3.176867	3.535310	15.612639	308865
random_08	3.176320	3.540394	15.595585	250566
random_09	3.183063	3.590174	15.640497	239311
random_10	3.182613	3.583461	15.649102	268714
random_11	3.182207	3.591320	15.645133	254247
random_12	3.177771	3.547417	15.591720	239201
random_13	3.176438	3.535214	15.586757	275999
random_14	3.173836	3.516588	15.548478	300144
random_15	3.181299	3.579398	15.644080	253540
random_16	3.182874	3.587844	15.655211	277529
random_17	3.182498	3.587336	15.670631	250445
random_18	3.182478	3.595176	15.660799	262180
random_19	3.182590	3.587091	15.668440	239160
exact_GWN_transfer	3.170488	3.478242	15.523990	267298
Linear_gradient	3.171591	3.493953	15.485573	276119
exact_reference	3.168144	3.462400	15.504834	264547

The primary interval uses 10,000 paired circular two-date resamples with seed 2026102242. Adjacent date positions wrap around the twelve-date sequence. An ordinary-date sensitivity uses 10,000 draws with seed 2026102243. Every draw shares its date indices and sampled valid-target denominators across policies and both reference configurations. DCRNN-minus-GWN changes subtract paired within-draw gains before taking percentiles. GWN results below use a new common-index reaggregation of archived daily sufficient statistics; the historical intervals in S1–S19 are retained separately. The unadjusted percentile intervals condition on the saved models and shared observations. The later blockwise label-exchange analysis is reported in S27.

Table S28. Paired contrasts in mph, using the saved subtraction order. Positive variance-minus-Linear favors Linear. Negative exact-DCRNN-minus-Linear favors the exact-DCRNN policy; negative exact-DCRNN-minus-inherited-exact favors the new exact-DCRNN set. The architecture-change row compares DCRNN and GWN fitted configurations by subtracting their variance-minus-Linear gains.

Contrast	Point	Circular2 conditional 95%	Ordinary-date sensitivity 95%
variance minus Linear	0.000377	-0.000849 to 0.001785	-0.001037 to 0.001933
DCRNN exact minus Linear	-0.001564	-0.003401 to 0.000331	-0.003755 to 0.000394
DCRNN exact minus transferred GWN exact	-0.002344	-0.003558 to -0.001159	-0.003710 to -0.001063
GWN variance minus Linear, common new indices	0.000868	-0.000450 to 0.002497	-0.000512 to 0.002461
DCRNN minus GWN change in variance-minus-Linear	-0.000491	-0.002002 to 0.000792	-0.001825 to 0.000715

Table S29. Every observed-date point and leave-one-date-out result for the three within-DCRNN contrasts, in mph. Date labels are frozen first-target date codes. V−L is variance minus Linear, E−L is DCRNN exact minus Linear, and E−G is DCRNN exact minus inherited GWN exact. Omission removes only that date's saved numerator and target count; it does not refit or reselect. Columns retain the same subtraction sign as Table S28.

Date code	V−L	Omit V−L	E−L	Omit E−L	E−G	Omit E−G
15507	0.002215	0.000257	0.002332	-0.001818	-0.000945	-0.002435
15508	0.005379	0.000248	-0.005232	-0.001470	0.003106	-0.002484
15509	-0.002006	0.000514	-0.006072	-0.001306	-0.003494	-0.002278
15510	0.001394	0.000267	-0.004431	-0.001253	-0.006794	-0.001860
15511	-0.000575	0.000481	0.000009	-0.001735	-0.000092	-0.002589
15512	0.000892	0.000322	0.002216	-0.001970	-0.000527	-0.002539
15513	0.005715	-0.000196	0.002315	-0.001980	-0.000881	-0.002501
15514	-0.001508	0.000574	-0.008934	-0.000795	-0.002879	-0.002288
15515	-0.002257	0.000654	0.001472	-0.001884	-0.004358	-0.002132
15516	0.001633	0.000247	-0.000670	-0.001656	-0.000041	-0.002581
15517	-0.003477	0.000781	-0.004449	-0.001262	-0.004658	-0.002101
15518	0.000317	0.000383	-0.000702	-0.001650	-0.002527	-0.002325

Table S30. All five mask-specific point contrasts in mph. Values pool each mask's full later-window error numerators and valid-target counts. Signs follow Table S28. These are repeated missingness conditions under the same fitted model and fixed sets, not five independent training replications.

Mask	Variance minus Linear	DCRNN exact minus Linear	DCRNN exact minus inherited exact
1001	-0.000316	-0.002112	-0.002707
1002	0.000386	-0.001961	-0.002224
1003	0.000486	-0.000760	-0.001378
1004	0.001905	0.000119	-0.002783
1005	-0.000574	-0.003107	-0.002628

Table S31. All observed-date and leave-one-date-out architecture-change points in mph. Each entry is the DCRNN-minus-GWN change in the variance-minus-Linear contrast. Dates, mask identities and valid-target counts are matched. This comparison conditions on the two fitted references and fixed policy sets, without isolating architecture from fitted weights or predictive accuracy.

Date code	DCRNN minus GWN change	Omit-date change
15507	-0.005468	-0.000166
15508	0.004249	-0.000613
15509	-0.001609	-0.000427
15510	0.000157	-0.000562
15511	-0.000005	-0.000544
15512	0.002822	-0.000848
15513	0.001126	-0.000665
15514	0.000751	-0.000621
15515	-0.000748	-0.000464
15516	0.000363	-0.000580
15517	-0.003999	-0.000124
15518	-0.002503	-0.000290

The Linear–variance increment is small and uncertain: 0.000377 mph, with the primary interval crossing zero. Three of five masks and seven of twelve date points favor Linear; removing one date reverses its pooled direction. The observed ordering varies across date omissions, with the temporal interval spanning zero. DCRNN-specific exact selection has lower point error than Linear, but its conditional gain interval also spans zero. Against the inherited GWN exact set, its lower error persists across all date omissions and the primary conditional interval excludes zero in the favorable direction. These are different comparator-specific conclusions. These comparisons retain policy-specific source identities and returned-entry counts. The paired DCRNN-minus-GWN change also has an interval spanning zero.

All five new-directory same-environment CUDA re-inferences were completed. Decompressed per-window content matched each original output exactly, including identity and count fields. Independently reaggregating saved daily numerators and direct sampled-date indices reproduced the reported intervals and omission points within 1e-10 mph. The replay verifies the recorded inference outputs and arithmetic within the same installed environment. Machine-readable CORE_RESULTS.json retains full precision, all sensitivities and the draw indices; ALL_28_POLICIES_ABSOLUTE.csv and MASK_CONTRASTS.csv provide every policy and mask contrast.

S22. Detailed tables retained from the main manuscript

These tables retain their original numerical precision and scientific scope. Moving them changes presentation only. Full-precision machine-readable source outputs remain unchanged.

Table S32. Separate METR-LA reference-42 DCRNN transfer under the original 10% graph-block fault and 21-source eligibility budget. Global and failed-source MAE are in mph. Rows pool the same 3,042 windows and five masks, with 35,845,775 valid targets per policy. Inherited exact-GWN and new exact-DCRNN selections use their respective earlier deletion scores. Restored occurrences count returned inputs across overlapping windows and masks; they do not measure independent acquisitions or physical recovery cost.

Policy	Global MAE	Failed-source MAE	Restored occurrences
No restoration	3.196830	3.703198	0
Variance	3.170085	3.478618	266,610
Linear	3.169708	3.473992	264,633
Bounded	3.169461	3.471483	255,881
Inherited GWN exact	3.170488	3.478242	267,298
DCRNN exact	3.168144	3.462400	264,547

Table S33. Original retrospective data roles and comparison units. Window counts precede expansion over the five shared masks. Both datasets had previous development exposure. The two METR-LA reference initializations share the original proxy fits and selected policies.

Dataset / measured quantity	Published input columns	Single-source analysis windows	Earlier selection windows	Later restoration windows / dates	Fixed GWN initializations	Primary fault / eligible sources
METR-LA / speed, mph	207	6,313	3,260	3,042 / 12	42, 43	21 / 21
PEMS08 / flow, vehicles per 5 min	170	3,561	1,545	2,005 / 7	42	17 / 17

Table S34. Information used to define restoration eligibility in the original GWN cases. All evaluated sets are fixed before the later comparison and all later prediction weights remain unchanged. Offline deletion and gradient scores require earlier target labels; inexpensive scoring is not necessarily label-free acquisition. The group constraint changes eligibility, without removing published columns from the predictor.

Policy	Ranking or selection input	Information period	Additional qualification
Linear / Bounded	Signed proxy complete-history deletion loss	Earlier selection period	Fixed dataset-specific proxy
Variance	Variance of valid source readings	Parameter-fitting prefix	No prediction target labels
Exact-reference deletion	Signed deletion loss of the corresponding GWN	Earlier selection period	Different exact sets for METR42 and METR43
Proxy-gradient magnitude	Value-channel Linear loss-gradient magnitude	Earlier selection period	Availability mask held fixed while scoring
Reference-gradient	Value-channel GWN loss gradient	Earlier selection period	Auxiliary policy in the original branch
Group-aware variance / Linear	Original score order, at most one column per exact-sequence group	Variance order and groups: fitting prefix; Linear order: earlier selection labels	Stops at 17 eligible columns; Linear is unchanged
Random	Prefix of each of 20 fixed random orders	Fixed before later evaluation	Mean and individual policies retained
Joint greedy, HJ	Actual joint Linear predictions under selection faults	Earlier METR-LA period	Evaluated only for the original primary reference-42 branch

S23. Additional METR-LA references and training-seed sensitivity

Two additional DCRNN reference-training seeds (43 and 44) and one non-graph LSTM reference (seed 42) use the original METR-LA reference-42 roles. No proxy is refitted. Earlier selection uses 3,260

windows, final evaluation 3,042 windows, masks 1001–1005 and a 21/207 eligibility budget. Each of 28 labels has 35,845,775 valid prediction targets and 3,622,771 fault-source targets across 15,210 window–mask cases. The 27 inherited sets are unchanged. All three new references and their own-exact sets were jointly frozen before any new final values were decoded.

DCRNN seeds change initialization, training order and training missingness together. The original seed 42 fit is retained. Three-seed means weight the saved fits equally in every shared temporal draw. LSTM uses two 128-unit unidirectional layers,828 inputs and a direct 12×207 output head, with 943,028 parameters; DCRNN has 374,273. Training settings, information boundaries, endpoints and model differences are described in Methods. No graph enters LSTM forecasting, although the same graph defines the synthetic fault.

Table S35. New reference configurations and selected training endpoints. All fits stopped by the original patience rule.

Reference	Seed	Parameters	Complete epochs	Selected epoch
LSTM	42	943028	47	32
DCRNN	43	374273	99	84
DCRNN	44	374273	93	78

Table S36. Complete saved policy outcomes for LSTM seed 42. Errors are in mph; returned occurrences count restored source-by-history-time entries and may repeat across windows. The exact_GWN_transfer label denotes the inherited GWN exact set, while exact_reference denotes the new fitted reference's own set.

Policy	Global MAE	Fault-source MAE	Tail MAE	Returned entries
none	3.65941006464	3.80863566417	17.5411625112	0
Linear	3.64552781856	3.7221190544	17.4447484091	264633
variance	3.64536274889	3.72799324262	17.4468370407	266610
GWN_gradient_transfer	3.64834336884	3.73880888254	17.4594432858	256471
Bounded	3.64600732778	3.72343923222	17.4434595671	255881
random_00	3.65077281522	3.77532845197	17.4915381051	252675
random_01	3.65248662424	3.773993354	17.5044372528	271555
random_02	3.65000335308	3.76905821978	17.4862346314	266273
random_03	3.65236238262	3.77386133779	17.4975975657	237877
random_04	3.64785093011	3.75848888851	17.4667774252	277989
random_05	3.65319170891	3.77834831041	17.5101983959	237602
random_06	3.65338713855	3.78960461203	17.5252252224	271138
random_07	3.64957082012	3.76868727152	17.4874134139	308865
random_08	3.6492878673	3.75501982363	17.4772138091	250566
random_09	3.65273482461	3.77147413133	17.5033700954	239311
random_10	3.6539479573	3.77729240102	17.506001079	268714
random_11	3.65271967175	3.78097851174	17.5013248013	254247
random_12	3.65203579418	3.76432233708	17.4888678409	239201
random_13	3.6494034138	3.75847420043	17.4864021478	275999
random_14	3.64875677249	3.74538730929	17.4740522951	300144
random_15	3.65326563994	3.77928767726	17.5120209574	253540
random_16	3.65191991763	3.77194375012	17.5017244281	277529
random_17	3.65385834034	3.78236711283	17.525251578	250445
random_18	3.65060451646	3.7799336713	17.4996491644	262180
random_19	3.65320484215	3.77710180329	17.5071722928	239160
exact_GWN_transfer	3.64396891582	3.71926121964	17.4496781255	267298
Linear_gradient	3.64893965509	3.73581300999	17.4634640211	276119
exact_reference	3.64498358095	3.73551371782	17.4586925103	251321

Table S37. Complete saved policy outcomes for DCRNN seed 43. Errors are in mph; returned occurrences count restored source-by-history-time entries and may repeat across windows. The exact_GWN_transfer label denotes the inherited GWN exact set, while exact_reference denotes the new fitted reference's own set.

Policy	Global MAE	Fault-source MAE	Tail MAE	Returned entries
none	3.19544315058	3.69871641957	15.756154692	0
Linear	3.16920679512	3.47479467698	15.4823170603	264633
variance	3.16927920978	3.47686370797	15.4822421387	266610
GWN_gradient_transfer	3.17264507402	3.51093095298	15.5060254602	256471
Bounded	3.16893321786	3.4721761407	15.4749332833	255881
random_00	3.17954347445	3.57259382821	15.6448913476	252675
random_01	3.18155321986	3.5943470652	15.6567022677	271555
random_02	3.17716894646	3.55255983322	15.6174434057	266273
random_03	3.18187639499	3.58312442962	15.6393115036	237877
random_04	3.17528433498	3.52984644871	15.5776670604	277989
random_05	3.18030540185	3.58111627779	15.6511336715	237602
random_06	3.18174157449	3.59186262808	15.6580468394	271138
random_07	3.17585040311	3.53633177114	15.6027722827	308865
random_08	3.17527866017	3.5384691903	15.5810928275	250566
random_09	3.18147091059	3.58544078459	15.6233851271	239311
random_10	3.18158259827	3.58309481975	15.6401188403	268714
random_11	3.18083503976	3.58637612798	15.6339990708	254247
random_12	3.17658878609	3.54401452669	15.5839242485	239201
random_13	3.17377921948	3.51745562093	15.5598008384	275999
random_14	3.17346373463	3.51804662939	15.5380817688	300144
random_15	3.18019342801	3.57491075921	15.6303350353	253540
random_16	3.18148680586	3.58548321881	15.644147495	277529
random_17	3.18110983007	3.58392496625	15.6587707862	250445
random_18	3.18116590784	3.58961883401	15.6494885243	262180
random_19	3.18081443856	3.58419501485	15.6507289615	239160
exact_GWN_transfer	3.16976738091	3.48266489757	15.5197245389	267298
Linear_gradient	3.17135153898	3.49609224471	15.4875811452	276119
exact_reference	3.16802262497	3.46687573235	15.4940803249	243900

Table S38. Complete saved policy outcomes for DCRNN seed 44. Errors are in mph; returned occurrences count restored source-by-history-time entries and may repeat across windows. The exact_GWN_transfer label denotes the inherited GWN exact set, while exact_reference denotes the new fitted reference's own set.

Policy	Global MAE	Fault-source MAE	Tail MAE	Returned entries
none	3.18726365722	3.69464447871	15.6275603815	0
Linear	3.16066179622	3.46679925448	15.361481326	264633
variance	3.16099189967	3.47211888448	15.3607446378	266610
GWN_gradient_transfer	3.16332625205	3.49881466919	15.3703228686	256471
Bounded	3.1599822273	3.46227262511	15.3513596189	255881
random_00	3.16974770698	3.55630347162	15.4958159524	252675
random_01	3.17351831832	3.59371249629	15.5424772926	271555
random_02	3.1685237622	3.54389135882	15.4911984991	266273
random_03	3.17336632423	3.57595921437	15.5124875315	237877
random_04	3.16605808722	3.51770552187	15.441262889	277989
random_05	3.17192321477	3.57560861489	15.5174084965	237602
random_06	3.17336496006	3.58350040387	15.5231299037	271138
random_07	3.16747121326	3.53115014704	15.478878998	308865
random_08	3.1663387084	3.52736220653	15.4538114206	250566
random_09	3.17327052568	3.5815037073	15.5013936117	239311
random_10	3.17348138858	3.57924188951	15.5115671644	268714
random_11	3.17218925412	3.57578779924	15.5024909638	254247
random_12	3.16832432153	3.54022980864	15.4590010736	239201
random_13	3.16550556286	3.51365046744	15.4336946322	275999
random_14	3.16539808724	3.51398937012	15.4207097169	300144
random_15	3.17270232986	3.58033552144	15.5141285824	253540
random_16	3.17251584178	3.57218799952	15.5072883804	277529
random_17	3.17304296692	3.58183553455	15.5303960563	250445
random_18	3.17345635036	3.59011505366	15.5251627391	262180
random_19	3.17159959005	3.5703498076	15.5119574601	239160
exact_GWN_transfer	3.16055883686	3.46569424827	15.3800003352	267298
Linear_gradient	3.1623443553	3.48704918382	15.362161544	276119
exact_reference	3.16175281006	3.47402569075	15.3915659057	257458

Table S39. Paired contrasts for LSTM seed 42 in mph. Intervals use 10,000 common circular two-date draws or ordinary-date draws over twelve observed dates. Negative exact-minus-Linear favors exact; positive variance-minus-Linear favors Linear. These intervals condition on the saved fit, source sets and dates.

Contrast	Point	Circular 2 lower	Circular 2 upper	Ordinary lower	Ordinary upper
variance minus Linear	-0.00016506966	-0.0011105222	0.00095490456	-0.00100833	0.00076612196
Own exact minus Linear	-0.00054423761	-0.0039877141	0.0029356681	-0.0034953457	0.0023112956
Own exact minus GWN exact	0.0010146651	-0.0027068422	0.0052511912	-0.0022347621	0.0045495784
Change in variance-minus-Linear vs original GWN42	-0.0010334847	-0.0019642926	-7.9965431e-05	-0.0020729592	-1.8952374e-05

Table S40. Paired contrasts for DCRNN seed 43 in mph. Intervals use 10,000 common circular two-date draws or ordinary-date draws over twelve observed dates. Negative exact-minus-Linear favors exact; positive variance-minus-Linear favors Linear. These intervals condition on the saved fit, source sets and dates.

Contrast	Point	Circular 2 lower	Circular 2 upper	Ordinary lower	Ordinary upper
variance minus Linear	7.2414654e-05	-0.0010697347	0.0018654379	-0.0011663331	0.0017699059
Own exact minus Linear	-0.0011841701	-0.0025306006	8.3951748e-06	-0.0025492617	0.0001234808
Own exact minus GWN exact	-0.0017447559	-0.0026544992	-0.00042913881	-0.0028023722	-0.00037689733
Change in variance-minus-Linear vs original GWN42	-0.00079600042	-0.0015419359	2.0862147e-05	-0.0017262107	0.00014837677

Table S41. Paired contrasts for DCRNN seed 44 in mph. Intervals use 10,000 common circular two-date draws or ordinary-date draws over twelve observed dates. Negative exact-minus-Linear favors exact; positive variance-minus-Linear favors Linear. These intervals condition on the saved fit, source sets and dates.

Contrast	Point	Circular 2 lower	Circular 2 upper	Ordinary lower	Ordinary upper
variance minus Linear	0.00033010345	-0.0012769578	0.0029393132	-0.0014095098	0.0025244821
Own exact minus Linear	0.0010910138	-0.0014118917	0.004659281	-0.001379494	0.0042869358
Own exact minus GWN exact	0.0011939732	-0.00058108038	0.0031481238	-0.00040432024	0.0029234084
Change in variance-minus-Linear vs original GWN42	-0.00053831162	-0.0018003237	0.0010449323	-0.0017683757	0.00087730581

Table S42. Equal-weight DCRNN means over reference-training seeds 42–44, in mph. Shared date draws preserve policy and seed alignment.

Contrast	Point	Circular 2 lower	Circular 2 upper	Ordinary lower	Ordinary upper
variance minus Linear	0.00025987833	-0.00094148136	0.0019908775	-0.0011147542	0.0018764469
Own exact minus Linear	-0.00055238156	-0.0020601419	0.0012023808	-0.0021714538	0.001071645
Own exact minus GWN exact	-0.00096483831	-0.0017726819	0.00010062692	-0.0018273789	-8.5740686e-05

The primary three-seed mean is positive on six of the twelve dates; leave-one-date-out points range from -0.00020237273 to 0.00059975408 mph. Exact-minus-Linear changes direction across individual seeds: negative for 42/43, positive for 44. The exact-versus-transferred-GWN-exact mean has a circular 2 interval spanning zero, while its ordinary-date sensitivity excludes zero. Both resampling schemes are reported. LSTM likewise gives small policy differences and a reversal between the own-exact versus Linear and own-exact versus transferred-GWN-exact comparisons.

Machine-readable tables retain every absolute row, every new seed/mask contrast, each date and each leave-one-date-out point: ALL_84_NEW_CASE_POLICY_OUTCOMES.csv, MASK_CONTRASTS.csv, DCRNN_THREE_SEED_POINTS.csv, DCRNN_THREE_SEED_ABSOLUTE.csv and ALL_DATES_AND_OMISSIONS.csv. CORE_RESULTS.json retains all contrasts and sensitivities, and SHARED_RESAMPLING_INDICES.npz retains the actual date-index draws. These files are included under evidence/additional_references_20261023/statistics in the manuscript package. Fifteen separate-

directory CUDA inference outputs match the corresponding original outputs after decompression. The numerical consistency check is distinct from independent third-party inference or training-population uncertainty.

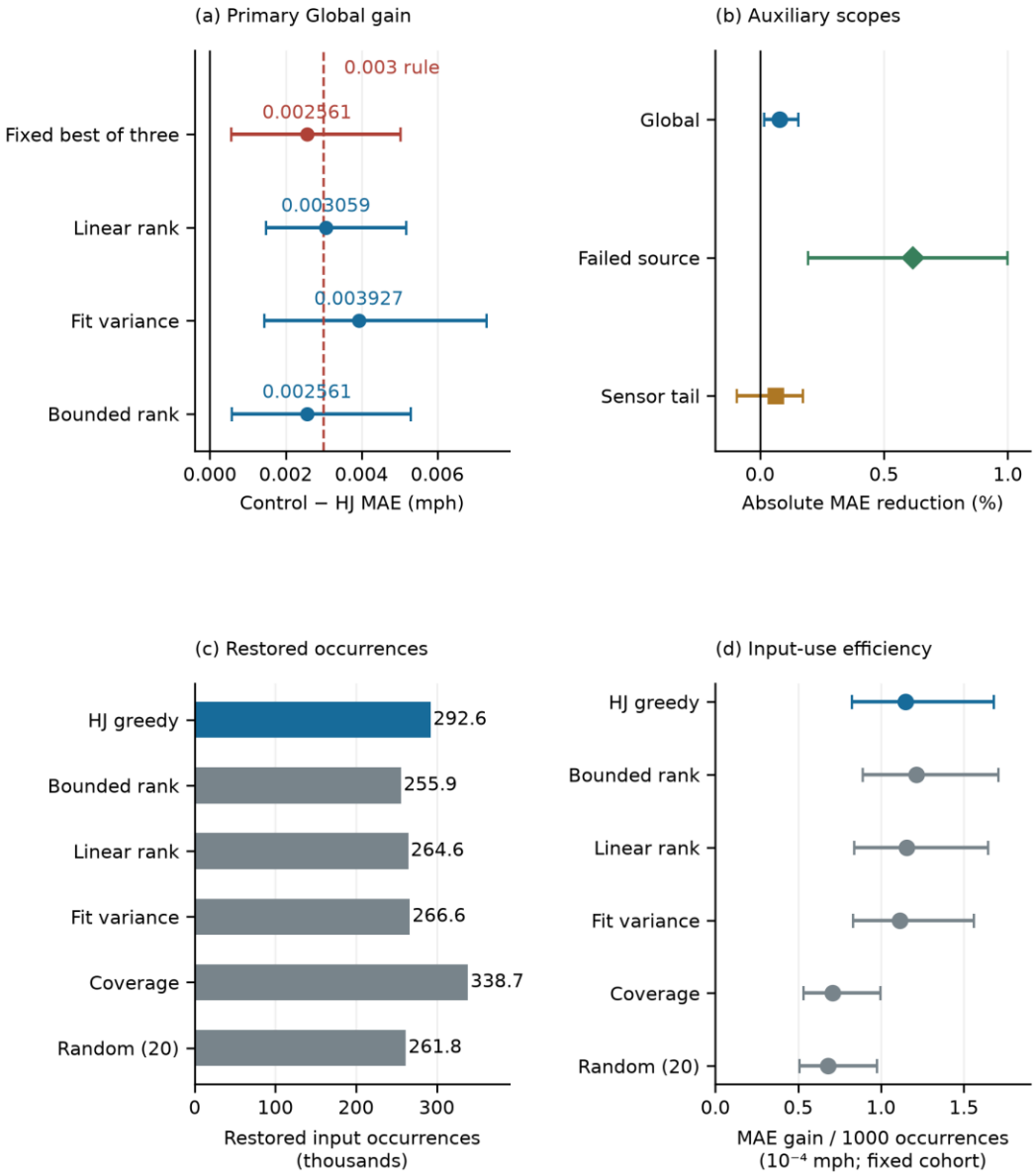


Figure S1. Joint greedy primary gain and original minimum-effect decision. The reused F20 display includes Global gain versus the original control family, secondary error scopes, restored occurrences and a descriptive gain-per-occurrence ratio. The Global interval is positive but the point gain is below 0.003 mph. The ratios describe each policy's realized returned entries; HJ conditions without evaluation are marked not run. Model, windows and masks match the original METR reference-42 primary experiment.

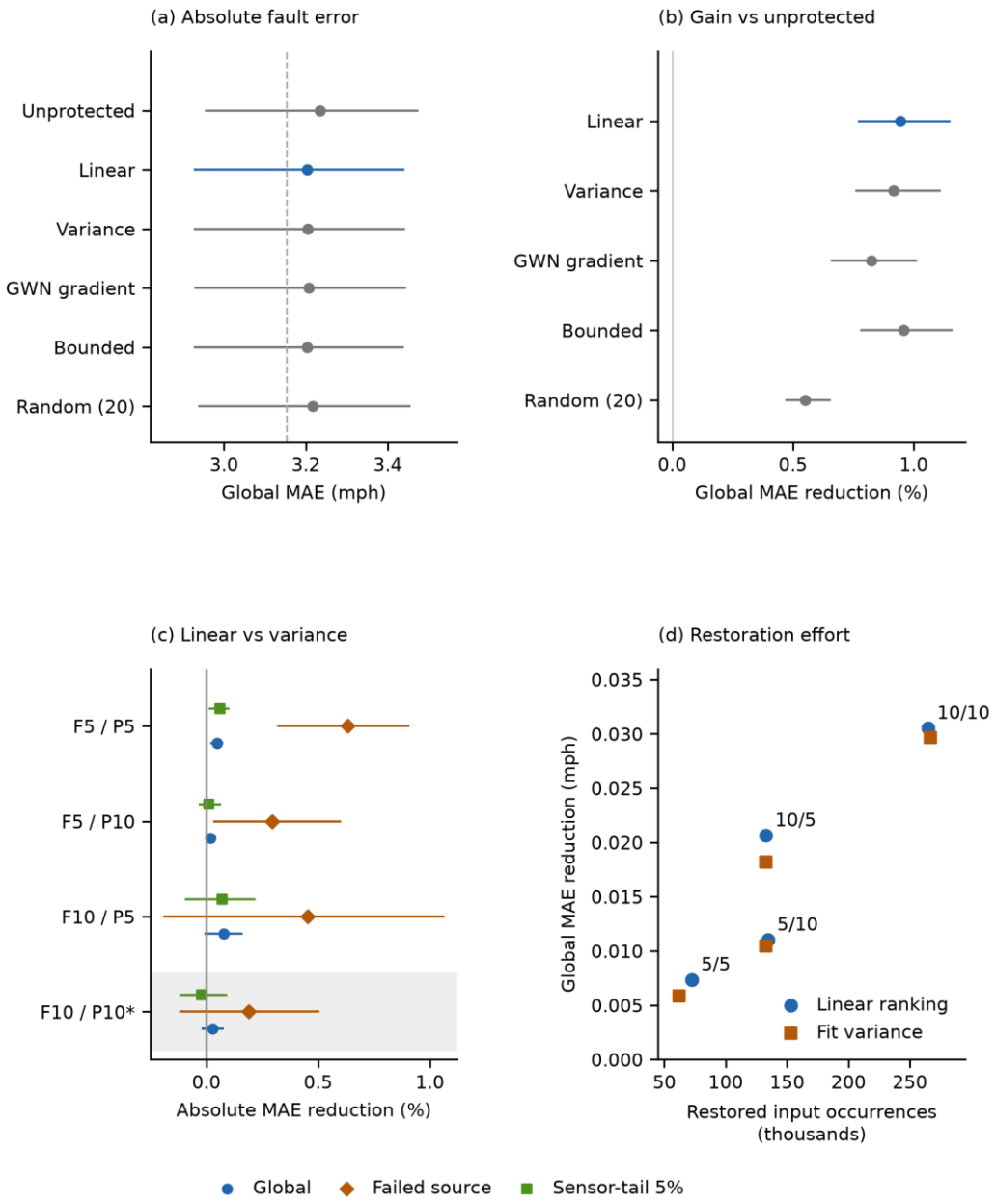


Figure S2. METR joint restoration under shared graph-block faults. Reused F19 shows absolute primary errors, gains relative to no restoration, all four fault/budget conditions and three error scopes, and restoration occurrences. The original Linear-versus-variance primary interval spans zero. The 20 random orders vary the eligible source set under fixed model weights. Returned-reading counts vary across equal-size source sets. The nominal point evaluates the base mask before the additional block fault.

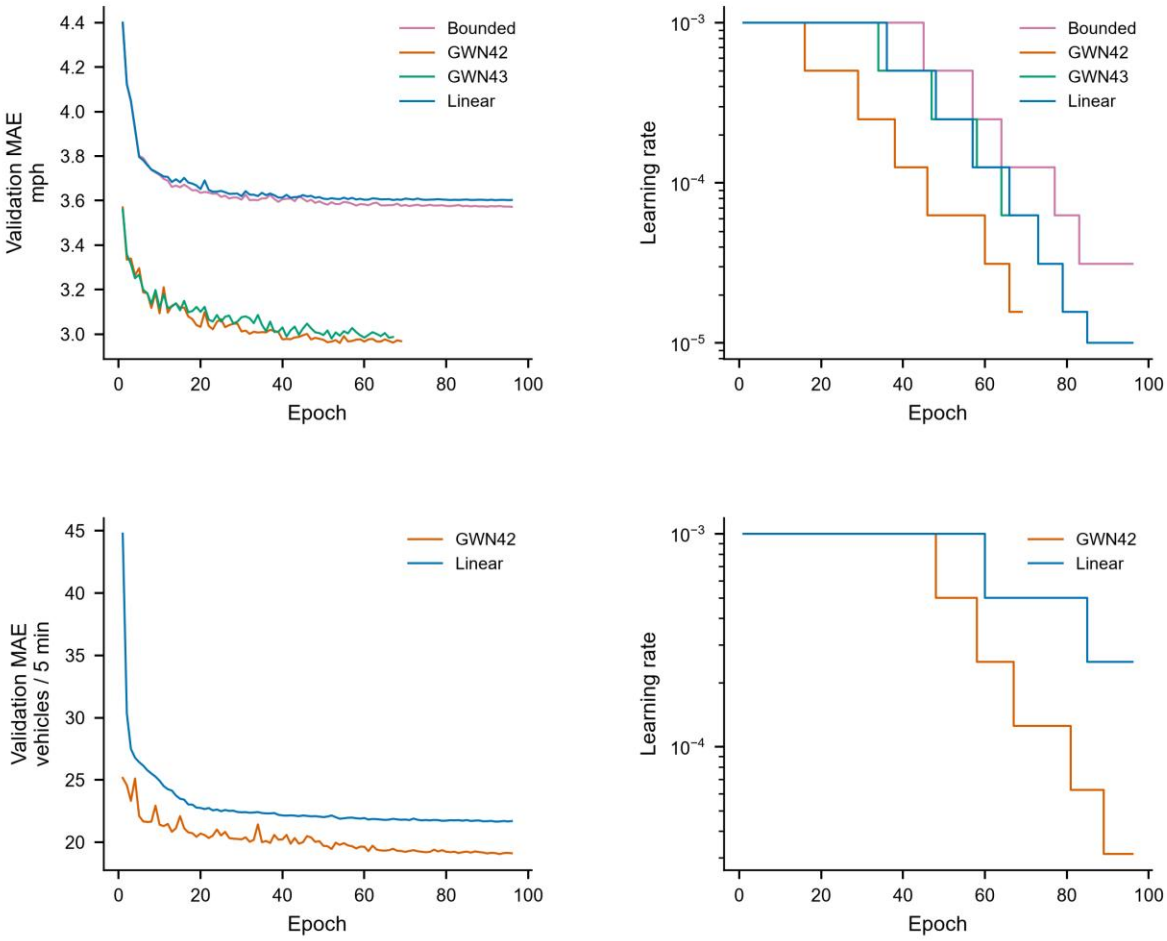


Figure S3. Logged validation trajectories and learning-rate schedules for the six relevant fitted predictors. The top row is METR and the bottom row PEMS, with dataset-specific MAE units. Learning rates are displayed separately on a log scale. Curves use archived epoch records without new training. Proxy training objectives, where available, are retained separately in CSV; a matching GWN training-loss history was not logged. The best/last checkpoint analysis in S11 and Table S8 evaluates ranking and restoration sensitivity separately from the validation trajectories.

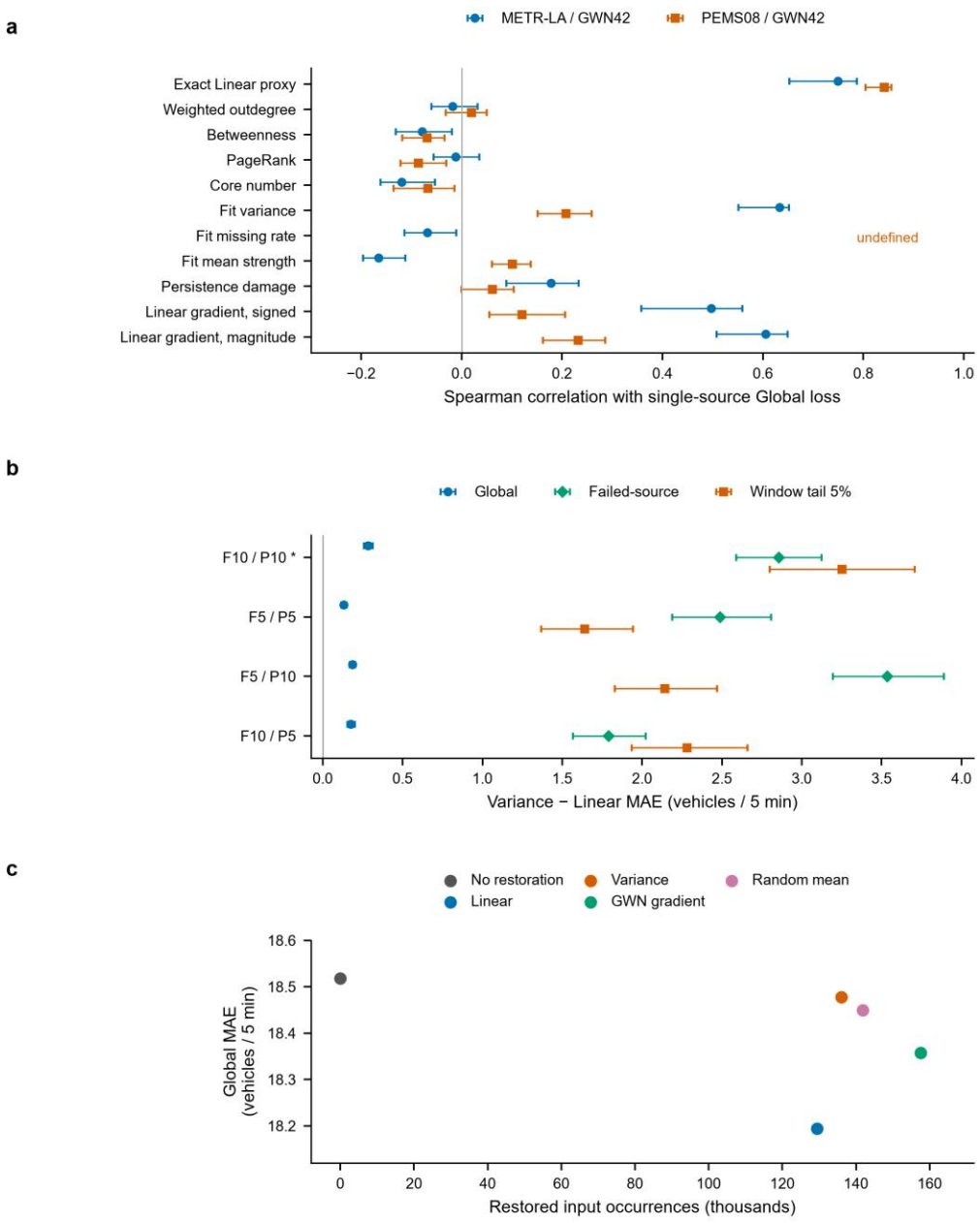


Figure S4. Ranking associations in both datasets and PEMS08 restoration utility. Panel a retains the original full-period correlations for the finite inexpensive-score family and Linear proxy; it is not the early-to-late evaluation in Supplementary Figure S9. PEMS08's constant missing-rate score is undefined. Panel b shows all four PEMS fault/budget conditions and three error scopes; positive values favor Linear over variance, with the primary condition marked. Panel c plots absolute Global MAE against cumulative restored input occurrences, with no gain-per-reading ratio or second y-axis. Counts represent repeated returned-input occurrences within one fixed cohort. Source values and original conditional intervals are unchanged. Supplementary Figure S9 and Supplementary Table S12 provide the temporally aligned diagnostics; Figure S6 shows every observed date contrast.

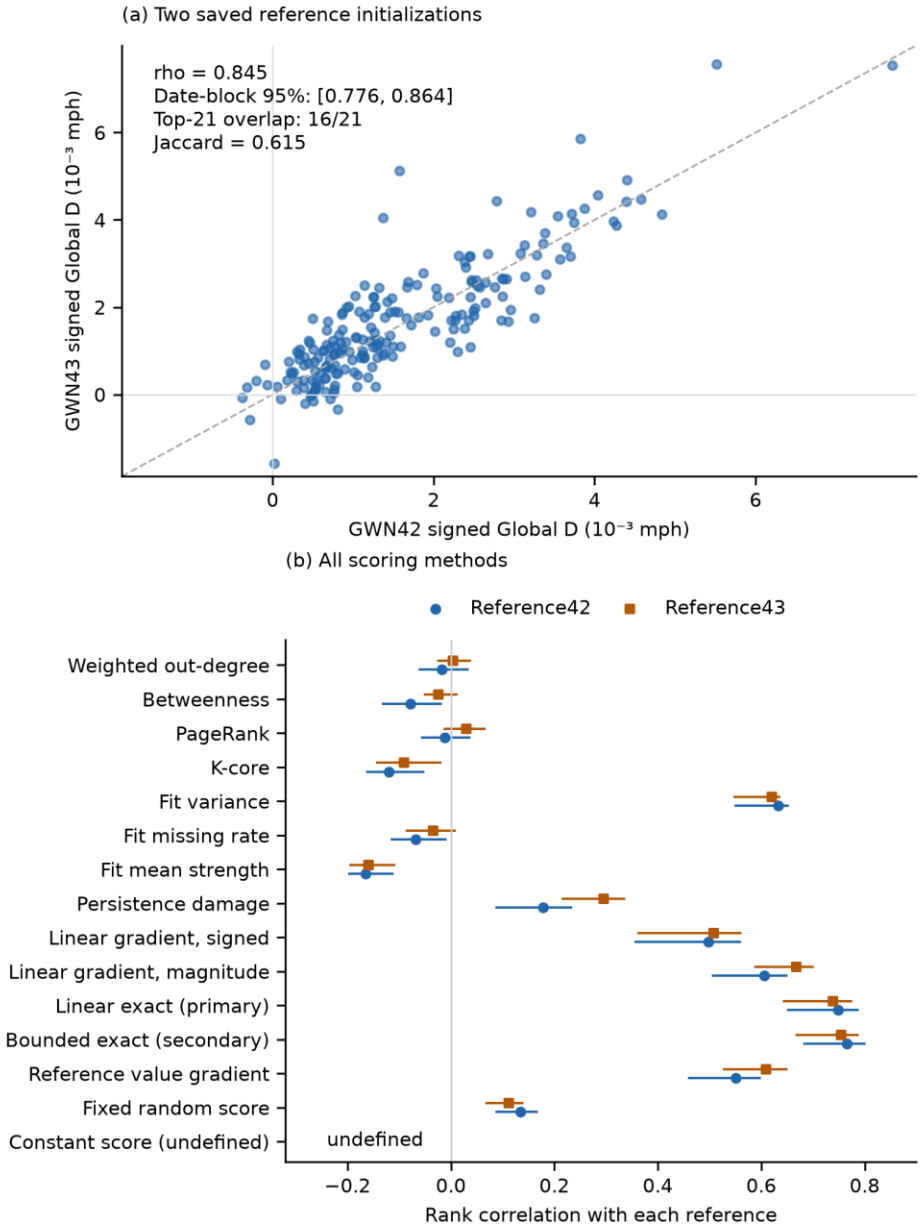


Figure S5. Signed single-source loss and ranking consistency on METR-LA. Panel a compares Global deletion-loss changes for the two saved Graph WaveNet references over 207 sources; negative values remain in the analysis and axes use 0.001 mph. Panel b retains all primary inexpensive scores and auxiliary references with 2,000 paired-date percentile intervals. Constant-score correlation is undefined. Linear is a nonlinear recurrent proxy with an additive graph correction. The cohort comprises 6,313 windows and five masks; intervals condition on the saved fitted weights. Source data: Fig1_artist_data.json preserves both panels; all_methods_intervals.csv gives original reference-42 diagnostics; S2_all_source_scores.csv records the second-reference source comparison.

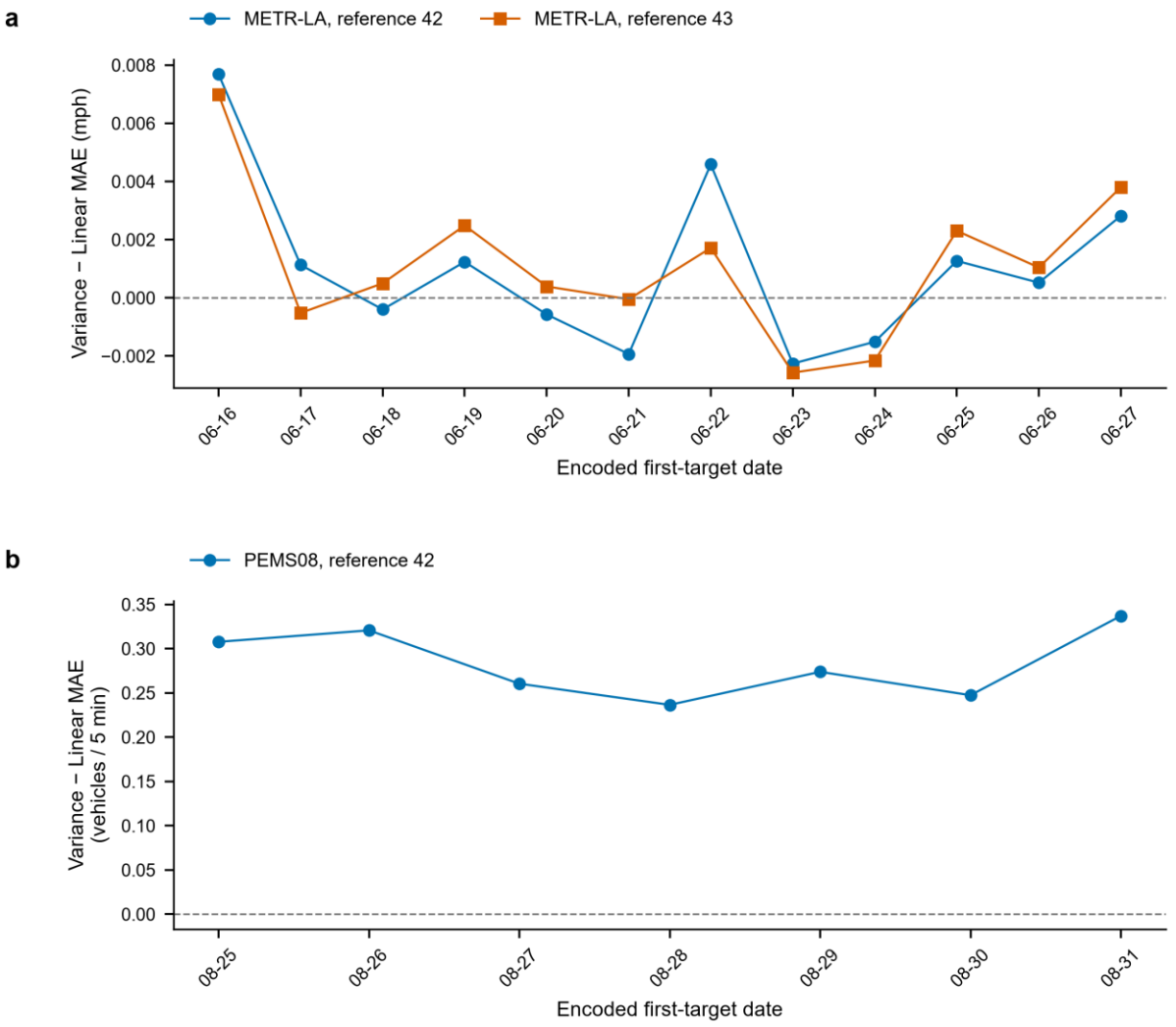


Figure S6. Every observed later-date Linear-versus-variance Global MAE contrast. Panel a displays the 12 METR dates under the two fixed references; panel b displays all seven PEMS dates in its own flow unit. Positive values favor Linear. Lines connect the observed encoded dates for readability. The figure uses the original per-date numerator/count pairs and reports point contrasts. The supplied DATE_WISE_PRIMARY_GAINS.csv includes each date's valid-target count. Shared mask and overlapping-window dependence remain. All PEMS date contrasts are positive, so their positive-weight resampled sums retain a positive sign.

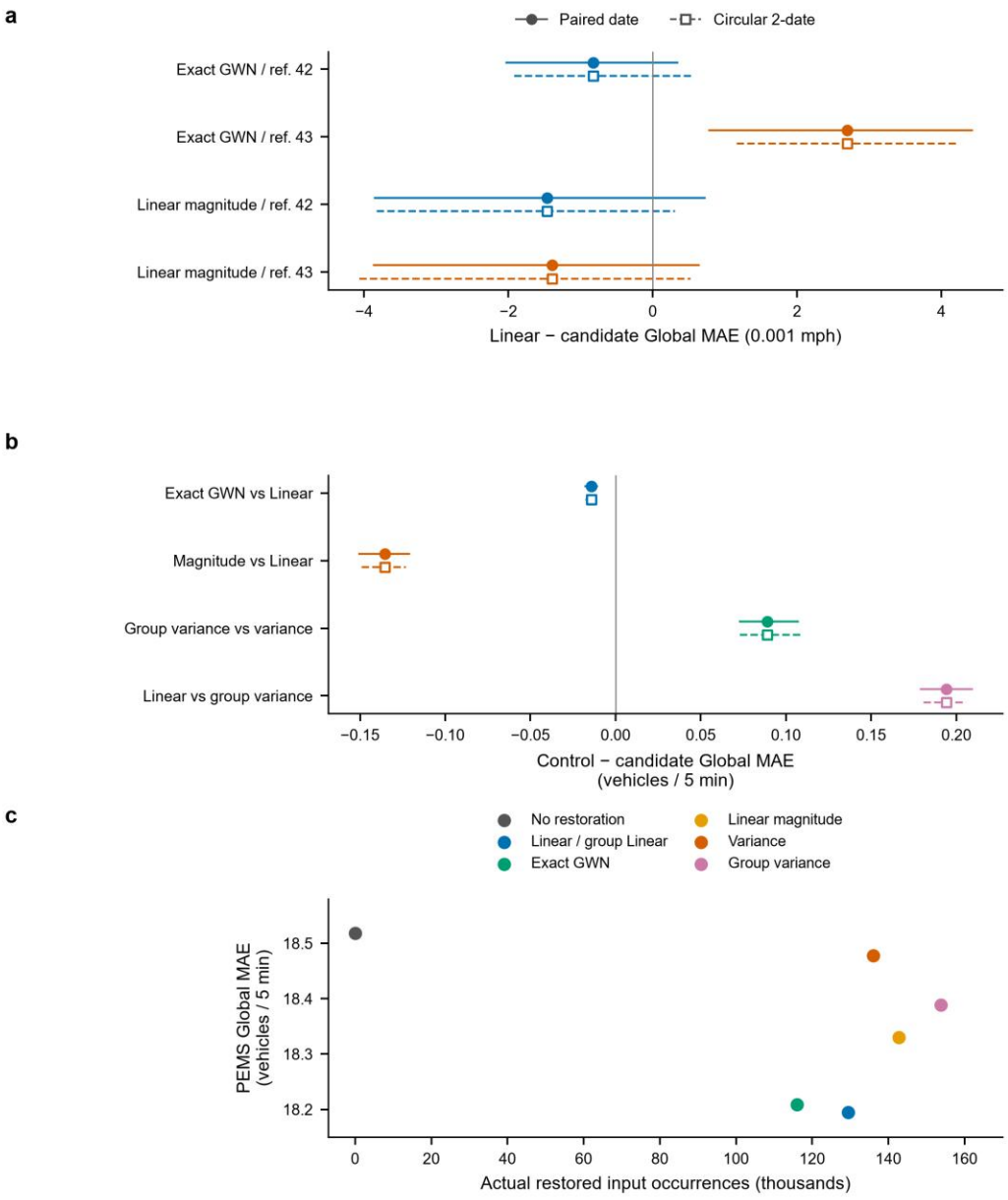


Figure S7. Direct restoration controls under the primary fault/budget condition. Panel a reports gains of the early exact-GWN and Linear-gradient-magnitude policies over Linear under both METR references; panel b gives the corresponding PEMS control and group-aware contrasts in flow units. Positive values favor the named candidate. Filled points/solid intervals are paired-date estimates; open points/dashed intervals are circular-two-date sensitivities from the same observed dates. Panel c shows absolute PEMS Global MAE against actual restored occurrences for original and new policies. All eligibility sets use earlier scores or fitting-only groups. Exact-GWN policies differ by reference. Group-aware Linear has the same eligible set and outcomes as Linear and is shown as an alias. Full values are in ABSOLUTE_PRIMARY.csv and PAIRED_PRIMARY_CONTRASTS.csv.

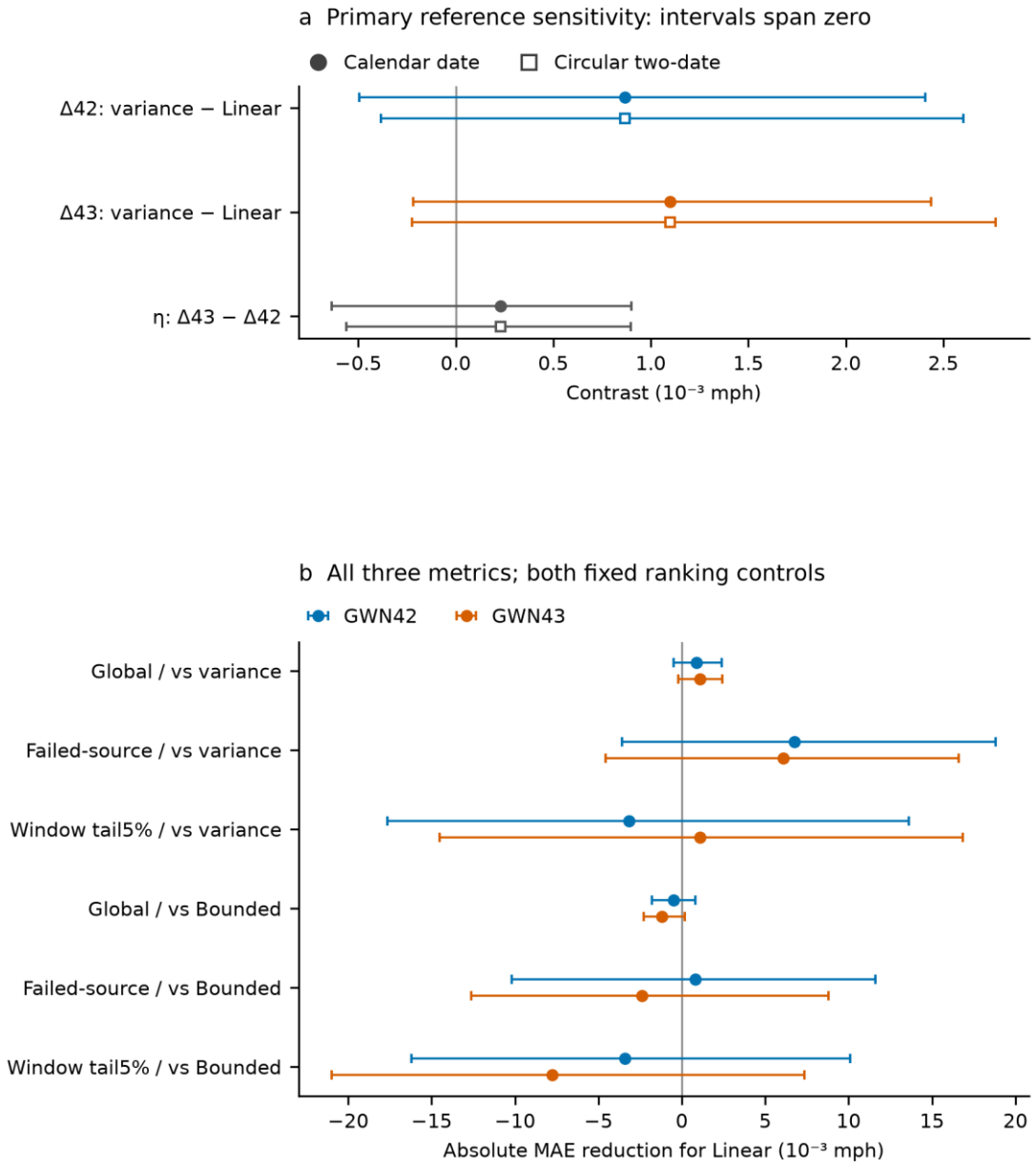


Figure S8. METR-LA primary restoration contrasts under two fixed references. Panel a shows Linear-versus-variance gains and their between-reference difference using common-date resampling. Filled circles and open squares show calendar-date and circular-two-date intervals, respectively. Panel b retains Global, original-failed-source and window sensor-tail contrasts against variance and Bounded. Axes use 0.001 mph. The two Linear-versus-variance Global intervals and their between-reference contrast span zero; HJ and PEMS08 outcomes are reported separately in their own comparisons. The common-date reference-42 intervals differ from the original draws and are labeled separately. Four policies, 3,042 later windows, five masks and 12 dates are retained. Source data: ALL_24_ABSOLUTE_ROWS.csv, ALL_36_PAIRED_CONTRASTS.csv and METR_PRIMARY_DATE_INPUTS.npz.

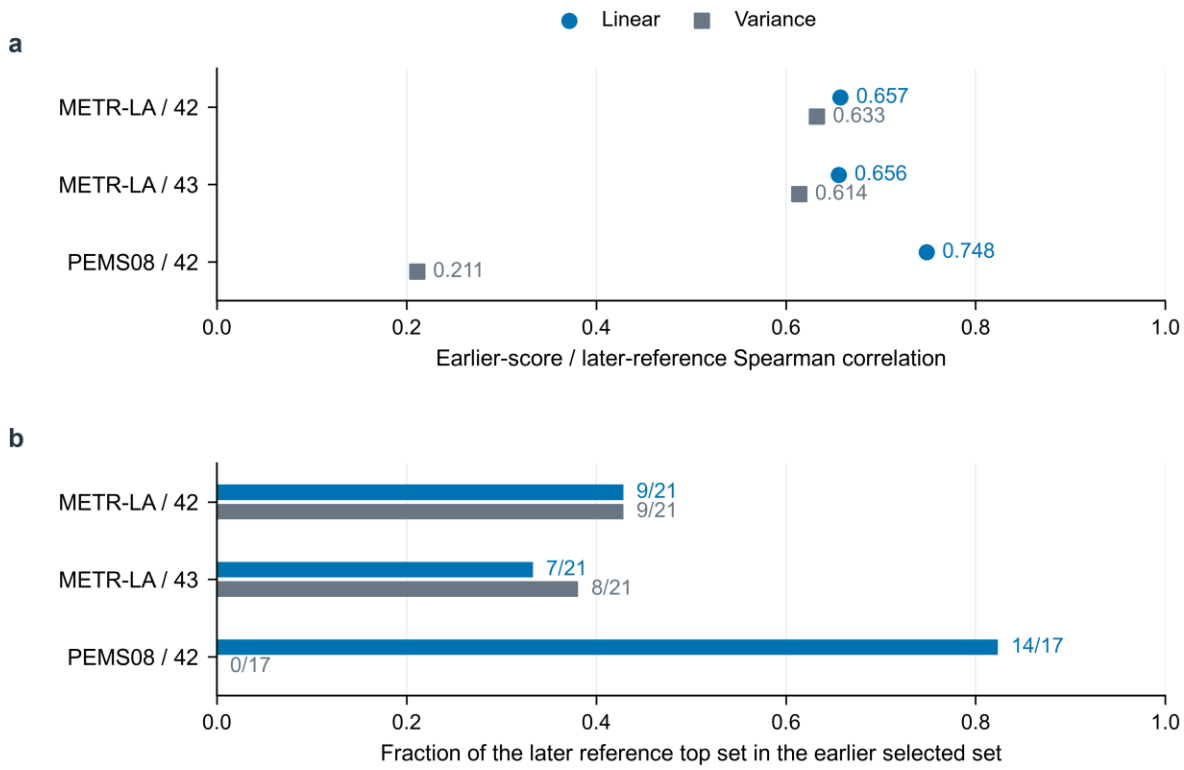


Figure S9. Earlier rankings and budget-relevant source correspondence. Panel a compares earlier Linear and fitting-variance scores with later GWN deletion losses. Panel b shows the fraction of the later reference top set contained in each earlier selected set, with intersection counts and set sizes labeled. Primary eligibility sizes are 21 for METR-LA and 17 for PEMS08. Later reference sets are diagnostic only and never select an evaluated restoration policy. Points and bars report the saved descriptive estimates.

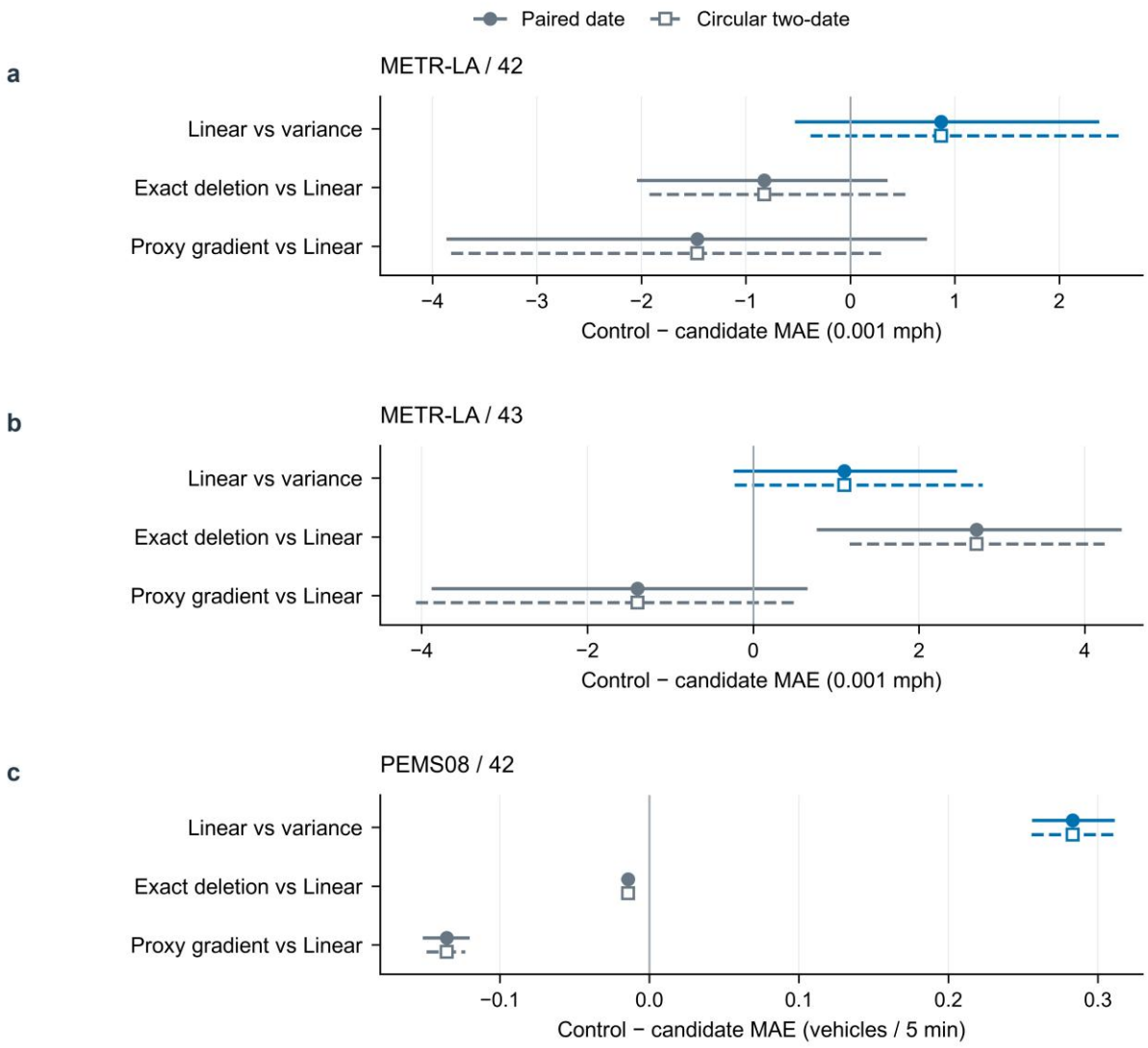


Figure S10. Direct primary restoration-policy contrasts. Panels a and b show METR-LA references 42 and 43 in units of 0.001 mph; panel c shows PEMS08 in vehicles per five minutes. Each row names candidate versus control; positive gain favors the candidate. Exact deletion denotes exact-reference selection, and proxy gradient denotes proxy-gradient selection. Filled circles and solid intervals use paired-date resampling; open squares and dashed intervals use circular two-date resampling. The two styles summarize the same 12 METR-LA or seven PEMS08 dates and fixed models. The original METR42 Linear–variance date interval is retained, while its separately labeled circular sensitivity uses common-date reaggregation. Other primary Linear–variance values use their original analysis, and added-policy rows use the added-control draws. Each policy retains its own source identities and realized returned-entry count.

S24. Original absolute restoration outcomes

Table S43. Absolute Global mean absolute error (MAE) and restored input occurrences under the primary 10% fault and eligibility budgets. METR-LA errors are in mph; PEMS08 errors are in vehicles per five minutes. A dash denotes an unevaluated branch. Occurrences include repeated entries across overlapping windows and masks; the random row is the mean of 20 fixed policies. Reference-gradient is the auxiliary GWN value-gradient policy; proxy-gradient uses Linear value-gradient magnitude. Exact-reference selection uses each predictor's earlier deletion scores. Contrasts in the text use unrounded errors; rounded MAEs may coincide. Group-aware Linear is an alias of Linear. Paired METR entry counts are ordered as reference 42 / reference 43; a single count applies to both references when the selected set is identical. For METR-LA reference 42, the MAE reduction from no restoration to Linear is 0.0305246884098298 mph, with original paired-date 95% endpoints 0.0243985695672933 and 0.0379728244170909; the provenance paragraph in this section identifies the source record and 2,000-draw calculation.

Exact METR-LA reference-42 provenance for the Linear-versus-no-restoration result: no-restoration Global MAE 3.23420746036031 mph, Linear Global MAE 3.20368277195048 mph, and unrounded gain 0.0305246884098298 mph. The original 2,000-draw paired-date 95% percentile interval is [0.0243985695672933, 0.0379728244170909] mph. These are the original retained draws, rather than the later common-date sensitivity. Source: the Zenodo release path original_release_1.6.1/original_gwn/supporting_records/S3_SUMMARY.json, primary fault_rate = protection_rate = 0.1 grid, metric_rows with scope GlobalMAE and policies linear/unprotected; the interval is gain_vs_unprotected_mph_temporal95. The accompanying local audit extracts this record and its SHA-256 identity.

Policy	**METR-LA ref. 42 MAE**	**METR-LA ref. 43 MAE**	**METR-LA restored entries**	**PEMS08 ref. 42 MAE**	**PEMS08 restored entries**
No restoration	3.23	3.27	0	18.5	0
Linear	3.2	3.24	264,633	18.2	129,432
Variance	3.2	3.25	266,610	18.5	136,032
Bounded	3.2	3.24	255,881	—	—
Exact-reference	3.2	3.24	267,298 / 275,405	18.2	115,980
Proxy-gradient	3.21	3.25	276,119	18.3	142,692
Reference-gradient	3.21	—	256,471	18.4	157,560
Random mean	3.22	—	261,750.5	18.4	141,771.6
Group-aware variance	—	—	—	18.4	153,804

S25. Post hoc correlation-based relevance–redundancy selection

This extension evaluates six fitting-only source-set pairs on 18 saved predictors, without retraining. It adds two policies under five masks per reference (180 new policy–mask outcomes). The original three GWN descriptive cases are outside this extension. The selection criterion, valid-pair handling, budgets and conditional resampling are specified in the main Methods. Correlation-based mRMR-style selection is a task-specific approximation, not mutual-information MID/MIQ. All outcomes, including worsening under redundancy penalization, are reported. Supplementary Data 1 contains every new policy–mask outcome; the additions component of the unified public archive includes per-window outputs, frozen sets, scores and source-code mappings.

Table S44. Absolute Global MAE in the post hoc comparison. PEMS03 and METR-LA DCRNN average three reference MAEs equally; rolling PEMS08 pools target numerators and counts. LSTM uses one reference. Flow units are vehicles per five minutes; speed units are mph.

Reference / cohort	None	Linear	Variance	Relevance only	Corr. mRMR	Exact
GWN / PEMS03	17.31801	17.1877	17.23977	17.25055	17.23999	17.15831
DCRNN / PEMS03	25.2597	24.83913	24.50919	24.15197	24.80069	23.83705
GWN / PEMS08	18.86433	18.60376	18.81329	18.85026	18.77016	18.60024
DCRNN / PEMS08	18.90081	18.66427	18.8192	18.85366	18.80139	18.64993
DCRNN / METR-LA	3.193179	3.166526	3.166786	3.173173	3.183892	3.165973
LSTM / METR-LA	3.65941	3.645528	3.645363	3.648962	3.654509	3.644984

Table S45. Pooled conditional contrasts for the new correlation-based policy. All differences are correlation-based mRMR-style MAE minus the named control; negative favors the new policy. Intervals use the new fixed shared 10,000 circular-block draws and condition on saved predictors and sets. Original paper intervals are unchanged. PEMS units are vehicles per five minutes; METR-LA units are mph.

Reference / cohort	Control	Difference	Lower	Upper
GWN / PEMS03	Linear	0.05228757	0.04757753	0.05708208
GWN / PEMS03	variance	0.0002166065	-0.01098515	0.01227733
GWN / PEMS03	relevance_only	-0.01056187	-0.01976079	0.004100017
GWN / PEMS03	group_variance	0.0002166065	-0.01098515	0.01227733
GWN / PEMS03	exact_reference	0.08168495	0.07757344	0.08551908
GWN / PEMS08	Linear	0.1664026	0.1493221	0.1839168
GWN / PEMS08	variance	-0.043126	-0.06018204	-0.02703762
GWN / PEMS08	relevance_only	-0.0800992	-0.09823893	-0.06278674
GWN / PEMS08	group_variance	0.02992252	0.01359327	0.04658326
GWN / PEMS08	exact_reference	0.1699197	0.153842	0.1867308
DCRNN / PEMS03	Linear	-0.0384448	-0.04880616	-0.02727179
DCRNN / PEMS03	variance	0.2915023	0.2655897	0.3226653
DCRNN / PEMS03	relevance_only	0.6487212	0.6077758	0.6825248
DCRNN / PEMS03	group_variance	0.2915023	0.2655897	0.3226653
DCRNN / PEMS03	exact_reference	0.9636406	0.8944094	1.024467
DCRNN / PEMS08	Linear	0.1371161	0.1203547	0.1540744
DCRNN / PEMS08	variance	-0.0178076	-0.0321937	-0.003922739
DCRNN / PEMS08	relevance_only	-0.05227241	-0.06572708	-0.03875863
DCRNN / PEMS08	group_variance	0.02091068	0.008347801	0.03366572
DCRNN / PEMS08	exact_reference	0.1514537	0.134786	0.1682384
DCRNN / METR-LA	Linear	0.01736602	0.0123171	0.02506145
DCRNN / METR-LA	variance	0.01710615	0.01272211	0.02326591
DCRNN / METR-LA	relevance_only	0.01071901	0.007851636	0.01495218
DCRNN / METR-LA	exact_reference	0.01791841	0.01287289	0.02479954
LSTM / METR-LA	Linear	0.008981654	0.00658902	0.01321597
LSTM / METR-LA	variance	0.009146723	0.007006965	0.01245853
LSTM / METR-LA	relevance_only	0.005547137	0.003269377	0.008481902
LSTM / METR-LA	exact_reference	0.009525891	0.006333759	0.01326689

Table S46. Every saved-reference point comparison. MAE units follow each cohort. Differences are correlation-based selection minus control; negative favors correlation-based selection. Returned entries count window–mask input occurrences. L/V overlap gives intersection counts with frozen Linear/variance sets.

Reference case	MAE	Minus Linear	Minus variance	Minus relevance	Entries	L/V overlap
GWN_PEMS03_s42	17.28738	0.05042793	-0.0001367528	-0.008975236	1057644	12/3
DCRNN_PEMS03_s42	24.87439	-0.04474551	0.3008719	0.6488576	1057644	12/3
GWN_PEMS03_s43	17.18487	0.07074464	0.006396293	-0.003756841	1057644	10/3
DCRNN_PEMS03_s43	24.72943	0.09346867	0.2785154	0.6485434	1057644	10/3
GWN_PEMS03_s44	17.24773	0.03569013	-0.005609721	-0.01895354	1057644	11/3
DCRNN_PEMS03_s44	24.79825	-0.1640576	0.2951195	0.6487627	1057644	11/3
GWN_PEMS08_F1_s42	18.90211	0.1283806	-0.05914434	-0.08560552	155472	3/1
DCRNN_PEMS08_F1_s42	19.12437	0.09755654	-0.03907542	-0.05910301	155472	3/1
GWN_PEMS08_F2_s42	19.36767	0.148258	-0.08540468	-0.1221426	158508	1/2
DCRNN_PEMS08_F2_s42	19.41521	0.1367853	-0.04687157	-0.07465259	158508	1/2
GWN_PEMS08_F3_s42	18.95267	0.2221848	-0.03260182	-0.06986068	168048	4/2
DCRNN_PEMS08_F3_s42	18.90561	0.1569495	-0.01365573	-0.05008028	168048	4/2
GWN_PEMS08_F4_s42	17.8582	0.1667869	0.004646853	-0.04278797	164268	2/1
DCRNN_PEMS08_F4_s42	17.76037	0.157173	0.02837233	-0.02525374	164268	2/1
DCRNN_METRLA_ref42_s42	3.18715	0.01744177	0.01706465	0.009674744	224660	1/1
DCRNN_METRLA_ref42_DCRNN_s43	3.185565	0.01635777	0.01628536	0.01068522	224660	1/1
DCRNN_METRLA_ref42_DCRNN_s44	3.17896	0.01829853	0.01796843	0.01179707	224660	1/1
LSTM_METRLA_ref42_LSTM_s42	3.654509	0.008981654	0.009146723	0.005547137	224660	1/1

Table S47. Frozen source selections using zero-based column identifiers. The same fitting-role pair is shared across its reference families and seeds. No restoration outcome was used to revise these sets.

Fitting role	Correlation-based mRMR-style	Relevance only
METRLA_ref42	73, 44, 186, 139, 8, 113, 184, 36, 204, 6, 108, 202, 17, 150, 26, 170, 104, 18, 174, 92, 93	73, 93, 81, 47, 141, 6, 137, 204, 52, 41, 149, 60, 94, 187, 38, 63, 174, 56, 164, 160, 191
PEMS03	141, 268, 132, 61, 101, 104, 150, 13, 168, 41, 125, 134, 334, 192, 37, 120, 23, 244, 180, 32, 170, 116, 198, 60, 313, 337, 205, 126, 9, 153, 344, 40, 20, 121, 14, 29	141, 238, 272, 241, 267, 244, 3, 246, 273, 213, 203, 223, 282, 257, 205, 2, 206, 247, 214, 199, 239, 243, 172, 167, 354, 319, 56, 315, 220, 287, 102, 201, 325, 180, 216, 317
PEMS08_F1	12, 154, 134, 19, 132, 56, 145, 117, 168, 82, 8, 87, 21, 63, 138, 129, 127	12, 28, 17, 62, 109, 149, 48, 61, 80, 121, 122, 123, 141, 142, 55, 70, 73
PEMS08_F2	12, 154, 134, 19, 132, 145, 57, 117, 168, 45, 63, 8, 34, 99, 112, 127, 21	12, 28, 55, 70, 73, 81, 83, 84, 159, 62, 109, 149, 48, 61, 80, 121, 122
PEMS08_F3	12, 154, 134, 19, 132, 168, 145, 117, 57, 45, 63, 8, 34, 99, 138, 13, 129	12, 28, 55, 70, 73, 81, 83, 84, 159, 62, 109, 149, 48, 61, 80, 121, 122
PEMS08_F4	12, 154, 132, 145, 63, 168, 45, 57, 117, 134, 99, 152, 138, 129, 13, 87, 8	12, 28, 55, 70, 73, 81, 83, 84, 159, 62, 109, 149, 48, 61, 80, 121, 122

S26. PEMS03 predictive accuracy and stratified policy comparisons

This post hoc analysis uses the saved forecasting outputs. The DCRNN/GWN final no-restoration MAEs are 25.2596968353/17.3180144889 vehicles per five minutes, a 45.8579% difference. Mean selected-checkpoint validation MAEs are 18.9784457584/15.3531475757, a 23.6127% difference. The six histories contain finite logged validation MAEs and minibatch objectives. The synthetic packed-input comparison returned maximum absolute difference zero. Validation uses the persistent missingness scenario; final evaluation adds graph-block faults on different dates. The comparisons retain both the fitted-accuracy differences and the changed dates and fault conditions.

Table S48. PEMS03 saved training histories. All MAEs are validation raw-unit Global MAE. The table reports the recorded stopping rules and validation trajectories for the two fitted reference families.

Reference	Seed	Epochs	Best epoch	First MAE	Best MAE	Last MAE
GWN	42	129	114	23.694873	15.340202	15.403478
GWN	43	108	93	24.857171	15.309924	15.325060
GWN	44	120	105	25.254047	15.409317	15.417596
DCRNN	42	147	132	31.861430	19.021818	19.045031
DCRNN	43	134	119	30.043078	18.940684	18.958020
DCRNN	44	161	146	31.388015	18.972835	19.008410

Table S49. Common window–mask strata defined by GWN no-restoration error quartiles, shared with DCRNN. Point MAEs average saved seed numerators before pooling. These descriptive post hoc strata condition on observed GWN error while retaining each original predictor configuration. Positive variance minus Linear favors Linear. Units: vehicles per five minutes.

Reference	GWN-error quartile	Occurrences	No-restoration MAE	Variance minus Linear
DCRNN	1	10052	11.266568	-0.104231
DCRNN	2	10051	22.134781	-0.295144
DCRNN	3	10051	29.984561	-0.390441
DCRNN	4	10051	37.654270	-0.529995
GWN	1	10052	8.500730	0.050565
GWN	2	10051	14.913948	0.045142
GWN	3	10051	19.742678	0.066690
GWN	4	10051	26.115579	0.045886

Table S50. Fault-anchor omission and realized-return-count strata on PEMS03. Contrasts are variance minus Linear; positive favors Linear. Counts in returned-entry strata include three saved seeds sharing windows and masks. Every one of 358 anchor omissions preserves the pooled sign, although individual anchor contrasts are heterogeneous. Returned-count strata condition the comparison on realized information access, with policy-specific source identities and unspecified physical costs. Full strata are in Supplementary Data 2.

Reference	Stratum	Anchor/occurrence count	Contrast	Additional summary
GWN	Omit one anchor	358	0.051065 to 0.052545	163/358 anchor contrasts negative
GWN	Linear_more	42955	0.189107	Omit stratum: -0.023726
GWN	equal	21923	0.026979	Omit stratum: 0.057645
GWN	variance_more	55737	-0.043669	Omit stratum: 0.134322
DCRNN	Omit one anchor	358	-0.332261 to -0.323930	207/358 anchor contrasts negative
DCRNN	Linear_more	42955	0.259590	Omit stratum: -0.656029
DCRNN	equal	21923	-0.107967	Omit stratum: -0.379257
DCRNN	variance_more	55737	-0.871598	Omit stratum: 0.135388

The saved per-window errors aggregate over prediction horizons and receiver sources. Returned-entry counts and separately supplied eligibility sets describe information access at that resolution. Horizon- or receiver-specific analysis would require additional inference. Validation and final evaluation differ in both dates and fault conditions.

S27. Post hoc blockwise policy-label exchange

The contrast variance minus Linear was fixed prospectively for the added DCRNN outputs after the GWN results were known. The blockwise exchange analysis itself is a post hoc addition. All three saved fits and five masks are exchanged synchronously within each block. The observed statistic remains -0.329947057636108 vehicles per five minutes. The seven-day reference uses four blocks; its smallest attainable two-sided p is 2/16 = 0.125. The exchange unit is the temporal block; three times five fit–mask combinations share its sign in each draw.

Table S51. Blockwise sign-exchange reference values for the prespecified primary PEMS03 DCRNN contrast. The post hoc exchange reference conditions on block sign invariance. The seven-day scale follows the original dependence sensitivity; shorter blocks require stronger assumptions. Values are unadjusted, and no significance threshold was prespecified.

Block days	Blocks	Sign patterns	Calculation	Two-sided reference p
7	4	16	Enumeration	0.125
14	2	4	Enumeration	0.5
3	10	1024	Enumeration	0.001953125
1	28	99999	MC +1	1e-05