LEARNING AROUND ANALYTICAL SOLUTIONS: RISK-CALIBRATED CORRECTIONS FOR SPARSE GRAPH SENSING

Anonymous authors
Paper under double-blind review

ABSTRACT

Analytical solutions can provide useful anchors for sparse graph sensing, yet learned flexibility may improve pooled error while worsening local outcomes. We ask when learning should modify or preserve an analytical anchor, whether reconstruction should correct or replace it, and whether pooled gains imply local benefits. We study two settings: a policy that makes one-sensor exchanges around a QR-selected set or retains it, and a spatiotemporal graph network that predicts a residual over Ridge reconstruction. At the smallest river budget, forced swapping improved the whole-period error point estimate but worsened 50.91% of local time-region blocks; allowing no-swap reduced harmful-block frequency by 14.95 percentage points across budgets. Under fixed QR masks, residual prediction reduced pooled unselected-field error by 18.5%–23.0% relative to direct spatiotemporal prediction, yet equal-weight local block gains remained negative on average. A three-domain comparison provided supporting evidence across river, traffic, and ocean graphs. Together, these findings show that intervention choice, correction form, and evaluation scale determine when learning improves on an analytical anchor.

1 INTRODUCTION

Analytical solutions are often treated as baselines for learned replacements. Yet when an anchor already captures useful structure, added flexibility does not make every learned intervention beneficial. At the smallest budget in our experiments, a forced sensor swap improved the whole-period error point estimate while worsening 50.91% of local time-region blocks. This tension raises a question beyond whether learning can beat a fixed baseline: when should it change a useful analytical solution, and when should it leave it intact?

Many environmental and infrastructure systems are observed as fields over connected locations, from rivers and roads to ocean grids. Measuring every location is often impractical. Sparse sensing therefore selects a limited set of sites and reconstructs conditions elsewhere. Its success depends on both sensor placement and reconstruction. Analytical methods provide useful anchors, or feasible starting points. For placement, pivoted QR factorization decomposes a data-informed matrix into orthogonal Q and upper-triangular R factors, using column pivoting to rank informative sites under a fixed sensor budget. Linear and Ridge models infer the full field from selected observations. Yet learned flexibility may still harm local outcomes.

In our smallest-budget river experiment, a forced sensor swap lowered the whole-period error point estimate but worsened 50.91% of local time-region blocks. This shows how pooled accuracy can hide harm in specific places or periods. We therefore ask when learning should alter a useful anchor and when it should preserve it.

Prior work covers one-for-one sensor exchange, QR with neural decoders, learned placement, joint placement and reconstruction, and residual correction (Joshi and Boyd, 2009; Williams et al., 2022; Li et al., 2025; Chen and Xiu, 2021). Yet it remains unclear how learning should be constrained around a useful anchor, whether reconstruction should replace or correct it, and whether pooled gains imply local benefits.

Reviewers: please read the Reviewer Guidelines (iclr.cc/Conferences/2027/ReviewerGuidelines) and the AI Policy for Reviewers (iclr.cc/Conferences/2027/AIPolicyForReviewers). If you used AI to expand, edit, or polish your review, please provide the input text to the LLM. Better still, consider skipping the LLM and submitting your original text: we, and the authors, are much more interested in your unedited thoughts than in what an LLM has to say. AI-assisted or not, you are putting your name and reputation behind your review: LLM-generated falsehoods, hallucinations or misrepresentations are subject to disciplinary action, which may include desk-rejecting all papers you have authored.

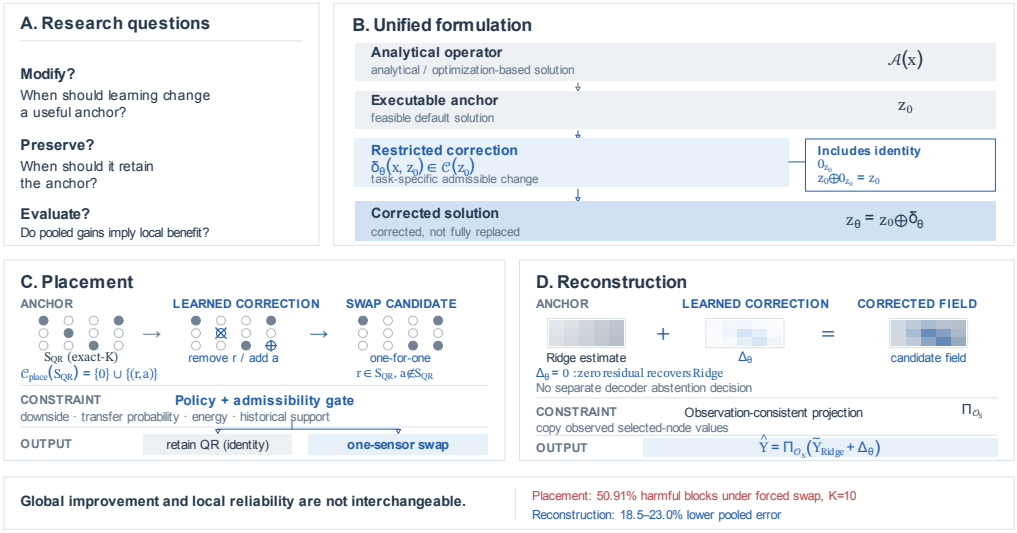

Figure 1: Learning around analytical solutions. An analytical method supplies an executable anchor; learning is restricted to task-specific corrections that include an identity element. For placement, the policy may retain the QR set or make one sensor exchange subject to an empirical admissibility gate. For reconstruction, an STGNN predicts an additive residual over Ridge, and observed selected-node values are copied into the final output. Zero residual recovers the Ridge anchor mathematically but is not a separate decoder abstention decision. Pooled gains and local reliability are evaluated separately.

We address these questions in two settings. A placement policy may swap one sensor around an exact-budget pivoted-QR set or retain it when no candidate passes an empirical gate. With masks fixed, a spatiotemporal graph network predicts an additive residual over Ridge; we compare direct prediction and Ridge-feature conditioning. We evaluate pooled and local outcomes, then test temporal and cross-domain behavior across river, traffic, and ocean data.

2 Related Work

2.1 Analytical and subspace-based sensor placement

Placement methods use information-theoretic design, subspace reconstruction, and graph sampling. Mutual-information objectives support greedy selection under Gaussian-process assumptions (Krause et al., 2008). Pivoted QR provides practical subspace-based selection (Manohar et al., 2018; Ranieri et al., 2014), with related methods for bandlimited graphs, smoothness and noise priors, localized selection, and accelerated sampling (Kim, 2021; Tanaka and Eldar, 2020; Wang et al., 2018; Bai et al., 2020; Sakiyama et al., 2019; Hashemi et al., 2022). Other work studies adjacent-node exchanges and robustness to model mismatch (Nomura et al., 2024; Pakiyarajah et al., 2025; Li and Cai, 2025). These objectives encode different assumptions and do not directly address when to preserve a feasible sensor set. Joshi and Boyd (2009, §III-C) establish one-for-one exchange search with a 2-opt stopping rule for parameter estimation. We use QR as an exact-budget executable anchor, learn cross-period utility, and apply an empirical admissibility gate.

2.2 Learned placement and sparse-observation reconstruction

Learned placement includes Concrete autoencoders and entropy-initialized geophysical sensing (Balın et al., 2019; Turko et al., 2022), joint graph-based placement and reconstruction with a directional GNN and constrained PPO (Li et al., 2025), and graph methods for water-network sensing (Menapace et al., 2024; Sirothia and Narasimhan, 2024; Truong et al., 2024). These approaches dif-

fer in whether they learn feature subsets, graph policies, or estimators for a fixed network. Sparse-field reconstruction includes graph imputation, Voronoi-assisted learning, attention-based decoders, and conditional spatiotemporal imputation (Marisca et al., 2022; Fukami et al., 2021; Santos et al., 2023; Liu et al., 2023). These methods condition on available observations but do not necessarily optimize a fixed-budget sensor set.

2.3 Hybrid analytical and learned correction

Prior work pairs a POD basis and QR placement with a shallow decoder (Williams et al., 2022), corrects prior or reduced models with residual networks (Chen and Xiu, 2021; Györök et al., 2025), and jointly optimizes placement and reconstruction through differentiable objectives (Liu et al., 2025). These studies establish that analytical and learned components can be combined, but address different tasks and constraints. Our focus is how correction is constrained around explicit anchors. Safe policy improvement also studies fallback to a baseline under uncertainty (Laroche et al., 2019), but its batch-MDP guarantees do not transfer to our empirical placement gate. Our policy permits one exchange or no-swap; reconstruction adds a graph residual over Ridge while copying available observations.

3 Learning Around Analytical Anchors

Let $x \in \mathcal{X}$ denote the information available at decision time. An analytical operator $\mathcal{A}$ produces a feasible anchor $z_0 \in \mathcal{Z}$:

$$z_0 = \mathcal{A}(x).$$

The learned correction belongs to a task-specific set $\mathcal{C}(z_0)$:

$$\delta_\theta(x, z_0) \in \mathcal{C}(z_0), \qquad z_\theta = z_0 \oplus \delta_\theta(x, z_0).$$

The correction set contains an identity element 0_{z_0} such that $z_0 \oplus 0_{z_0} = z_0$. Together, the anchor, correction set, and identity define which changes are available. The two implementations constrain and train their corrections differently.

3.1 Discrete correction of sensor placement

For a graph $G = (V, E)$ and sensor budget K, pivoted QR produces an exact-K anchor $S_{\mathrm{QR}} \subset V$. The permitted correction set is

$$\mathcal{C}_{\mathrm{place}}(S_{\mathrm{QR}}) = \{0\} \cup \{(r, a) : r \in S_{\mathrm{QR}},\ a \in V \setminus S_{\mathrm{QR}}\}.$$

A swap gives $S_{\mathrm{QR}} \oplus (r, a) = (S_{\mathrm{QR}} \setminus \{r\}) \cup \{a\}$, while the identity gives $S_{\mathrm{QR}} \oplus 0 = S_{\mathrm{QR}}$. Each action preserves the sensor budget. The learned policy ranks candidate swaps and applies a deterministic admissibility gate; if no candidate qualifies, it retains the QR set.

3.2 Continuous correction of field reconstruction

For a selected set S, a Ridge model fitted on historical data produces a full-field estimate $\widetilde{Y}_{\mathrm{Ridge}}$. At prediction time, the model receives a causal history of selected-node observations and the current measurements available at S. Let $\Pi_{\mathcal{O}_S}$ copy those available current measurements into the output. The observation-consistent anchor is $z_0 = \Pi_{\mathcal{O}_S}(\widetilde{Y}_{\mathrm{Ridge}})$. A spatiotemporal graph network predicts a residual Δ_θ:

$$\widehat{Y} = \Pi_{\mathcal{O}_S}(\widetilde{Y}_{\mathrm{Ridge}} + \Delta_\theta).$$

The projection copies only real selected-node measurements and does not use unselected targets. Setting $\Delta_\theta = 0$ recovers the Ridge anchor mathematically; the decoder has no separate abstention

decision that selects this case. The two implementations share an analytical anchor, a restricted correction, and an identity element. Placement constrains corrections with an explicit intervention gate; reconstruction relies on residual parameterization and observation consistency.

4 METHOD

4.1 QR CANDIDATES AND GRAPH SCORER

For each allowed fitting context and budget, pivoted QR forms the exact-K sensor anchor. We enumerate all $K(N-K)$ remove-add pairs. Candidate representations combine removed- and added-node embeddings with budget, QR-rank, signal-energy, redundancy, spatial-coverage, and graph-context features. All time-varying feature statistics are computed from permitted historical data, and canonical node indices resolve ties deterministically.

For a historical period p, let $u_p = 1 - E_{\mathrm{swap},p}/E_{\mathrm{QR},p}$ be the relative reconstruction gain of a candidate swap under the same linear decoder. For a label period p and subsequent transfer period q, expected utility targets $(u_p+u_q)/2$, and downside utility targets $\min(u_p, u_q)$. Positive-transfer probability predicts whether both gains are positive. The graph scorer combines three relation-specific message-passing representations into a candidate embedding and estimates these three targets with a budget-conditioned head. They summarize historical transfer behavior rather than guaranteeing improvement in a future period.

4.2 ADMISSIBILITY-GATED SWAP OR NO-SWAP

A candidate is admissible when its ensemble downside estimate is positive, predicted positive-transfer probability is at least 0.5, fitting-period energy margin is nonnegative, and add-role support appears in at least two of three historical folds. The sign conditions reject predicted downside and energy loss; 0.5 is the probability majority cutoff, and two folds require repeated historical support. These thresholds were frozen in the station-selection rule before held-out cross-domain evaluation. The policy selects the admissible candidate with the largest predicted downside, using canonical indices to break ties; otherwise it retains QR. The empirical gate provides no formal safe-improvement guarantee, and its joint application does not isolate any one condition's effect.

4.3 RESIDUAL-RIDGE SPATIOTEMPORAL RECONSTRUCTION

The Ridge path maps selected-node observations to a full-field estimate using fitting data only. A causal temporal encoder processes selected-node histories, relational graph blocks propagate information over the fixed graph, and global slots provide nonlocal interactions. An output head predicts the residual added to Ridge. The final projection copies available current observations at selected nodes. The residual parameterization is related to prior model-correction methods (Chen and Xiu, 2021), while the task-specific graph architecture and comparison are evaluated here.

5 EXPERIMENTAL PROTOCOL

5.1 TASKS, DATASETS, AND TEMPORAL ROLES

We study static, exact-K placement for contemporaneous field reconstruction, not future forecasting. At time t, the decoder receives the preceding 13 native time steps and available current values at selected nodes. It predicts the field at all nodes at time t. Missing targets are excluded from scoring, and each selected set remains static throughout its evaluation period.

The N264 USGS river network supports the mechanism and temporal studies. The heterogeneous-domain evaluation uses a 512-station USGS network, 716 San Diego traffic sensors from LargeST, and a 512-cell Kuroshio region from NOAA OISST. The ocean task uses an interpolated gridded analysis field and therefore represents sparse sampling simulation, not physical buoy deployment.

Full temporal splits are provided in Appendix Table S2.

Table 1: Datasets and experimental roles. Budgets are exact sensor counts. The two USGS graphs serve different analyses: N264 supports mechanism and temporal studies, whereas N512 is the river domain in the cross-domain comparison.

Dataset	Graph size	Temporal resolution	Budgets (K)	Role
USGS N264	264	Daily	10, 25, 50	Mechanism + temporal
USGS N512	512	Daily	10, 26, 51	Cross-domain
San Diego (LargeST)	716	Hourly	14, 36, 72	Cross-domain
Kuroshio (OISST)	512	Daily	10, 26, 51	Cross-domain

LargeST five-minute flow observations are aggregated to hourly means using the frozen availability rule (Liu et al., 2023). OISST contains daily sea-surface-temperature anomalies from the NOAA daily analysis product (Huang et al., 2021). Models are fitted separately by domain; the cross-domain analysis tests a shared method recipe, not zero-shot transfer.

5.2 Comparisons and evaluation targets

For placement risk, four policies act on the same frozen candidate scores: forced expected-utility selection, forced expected-plus-downside selection, gate-guided forced intervention, and gate-guided swap or no-swap. The graph scorer averages five fixed selector seeds before action selection. A same-feature MLP provides the non-graph ranking comparison on the same candidate population.

For reconstruction, we compare Linear and Ridge references with three neural variants on identical fixed QR masks. Direct STGNN predicts the full field; Ridge-feature STGNN receives Ridge as an input but predicts the full field without an additive skip. Residual-Ridge predicts an additive correction. The neural variants share their encoder, graph blocks, global slots, selected-value projection, training conditions, and paired mask schedules. Five paired seeds describe training variability, not independent observational replicates.

The cross-domain complete-system study compares Proposed with Li-adapted and CAE-adapted learned systems (Li et al., 2025; Turko et al., 2022). All systems share evaluation periods, exact-K budgets, selected histories, observation availability, target masks, and selected-observation copy rules. Their architectures, optimization schedules, and tuning procedures differ, so this is neither a compute-matched nor a capacity-matched comparison. The adaptations are task-aligned rather than verbatim reproductions; their configurations and deviations belong in Supplementary Table S1.

5.3 Outcomes and statistical analysis

For method m and scoring mask M, relative Frobenius error is

$$E_m(M) = \sqrt{\frac{\sum_{t,j} M_{t,j}(Y_{t,j} - \widehat{Y}^{(m)}_{t,j})^2}{\sum_{t,j} M_{t,j} Y^2_{t,j}}}.$$

The cross-domain complete-system primary endpoint uses all observed cells after selected-observation copy; it measures the performance of the sensing system as a whole. We also report a descriptive common-target error on observed cells outside the union of QR, Proposed, Li-adapted, and CAE-adapted sensor sets within each domain-budget setting. This shared mask isolates locations that every complete system must infer. Fixed-mask decoder comparisons instead use unselected-only cells, and selector attribution uses its specified common-target mask. A positive relative gain $(E_{\text{reference}} - E_m)/E_{\text{reference}}$ favors the evaluated method.

We also report equal-weight block relative gain, computed within prespecified time-by-region blocks. This estimand asks whether benefit is distributed locally, whereas whole-period relative Frobenius pools squared errors before taking a ratio. For N264 placement risk, blocks cross 90-day windows with HUC8 regions; decoder blocks cross 30-day windows with HUC8 regions. Cross-domain blocks use 30-day by HUC8 river blocks, 168-hour by freeway-direction traffic blocks, and 30-day by 4-by-4 OISST tiles. Confirmatory all-site inference uses 10,000 paired block-bootstrap

replicates with Holm correction within the frozen comparison families; common-target summaries are descriptive and do not inherit those intervals or decisions.

For each learned comparator, the original frozen decision required positive gains in at least seven of nine cells and at least two budgets per dataset. It also required no cell below -3%, a positive aggregate confidence-interval lower bound, and adjusted significance.

6 RESULTS

6.1 ARE THE ANALYTICAL ANCHORS STRONG ENOUGH TO JUSTIFY CORRECTION?

Ridge is a strong starting point, but it can still be improved. It accounts for 86.5%–93.4% of the target energy across the three sensor budgets. With sensor locations fixed, Residual-Ridge still reduces pooled error by 18.5%–23.0% relative to Direct STGNN.

QR is a useful default, but the value of sensor swapping depends on the budget and the evaluation scale. At the smallest budget, retaining QR produced no harmful blocks, whereas forced swapping harmed 50.91% of local blocks. At the same time, the whole-period error point estimate under forced swapping still improved by 4.17%, showing that "reducing local harm" and "lowering overall error" are not the same objective.

Analytical anchors are sufficiently strong to serve as stable, deployable starting points. Learned corrections can provide gains under specific conditions and evaluation objectives, but **whether to correct, how to correct, and how to evaluate the correction must each be validated separately**.

6.2 WHEN DOES SELECTIVE DECISION CORRECTION HELP?

In our experiments, **selective decision correction was most helpful at the smallest sensor budget, when the policy was allowed to choose not to swap**. At $K = 10$, forced swapping harmed 50.91% of local blocks, whereas retaining the QR set produced no harmful blocks. When aggregated across budgets, the threshold-based policy with a no-swap option reduced the harmful-block rate by 14.95 percentage points.

However, this benefit primarily reflects **lower local intervention risk** and does not imply lower aggregate error. At $K = 10$, forced swapping improved the whole-period error point estimate by 4.17%, whereas choosing no-swap yielded a 0% change relative to QR. At $K = 25$ and $K = 50$, the policy executed swaps, and the harmful-block fraction increased slightly. Therefore, the gate cannot be said to help across all budgets or all evaluation metrics.

Another limitation is that expected-utility ranking and expected-plus-downside ranking selected the same actions. As a result, the experiment does not establish an additional benefit from the downside-utility prediction head.See Figure 2.

6.3 DOES RESIDUAL PREDICTION CORRECTION IMPROVE RECONSTRUCTION?

Under the fixed QR sensor masks, Residual-Ridge achieved whole-period relative Frobenius errors on unselected nodes of 0.329, 0.304, and 0.235 at $K = 10, 25$, and 50, respectively. Compared with Direct STGNN, whose corresponding errors were 0.404, 0.388, and 0.305, these represent reductions of 18.5%, 21.7%, and 23.0%. Relative to Ridge alone, Residual-Ridge also reduced error by 10.4%, 9.3%, and 8.3%, respectively.

In addition, the STGNN variant that used the Ridge prediction only as an input feature achieved errors of 0.340, 0.315, and 0.258 across the three budgets, all of which were higher than those of Residual-Ridge. Overall, under the pooled-error metric used in this experiment, explicitly predicting the residual over Ridge outperformed both direct full-field prediction and using Ridge only as an input feature.See Figure 3.

6.4 DOES THE CORRECTION PERSIST OVER TIME?

Across two non-overlapping future periods on N264, the complete system achieved positive results, although different selector versions were used. The 2021–2022 test used the then-frozen v5 selector

Table 2: Complete-system cross-domain comparison on the all-site endpoint. Entries under method names are relative Frobenius errors (lower is better); gains are $(E_{\text{comparator}} - E_{\text{Proposed}})/E_{\text{comparator}}$, expressed as percentages. The mean row averages the nine cell-level gains equally, rather than pooling errors across domains.

Dataset	K	Proposed	Li-adapted	Gain vs Li	CAE-adapted	Gain vs CAE
USGS N512	10	0.03158	0.03917	**+19.37%**	0.03802	**+16.95%**
	26	0.01705	0.02761	**+38.23%**	0.02932	**+41.83%**
	51	0.01214	0.02608	**+53.43%**	0.01926	**+36.93%**
San Diego	14	0.12267	0.46874	**+73.83%**	0.11947	**-2.676%**
	36	0.11022	0.43563	**+74.70%**	0.11148	**+1.13%**
	72	0.09935	0.17258	**+42.44%**	152.45303	**+99.93%**
Kuroshio	10	0.34111	0.75529	**+54.84%**	0.49224	**+30.70%**
	26	0.21904	0.64512	**+66.05%**	0.43804	**+50.00%**
	51	0.10762	0.54731	**+80.34%**	0.32499	**+66.88%**
Mean gain				**+55.91%**		**+37.96%**

with the v4 decoder. Relative to QR+Linear, its all-site relative error improved by 4.59%, 8.00%, and 4.24% at $K = 10$, 25, and 50, respectively. The subsequently frozen v6+v4 system improved over QR+Linear by 4.30%, 1.21%, and 6.79%, respectively, during the independent 2023–2025 evaluation period. The common-target HUC8 intervals excluded zero at $K = 10$ and $K = 50$, but crossed zero at $K = 25$. Applying v6+v4 retrospectively to the previously used 2021–2022 period yielded gains of 4.275%, 1.360%, and 3.761%, respectively. These results show positive complete-system gains across the evaluated periods, while differences in selector versions preclude a claim of time-invariant swap utility.

6.5 Does the complete system transfer across graph domains?

Across USGS river, San Diego traffic, and Kuroshio temperature data, Proposed combined QR-anchored placement with Residual-Ridge reconstruction. The complete-system comparators used Li-adapted placement with DTA-GNN and CAE-adapted placement with its learned decoder. The confirmatory all-site endpoint favored Proposed over Li-adapted in all nine domain-budget cells, with a mean point gain of 55.91% (95% bootstrap CI, 48.96–60.98; Holm-adjusted $p = 0.00020$). Against CAE-adapted, the mean gain was 37.96% (95% CI, 20.55–42.71; Holm-adjusted $p = 0.00020$), with eight of nine cells positive.

The San Diego $K = 72$ CAE-adapted error is the retained frozen result; its extreme value was traced to a finite decoder extrapolation in one block. The displayed mean includes this cell.

On the descriptive common target, which excludes nodes selected by any of the QR or learned systems, the nine-cell mean point gains were 49.10% over Li-adapted and 33.40% over CAE-adapted. Gains were positive in nine cells, respectively. These values are arithmetic summaries of the archived common-target errors, without a separately audited bootstrap or formal decision rule. All-site error evaluates the deployed sensing system; common-target error evaluates locations every system must infer.

For the San Diego $K = 14$ all-site cell, the **relative Frobenius error of Proposed is approximately 2.676% higher than that of CAE-adapted**, indicating slightly worse performance. However, the rule requires the worst-cell gain to be no lower than -3%. Since -2.676% is within the allowed degradation range, this cell is classified as **Go**. In other words, it falls within the predefined tolerance.

Common-decoder attribution favored Proposed's selector across all three datasets. Its mean point gains over the Li selector were 35.86% with Linear and 36.40% with the Proposed decoder; corresponding gains over CAE were 26.23% and 25.77%. The strongest non-river result was on OISST, with gains of 67.07% over Li-adapted and 49.19% over CAE-adapted. On San Diego, the 32.80% point gain over CAE-adapted had an interval crossing zero. The worst observed active-swap utility was -1.883%, within the original -3% safety tolerance. All nine Proposed cells executed a swap, so the confirmatory matrix did not exercise no-swap.See Figure 4.

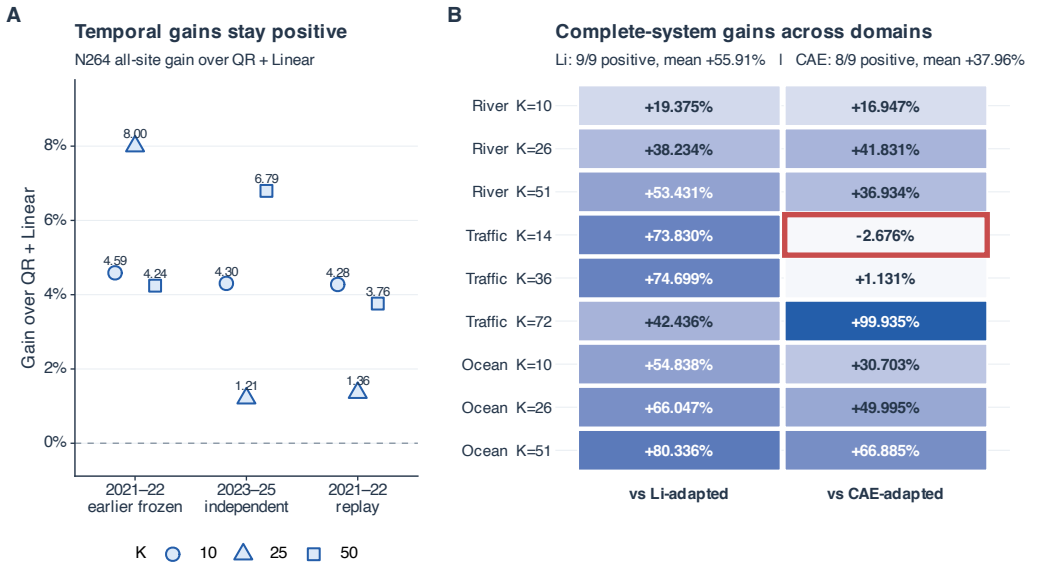

Figure 4: Temporal and cross-domain performance of analytical-anchor correction. (A) N264 all-site gains over QR+Linear in the frozen 2021–2022 test, the independent 2023–2025 test of a later frozen system, and its retrospective 2021–2022 replay; selector versions differ. (B) All-site point gains against Li-adapted (9/9 positive; mean +55.91%) and CAE-adapted (8/9 positive; mean +37.96%) across three graph domains and three budgets. San Diego K=14 is -2.676% versus CAE-adapted. Systems were trained separately; the comparison is task-adapted and lacks a standalone QR+Linear/QR+Ridge complete-system baseline. Cellwise confidence intervals are omitted.

7 Discussion and Limitations

7.1 What does learning around an anchor add?

The results support a conditional role for learning when an analytical anchor already captures useful structure. Ridge explained a large fraction of target energy, and a residual network improved pooled reconstruction error without predicting the entire field from scratch. In placement, the QR set provided an exact-budget default around which a model could make a narrow, auditable change.

7.2 Why do global and local metrics differ?

Whole-period relative Frobenius error pools squared errors, so large-error cells have greater influence. Equal-weight block gain instead averages relative gains across time-region blocks; the two estimands can therefore disagree. Blocks with low reference error or target energy can amplify negative ratios. In our decoder analysis, removing the lowest reference-error decile did not eliminate negative mean block gain, so small denominators alone do not explain the local tail.

7.3 Scope and limitations

The N264 temporal findings involve successive frozen selector versions. The 2021–2022 v5+v4 test and 2023–2025 v6.1+v4 test are independent periods, but they are not repeated tests of one unchanged system. The v6.1 replay on the earlier period is retrospective. The K50 selector-only sign change further limits claims of time-invariant intervention utility.

The three-domain study uses task-aligned adaptations of two learned systems, rather than verbatim reproductions. The Li adaptation retains a PPO selector and the CAE adaptation retains a differentiable Concrete-AE selector, but both are configured for this paper's static exact-K task and are

not replications of the source papers' full pipelines. Their fidelity is therefore limited to the task interface and method family; the supplement should document each source-to-adaptation deviation. The adaptations share evaluation periods, budgets, input availability, and output-copy semantics, but not model capacity or tuning budget. The primary matrix has no standalone QR+Linear/QR+Ridge arm, so the large cross-domain gains do not establish superiority over an analytical complete-system baseline. Common-target gains are descriptive and have no confirmatory interval. The ocean task samples an interpolated analysis grid, not physical buoy deployment. The confirmatory matrix also contained no observed no-swap event.

8 Conclusion

We studied sparse graph sensing through a question that arises whenever learning is introduced on top of an existing analytical solution: **should the learned model replace the analytical solution, modify it, or sometimes leave it unchanged?** Rather than treating analytical methods only as baselines, we used QR placement and Ridge reconstruction as executable anchors and examined restricted learned corrections around them.

The experiments reveal two complementary behaviors. In placement, making learned intervention optional can reduce local intervention harm even when forced intervention improves pooled error, showing that whether a correction is desirable depends on the evaluation objective. In reconstruction, predicting an explicit residual over Ridge consistently improves pooled unselected-field error relative to direct neural prediction and feature-level use of the Ridge estimate, but the corresponding equal-weight local block gains remain heterogeneous and can be negative on average. These results show that aggregate accuracy and local reliability are not interchangeable measures of correction quality.

Across three graph domains, the same correction recipe also compares favorably with the evaluated task-adapted learned systems, providing evidence that the observed behavior is not confined to a single river setting. At the same time, these experiments do not establish that learned correction universally dominates analytical complete systems, nor that abstention will be equally important in every domain. The main conclusion is therefore narrower: **when analytical solutions already provide useful structure, the value of learning depends not only on how much flexibility is added, but also on whether intervention is optional, whether learning corrects or replaces the analytical predictor, and which scale of performance is used to judge success.**

References

Bai, Y., Wang, F., Cheung, G., Nakatsukasa, Y., and Gao, W. (2020). Fast graph sampling set selection using Gershgorin disc alignment. *IEEE Transactions on Signal Processing*, 68, 2419–2434. https://doi.org/10.1109/TSP.2020.2981202

Balın, M. F., Abid, A., and Zou, J. (2019). Concrete autoencoders: Differentiable feature selection and reconstruction. *Proceedings of the 36th International Conference on Machine Learning*, PMLR 97, 444–453. https://proceedings.mlr.press/v97/balin19a.html

Chen, Z., and Xiu, D. (2021). On generalized residual network for deep learning of unknown dynamical systems. *Journal of Computational Physics*, 438, 110362. https://doi.org/10.1016/j.jcp.2021.110362

Fukami, K., Maulik, R., Ramachandra, N., Fukagata, K., and Taira, K. (2021). Global field reconstruction from sparse sensors with Voronoi tessellation-assisted deep learning. *Nature Machine Intelligence*, 3, 945–951. https://doi.org/10.1038/s42256-021-00402-2

Györök, B. M., Hoekstra, J. H., Kon, J., Péni, T., Schoukens, M., and Tóth, R. (2025). Orthogonal projection-based regularization for efficient model augmentation. *Proceedings of the 7th Annual Learning for Dynamics & Control Conference*, PMLR 283, 166–178. https://proceedings.mlr.press/v283/gyorok25a.html

Hashemi, A., Shafipour, R., Vikalo, H., and Mateos, G. (2022). Towards accelerated greedy sampling and reconstruction of bandlimited graph signals. *Signal Processing*, 195, 108505. https://doi.org/10.1016/j.sigpro.2022.108505

Huang, B., Liu, C., Banzon, V., Freeman, E., Graham, G., Hankins, B., Smith, T., and Zhang, H.-M. (2021). Improvements of the Daily Optimum Interpolation Sea Surface Temperature (DOISST) Version 2.1. *Journal of Climate*, 34(8), 2923–2939. https://doi.org/10.1175/JCLI-D-20-0166.1

Joshi, S., and Boyd, S. (2009). Sensor selection via convex optimization. *IEEE Transactions on Signal Processing*, 57(2), 451–462. https://doi.org/10.1109/TSP.2008.2007095

Kim, Y. H. (2021). QR factorization-based sampling set selection for bandlimited graph signals. *Signal Processing*, 179, 107847. https://doi.org/10.1016/j.sigpro.2020.107847

Krause, A., Singh, A., and Guestrin, C. (2008). Near-optimal sensor placements in Gaussian processes: Theory, efficient algorithms and empirical studies. *Journal of Machine Learning Research*, 9(8), 235–284. https://www.jmlr.org/papers/v9/krause08a.html

Laroche, R., Trichelair, P., and Tachet des Combes, R. (2019). Safe policy improvement with baseline bootstrapping. *Proceedings of the 36th International Conference on Machine Learning*, PMLR 97, 3652–3661. https://proceedings.mlr.press/v97/laroche19a.html

Li, J., and Cai, B. (2025). Sensor placement method for water distribution networks based on sampling of non-bandlimited graph signals. *Digital Signal Processing*, 156A, 104809. https://doi.org/10.1016/j.dsp.2024.104809

Li, Ruoyan, Wan, Frank, Huang, Zijie, Liu, Zixiao, Wang, Haixin, Luo, Xiao, Wang, Wei, and Sun, Yizhou. (2025). Flow field reconstruction with sensor placement policy learning. *Advances in Neural Information Processing Systems 38*. https://papers.nips.cc/paper_files/paper/2025/hash/9708c7d3a0fef3710f33ba05a74e10b3-Abstract-Conference.html

Liu, M., Huang, H., Feng, H., Sun, L., Du, B., and Fu, Y. (2023). PriSTI: A conditional diffusion framework for spatiotemporal imputation. In *2023 IEEE 39th International Conference on Data Engineering (ICDE)*, 1927–1939. https://doi.org/10.1109/ICDE55515.2023.00150

Liu, X., Xia, Y., Liang, Y., Hu, J., Wang, Y., Bai, L., Huang, C., Liu, Z., Hooi, B., and Zimmermann, R. (2023). LargeST: A benchmark dataset for large-scale traffic forecasting. arXiv:2306.08259. https://arxiv.org/abs/2306.08259

Liu, X., Peng, W., Zhang, X., Zhao, X., Zhou, W., Yao, W., and Chen, X. (2025). Enhancing deep learning-based field reconstruction with a differentiable learning framework. *Nature Machine Intelligence*, 7, 1129–1140. https://doi.org/10.1038/s42256-025-01063-1

Manohar, K., Brunton, B. W., Kutz, J. N., and Brunton, S. L. (2018). Data-driven sparse sensor placement for reconstruction: Demonstrating the benefits of exploiting known patterns. *IEEE Control Systems Magazine*, 38(3), 63–86. https://doi.org/10.1109/MCS.2018.2810460

Marisca, I., Cini, A., and Alippi, C. (2022). Learning to reconstruct missing data from spatiotemporal graphs with sparse observations. *Advances in Neural Information Processing Systems 35*. https://proceedings.neurips.cc/paper_files/paper/2022/hash/cf70320e93c08b39b1b29a348097a376-Abstract-Conference.html

Menapace, A., Zanfei, A., Herrera, M., and Brentan, B. (2024). Graph neural networks for sensor placement: A proof of concept towards a digital twin of water distribution systems. *Water*, 16(13), 1835. https://doi.org/10.3390/w16131835

Nomura, S., Hara, J., Higashi, H., and Tanaka, Y. (2024). Dynamic sensor placement based on sampling theory for graph signals. *IEEE Open Journal of Signal Processing*, 5, 1042–1051. https://doi.org/10.1109/OJSP.2024.3466133

Pakiyarajah, D., Pavez, E., and Ortega, A. (2025). Graph-based signal sampling with adaptive subspace reconstruction for spatially-irregular sensor data. In *2025 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, 1–5. https://doi.org/10.1109/ICASSP49660.2025.10890658

Ranieri, J., Chebira, A., and Vetterli, M. (2014). Near-optimal sensor placement for linear inverse problems. *IEEE Transactions on Signal Processing*, 62(5), 1135–1146. https://doi.org/10.1109/TSP.2014.2299518

Sakiyama, A., Tanaka, Y., Tanaka, T., and Ortega, A. (2019). Eigendecomposition-free sampling set selection for graph signals. *IEEE Transactions on Signal Processing*, 67(10), 2679–2692. https://doi.org/10.1109/TSP.2019.2908129

Santos, J. E., Fox, Z. R., Mohan, A., O'Malley, D., Viswanathan, H., and Lubbers, N. (2023). Development of the Senseiver for efficient field reconstruction from sparse observations. *Nature Machine Intelligence*, 5, 1317–1325. https://doi.org/10.1038/s42256-023-00746-x

Sirothia, A., and Narasimhan, S. (2024). Sensor placement in water distribution networks using graph neural networks. *IFAC-PapersOnLine*, 57, 268–273. https://doi.org/10.1016/j.ifacol.2024.05.046

Tanaka, Y., and Eldar, Y. C. (2020). Generalized sampling on graphs with subspace and smoothness priors. *IEEE Transactions on Signal Processing*, 68, 2272–2286. https://doi.org/10.1109/TSP.2020.2982325

Truong, H., Tello, A., Lazovik, A., and Degeler, V. (2024). Graph neural networks for pressure estimation in water distribution systems. *Water Resources Research*, 60, e2023WR036741. https://doi.org/10.1029/2023WR036741

Turko, N., Lobashev, A., Ushakov, K., Kaurkin, M., and Ibrayev, R. (2022). Information entropy initialized concrete autoencoder for optimal sensor placement and reconstruction of geophysical fields. In *Supercomputing: 8th Russian Supercomputing Days*, Lecture Notes in Computer Science, 167–184. https://doi.org/10.1007/978-3-031-22941-1_12

Wang, F., Wang, Y., and Cheung, G. (2018). A-optimal sampling and robust reconstruction for graph signals via truncated Neumann series. *IEEE Signal Processing Letters*, 25(5), 680–684. https://doi.org/10.1109/LSP.2018.2818062

Williams, J., Zahn, O., and Kutz, J. N. (2022). Data-driven sensor placement with shallow decoder networks. arXiv:2202.05330. https://doi.org/10.48550/arXiv.2202.05330

AI USE STATEMENT

Generative AI tools were used to assist with manuscript editing and organization, translation, literature search and summarization, figure design and presentation, and feedback on research methodology and experimental design. They were also used to assist in discussing and interpreting experimental results and in improving the clarity of the paper's presentation. All research decisions, experimental results, quantitative analyses, and final claims were reviewed and verified by the authors. The authors take full responsibility for the content of this work, including any text or artifacts produced with the assistance of generative AI. Generative AI tools were additionally used to assist with software implementation and debugging; AI-assisted code was reviewed and tested by the authors before use in the reported experiments.

A Additional Figures

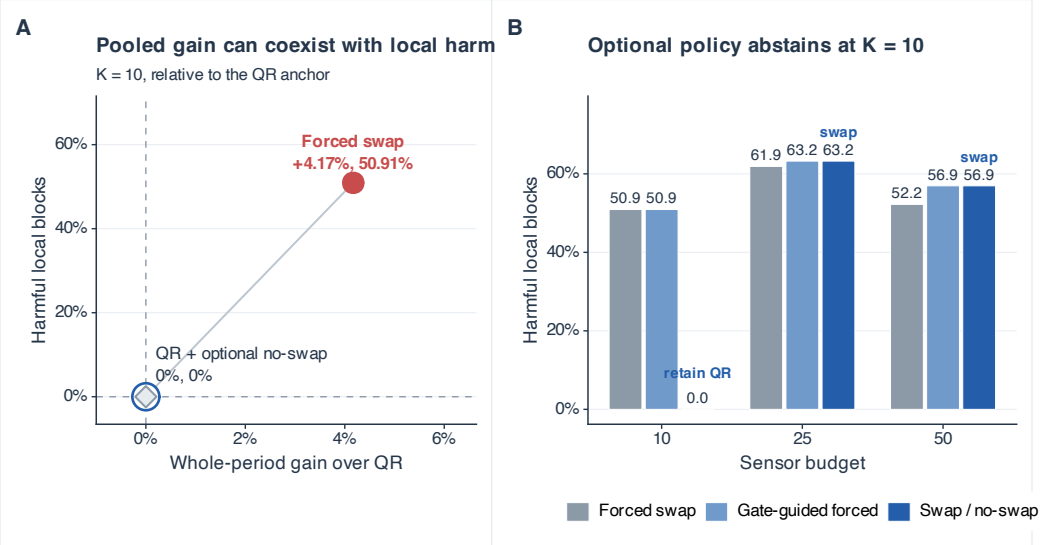

Figure 2: Selective intervention in sensor placement. (A) At K=10, the forced swap improved whole-period error over QR by 4.17% but harmed 50.91% of local time-by-region blocks. The optional-intervention policy retained QR, producing zero gain and zero harmful blocks relative to QR. (B) Harmful-block rates for the expected-utility forced swap, gate-guided forced swap, and gate-guided swap/no-swap policies under the same frozen candidate scores. Across equally weighted budgets, the last policy reduced harmful-block rate by 14.95 percentage points relative to expected-utility forced swapping. It retained QR at K=10 and swapped at K=25 and 50; the latter two budgets did not individually show a harm-rate reduction.

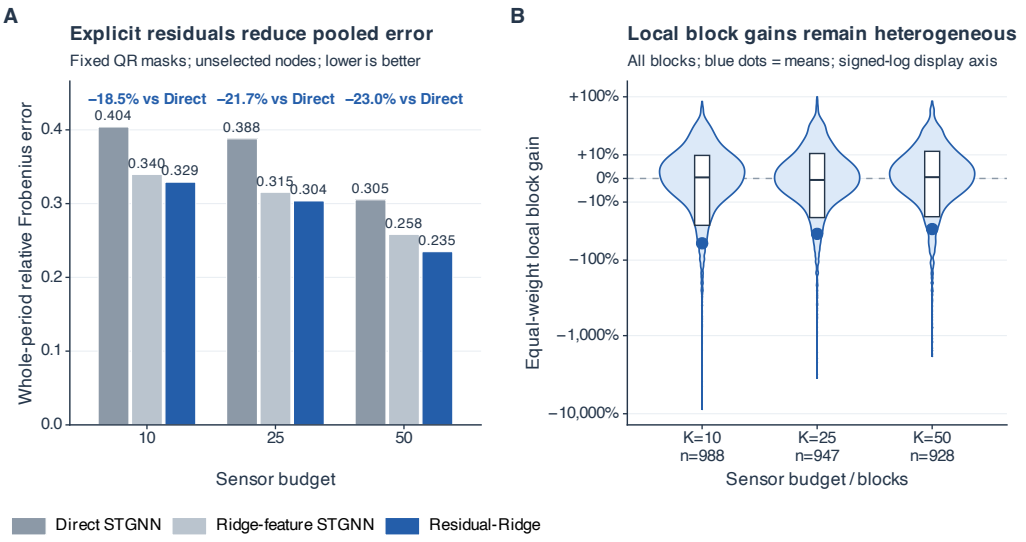

Figure 3: Residual correction of field reconstruction under fixed QR masks. (A) Whole-period relative Frobenius error on unselected nodes for Direct STGNN, Ridge-feature STGNN, and Residual-Ridge. Residual-Ridge reduced pooled error versus Direct STGNN by 18.52%, 21.71%, and 23.05% at K=10, 25, and 50. (B) Full distributions of equal-weight local block relative gains for Residual-Ridge versus Direct STGNN. For each block, $g_b = 1 - E_{R,b}/E_{D,b}$, where $E_{D,b} = \sqrt{\mathrm{SSE}_{D,b}}$ and $E_{R,b} = \sqrt{\mathrm{SSE}_{R,b}}$. The violins, quartile boxes, and blue mean dots use all 988, 947, and 928 blocks, respectively. Kernel density is estimated on the raw gain scale; the signed-log display axis shows the full negative tail without discarding blocks. Full-block means are -57.16%, -40.87%, and -34.20%.

B SUPPLEMENTARY TABLES

Table S1: Cross-domain system configurations and task-specific adaptations. All three systems used the same held-out evaluation periods, exact-K budgets (USGS and OISST: 10, 26, 51; San Diego: 14, 36, 72), selected-node histories, observation availability, scoring masks, and selected-current-observation copy rule. Selector and decoder training resources were method-specific; model capacity and compute were not matched. The Li-adapted and CAE-adapted systems are task-aligned implementations, not verbatim reproductions of their source papers.

Aspect	Proposed	Li-adapted	CAE-adapted
Method family and selector	QR Top-K anchor with single-swap or no-swap decision.	Static constrained PPO actor with exact-K Gumbel Top-K selection; actor reads graph and static node features, not the complete current field.	Entropy-initialized Concrete selector with straight-through exact Top-K; separate search for each dataset and budget.
Decoder	Ridge analytical estimate plus a 14-step temporal graph residual decoder; observed selected current values are copied to output.	Directional transport-aware GNN with six processing layers and latent width 64; selected current values are copied.	Low-rank masked temporal decoder for USGS and San Diego; small U-Net for OISST; selected current values are copied.
Selector training and selection	Five seeds (17, 29, 43, 59, 71); P fold training, A fold early stopping, B fold development report; maximum 180 epochs, patience 30.	Five actor seeds (17, 29, 43, 59, 71); P fold training and A fold checkpoint selection, B fold forward report; 480,000 warm and 160,000 constrained environment steps, PPO clip 0.2; selected actor checkpoint frozen without refit.	Five selector seeds (17, 29, 43, 59, 71); fit training and development-only epoch selection; maximum 120 epochs, 10 warm-up epochs, patience 20; temperature 1.5 to 0.2.
Decoder fitting and final refit	Three decoder seeds (17, 29, 43); random exact-K training masks mixed across the three budgets; from-scratch pre-test refit at development-selected fixed epochs.	Three DTA-GNN seeds (17, 29, 43); random exact-K training masks mixed across budgets; from-scratch pre-test refit at development-selected fixed epochs.	Three decoder seeds (17, 29, 43) per dataset and budget; from-scratch pre-test refit with the frozen CAE mask at development-selected fixed epochs.
Training objective and schedule	Selector uses expected/downside regression, positive-swap classification, and within-budget ranking; decoder uses global and site-balanced reconstruction loss. Decoder maximum 120 epochs, patience 20.	PPO reward is negative unselected-only observed-cell SSE/energy from the DTA ensemble; initial DTA decoder maximum 120 epochs, patience 20.	Search loss is unselected-only normalized reconstruction MSE plus 0.05 times all-site MSE; initial search maximum 120 epochs, patience 20.
Source-to-task adaptation	Analytical-anchor method defined in this study.	Based on the Li et al. placement/reconstruction family. Replaces sample-dependent access to a complete current field with a static, outcome-blind exact-K policy; adapts directional transport to scalar fields; uses a budgeted 640,000-step PPO schedule.	Based on the entropy-initialized Concrete-AE family of Turko et al. Replaces conditional PixelCNN entropy initialization with a fit-only Gaussian Ridge-residual proxy; uses hard exact-K selection and domain-specific decoders.

Table S2: Temporal splits and evaluation status. The N264 rows distinguish the 2020 mechanism study, two frozen-system evaluations using different selector versions, and an inference-only replay of an already-consumed period. Fitting and development periods are reported at the level documented for each experiment; a freeze does not imply retraining on the evaluation period.

Setting	Fitting/development period	Evaluation period and status	Exact-K budgets
USGS N264 mechanism study	2015–2018 fit; 2019 validation; 2020 development	2020 development evaluation	10, 25, 50
N264 earlier frozen system (v5+v4)	Frozen before its test; pre-2021 development	2021–2022 frozen test	10, 25, 50
N264 final frozen system (v6.1+v4)	Frozen before its test; developed after the earlier 2021–2022 test had been consumed	2023–2025 independent future evaluation	10, 25, 50
N264 final-system retrospective replay	Same frozen v6.1+v4 system; no retraining or tuning	Previously consumed 2021–2022; retrospective replay	10, 25, 50
USGS N512 cross-domain study	2015–2020 development	2021–2022 evaluation	10, 26, 51
LargeST San Diego cross-domain study	2017–2019 fitting; 2020 development	2021 evaluation	14, 36, 72
OISST Kuroshio cross-domain study	2016–2018 fitting; 2019 development	2020 evaluation	10, 26, 51