UNCERTAINTY-CALIBRATED LABEL DENSITY REGULARIZATION FOR FEW-LABELED SOURCE-FREE REGRESSION DOMAIN ADAPTATION

Anonymous authors
Paper under double-blind review

ABSTRACT

Few-labeled source-free regression adapts a pretrained regressor using scarce labeled target data and abundant unlabeled target inputs, without accessing source samples. Observed responses provide verified but sparse supervision, whereas predictions offer broader coverage but may be biased under domain shift. We propose Uncertainty-Calibrated Label Density Regularization (UCLDR), which uses predictions as uncertainty-regulated response evidence rather than directly fitting them as hard point-wise pseudo-labels. Uncertainty controls the admission, probability mass, and spread of predicted responses. The mixture supplies global mean–variance constraints, while its constituent assignments connect unlabeled features to labeled prototypes in overlapping response regions. This shared representation supports both forms of supervision without a separate trainable density model. At ten target labels, UCLDR achieves the lowest mean MSE on three of four benchmarks and the best average rank of 1.25 among the evaluated adaptation methods. On Cycle, it reduces mean MSE relative to CRAFT by 3.45–10.36% across five label budgets, with lower error on 7 to 11 of 12 transfers at each budget.

1 INTRODUCTION

Source-free regression adaptation transfers a pretrained model to a new domain when source samples are unavailable. A few target labels can help recover accuracy, but collecting continuous responses is often costly (Li et al., 2024; Liang et al., 2020; Li et al., 2021; Biswas et al., 2025). Observed target responses establish verified input–response correspondences, but cover only a small subset of target inputs. Extending supervision through point-wise pseudo-labels preserves input-specific response assignments, but can propagate prediction errors under domain shift. In contrast, marginal response statistics capture population-level structure without identifying which response belongs to which input. The central challenge is to exploit these complementary forms of supervision without treating uncertain predictions as verified labels.

This limitation matters for regression: identical response moments can arise from different input–response assignments, leading to substantially different prediction errors. CRAFT combines scarce labels with a marginal-aware objective (Biswas et al., 2025), TASFAR uses a confidence-derived label-density map for pseudo-label generation (He et al., 2024), and MERCI models instance-dependent conditional response densities (Xu et al., 2026). These approaches provide different forms of response supervision. Here, we investigate how uncertain prediction-derived evidence can support both population-level constraints and input-related supervision without being treated as verified point targets.

We propose *Uncertainty-Calibrated Label Density Regularization* (UCLDR). Each observed response or admitted prediction contributes a kernel on a common response grid. For prediction-derived evidence, predictive-mean uncertainty serves three coordinated roles: admission control determines whether a prediction contributes, mass allocation determines its probability mass, and spread modulation determines its response-space precision. Per-kernel normalization keeps the assigned probability mass separate from its spread: broadening a kernel changes where its evidence is placed without changing its total contribution. The labeled and predicted components have explicit mixture masses, so the amount of unlabeled data does not by itself determine their relative influence.

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.

The key is to preserve both the marginal statistics and the constituent soft assignments of a shared response representation. The former constrain the overall location and scale of predictions, while the latter enable local feature supervision through label-supported prototypes within overlapping response regions. By retaining these complementary forms of information, UCLDR combines global moment regularization with input-related supervision anchored by observed target labels, without fitting uncertain predictions as hard point targets.

At ten target labels, UCLDR achieves the lowest mean MSE on Bike, Cycle, and Housing and the second-lowest on Kin8xy, giving an average rank of 1.25. On Housing, the reported MSEs correspond to a 59.35% reduction relative to CRAFT. The Cycle comparison extends across all five label budgets from 10 to 50: mean MSE is 3.45–10.36% lower than CRAFT, with gains on a majority of the 12 directed transfers at each budget. Component comparisons show that adding local supervision improves the safeguarded density objective at every Cycle budget, while a separate replacement study measures the contribution of uncertainty-based mass allocation on Bike and Kin8xy.

The paper makes two contributions:

- A shared, uncertainty-regulated response representation that integrates global statistical constraints with label-anchored local supervision. Normalized response kernels preserve individual soft assignments alongside marginal moments, allowing both forms of supervision to be derived from the same evidence without fitting uncertain predictions as hard point targets.
- Evidence across four benchmarks, with the best average rank at ten target labels and a Cycle advantage over five label budgets, supported by transfer-level comparisons and controlled studies of local supervision and uncertainty-based mass allocation.

2 Related Work

Source-free regression adaptation. Domain adaptation methods align feature distributions through adversarial, moment-matching, correlation-based, optimal-transport, and related objectives (Ganin & Lempitsky, 2015; Long et al., 2019; Sun & Saenko, 2016; Courty et al., 2017; Gretton et al., 2012; Chen et al., 2020). Regression-specific approaches further exploit output-relevant subspaces, distribution information, conditional dependence, and inverse-Gram structure (Cortes & Mohri, 2011; Xia et al., 2022; Chen et al., 2021; Wu et al., 2022; Chen et al., 2026a;b; Nejjar et al., 2023). In the source-free setting, adaptation retains the pretrained model but cannot revisit its source samples (Li et al., 2024; Liang et al., 2020; Yang et al., 2021a; Cai et al., 2023; Chen et al., 2025; Hou et al., 2025). TASFAR builds a label-density map from confident predictions (He et al., 2024); BRR combines reliability evaluation, bias reduction, and pseudo-label refinement (Zhan et al., 2025); and SSA aligns feature distributions in an output-relevant subspace (Adachi et al., 2025). CRAFT is closest to our supervision setting, using scarce labeled and abundant unlabeled target data (Biswas et al., 2025). UCLDR's specific design is to mix observed and predicted responses with explicit probability masses, then reuse their response assignments for local feature supervision alongside marginal moment constraints.

Uncertainty and distributional supervision. Predictive uncertainty can be estimated using Bayesian approximations and ensembles (Gal & Ghahramani, 2016; Lakshminarayanan et al., 2017). Another approach, deep evidential regression, parameterizes a Normal–Inverse-Gamma distribution (Amini et al., 2020; Meinert et al., 2023; Wu et al., 2024; Ye et al., 2024), whose reliability can change under dataset shift (Ovadia et al., 2019). UGA uses evidential uncertainty for feature and posterior alignment (Nejjar et al., 2026). UCLDR instead uses predictive-mean uncertainty to regulate the contribution of prediction-derived response evidence, without assuming that absolute uncertainty calibration is preserved under domain shift (Litrico et al., 2023; Lee & Lee, 2023; Lee et al., 2025; Xu et al., 2025a;b). Label distribution learning and regression smoothing also encode structure around continuous responses (Geng, 2016; Yang et al., 2021b; Chen & Ji, 2018; Jeon & Taylor, 2012). MERCI learns an instance-dependent discretized conditional density (Xu et al., 2026). UCLDR instead forms a target-level mixture from observed labels and prediction kernels, using its moments and assignments directly in the adaptation loss. The construction adds no trainable density network.

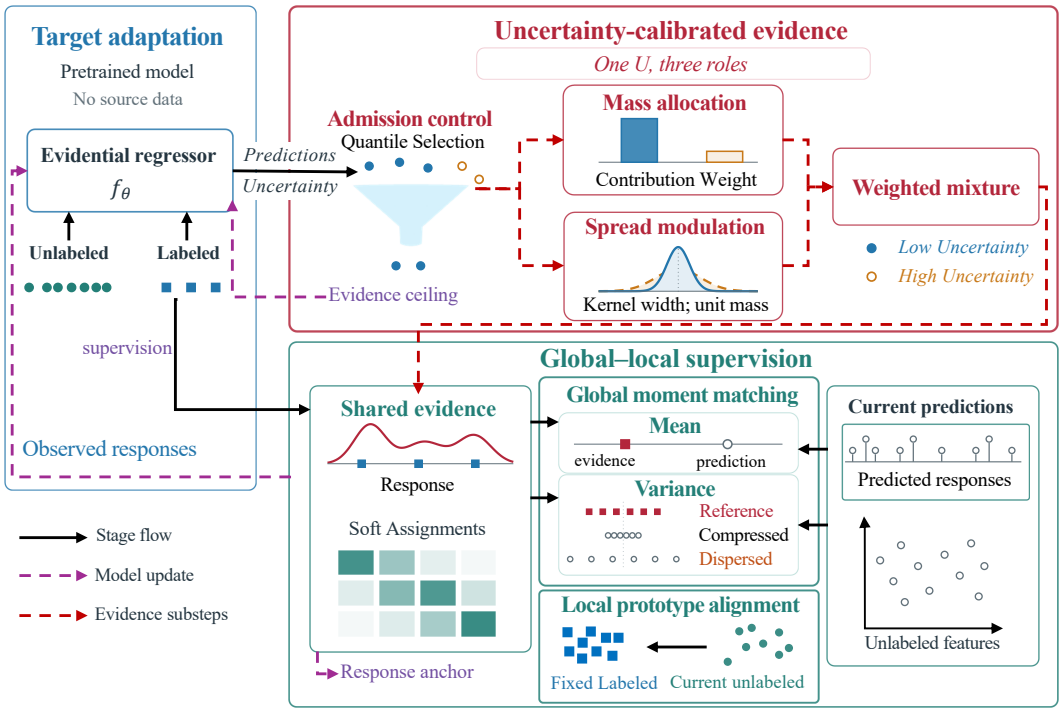

Figure 1: **UCLDR couples response-moment constraints with local feature supervision.** Observed labels and uncertainty-regulated predictions contribute normalized response kernels. Their mixture defines reference moments, while their assignments group labeled and unlabeled features into response-conditioned prototypes. An evidence ceiling and response anchor constrain optimization.

3 Method

UCLDR uses one set of response kernels to connect global output statistics and local feature supervision (Figure 1). The construction has three steps: represent labeled and predicted responses with explicit probability masses, derive moments and response-conditioned assignments from those kernels, and optimize predictions and features against the cached evidence.

3.1 Setting and Predictive-Mean Uncertainty

Let f_{θ_s} be a source-pretrained evidential regressor. Target adaptation has access to this model, a labeled set $\mathcal{D}_t^l = \{(x_i^l, y_i^l)\}_{i=1}^{N_l}$, and an unlabeled set $\mathcal{D}_t^u = \{x_j^u\}_{j=1}^{N_u}$, with $N_l \ll N_u$. The target subsets are sampled from the same target population; source and target share input and response semantics. Source samples are unavailable during adaptation, and target-test responses are reserved for evaluation.

The Normal–Inverse-Gamma (NIG) output is $f_\theta(x) = (\gamma_\theta(x), \nu_\theta(x), \alpha_\theta(x), \beta_\theta(x))$, with $\nu_\theta, \beta_\theta > 0$, $\alpha_\theta > 1$, and predictive mean $h_\theta(x) = \gamma_\theta(x)$. We use the variance of that mean and its standard deviation:

$$U_\theta(x) = \frac{\beta_\theta(x)}{\nu_\theta(x)(\alpha_\theta(x) - 1)}, \qquad S_\theta(x) = \sqrt{U_\theta(x)}. \tag{1}$$

Here $U_\theta(x)$ measures uncertainty about the response estimate used as a kernel center. We initialize the target model at θ_s and warm it up on the labeled targets using

$$\mathcal{L}_{\text{sup}} = \frac{1}{N_l} \sum_{i=1}^{N_l} \left[\mathcal{L}_{\text{NLL}}(f_\theta(x_i^l), y_i^l) + \lambda_R |y_i^l - \gamma_\theta(x_i^l)| (2\nu_\theta(x_i^l) + \alpha_\theta(x_i^l)) \right], \tag{2}$$

where $\mathcal{L}_{\text{NLL}}$ is the NIG negative log marginal likelihood (Amini et al., 2020). Warm-up adapts predictions and uncertainty to the observed target responses. Subsequent refreshes use this adapted

model. Here, *uncertainty-calibrated* refers to regulating how prediction-derived evidence enters the shared response representation: uncertainty controls its admission, probability mass, and response-space spread. It does not denote post-hoc calibration or assume that absolute uncertainty estimates remain calibrated under domain shift. We assess the uncertainty–error relationship empirically in Section 4.4.

3.2 Constructing Shared Response Evidence

Admission, mass, and spread. At refresh r, cache $\widehat{y}_j^{u,(r)} = \mathrm{sg}[h_{\theta_r}(x_j^u)]$ and $U_j^{(r)} = \mathrm{sg}[U_{\theta_r}(x_j^u)]$, where sg stops gradients. With $S_j^{(r)} = \sqrt{U_j^{(r)}}$, define

$$\begin{aligned} &\bar{\tau}^{(r)} = Q_q(\{S_j^{(r)}\}_{j=1}^{N_u}), \\ &\tau^{(r)} = \begin{cases} \bar{\tau}^{(r)}, & r = 1, \\ \rho\tau^{(r-1)} + (1-\rho)\bar{\tau}^{(r)}, & r > 1, \end{cases} \qquad \mathcal{C}^{(r)} = \{j : S_j^{(r)} \le \tau^{(r)}\}. \end{aligned} \tag{3}$$

The quantile q controls nominal coverage and the exponential moving average (EMA) coefficient ρ smooths the admission threshold. The retained fraction can differ from q as the model and threshold evolve.

For each admitted prediction,

$$\omega_j^{(r)} = \frac{(U_j^{(r)} + \epsilon)^{-1}}{\sum_{k \in \mathcal{C}^{(r)}} (U_k^{(r)} + \epsilon)^{-1}}, \qquad w_j^{u,(r)} = m_u \omega_j^{(r)}, \qquad \sigma_j^{u,(r)} = \sqrt{U_j^{(r)} + \epsilon_\sigma}. \tag{4}$$

Admission control makes a coarse reliability decision about which predictions can enter the representation. Within the admitted set, mass allocation differentiates their contributions, while spread modulation expresses their remaining response-space imprecision. Thus, uncertainty controls whether, how much, and how precisely prediction-derived evidence enters the shared representation.

Mass-preserving mixture. A shared grid $\{c_b^{(r)}\}_{b=1}^B$ spans the observed labels and admitted prediction centers, with a boundary margin and spacing $\Delta_c^{(r)}$ (Appendix A.1). Each observed label has mass m_l/N_l and bandwidth $\sigma_l^{(r)} = \kappa_\sigma \Delta_c^{(r)}$. Let $m_l + m_u = 1$, with $m_l = m_u = 1/2$ by default. Define

$$\bar{\kappa}_\sigma(c_b, z) = \frac{\exp[-(c_b - z)^2/(2\sigma^2)]}{\sum_{b'=1}^B \exp[-(c_{b'} - z)^2/(2\sigma^2)]}. \tag{5}$$

The shared response representation is

$$\widehat{p}_{t,b}^{(r)} = \frac{m_l}{N_l} \sum_i \bar{\kappa}_{\sigma_l^{(r)}}(c_b^{(r)}, y_i^l) + m_u \sum_{j \in \mathcal{C}^{(r)}} \omega_j^{(r)} \bar{\kappa}_{\sigma_j^{u,(r)}}(c_b^{(r)}, \widehat{y}_j^{u,(r)}). \tag{6}$$

Every kernel sums to one on the finite grid, so $\widehat{p}_{t,b}^{(r)}$ is probability mass and $\sum_b \widehat{p}_{t,b}^{(r)} = 1$. Broadening a kernel redistributes its assigned mass, including near the grid boundary, without changing its total contribution. Discrete moments are computed from these normalized grid masses. If no prediction is admitted, we use $m_l = 1$, $m_u = 0$ and the labeled-only representation.

3.3 Global and Local Adaptation

Global response location and scale. The representation in equation 6 supplies the mean $\mu_p^{(r)} = \sum_b c_b^{(r)} \widehat{p}_{t,b}^{(r)}$ and variance $v_p^{(r)} = \sum_b (c_b^{(r)} - \mu_p^{(r)})^2 \widehat{p}_{t,b}^{(r)}$. For an unlabeled mini-batch $\mathcal{B}_u$ of size M_u, let

$$\mu_h = \frac{1}{M_u} \sum_{x_j^u \in \mathcal{B}_u} h_\theta(x_j^u), \qquad v_h = \frac{1}{M_u} \sum_{x_j^u \in \mathcal{B}_u} (h_\theta(x_j^u) - \mu_h)^2. \tag{7}$$

We regularize these prediction statistics with

$$\mathcal{L}_{\mathrm{dist}} = |\mu_h - \mu_p^{(r)}| + \eta_{\mathrm{var}} |v_h - v_p^{(r)}|. \tag{8}$$

The mean term constrains response offset; the variance term constrains compression or dispersion relative to the response evidence. Both variances use uncorrected second central moments. The evidence statistics are cached, while gradients pass through current predictions. Moment regularization constrains the marginal location and scale of current predictions, but cannot by itself determine which response belongs to each input. The full grid is therefore retained not for binwise density matching, but for the response assignments used by the local branch below.

Local response-conditioned supervision. The constituent kernels also define soft assignments

$$a_{ib}^{l,(r)} = \bar{\kappa}_{\sigma_l^{(r)}}(c_b^{(r)}, y_i^l), \qquad a_{jb}^{u,(r)} = \bar{\kappa}_{\sigma_j^{u,(r)}}(c_b^{(r)}, \widehat{y}_j^{u,(r)}). \tag{9}$$

Let ϕ_θ denote the features before the regression head and $\mathcal{C}_{\mathcal{B}}^{(r)} = \{j \in \mathcal{C}^{(r)} : x_j^u \in \mathcal{B}_u\}$. Renormalize the reliability weights within this set as $\widetilde{\omega}_j^{(r)} = \omega_j^{(r)} / \sum_{k \in \mathcal{C}_{\mathcal{B}}^{(r)}} \omega_k^{(r)}$. For nonempty $\mathcal{C}_{\mathcal{B}}^{(r)}$, define

$$\mathbf{m}_b^{l,(r)} = \mathrm{sg}\left[\frac{\sum_i a_{ib}^{l,(r)} \phi_{\theta_r}(x_i^l)}{\sum_i a_{ib}^{l,(r)} + \epsilon}\right], \qquad \mathbf{m}_b^{u,(r)} = \frac{\sum_{j \in \mathcal{C}_{\mathcal{B}}^{(r)}} \widetilde{\omega}_j^{(r)} a_{jb}^{u,(r)} \phi_\theta(x_j^u)}{\sum_{j \in \mathcal{C}_{\mathcal{B}}^{(r)}} \widetilde{\omega}_j^{(r)} a_{jb}^{u,(r)} + \epsilon}. \tag{10}$$

The assignments retain which inputs contributed to each response region, information that the mixture moments alone do not retain. Labeled prototypes and assignments are cached; unlabeled prototypes use current features and retain gradients. Define the component masses $p_b^{l,(r)} = N_l^{-1} \sum_i a_{ib}^{l,(r)}$ and $p_b^{u,(r)} = \sum_{j \in \mathcal{C}^{(r)}} \omega_j^{(r)} a_{jb}^{u,(r)}$. Their overlap weights emphasize regions supported by both evidence sources:

$$\pi_b^{(r)} = \frac{\min(p_b^{l,(r)}, p_b^{u,(r)})}{\sum_{b'} \min(p_{b'}^{l,(r)}, p_{b'}^{u,(r)}) + \epsilon}, \qquad \mathcal{L}_{\mathrm{semi}} = \sum_b \pi_b^{(r)} \|\mathbf{m}_b^{u,(r)} - \mathbf{m}_b^{l,(r)}\|_1. \tag{11}$$

The overlap weights use the full admitted target set and remain fixed between refreshes. We set $\mathcal{L}_{\mathrm{semi}} = 0$ for an empty admitted mini-batch. This branch transfers feature structure within label-supported response regions rather than fitting unlabeled predictions as point targets.

Output statistics and response assignment. For the population variables $H = h_\theta(X)$ and target response Y, let μ_h, v_h and μ_t, v_t denote their respective means and positive variances. Squared risk decomposes as

$$\mathbb{E}(H - Y)^2 = (\mu_h - \mu_t)^2 + (\sqrt{v_h} - \sqrt{v_t})^2 + 2(1 - \mathrm{Corr}(H, Y))\sqrt{v_h v_t}. \tag{12}$$

The first two terms measure errors in response location and scale. The last term depends on which response is assigned to each input. For example, for two equally likely inputs with responses $-a$ and a, $a > 0$, predicting the responses in the correct order or in reverse gives the same output mean and variance. The respective MSEs are 0 and $4a^2$. This distinction motivates pairing moment regularization with labeled supervision and response-conditioned feature constraints. Appendix B derives the identity and its conditional risk bound.

3.4 Safeguards and Optimization

Two safeguards regulate optimization. An *evidence ceiling* penalizes $e_\theta(x) = 2\nu_\theta(x) + \alpha_\theta(x)$ on unlabeled inputs only above a reference formed from the median and median absolute deviation (MAD) of labeled-target evidence. A *response anchor* penalizes discrepancies between the mean labeled prediction and the evidence mean outside a label-dependent tolerance; its weight attenuates when the evidence mean is incompatible with the observed labels. Their exact definitions are in Appendix A.2. The complete objective is

$$\mathcal{L}_{\mathrm{UCLDR}} = \mathcal{L}_{\mathrm{sup}} + \lambda_{\mathrm{dist}}\mathcal{L}_{\mathrm{dist}} + \lambda_{\mathrm{semi}}\mathcal{L}_{\mathrm{semi}} + \lambda_{\mathrm{ev}}\mathcal{L}_{\mathrm{ev}} + \lambda_{\mathrm{anc}}\mathcal{L}_{\mathrm{anc}}. \tag{13}$$

The first E_w epochs optimize $\mathcal{L}_{\mathrm{sup}}$. Evidence is first constructed before epoch $E_w + 1$ and refreshed every R epochs. Between refreshes, cached predictions, assignments, moments, labeled prototypes, overlap weights, and safeguard references remain fixed. Current predictions and unlabeled features are optimized through equation 13. This schedule reuses the evidence without a separate trainable density model. Appendix A.3 gives the full update sequence.

4 EXPERIMENTS

We evaluate UCLDR on four regression benchmarks under scarce target supervision. The main comparison examines performance across benchmarks at ten labeled target samples, while a label-budget study on Cycle examines how the results change as more target labels become available. Component studies investigate the contribution of local response-conditioned supervision and uncertainty-regulated evidence, and response-level diagnostics examine the reliability and coverage of the constructed evidence.

4.1 EXPERIMENTAL SETUP

Benchmarks and comparisons. We evaluate Bike, Cycle, Housing, and Kin8xy, each with four domains and all 12 directed source–target transfers. Bike uses seasonal domains, Housing uses the four non-island `ocean_proximity` categories, and Kin8xy uses `kin-8fh`, `kin-8fm`, `kin-8nh`, and `kin-8nm`. Cycle is the Combined Cycle Power Plant benchmark, with four domains defined by ambient-temperature quartiles. Its inputs are ambient temperature, exhaust vacuum, ambient pressure, and relative humidity; the response is electrical energy output. The primary budget is $N_l = 10$; Cycle additionally covers $N_l \in \{20, 30, 40, 50\}$.

The baselines are pseudo-label self-training, uncertainty-filtered self-training, CRAFT (Biswas et al., 2025), SSA (Adachi et al., 2025), and MERCI (Xu et al., 2026). The first two fit point predictions, with the latter restricting pseudo-label use by uncertainty. CRAFT supplies the closest comparison in supervision setting; SSA and MERCI supply subspace- and density-based source-free comparisons. Source Only and fully supervised Target Only use different supervision protocols and serve as contextual references. Neither enters the adaptation ranking.

Training and evaluation. The UCLDR main benchmark and uncertainty-role study use the separate protocols specified in Appendix C. Within the uncertainty-role study, each replacement is paired with its canonical configuration on the same transfer, label budget, split, and seed. Main UCLDR runs use Adam, learning rate 10^{-4}, batch size 32, up to 200 epochs, and early-stopping patience 30. We use $E_w = 10$, $R = 20$, $q = 0.4$, $\rho = 0.9$, $B = 100$, $(m_l, m_u) = (0.5, 0.5)$, and $(\lambda_{\text{dist}}, \lambda_{\text{ev}}, \lambda_{\text{anc}}, \lambda_{\text{semi}}) = (0.5, 0.1, 0.4, 1.0)$. Additional numerical settings and the computational profile appear in Appendix C. The UCLDR checkpoint minimizes the epoch-averaged training objective, with selection beginning after the first density refresh. Early stopping counts epochs without improvement in that objective. Target-test responses are excluded from optimization, early stopping, and model selection. Response-dependent diagnostics are evaluated after checkpoint selection.

Target-test MSE is the primary metric. For UCLDR, it is the mean squared difference between predictions and responses in the recorded evaluation coordinates, with no additional reporting scale factor. We first average three seeds within each matched transfer–budget condition, then average the 12 transfer values within each dataset. Cross-dataset comparisons use ranks or relative error changes rather than pooling raw errors across response scales. The reported means, transfer win counts, and uncertainty bands are descriptive summaries. Directed transfers reuse source and target domains, so the transfer counts are not independent replication counts. The scope of each diagnostic is specified separately.

4.2 PREDICTIVE PERFORMANCE ACROSS BENCHMARKS AND LABEL BUDGETS

On Cycle, UCLDR obtains lower mean MSE than CRAFT at every evaluated budget from 10 to 50 target labels (Figure 2). Its mean MSE decreases from 1.445 to 0.781 over this range, with relative reductions of 3.45–10.36% against CRAFT. The advantage therefore persists as additional labels become available.

The transfer-level results show where these mean gains occur (Table 1). UCLDR improves 7 of 12 directed transfers at ten labels and 11 of 12 at fifty labels. Median paired reductions are positive at every budget, ranging from 4.80% to 13.64%. Thus, the mean advantage coincides with gains across a majority of transfers at each budget. The full budget-level MSEs appear in Appendix D.

Results across benchmarks. At ten target labels, UCLDR has the lowest mean MSE on Bike, Cycle, and Housing and the second-lowest on Kin8xy (Table 2). Its average rank is 1.25, followed

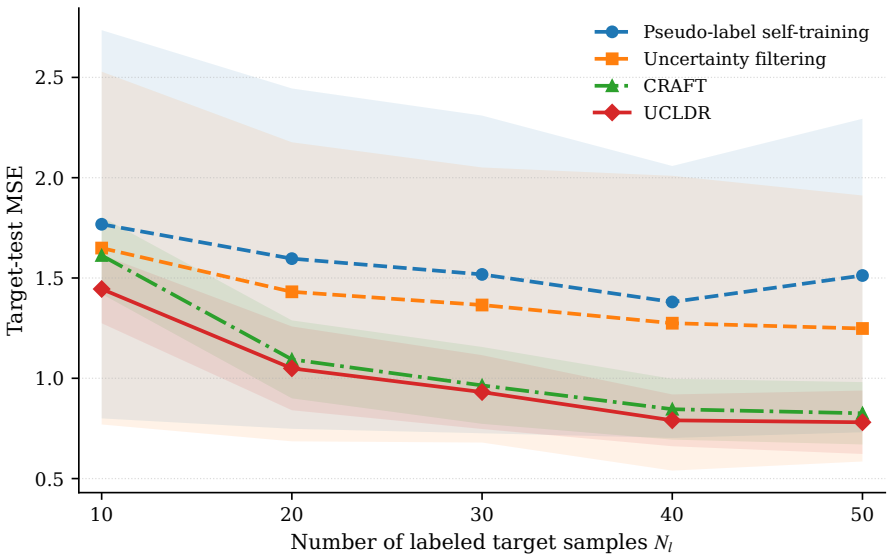

Figure 2: **The Cycle advantage persists across all evaluated label budgets.** Points are mean target-test MSE across 12 transfers after seed aggregation; bands are descriptive 95% intervals across transfer-level aggregates. The plot compares UCLDR, CRAFT, and the two self-training baselines.

Table 1: **Cycle gains extend across directed transfers.** Comparisons use three-seed means within each transfer. Mean MSE reduction compares the two methods' transfer averages; median reduction is computed from the 12 paired relative reductions. Counts and medians are descriptive summaries.

Target labels	10	20	30	40	50
Mean MSE reduction vs. CRAFT (%)	10.36	4.09	3.45	6.57	5.43
Transfers with lower MSE / 12	7	7	8	10	11
Median paired reduction (%)	13.64	4.80	6.14	5.83	7.66

by CRAFT at 1.75. The largest reported separation is on Housing: MSE is 1.449 for UCLDR and 3.565 for CRAFT, a 59.35% reduction. On Cycle, the ten-label reduction is 10.36%, with the budget study above showing that this advantage persists as more labels become available. Together, the cross-benchmark and budget comparisons characterize the predictive value of the shared response representation under scarce target supervision.

Table 2: **UCLDR leads on three benchmarks and has the best average rank at ten target labels.** Mean target-test MSE over 12 transfers after averaging three seeds within each transfer. Bold and underlined entries denote the best and second-best adaptation results. Average rank is computed across the four datasets. Reference models are excluded from ranking.

Method	Bike	Cycle	Housing	Kin8xy	Avg. rank
Reference models					
Target Only (full supervision)	1.483	1.250	6.997	1.351	–
Source Only	2.261	2.447	5.287	4.473	–
Adaptation methods					
Pseudo-label self-training	2.123	1.768	6.167	3.636	5.25
Uncertainty filtering only	1.908	1.649	4.808	3.627	3.50
CRAFT	<u>1.623</u>	<u>1.612</u>	<u>3.565</u>	**0.988**	<u>1.75</u>
SSA	1.986	11.096	4.596	4.077	4.50
MERCI	2.019	26.943	5.096	2.663	4.75
UCLDR	**1.611**	**1.445**	**1.449**	<u>1.419</u>	**1.25**

4.3 Component and Uncertainty-Control Analyses

Adding the local branch. Adding response-conditioned local supervision improves the safeguarded density objective at each Cycle label budget. At ten labels, mean MSE changes from 2.503 to 1.445. This contrast holds density-moment regularization and both safeguards active, isolating the addition of $\mathcal{L}_{\text{semi}}$ within that objective. The feature branch uses the density kernels to group inputs by response and align them with labeled prototypes. Its measured benefit supports this local use of the shared representation. Appendix D gives the component map and all five budget comparisons.

Sensitivity to uncertainty controls. We replace one construction choice at a time: uncertainty-based admission ordering with random ordering, the EMA boundary with a fixed threshold, inverse-variance masses with uniform weights, or uncertainty-dependent widths with a fixed bandwidth. This separate controlled study uses a fixed target split and frozen source-checkpoint normalization (Appendix C). It covers Bike, Housing, and Kin8xy, all 12 transfers, budgets $N_l \in \{10, 20, 30, 50\}$, and three seeds. We first average RMSE over the three seeds within each matched transfer–budget condition, then compute the relative change between the replacement and canonical configurations:

$$\Delta\text{RMSE} = 100\frac{\text{RMSE}_{\text{replacement}} - \text{RMSE}_{\text{canonical}}}{\text{RMSE}_{\text{canonical}}}. \tag{14}$$

Each dataset contributes 48 matched conditions. With complete UCLDR, uniform weighting increases mean relative RMSE by 1.79% across datasets, including 3.56% on Bike and 1.80% on Kin8xy. The corresponding average increase is 0.53% when the local branch is disabled. These results identify a predictive contribution from uncertainty-based mass allocation and show that its measured effect depends on the downstream objective. Appendix E reports every replacement in both contexts.

The effects of admission and bandwidth replacements vary by dataset. Housing entries are 0.00% at the displayed precision; saved canonical checkpoints have nearly uniform effective weights, making inverse-variance and uniform allocation effectively equivalent in those states. The complete replacement matrix and checkpoint diagnostics are reported in Appendix E.

4.4 Evidence Reliability and Coverage

Post-selection diagnostics show that uncertainty ordering depends on the adapted predictor and dataset. They use 36 saved configurations on Bike, Housing, and Kin8xy, spanning two transfers, two label budgets, and three seeds; target responses enter after checkpoint selection. Retaining the 40% lowest-uncertainty predictions changes RMSE from 1.026 to 0.769 on Bike and from 1.242 to 0.913 on Housing. The corresponding Kin8xy values are 0.875 and 0.982. Together with the replacement study, these results motivate assessing admission, mass allocation, and bandwidth separately (Appendices E and F.1).

Coverage also affects the amount of usable evidence. The $q \in \{0.1, 0.2, 0.4, 0.6, 0.8, 1.0\}$ study on Housing and Kin8xy tracks retained fraction, effective sample size, and prediction error together (Appendix F.3). Density-fidelity results distinguish 26 mixed-evidence configurations from 10 labeled-only fallbacks (Appendix F.2). These diagnostics distinguish the error of individual predictions from the response coverage of the retained pool.

5 Conclusion

UCLDR adapts a source-pretrained regressor by representing scarce observed responses and uncertain predictions with shared response kernels. Their normalized masses supply global output constraints, while their assignments connect unlabeled features to labeled prototypes. At ten target labels, UCLDR achieves the lowest mean MSE on three of four benchmarks and the best average rank of 1.25. On Cycle, its advantage over CRAFT persists across five label budgets and a majority of directed transfers at each budget. Together with the local-branch comparisons, these results support using shared response evidence to connect population statistics with label-supported feature structure.

AI USE STATEMENT

Generative AI tools were used to aid and polish manuscript writing, assist with literature retrieval and discovery, and help draft portions of the paper. They were also used during research execution to assist with method implementation and code development, and to assist with parts of the mathematical derivations and proofs presented in the paper. All AI-assisted code was reviewed and tested by the authors, retrieved references were verified against their original sources, and AI-assisted mathematical arguments were independently checked by the authors. The authors take full responsibility for the final manuscript, including its implementation, citations, mathematical arguments, empirical results, and claims.

REPRODUCIBILITY STATEMENT

Section 3 specifies the response representation and adaptation objectives. Appendix A gives the dynamic grid, safeguards, and update sequence; Appendix B provides derivations; Appendix C records numerical settings and the computational profile. The remaining appendices give complete objective comparisons, both uncertainty-replacement contexts, and response-level and seed-level diagnostics with their aggregation scopes.

REFERENCES

Kazuki Adachi, Shin'ya Yamaguchi, Atsutoshi Kumagai, and Tomoki Hamagami. Test-time adaptation for regression by subspace alignment. In *International Conference on Learning Representations*, volume 2025, pp. 77850–77874, 2025. URL https://proceedings.iclr.cc/paper_files/paper/2025/file/c19c4def293b6a63db1ff27143dd4f10-Paper-Conference.pdf.

Alexander Amini, Wilko Schwarting, Ava Soleimany, and Daniela Rus. Deep evidential regression. In *Advances in Neural Information Processing Systems*, volume 33, pp. 14927–14937, 2020. URL https://proceedings.neurips.cc/paper/2020/hash/aab085461de182608ee9f607f3f7d18f-Abstract.html.

Mainak Biswas, Ambedkar Dukkipati, and Devarajan Sridharan. Semi-supervised deep transfer for regression without domain alignment. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pp. 827–836, 2025. doi: 10.1109/ICCV51701.2025.00085.

Xin Cai, Jiabei Zeng, Shiguang Shan, and Xilin Chen. Source-free adaptive gaze estimation by uncertainty reduction. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 22035–22045, 2023. doi: 10.1109/CVPR52729.2023.02110.

Chao Chen, Zhihang Fu, Zhihong Chen, Sheng Jin, Zhaowei Cheng, Xinyu Jin, and Xian-Sheng Hua. HoMM: Higher-order moment matching for unsupervised domain adaptation. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 34, pp. 3422–3429, 2020. doi: 10.1609/aaai.v34i04.5745.

Hongrui Chen, Yiming Zhai, Mingjun Pan, and You-Wei Luo. BiCD: Learning conditional dependence for continuous dataset shift in regression. In *Proceedings of the ACM SIGKDD Conference on Knowledge Discovery and Data Mining*, pp. 71–81, 2026a. doi: 10.1145/3770854.3780210.

Lisha Chen and Qiang Ji. Kernel density network for quantifying regression uncertainty in face alignment. In *Third Workshop on Bayesian Deep Learning at NeurIPS*, 2018.

Xinyang Chen, Sinan Wang, Jianmin Wang, and Mingsheng Long. Representation subspace distance for domain adaptation regression. In *Proceedings of the 38th International Conference on Machine Learning*, volume 139 of *Proceedings of Machine Learning Research*, pp. 1749–1759, 2021. URL https://proceedings.mlr.press/v139/chen21u.html.

Zhuohang Chen, Jinglong Chen, Tongyang Pan, and Jingsong Xie. Uncertainty estimation pseudo-label-guided source-free domain adaptation for cross-domain remaining useful life prediction in IIoT. *IEEE Internet of Things Journal*, 12(1):236–249, 2025. doi: 10.1109/JIOT.2024.3464854.

Zi-Ying Chen, Chuan-Xian Ren, and Hong Yan. Label-guided optimal transport for domain adaptation regression. *Pattern Recognition*, 179:113697, 2026b. doi: 10.1016/j.patcog.2026.113697.

Corinna Cortes and Mehryar Mohri. Domain adaptation in regression. In *Algorithmic Learning Theory*, pp. 308–323, 2011. doi: 10.1007/978-3-642-24412-4_25.

Nicolas Courty, Rémi Flamary, Devis Tuia, and Alain Rakotomamonjy. Optimal transport for domain adaptation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 39(9):1853–1865, 2017. doi: 10.1109/TPAMI.2016.2615921.

Yarin Gal and Zoubin Ghahramani. Dropout as a bayesian approximation: Representing model uncertainty in deep learning. In *Proceedings of the 33rd International Conference on Machine Learning*, volume 48 of *Proceedings of Machine Learning Research*, pp. 1050–1059, 2016. URL https://proceedings.mlr.press/v48/gal16.html.

Yaroslav Ganin and Victor Lempitsky. Unsupervised domain adaptation by backpropagation. In *Proceedings of the 32nd International Conference on Machine Learning*, volume 37, pp. 1180–1189. PMLR, 2015. URL https://proceedings.mlr.press/v37/ganin15.html.

Xin Geng. Label distribution learning. *IEEE Transactions on Knowledge and Data Engineering*, 28 (7):1734–1748, 2016. doi: 10.1109/TKDE.2016.2545658.

Arthur Gretton, Karsten M. Borgwardt, Malte J. Rasch, Bernhard Schölkopf, and Alexander Smola. A kernel two-sample test. *Journal of Machine Learning Research*, 13(25):723–773, 2012. URL https://www.jmlr.org/papers/v13/gretton12a.html.

Tianlang He, Zhiqiu Xia, Jierun Chen, Haoliang Li, and S.-H. Gary Chan. Target-agnostic source-free domain adaptation for regression tasks. In *IEEE 40th International Conference on Data Engineering*, pp. 1464–1477, 2024. doi: 10.1109/ICDE60146.2024.00121.

Yubo Hou, Yucheng Wang, Min Wu, Chee-Keong Kwoh, Xiaoli Li, and Zhenghua Chen. Source-free calibrated uncertainty for RUL adaptation with incomplete degradation. *Mechanical Systems and Signal Processing*, 233:112718, 2025. doi: 10.1016/j.ymssp.2025.112718.

Jooyoung Jeon and James W. Taylor. Using conditional kernel density estimation for wind power density forecasting. *Journal of the American Statistical Association*, 107(497):66–79, 2012. doi: 10.1080/01621459.2011.643745.

Balaji Lakshminarayanan, Alexander Pritzel, and Charles Blundell. Simple and scalable predictive uncertainty estimation using deep ensembles. In *Advances in Neural Information Processing Systems*, volume 30, pp. 6402–6413, 2017. URL https://proceedings.neurips.cc/paper/2017/hash/9ef2ed4b7fd2c810847ffa5fa85bce38-Abstract.html.

JaeYun Lee, Jae Hyeon Park, Gyoomin Lee, Bogyeong Kim, Min Hee Cha, Hyeok Nam, Joo Jeon, Hyunse Lee, and Sung In Cho. DUET: Dual-perspective pseudo labeling and uncertainty-aware exploration & exploitation training for source-free domain adaptation. In *Advances in Neural Information Processing Systems*, volume 38, pp. 74927–74954, 2025. doi: 10.52202/085713-2516. URL https://proceedings.neurips.cc/paper_files/paper/2025/file/6c8dfbbd1ba3e22339e58a336cbed52b-Paper-Conference.pdf.

JoonHo Lee and Gyemin Lee. Feature alignment by uncertainty and self-training for source-free unsupervised domain adaptation. *Neural Networks*, 161:682–692, 2023. doi: 10.1016/j.neunet.2023.02.009.

Bo Li, Yezhen Wang, Shanghang Zhang, Dongsheng Li, Kurt Keutzer, Trevor Darrell, and Han Zhao. Learning invariant representations and risks for semi-supervised domain adaptation. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 1104–1113, 2021. doi: 10.1109/CVPR46437.2021.00116.

Jingjing Li, Zhiqi Yu, Zhekai Du, Lei Zhu, and Heng Tao Shen. A comprehensive survey on source-free domain adaptation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 46(8):5743–5762, 2024. doi: 10.1109/TPAMI.2024.3370978.

Jian Liang, Dapeng Hu, and Jiashi Feng. Do we really need to access the source data? source hypothesis transfer for unsupervised domain adaptation. In *Proceedings of the 37th International Conference on Machine Learning*, volume 119, pp. 6028–6039. PMLR, 2020. URL https://proceedings.mlr.press/v119/liang20a.html.

Mattia Litrico, Alessio Del Bue, and Pietro Morerio. Guiding pseudo-labels with uncertainty estimation for source-free unsupervised domain adaptation. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 7640–7650, 2023. doi: 10.1109/CVPR52729.2023.00738.

Mingsheng Long, Yue Cao, Zhangjie Cao, Jianmin Wang, and Michael I. Jordan. Transferable representation learning with deep adaptation networks. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 41(12):3071–3085, 2019. doi: 10.1109/TPAMI.2018.2868685.

Nis Meinert, Jakob Gawlikowski, and Alexander Lavin. The unreasonable effectiveness of deep evidential regression. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 37, pp. 9134–9142, 2023. doi: 10.1609/aaai.v37i8.26096.

Ismail Nejjar, Qin Wang, and Olga Fink. DARE-GRAM: Unsupervised domain adaptation regression by aligning inverse gram matrices. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 11744–11754, 2023. doi: 10.1109/CVPR52729.2023.01130.

Ismail Nejjar, Gaetan Frusque, Florent Forest, and Olga Fink. Uncertainty-guided alignment for unsupervised domain adaptation in regression. *Reliability Engineering & System Safety*, 270: 112143, 2026. doi: 10.1016/j.ress.2025.112143.

Yaniv Ovadia, Emily Fertig, Jie Ren, Zachary Nado, D. Sculley, Sebastian Nowozin, Joshua Dillon, Balaji Lakshminarayanan, and Jasper Snoek. Can you trust your model's uncertainty? evaluating predictive uncertainty under dataset shift. In *Advances in Neural Information Processing Systems*, volume 32, pp. 13991–14002, 2019. URL https://proceedings.neurips.cc/paper/2019/hash/8558cb408c1d76621371888657d2eb1d-Abstract.html.

Baochen Sun and Kate Saenko. Deep CORAL: Correlation alignment for deep domain adaptation. In *Computer Vision–ECCV 2016 Workshops*, pp. 443–450, 2016. doi: 10.1007/978-3-319-49409-8_35.

Jun Wu, Jingrui He, Sheng Wang, Kaiyu Guan, and Elizabeth Ainsworth. Distribution-informed neural networks for domain adaptation regression. In *Advances in Neural Information Processing Systems*, volume 35, pp. 10040–10054, 2022. doi: 10.52202/068431-0729.

Yuefei Wu, Bin Shi, Bo Dong, Qinghua Zheng, and Hua Wei. The evidence contraction issue in deep evidential regression: Discussion and solution. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 38, pp. 21726–21734, 2024. doi: 10.1609/aaai.v38i19.30172.

Haifeng Xia, Pu Wang, Toshiaki Koike-Akino, Ye Wang, Philip Orlik, and Zhengming Ding. Adversarial bi-regressor network for domain adaptive regression. In *Proceedings of the Thirty-First International Joint Conference on Artificial Intelligence*, pp. 3608–3614, 2022. doi: 10.24963/ijcai.2022/501.

Gezheng Xu, Hui Guo, Li Yi, Charles Ling, Boyu Wang, and Grace Y. Yi. Revisiting source-free domain adaptation: A new perspective via uncertainty control. In *International Conference on Learning Representations*, volume 2025, pp. 92900–92939, 2025a. URL https://proceedings.iclr.cc/paper_files/paper/2025/file/e85454a113e8b41e017c81875ae68d47-Paper-Conference.pdf.

Gezheng Xu, Li Yi, Pengcheng Xu, Jiaqi Li, Ruizhi Pu, Changjian Shui, A. Ian McLeod, Boyu Wang, and Charles Ling. Unraveling the mysteries of label noise in source-free domain adaptation: Theory and practice. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 47 (5):4007–4023, 2025b. doi: 10.1109/TPAMI.2025.3536845.

Gezheng Xu, Qi Chen, Qiuhao Zeng, Charles Ling, and Boyu Wang. Discretized density-guided source-free adaptation for continuous targets. In *Proceedings of the 43rd International Conference on Machine Learning*, volume 306 of *Proceedings of Machine Learning Research*, 2026. URL https://openreview.net/forum?id=UQLGAjV1cf.

Shiqi Yang, Yaxing Wang, Joost van de Weijer, Luis Herranz, and Shangling Jui. Exploiting the intrinsic neighborhood structure for source-free domain adaptation. In *Advances in Neural Information Processing Systems*, volume 34, pp. 29393–29405, 2021a. URL https://proceedings.neurips.cc/paper/2021/hash/f5deaeeae1538fb6c45901d524ee2f98-Abstract.html.

Yuzhe Yang, Kaiwen Zha, Yingcong Chen, Hao Wang, and Dina Katabi. Delving into deep imbalanced regression. In *Proceedings of the 38th International Conference on Machine Learning*, volume 139 of *Proceedings of Machine Learning Research*, pp. 11842–11851, 2021b. URL https://proceedings.mlr.press/v139/yang21m.html.

Kai Ye, Tiejin Chen, Hua Wei, and Liang Zhan. Uncertainty regularized evidential regression. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 38, pp. 16460–16468, 2024. doi: 10.1609/aaai.v38i15.29583.

Qianshan Zhan, Xiao-Jun Zeng, and Qian Wang. Reducing bias in source-free unsupervised domain adaptation for regression. *Neural Networks*, 185:107161, 2025. doi: 10.1016/j.neunet.2025.107161.

A Additional Method Details

A.1 Dynamic Response Grid

At refresh r, collect response centers $\mathcal{Z}^{(r)} = \{y_i^l\}_{i=1}^{N_l} \cup \{\widehat{y}_j^{u,(r)} : j \in \mathcal{C}^{(r)}\}$ and let $z_-^{(r)} = \min \mathcal{Z}^{(r)}$, $z_+^{(r)} = \max \mathcal{Z}^{(r)}$. Define $\overline{U}_w^{(r)} = \sum_{j \in \mathcal{C}^{(r)}} \omega_j^{(r)} U_j^{(r)}$. The finite grid used in equation 5 is

$$\begin{aligned} d_z^{(r)} &= \max(z_+^{(r)} - z_-^{(r)}, \epsilon_g), \quad & \delta_g^{(r)} &= \kappa_g \max\left(d_z^{(r)}/B, \sqrt{\overline{U}_w^{(r)} + \epsilon_\sigma} \right), \\ a_g^{(r)} &= z_-^{(r)} - \delta_g^{(r)}, & b_g^{(r)} &= z_+^{(r)} + \delta_g^{(r)}, \\ \Delta_c^{(r)} &= (b_g^{(r)} - a_g^{(r)})/B, & c_b^{(r)} &= a_g^{(r)} + (b - 1/2)\Delta_c^{(r)}. \end{aligned} \tag{15}$$

The labeled bandwidth is $\sigma_l^{(r)} = \kappa_\sigma \Delta_c^{(r)}$. For an empty admitted set, set $\overline{U}_w^{(r)} = 0$ and construct the grid from labels alone. The span floor ϵ_g handles constant labeled responses. Kernels are evaluated with a numerically stable softmax over grid locations, preserving unit mass without denominator underflow.

A.2 Evidence Ceiling and Response Anchor

Evidence ceiling. Let $e_\theta(x) = 2\nu_\theta(x) + \alpha_\theta(x)$ and $E_i^{(r)} = \mathrm{sg}[e_{\theta_r}(x_i^l)]$. The labeled reference is

$$\widetilde{e}_l^{(r)} = \mathrm{Med}_i\, E_i^{(r)}, \qquad d_e^{(r)} = \mathrm{Med}_i\, |E_i^{(r)} - \widetilde{e}_l^{(r)}|, \qquad T_e^{(r)} = \widetilde{e}_l^{(r)} + \kappa_e(d_e^{(r)} + \epsilon). \tag{16}$$

With $[z]_+ = \max(z, 0)$, the one-sided penalty is

$$\mathcal{L}_{\mathrm{ev}} = \frac{1}{M_u} \sum_{x_j^u \in \mathcal{B}_u} [e_\theta(x_j^u) - T_e^{(r)}]_+^2. \tag{17}$$

It is active above the labeled reference, leaving evidence below that ceiling unpenalized.

Response anchor. Let $\bar{y}_l = N_l^{-1} \sum_i y_i^l$, $v_l = N_l^{-1} \sum_i (y_i^l - \bar{y}_l)^2$, and $\mu_a = N_l^{-1} \sum_i h_\theta(x_i^l)$. Define

$$\delta_l = \kappa_l \sqrt{v_l/N_l + \epsilon}, \qquad r_p^{(r)} = \mathrm{sg}\left[\exp\left(-\frac{[|\mu_p^{(r)} - \bar{y}_l| - \delta_l]_+}{\delta_l + \epsilon} \right) \right], \tag{18}$$

$$\mathcal{L}_{\mathrm{anc}} = r_p^{(r)} [|\mu_a - \mu_p^{(r)}| - \delta_l]_+^2. \tag{19}$$

The tolerance is a label-dependent trust region, not a confidence interval. The weight attenuates the anchor when the evidence mean is incompatible with observed target responses. If $v_l = 0$, including $N_l = 1$, the tolerance remains $\kappa_l \sqrt{\epsilon}$. For a singleton or constant-prediction unlabeled mini-batch, equation 7 gives $v_h = 0$ directly. Zero-assignment prototypes are zero vectors; zero-overlap bins contribute nothing to equation 11.

A.3 Update Sequence

The following sequence implements equation 13 without updating evidence through the current gradient step.

1. Initialize $\theta \leftarrow \theta_s$, with empty evidence cache and EMA threshold.
2. For epochs $e \le E_w$, optimize equation 2 on labeled target samples.
3. Before epoch $e = E_w + 1$, and subsequently whenever $(e - E_w - 1) \bmod R = 0$, evaluate the current model on target inputs without gradients. Apply equation 3, construct masses and bandwidths with equation 4, and update the grid and mixture using equation 15 and equation 6, respectively.
4. At each such refresh, cache the mixture moments, response assignments, labeled prototypes, full-set overlap weights, and safeguard references from equations 9–11 and equations 16–18.

5. For each subsequent training mini-batch, compute current predictions and unlabeled prototypes with gradients, and optimize equation 13. Use the labeled-only mixture if the full admitted set is empty, and set $\mathcal{L}_{\text{semi}} = 0$ if the admitted mini-batch is empty.
6. Select the checkpoint with the smallest epoch-averaged training objective after the first density refresh. Stop after 30 epochs without improvement or at the epoch limit.

B Analytical Motivation and Derivations

The following derivations establish probability-mass conservation, the propagation of component moment errors, and the squared-risk relation used in Section 3.3.

B.1 Mass Conservation and Mixed Moments

For fixed refresh r, equation 5 gives $\sum_b \bar{\kappa}_\sigma(c_b, z) = 1$. Since $\sum_j \omega_j^{(r)} = 1$ on a nonempty admitted set, the two terms of equation 6 have masses m_l and m_u, respectively. This remains true for any positive bandwidth and any finite grid. The labeled-only fallback also has mass one.

Let $\widehat{m}_{l,k}$ and $\widehat{m}_{u,k}$ be the kth raw moments of the separately normalized components, and let $m_{t,k} = \mathbb{E}[Y^k]$. Linearity gives

$$\widehat{m}_{p,k} = m_l \widehat{m}_{l,k} + m_u \widehat{m}_{u,k}, \qquad |\widehat{m}_{p,k} - m_{t,k}| \le m_l |\widehat{m}_{l,k} - m_{t,k}| + m_u |\widehat{m}_{u,k} - m_{t,k}|. \tag{20}$$

For fixed mixture masses, taking expectations yields the corresponding expected-error inequality. The bound specifies how component errors enter the mixture; its value depends on the quality of both components. Per-kernel normalization makes the coefficients in this relation equal to the intended evidence masses.

B.2 Moment Error and Squared Risk

Suppose $|\widehat{m}_{p,1} - m_{t,1}| \le \varepsilon_{p,1}$ and $|\widehat{m}_{p,2} - m_{t,2}| \le \varepsilon_{p,2}$. Writing $\mu_p = \widehat{m}_{p,1}$ and $v_p = \widehat{m}_{p,2} - \mu_p^2$, we obtain

$$|v_p - v_t| \le \varepsilon_v := \varepsilon_{p,2} + \varepsilon_{p,1}(2|\mu_t| + \varepsilon_{p,1}). \tag{21}$$

Indeed, $|v_p - v_t| \le |\widehat{m}_{p,2} - m_{t,2}| + |\mu_p - \mu_t||\mu_p + \mu_t|$, and $|\mu_p + \mu_t| \le 2|\mu_t| + \varepsilon_{p,1}$.

Proposition 1 (Squared-risk relation)**.** *Let $H = h_\theta(X)$ and Y have finite second moments, $v_h > 0$, $v_t > 0$, and $\mathrm{Corr}(H, Y) \ge \kappa_{\text{corr}} > 0$. Under the moment-error conditions above,*

$$\mathcal{R}_t(h) \le (|\mu_h - \mu_p| + \varepsilon_{p,1})^2 + |v_h - v_p| + \varepsilon_v + 2(1 - \kappa_{\text{corr}})\sqrt{v_h v_t}. \tag{22}$$

Proof. Expanding squared error gives

$$\mathcal{R}_t(h) = (\mu_h - \mu_t)^2 + v_h + v_t - 2\,\mathrm{Cov}(H, Y).$$

Substituting $\mathrm{Cov}(H, Y) = \mathrm{Corr}(H, Y)\sqrt{v_h v_t}$ and completing the variance square gives equation 12. For nonnegative a, b, $(\sqrt{a} - \sqrt{b})^2 \le |a - b|$. Therefore the variance-square term is bounded by $|v_h - v_p| + |v_p - v_t|$, and the mean-square term by $(|\mu_h - \mu_p| + \varepsilon_{p,1})^2$. Apply equation 21 and the lower bound on correlation to obtain equation 22. □

The bound separates output-moment discrepancies, evidence-moment error, and prediction–response association. It motivates combining population statistics with supervision that retains input identity. UCLDR supplies the latter through labeled regression and response-conditioned prototype alignment; Section 4.3 evaluates the predictive effect of adding the local branch. The bound is conditional on the stated moment-error and correlation assumptions.

B.3 Reliability, Coverage, and Statistical Support

The construction uses three different quantities: the admitted set $\mathcal{C}^{(r)}$, weights $\omega_j^{(r)}$, and spreads $\sigma_j^{u,(r)}$. Their effects can be monitored using

$$\overline{U}_w^{(r)} = \sum_j \omega_j^{(r)} U_j^{(r)}, \qquad n_{\text{eff}}^{(r)} = \left[\sum_j (\omega_j^{(r)})^2\right]^{-1}. \tag{23}$$

For n admitted items, $1 \le n_{\text{eff}} \le n$, with equality at n for uniform weights. This follows from $(\sum_j \omega_j)^2 \le n \sum_j \omega_j^2$ and $\sum_j \omega_j^2 \le 1$. Concentrating mass changes the statistical support of the evidence even when coverage is fixed. Narrower selection also changes response coverage, so lower retained uncertainty is not by itself a guarantee of a more accurate target marginal. Appendix F.3 examines these quantities together.

C Implementation Settings and Computational Profile

The experiments use Python 3.12.3, PyTorch 2.8.0, CUDA 12.8, and an NVIDIA RTX 4090. Adam uses $(\beta_1, \beta_2) = (0.9, 0.999)$, $\epsilon_{\text{Adam}} = 10^{-8}$, and zero weight decay. The supervised evidential coefficient is $\lambda_R = 1$, the variance coefficient is $\eta_{\text{var}} = 0.5$, and $\kappa_g = \kappa_\sigma = \kappa_e = \kappa_l = 1$. The stabilizers are $\epsilon = 10^{-8}$ and $\epsilon_g = \epsilon_\sigma = 10^{-6}$. All reported UCLDR configurations use $B = 100$ response-grid locations and stable softmax kernel normalization.

Main benchmark and controlled-study protocols. The main UCLDR runs use learning rate 10^{-4} and seeds $\{42, 2024, 2025\}$, with target sampling following each seed. Their recorded normalization policy fits adaptation standardization to the labeled target split. The uncertainty-role study in Appendix E uses learning rate 10^{-3}, a fixed split seed of 42, and the normalization stored with each source checkpoint. Its training seeds remain $\{42, 2024, 2025\}$. Each replacement is compared with its matched canonical configuration within that study. The computational profile below uses its complete-objective canonical runs. These protocols keep within-study comparisons paired while separating the role study from the main benchmark.

Cycle domains. The Combined Cycle Power Plant data contain 9,568 examples, divided into four equally sized temperature domains with 2,392 examples each. Ambient temperatures are ranked in input order to break ties before quartile assignment. The loader reserves 30% of each target domain, 717 examples, for adaptation and the remaining 1,675 for testing. It samples N_l labels from the adaptation pool and treats the remaining responses as unobserved during training.

Backbones and numerical uncertainty. Bike, Cycle, and Housing use two 64-channel one-dimensional convolutions with kernel size 3, stride 1, and padding 1, followed by fully connected layers of widths 128, 64, and 16. Each convolution is followed by batch normalization and ReLU; each fully connected layer uses ReLU. Kin8xy uses two 64-channel convolutions with kernel sizes 5 and 3, stride 1, and no padding, with a size-2 max-pooling layer after the second convolution. Its fully connected widths are 32, 16, and 8. Four scalar heads output the NIG parameters, with softplus enforcing positivity and an additional unit offset for α. For numerical stability, the implementation floors ν and β at 10^{-6} and α at $1 + 10^{-6}$. It bounds the aleatoric variance and then the resulting predictive-mean variance to $[10^{-6}, 10^4]$; the latter supplies the uncertainty used in the kernels. The checkpoint diagnostics in Appendix E describe a case at this numerical ceiling.

Table 3 reports the complete-UCLDR profile over 432 runs: three datasets, 12 transfers, four budgets $N_l \in \{10, 20, 30, 50\}$, and three seeds. Each dataset contributes 144 runs under the controlled-study protocol above. Executed epochs and training times reflect early stopping and are summarized within each dataset. The representation adds no trainable density network; full-set evidence is refreshed periodically and cached.

Table 3: Adaptation-side computational profile. Time and executed epochs are mean ± SD over 144 runs per dataset. Memory is recorded peak allocated GPU memory; parameters belong to the dataset-specific evidential backbone.

Dataset	Time (s)	Epochs	GPU allocation (GB)	Parameters (M)
Bike	55.53 ± 29.36	90.34 ± 43.82	0.0194	0.1207
Housing	128.67 ± 89.70	158.81 ± 38.57	0.0188	0.0879
Kin8xy	42.73 ± 13.67	134.53 ± 39.71	0.0183	0.0158

Table 4: **Objective contrasts isolate the addition of local supervision.** All rows retain $\mathcal{L}_{\mathrm{sup}}$ and the density construction. A 1 indicates that a loss is active at its fixed coefficient; a 0 indicates that it is disabled. Safeguards are $\mathcal{L}_{\mathrm{ev}}$ and $\mathcal{L}_{\mathrm{anc}}$. Results are Cycle mean target-test MSE at ten labels across 12 transfers.

Objective	$\mathcal{L}_{\mathrm{dist}}$	Safeguards	$\mathcal{L}_{\mathrm{semi}}$	MSE
Density pathway	1	0	0	2.462
Safeguarded density	1	1	0	2.503
Complete UCLDR	1	1	1	**1.445**

D Complete Objective and Label-Budget Comparisons

Table 5 compares three nested objectives on Cycle at all five label budgets. The density pathway uses $\mathcal{L}_{\mathrm{sup}} + \lambda_{\mathrm{dist}}\mathcal{L}_{\mathrm{dist}}$; safeguarded density adds $\lambda_{\mathrm{ev}}\mathcal{L}_{\mathrm{ev}} + \lambda_{\mathrm{anc}}\mathcal{L}_{\mathrm{anc}}$; and complete UCLDR adds $\lambda_{\mathrm{semi}}\mathcal{L}_{\mathrm{semi}}$. Active coefficients remain fixed. Table 4 identifies the active components. The contrast between safeguarded density and complete UCLDR measures the effect of adding the local branch while density-moment regularization remains active. These nested comparisons do not isolate removal of $\mathcal{L}_{\mathrm{dist}}$ from the complete objective.

Table 5: Complete Cycle budget comparison. Mean target-test MSE across 12 directed transfers. Partial UCLDR objectives are diagnostic configurations. Relative reductions use unrounded transfer aggregates; displayed MSEs are rounded. The last two rows compare UCLDR with CRAFT within each transfer.

Objective / method	$N_l = 10$	$N_l = 20$	$N_l = 30$	$N_l = 40$	$N_l = 50$
Density pathway	2.462	474.140	46.897	170.887	0.910
Safeguarded density	2.503	331.086	303.903	4.262	0.900
Complete UCLDR	**1.445**	**1.049**	**0.931**	**0.790**	**0.781**
CRAFT	1.612	1.094	0.964	0.846	0.826
Reduction over CRAFT (%)	10.36	4.09	3.45	6.57	5.43
Transfers with lower MSE / 12	7	7	8	10	11
Median paired reduction (%)	13.64	4.80	6.14	5.83	7.66

Complete UCLDR gives the lowest mean error among the nested configurations at every budget. At 20 labels, for example, the density and safeguarded-density configurations have mean MSEs of 474.140 and 331.086, compared with 1.049 for complete UCLDR. The comparison measures the effect of adding local supervision within this density-based objective.

E Full Uncertainty-Role Replacement Study

Table 6 gives both optimization contexts. Panel A uses complete UCLDR. Panel B disables only $\mathcal{L}_{\mathrm{semi}}$, retaining density regularization and both safeguards. Each context covers 144 seed-aggregated conditions: three datasets, 12 transfers, and four budgets. The fixed-threshold replacement uses $S = 0.05$, and the fixed-bandwidth replacement uses $\sigma = 0.02$ in the study's response coordinates. Relative RMSE changes are computed within matched conditions using equation 14, then averaged.

Uniform mass allocation increases average relative RMSE by 1.79% in Panel A and by 0.53% in Panel B. Admission and bandwidth replacements have effects of different signs across datasets. Housing entries are 0.00% in both panels at the reported precision. The replacement study therefore provides its clearest positive evidence for mass allocation on Bike and Kin8xy.

Housing checkpoint diagnostics. Across the 144 complete-objective canonical Housing runs, the saved checkpoint summaries report retained fraction 1.0 and mean selected uncertainty 10^4. The effective-sample-size ratio $n_{\mathrm{eff}}/|\mathcal{C}|$ differs from one by less than 10^{-6}. Thus, these saved states have nearly uniform reliability weights. This provides a concrete setting in which inverse-variance and

Table 6: All uncertainty-role replacement outcomes. Positive values mean the replacement increases RMSE. Each dataset contributes 48 matched conditions after seed aggregation.

Role	Replacement	Bike	Housing	Kin8xy	Mean
A: Complete UCLDR					
Admission ordering	Random selection	0.00	0.00	-1.00	-0.33
Admission boundary	Fixed threshold	-1.81	0.00	-1.39	-1.07
Mass allocation	Uniform weighting	$+3.56$	0.00	$+1.80$	$+1.79$
Spread modulation	Fixed bandwidth	$+1.64$	0.00	-2.99	-0.45
B: Only $\mathcal{L}_{\text{semi}}$ disabled					
Admission ordering	Random selection	0.00	0.00	$+1.32$	$+0.44$
Admission boundary	Fixed threshold	-0.55	0.00	-2.94	-1.16
Mass allocation	Uniform weighting	-0.07	0.00	$+1.66$	$+0.53$
Spread modulation	Fixed bandwidth	$+0.03$	0.00	$+0.59$	$+0.21$

uniform mass allocations are effectively the same. The observation concerns the saved checkpoints; the replacement errors in Table 6 summarize the full training runs.

F Response-Evidence Diagnostics

F.1 Uncertainty Ordering and Risk–Coverage Curves

The reliability and density-fidelity analyses use Bike, Housing, and Kin8xy, transfers $0 \to 1$ and $3 \to 2$, budgets $N_l \in \{10, 50\}$, and three seeds, for 36 saved configurations. Target responses enter these diagnostics only after checkpoint selection. For nominal coverage c, predictions are ranked by increasing predictive-mean uncertainty and the lowest-uncertainty fraction c is evaluated. This diagnostic coverage is distinct from the potentially lagged EMA training threshold.

Table 7 reports uncertainty–error correlations and pseudo-label RMSE, and Figure 3 shows how error changes with coverage. Bike and Housing have positive correlations and lower selected-subset error at $c = 0.4$. Kin8xy has a negative correlation and lower error at full coverage. The value of uncertainty-based admission consequently depends on the dataset and the adapted predictor.

Table 7: Dataset-stratified pseudo-label reliability. $r_S(U, e^2)$ is the Spearman correlation between uncertainty and squared error. Risk is RMSE at nominal coverage 0.1 and 1.0; selected/all compares coverage 0.4 with full coverage.

Dataset	$r_S(U, e^2)$	Risk RMSE $0.1 \to 1.0$	Selected / all RMSE
Bike	0.359	$0.498 \to 1.026$	$0.769/1.026$
Housing	0.226	$0.882 \to 1.242$	$0.913/1.242$
Kin8xy	-0.161	$1.149 \to 0.875$	$0.982/0.875$

F.2 Response-Density Fidelity and Fallback Cases

We compare the frozen mixture with a post-hoc target reference using absolute mean error, absolute variance error, and one-dimensional Wasserstein distance W_1 (Table 8). Figure 4 separates the 26 mixed-evidence configurations from the 10 labeled-only fallbacks. The latter measure approximation from the observed labels alone. These diagnostics measure the mixture's agreement with target responses after model selection.

On Housing, absolute mean error is 0.286 and absolute variance error is 17.772. The quantities have different units and are interpreted separately: the first measures displacement of the response mean, and the second measures mismatch in dispersion.

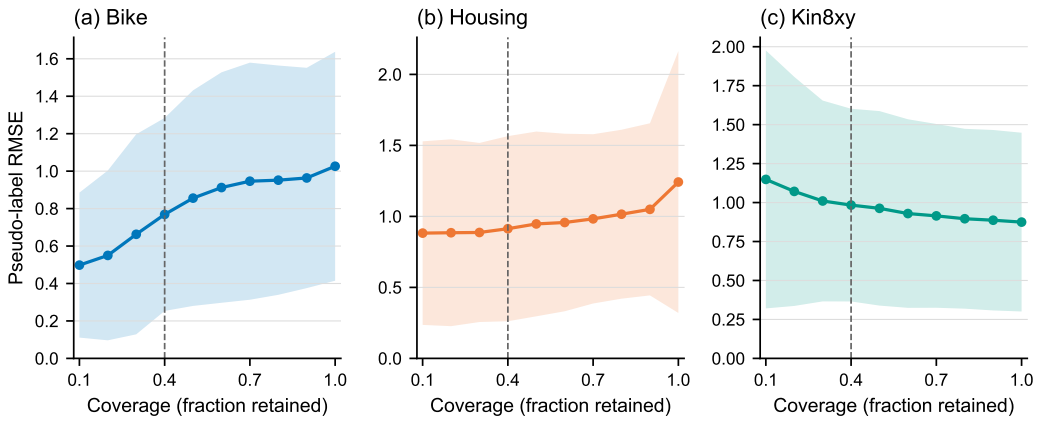

Figure 3: Pseudo-label risk–coverage curves by dataset. Lines are condition means and bands are descriptive 95% intervals after seed aggregation. The vertical line marks nominal coverage 0.4.

Table 8: Target response-density approximation errors. Scale-dependent values are reported separately by dataset.

Dataset	$\lvert\Delta\mu\rvert$	$\lvert\Delta v\rvert$	W_1
Bike	0.484	0.611	0.491
Housing	0.286	17.772	1.429
Kin8xy	0.745	0.662	0.761

F.3 Selection Quantile and Statistical Support

The quantile study uses Housing and Kin8xy at $N_l = 10$ and $q \in \{0.1, 0.2, 0.4, 0.6, 0.8, 1.0\}$. We record retained fraction, weighted uncertainty, effective sample size in equation 23, and within-transfer relative RMSE, $\mathrm{RMSE}_{\mathrm{task}}(q)/\mathrm{RMSE}_{\mathrm{task}}(0.4)$. Seeds are averaged within each transfer. Figure 5 shows Kin8xy $3 \to 2$. Reading prediction error alongside retained fraction and effective sample size distinguishes a small, concentrated evidence pool from broader response coverage.

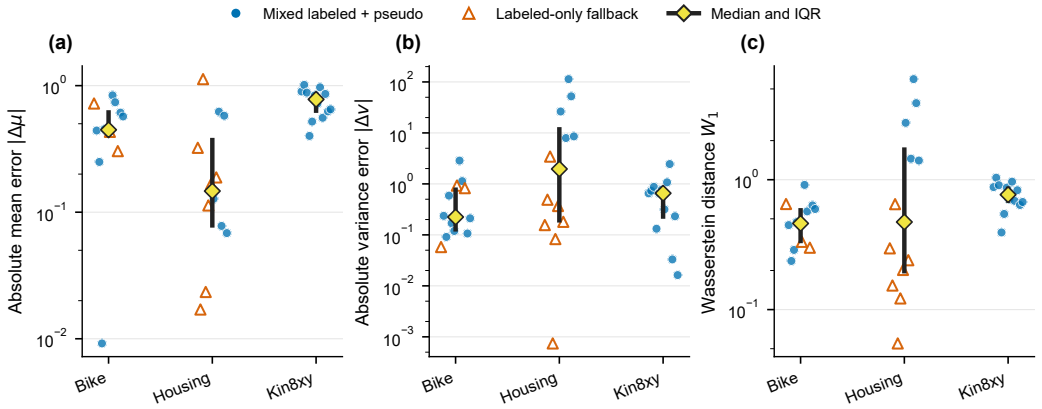

Figure 4: Run-level errors in response mean, variance, and W_1. Mixed-evidence and labeled-only fallback configurations are displayed separately, with median and interquartile summaries.

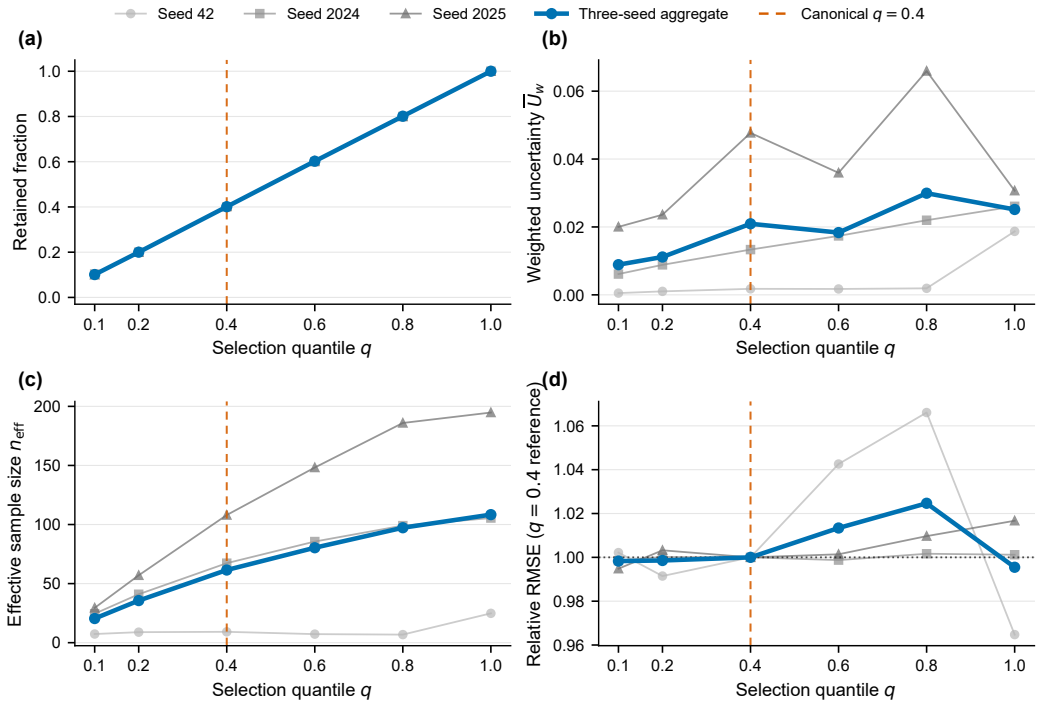

Figure 5: Coverage and statistical support on Kin8xy $3 \to 2$ at ten labels. Panels report retained fraction, weighted uncertainty, effective sample size, and relative RMSE across selection quantiles. The canonical $q = 0.4$ supplies the prediction-error reference.

G Optimization and Seed-Level Diagnostics

G.1 Safeguard Dynamics

The evidence ceiling and response anchor target particular optimization quantities. Figure 6 shows paired trajectories under fixed stress conditions: Kin8xy $0 \to 1$ at $N_l = 50$ for the evidence ceiling, and Kin8xy $3 \to 2$ at $N_l = 10$ for the anchor. Effects are localized to particular epochs. The anchor case has a median paired change in its gap diagnostic of -0.00249. A separate Housing ceiling stress case records a mean over-evidence fraction of 0.3385 with $\mathcal{L}_{\text{ev}}$ disabled; that scalar summary and the Kin8xy trajectory concern different conditions. The paired trajectories show when each penalty changes its targeted quantity. Predictive comparisons of the nested objectives appear in Appendix D.

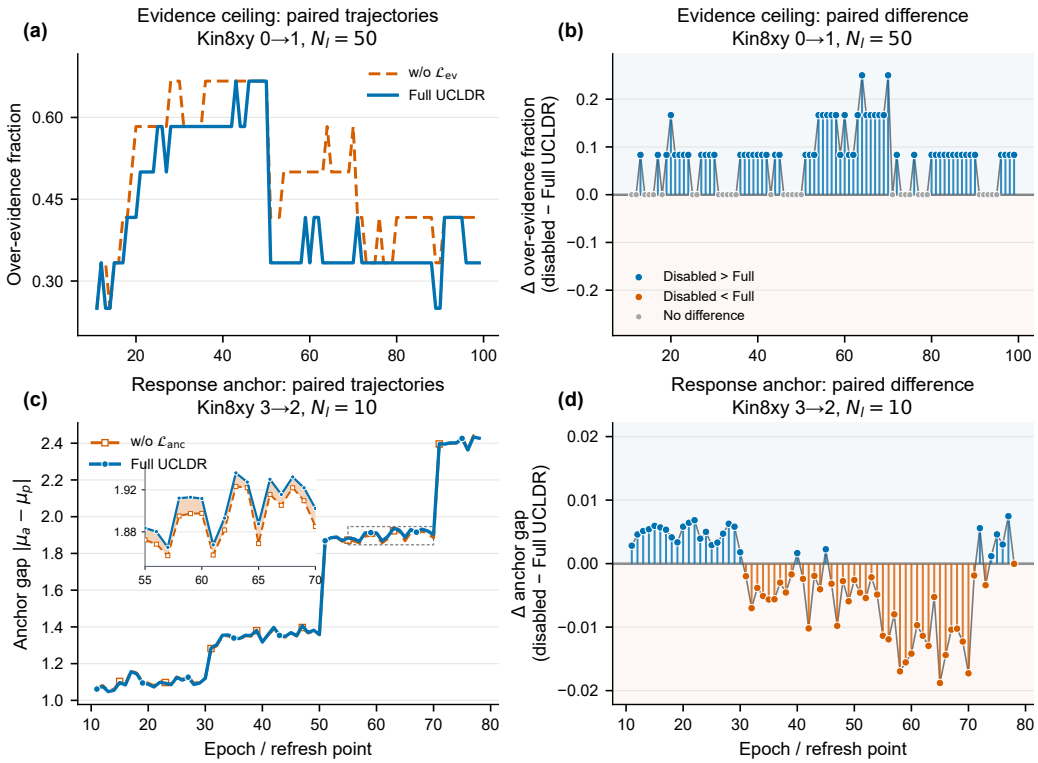

Figure 6: Safeguard trajectories and same-epoch differences. Panels (a)–(b): evidence ceiling on Kin8xy $0 \to 1$, $N_l = 50$. Panels (c)–(d): response anchor on Kin8xy $3 \to 2$, $N_l = 10$. Positive differences mean the safeguard-targeted quantity is larger when the corresponding term is disabled.

G.2 Variation Across Random Seeds

For the 12 dataset–transfer–budget conditions used in the reliability analysis, we compute the SD of test RMSE across three seeds, then summarize the four condition-level SDs within each dataset (Table 9). Figure 7 summarizes the seed distributions. The largest condition-level SD occurs at $3 \to 2$, $N_l = 10$ in all three datasets. These summaries measure repeatability within a condition, separately from variation across source–target transfers.

Table 9: Seed-level variation. Each dataset has four transfer–budget conditions, with three seeds per condition. Median, IQR, and maximum summarize the condition-level SDs of target-test RMSE.

Dataset	Conditions	Median SD	IQR	Maximum condition
Bike	4	0.135	$[0.070, 0.302]$	0.652 ($3 \to 2$, 10)
Housing	4	0.392	$[0.083, 0.831]$	1.301 ($3 \to 2$, 10)
Kin8xy	4	0.112	$[0.076, 0.153]$	0.184 ($3 \to 2$, 10)

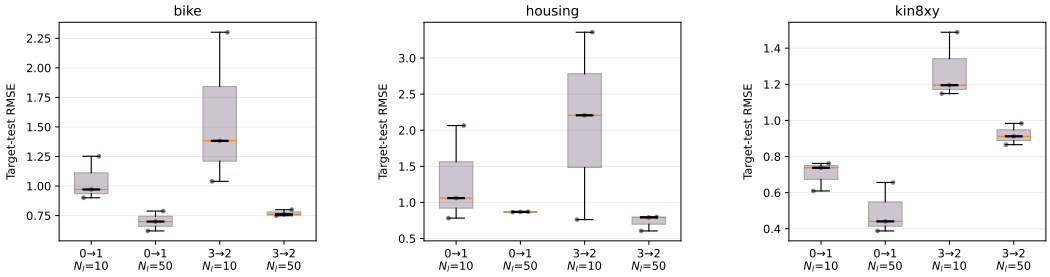

Figure 7: Target-test RMSE distributions across three seeds in each predefined dataset–transfer–budget condition. Boxplots summarize seed variation; orange horizontal lines indicate medians.