

Supplementary Material to “Measurement-Driven Path-Loss Modeling for Low-Altitude UAV Air-to-Ground Links Using Physics-Aware Ensemble Learning and Calibrated Uncertainty”

The Authors

September 13, 2026

This supplement contains the detailed sweeps, additional figures, and per-band visualizations referenced in the main paper. Section S1 reports the full input encoder dimensionality sweep (Section 4.7 of the main paper). Section S2 reports the feature-selection Pareto curves and group-attribution maps (Section 4.7). Section S3 shows the ID/OOD trade-off plane of an additional physics-regularized-loss study that complements the main paper. Section S4 provides the augmentation manifold projections and the Jensen–Shannon divergence table (Section 4.9). Section S5 reports the full per-dataset and per-protocol results of the outlier-rule sensitivity study summarized in Section 4.3. Section S6 lists the per-protocol results of the coordinate-free input variant summarized in Table 5 of Section 4.4. Section S7 lists the per-pair results of the out-of-distribution gate and adaptive recalibration summarized in Table 7 of Section 4.6. The 1500 MHz per-band performance panels are overlaid with the 433/740 MHz panels in the main-paper composite Figure 4 of Section 4.1.

S1 Input Encoder Dimensionality

This section accompanies Section 4.7 of the main paper and reports the full per-band cumulative effect of expanding the input encoder from 6 to 14 dimensions.

We sweep the input dimension from a 6-D coordinate baseline (transceiver positions $[lat_{tx}, lon_{tx}, h_{tx}, lat_{rx}, lon_{rx}, h_{rx}]$) to the full 14-D physics-aware vector by cumulatively adding $\log_{10}(d_{3D})$, $\log_{10}(d_h)$, Δh , θ_{pit} , θ_{azm} , $\log_{10}(f)$, R_{F1} , and θ_{inc} . The encoder output is held at 128 dimensions throughout via an adaptive embedding module. The cumulative improvement metric is

$$\Delta_{cum}^{(i)} = \text{RMSE}_{\text{base}} - \text{RMSE}^{(i)}, \quad i = 7, 8, \dots, 14, \quad (\text{S1})$$

with $\text{RMSE}_{\text{base}}$ the 6-D baseline.

Table S1 reports the per-dataset cumulative effect. The overwater 740 MHz dataset is the only band on which the cumulative RMSE drop from 6-D to 14-D is monotone and clearly positive (+0.450 dB, R^2 improvement from 0.681 to 0.798). For the other four datasets the cumulative RMSE drop from 6-D to 14-D is within ± 0.07 dB. Training-history inspection further shows that the convergence behavior of the baseline and the full-feature configuration is similar; the principal effect of adding physical features is to accelerate the initial 100 epochs of training rather than to lower the asymptotic loss.

This result is consistent with the optimistic-bias analysis of Section 4.2 of the main paper: under random in-distribution splits, the model is able to recover most of the predictive signal from coordinates alone, because the train and test points are drawn from overlapping spatial neighborhoods.

The practical implication is that the choice of feature set should be band-aware: a compact 6–10-dimensional vector is sufficient for most bands considered here, while bands dominated by Fresnel-zone effects (overwater 740 MHz) benefit from the full physics-aware encoding.

Table S1: Cumulative effect of progressively expanding the input feature dimension from 6-D to 14-D on test-set RMSE (dB) and R^2 under the random 70/30 split. The cumulative drop is $\Delta_{\text{cum}} = \text{RMSE}_{\text{base}} - \text{RMSE}_{\text{full}}$; positive values indicate that the 14-D set helps over the 6-D coordinate baseline.

Dataset	6-D RMSE	14-D RMSE	Δ_{cum}	R^2 (6→14)
W433	3.548	3.562	−0.014	0.510 → 0.506
S433	3.236	3.301	−0.065	0.363 → 0.337
W740	2.190	1.740	+0.450	0.681 → 0.798
M1500	2.998	2.985	+0.013	0.566 → 0.570
M3300	2.626	2.645	−0.020	0.343 → 0.333

Feature-subset Pareto: 4 selection criteria $\times K \in \{4, \dots, 14\}$

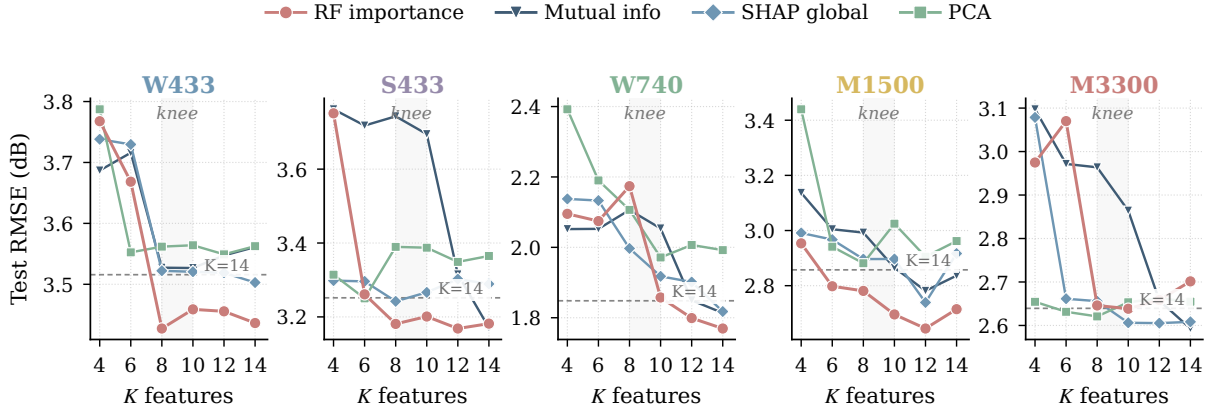


Figure S1: Feature-subset Pareto curves across the five frequency-scenario datasets. Each panel shows test-set RMSE versus subset size K for four selection criteria (random-forest impurity, mutual information, global SHAP, and PCA projection). The curves saturate at $K \approx 8-10$ for all four criteria, indicating that 4–6 of the 14 engineered features are redundant in the present regime.

S2 Feature Selection and Attribution

This section accompanies Section 4.7 of the main paper and provides the full feature-selection Pareto curves and feature-group attribution maps.

We conduct a feature-selection study that ranks the 14-D feature vector with four different criteria and then sweeps subset size $K \in \{4, 6, 8, 10, 12, 14\}$, training an adaptive variant of the deep backbone on each $(K, \text{ranking})$ configuration. The four ranking criteria are: (i) random-forest impurity-based feature importance, (ii) mutual-information regression, (iii) global SHAP value averaged over a 500-sample explainer set on a trained deep backbone, and (iv) principal-component analysis. Across the five datasets, the test-set RMSE saturates at $K \approx 8-10$ for all four criteria, with the full $K = 14$ configuration providing at most ± 0.05 dB improvement over the best $K = 8$. The set of top-5 features is largely consistent across criteria and datasets: Δh , h_{tx} , $\log_{10} d_{3D}$, θ_{pit} , and θ_{inc} dominate the rankings in random-forest and SHAP analyses on three of the five datasets, while θ_{azm} and R_{F1} rise to the top three for the mountainous datasets.

To quantify the contribution of each *group* of physical features rather than each individual feature, we further train a single pooled instance of the deep backbone on the union of the five datasets and replace, at inference time, each feature group with its training-set median, measuring the resulting RMSE inflation ΔRMSE on the test partition. Grouping the 14 features as *position* ($lat_{tx/rx}$, $lon_{tx/rx}$, $h_{tx/rx}$), *distance* ($\log_{10} d_{3D}$, $\log_{10} d_h$, Δh), *angle* (θ_{pit} , θ_{azm} , θ_{inc}), *frequency* ($\log_{10} f$), and *Fresnel* (R_{F1}), the pooled baseline RMSE of 4.19 dB inflates by +6.74 dB when position is ablated, +4.98 dB when distance is ablated, +7.09 dB when frequency is ablated, +3.10 dB when Fresnel is ablated, and +1.54 dB when angle is ablated. The dominance of position and frequency in the pooled ablation confirms that the model

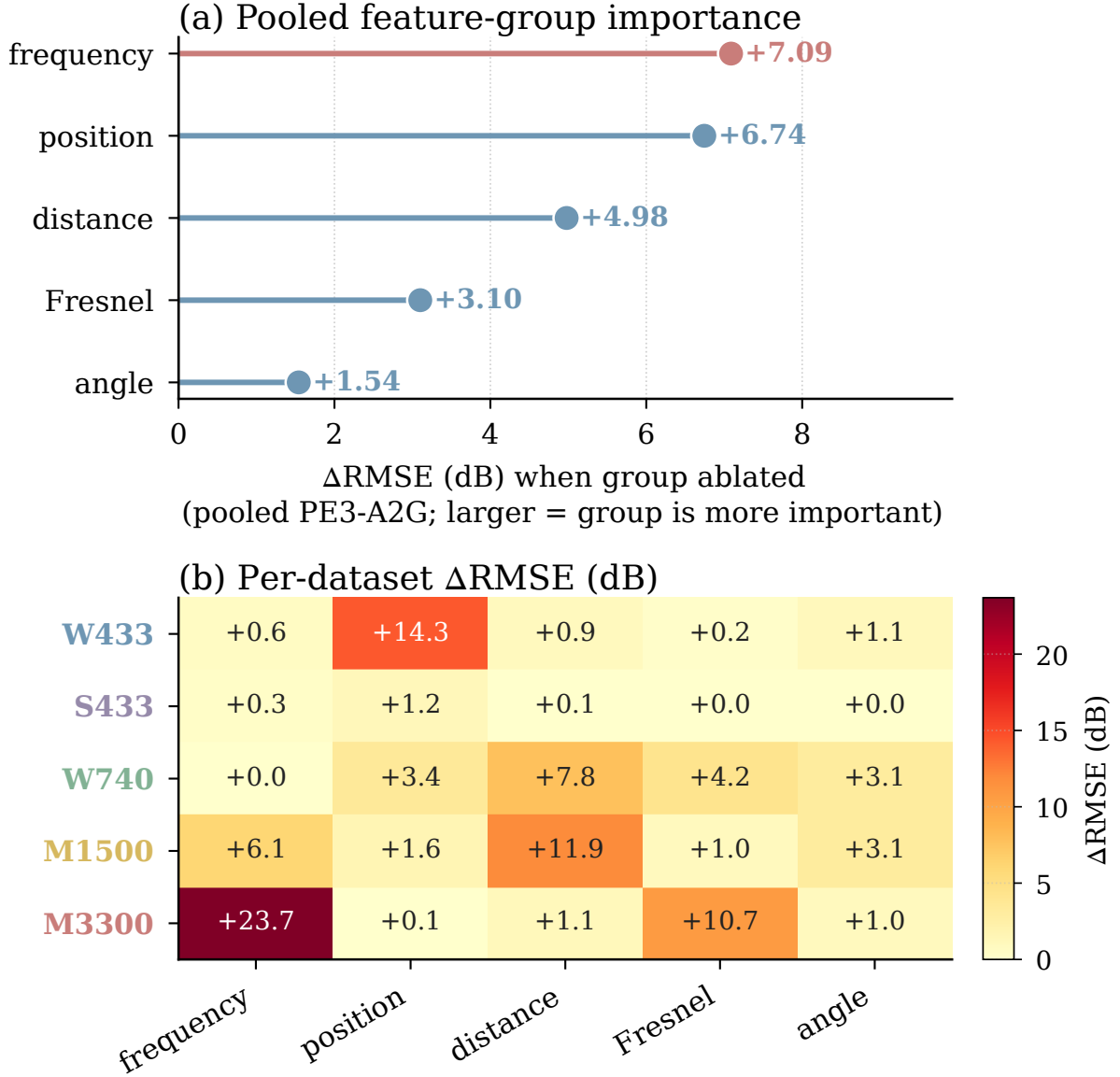


Figure S2: Feature-group attribution on a pooled deep backbone. (a) When each group is replaced by its training median at inference, the pooled-test RMSE inflates by +1.54 to +7.09 dB; the frequency and position groups dominate the pooled signal. (b) Per-dataset ΔRMSE map shows strong scenario specificity: M3300 depends almost entirely on the frequency feature, while W740 is distance-dominated.

conditions on scenario via position and on carrier via the frequency feature, which is the expected behavior for a model that has to disambiguate five heterogeneous bands. Per-dataset ablation reveals strong scenario specificity: for M3300 the frequency ablation alone inflates RMSE by +23.69 dB, while for W740 the dominant group is distance (+7.77 dB).

S3 Physics-Regularized Loss: ID/OOD Trade-off

This section reports an additional physics-regularized-loss experiment that complements the main paper. Figure S3 shows the full per-source ID/OOD trade-off plane for the physics-regularized loss across four source datasets, all transferring to M1500. Each marker corresponds to one $(\lambda_{\text{Friis}}, \lambda_{\text{mono}})$ configuration from the sweep in $\{0, 10^{-2}, 10^{-1}\}^2$.

Physics-regularized loss: in-distribution vs. out-of-distribution Pareto plane

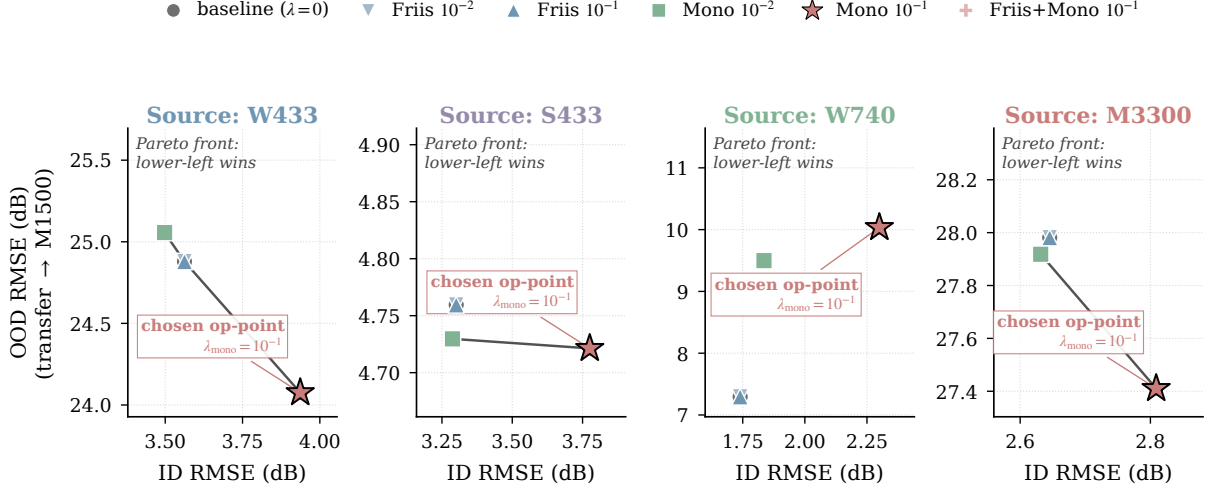


Figure S3: Physics-regularized loss: in-distribution (ID) versus out-of-distribution (OOD) test-set RMSE trade-off across four source datasets, all transferring to M1500. Each marker is one $(\lambda_{\text{Friis}}, \lambda_{\text{mono}})$ configuration. The monotonic-distance penalty with $\lambda = 10^{-1}$ trades a small ID penalty for an OOD improvement on two of four source datasets; the Friis-floor penalty is inactive in our regime.

S4 Augmentation Manifold Analysis

This section accompanies Section 4.8 of the main paper and provides the full manifold-divergence analysis of the physics-constrained augmentation pipeline.

Four augmentation variants are compared with the raw measurement distribution. The gaussian_only variant keeps only the physics-aware Gaussian perturbation step. The mixup_unconstrained variant removes the same-scenario and same-frequency constraints from the Mixup pairing rule. The mixup_physics variant corresponds to the protocol of Section 3.4 of the main paper. The full variant combines mixup_physics with Gaussian perturbation.

For each variant, the 14-D feature vectors of the resulting samples are projected into a common 2-D principal-component subspace fitted on the raw measurement set, partitioned into a shared 50×50 grid with Laplace smoothing, and scored by the Jensen–Shannon (JS) divergence between the variant histogram and the raw histogram. Smaller JS values indicate that the augmented distribution stays closer to the measurement manifold. The analysis is run both per dataset and on a pooled multi-scenario dataset; the pooled setting is the only configuration in which unconstrained Mixup crosses scenario and frequency boundaries, so it is the most discriminating regime for evaluating the physics constraints.

Table S2 and Figure S4 report the resulting rankings. Removing the same-scenario and same-frequency constraints from Mixup increases the pooled JS divergence to the raw distribution from 0.072 to 0.197, a factor of about 2.7, and mixup_physics is the closest of the four augmentation variants to the raw distribution in the pooled setting. Even so, this best variant remains an order of magnitude above the raw control, 0.072 against 0.005: even physics-constrained augmentation introduces measurable distributional drift relative to the raw measurement set, consistent with the in-distribution RMSE degradation reported in the component-ablation table of the main paper. Same-scenario and same-frequency constraints are degenerate within each individual dataset, so the 2.7-fold gap appears only on the pooled set.

S5 Sensitivity to the Outlier Rule: Full Results

This section accompanies Section 4.3 of the main paper. The interquartile-range rule of the main paper computes its fences on the path-loss labels of the whole dataset before any split, which also removes

Table S2: JS divergence between augmented and raw distributions in a shared 2-D PCA subspace, computed as per-dataset averages and on the pooled set.

Variant	JS (per-dataset mean)	JS (pooled)
raw (control)	≈ 0	0.0053
mixup_physics	0.117	0.072
mixup_unconstrained	0.117	0.197
gaussian_only	0.099	0.102
full (mixup+noise)	0.154	0.107

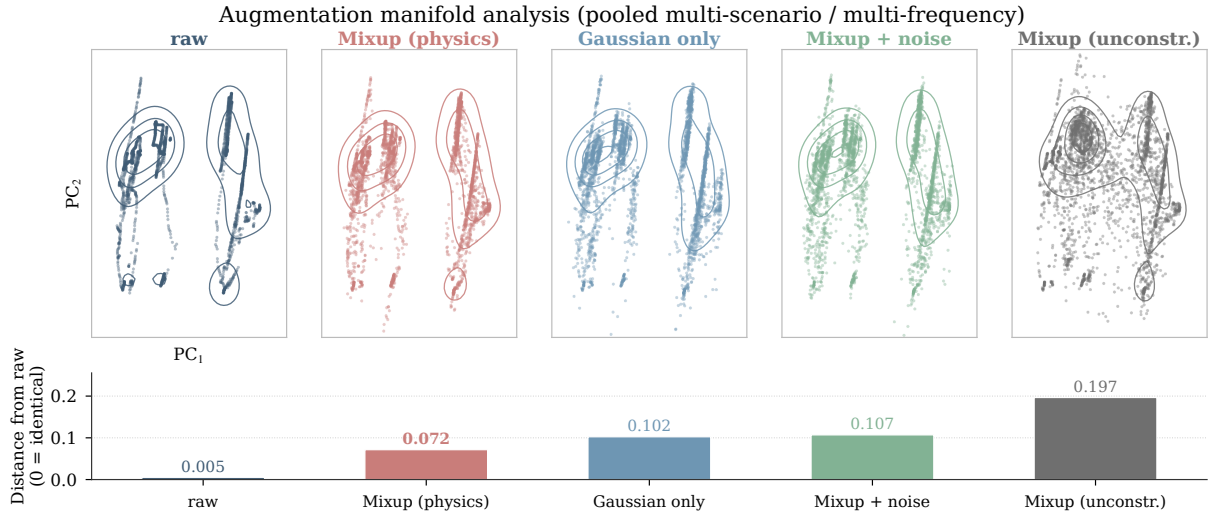


Figure S4: Augmentation manifold analysis on the pooled five-dataset set. The top row shows the 2-D principal-component projection of the raw measurement set and four augmentation variants: raw control, physics-constrained Mixup, Gaussian noise on the geometric features, Mixup combined with Gaussian noise, and unconstrained Mixup that allows cross-scenario and cross-frequency pairings. They are ordered left to right from closest to farthest from the raw distribution. The bottom row shows distribution distance from each variant to the raw measurement set, computed as Jensen–Shannon divergence between raw and variant 2-D PCA histograms on a common grid. A value of 0 means identical distributions, and larger values mean more drift.

3.9% of the test records and, in the transfer protocols, filters the target dataset with its own fences. Two stricter regimes were run with otherwise identical settings (500 epochs, seeds 42 to 44, no augmentation, the canonical hyperparameters). In the *train-fence* regime the data are split first, the fences are computed on the training partition only, and the same fences are applied to both partitions. In the *train-only* regime the training partition is filtered with its own fences and the test partition, or the target dataset, is left unfiltered. Table S3 reports every dataset and split protocol; Table S4 reports the nine transfer protocols. Spatial-block values are means over the five folds within a seed, then mean $\pm$ std over seeds; random and chronological values are mean $\pm$ std over seeds.

S6 Coordinate-Free Input Variant: Per-Protocol Results

This section accompanies Section 4.4 and Table 5 of the main paper. Every transfer protocol (Groups A to C), every leave-one-dataset-out protocol (Group D), and the in-distribution random 70/30 split were repeated with a coordinate-free 10-D input that drops $lat_{tx}, lon_{tx}, lat_{rx}, lon_{rx}$ and keeps both altitudes and the eight physical features, with an otherwise identical architecture, random forest, and blend (500 epochs, seeds 42 to 44, no augmentation). Table S5 lists test RMSE for the canonical 14-D vector and the 10-D variant side by side; Table 5 of the main paper reports the paired differences averaged per group.

Table S3: Outlier-rule sensitivity, in-distribution protocols. Test RMSE in dB, mean $\pm$ std over seeds 42 to 44 (spatial-block: mean over the five folds within a seed, then over seeds). Records is the mean number of scored test records. Fence regimes: manuscript computes the fences on the labels of the whole dataset before splitting; train, both computes them on the training partition and applies them to both partitions; train, test unfiltered filters the training partition only. The best of the three models in each row is bolded.

Dataset	Split	Fence regime	Records	Random Forest	Deep backbone	PACE-A2G
W433	random	manuscript	294	2.01 $\pm$ 0.08	2.37 $\pm$ 0.19	1.96$\pm$0.09
W433	random	train, both	296	2.10 $\pm$ 0.18	2.46 $\pm$ 0.22	2.07$\pm$0.15
W433	random	train, test unfiltered	318	6.37$\pm$0.29	7.90 $\pm$ 1.50	6.75 $\pm$ 0.62
W433	chronological	manuscript	294	9.65 $\pm$ 0.02	8.84$\pm$0.27	9.34 $\pm$ 0.09
W433	chronological	train, both	126	4.95$\pm$0.02	7.13 $\pm$ 0.64	5.27 $\pm$ 0.17
W433	chronological	train, test unfiltered	318	15.96$\pm$0.00	18.86 $\pm$ 0.73	16.73 $\pm$ 0.24
W433	spatial-block	manuscript	195	3.89 $\pm$ 0.31	4.65 $\pm$ 0.09	3.80$\pm$0.27
W433	spatial-block	train, both	179	3.85 $\pm$ 0.33	4.29 $\pm$ 0.12	3.68$\pm$0.25
W433	spatial-block	train, test unfiltered	211	5.40 $\pm$ 0.41	6.17 $\pm$ 0.25	5.37$\pm$0.37
S433	random	manuscript	259	2.35 $\pm$ 0.11	2.36 $\pm$ 0.08	2.19$\pm$0.09
S433	random	train, both	259	2.34 $\pm$ 0.13	2.49 $\pm$ 0.16	2.22$\pm$0.17
S433	random	train, test unfiltered	263	2.66 $\pm$ 0.20	2.72 $\pm$ 0.19	2.53$\pm$0.21
S433	chronological	manuscript	259	4.24$\pm$0.01	5.81 $\pm$ 0.63	4.54 $\pm$ 0.18
S433	chronological	train, both	253	4.88$\pm$0.09	5.65 $\pm$ 0.85	4.91 $\pm$ 0.34
S433	chronological	train, test unfiltered	263	5.30$\pm$0.07	5.93 $\pm$ 0.77	5.30$\pm$0.30
S433	spatial-block	manuscript	172	4.32 $\pm$ 0.05	4.38 $\pm$ 0.15	3.98$\pm$0.12
S433	spatial-block	train, both	173	4.34 $\pm$ 0.03	5.10 $\pm$ 0.08	4.15$\pm$0.07
S433	spatial-block	train, test unfiltered	175	4.45 $\pm$ 0.03	5.18 $\pm$ 0.08	4.26$\pm$0.07
W740	random	manuscript	281	1.03 $\pm$ 0.09	1.13 $\pm$ 0.04	1.00$\pm$0.08
W740	random	train, both	281	1.00 $\pm$ 0.09	1.13 $\pm$ 0.15	0.99$\pm$0.10
W740	random	train, test unfiltered	296	2.40$\pm$0.13	3.25 $\pm$ 0.87	2.63 $\pm$ 0.35
W740	chronological	manuscript	281	9.30 $\pm$ 0.08	5.20$\pm$0.83	7.78 $\pm$ 0.29
W740	chronological	train, both	247	10.52 $\pm$ 0.12	7.44$\pm$1.12	9.45 $\pm$ 0.28
W740	chronological	train, test unfiltered	296	14.11 $\pm$ 0.12	10.85$\pm$1.12	13.03 $\pm$ 0.30
W740	spatial-block	manuscript	187	3.83 $\pm$ 0.63	3.73 $\pm$ 0.43	3.58$\pm$0.53
W740	spatial-block	train, both	183	3.82 $\pm$ 0.69	3.93 $\pm$ 0.34	3.60$\pm$0.58
W740	spatial-block	train, test unfiltered	197	4.62 $\pm$ 0.76	4.73 $\pm$ 0.43	4.41$\pm$0.66
M1500	random	manuscript	222	2.07 $\pm$ 0.10	2.18 $\pm$ 0.12	1.95$\pm$0.07
M1500	random	train, both	222	2.04 $\pm$ 0.06	2.20 $\pm$ 0.23	1.94$\pm$0.07
M1500	random	train, test unfiltered	224	2.11 $\pm$ 0.10	2.29 $\pm$ 0.25	2.02$\pm$0.11
M1500	chronological	manuscript	222	5.63$\pm$0.06	6.57 $\pm$ 0.45	5.64 $\pm$ 0.14
M1500	chronological	train, both	223	5.75$\pm$0.06	6.44 $\pm$ 0.28	5.80 $\pm$ 0.12
M1500	chronological	train, test unfiltered	223	5.75$\pm$0.06	6.44 $\pm$ 0.28	5.80 $\pm$ 0.12
M1500	spatial-block	manuscript	148	4.14 $\pm$ 0.23	5.05 $\pm$ 0.93	4.06$\pm$0.35
M1500	spatial-block	train, both	146	4.07$\pm$0.23	4.95 $\pm$ 0.55	4.08 $\pm$ 0.26
M1500	spatial-block	train, test unfiltered	149	4.22$\pm$0.20	5.09 $\pm$ 0.51	4.24 $\pm$ 0.22
M3300	random	manuscript	158	1.75$\pm$0.09	2.06 $\pm$ 0.19	1.77 $\pm$ 0.12
M3300	random	train, both	158	1.70$\pm$0.16	1.99 $\pm$ 0.09	1.72 $\pm$ 0.14
M3300	random	train, test unfiltered	164	1.99$\pm$0.15	2.30 $\pm$ 0.10	2.02 $\pm$ 0.09
M3300	chronological	manuscript	158	4.83 $\pm$ 0.01	4.46$\pm$0.04	4.56 $\pm$ 0.14
M3300	chronological	train, both	138	4.28 $\pm$ 0.01	4.35 $\pm$ 0.53	4.01$\pm$0.09
M3300	chronological	train, test unfiltered	163	6.61 $\pm$ 0.02	5.90$\pm$0.39	6.22 $\pm$ 0.12
M3300	spatial-block	manuscript	105	2.98 $\pm$ 0.14	3.36 $\pm$ 0.14	2.89$\pm$0.13
M3300	spatial-block	train, both	97	2.80 $\pm$ 0.18	2.91 $\pm$ 0.22	2.68$\pm$0.05
M3300	spatial-block	train, test unfiltered	109	3.17 $\pm$ 0.11	3.46 $\pm$ 0.17	3.11$\pm$0.13

Table S4: Outlier-rule sensitivity, transfer protocols. Test RMSE in dB, mean $\pm$ std over seeds 42 to 44. In the manuscript regime the target dataset is filtered with its own fences; “train, both” applies the fences of the source to the target, which discards most target records in the cross-frequency pairs and all of them for M3300 $\rightarrow$ M1500 (no value); “train, test unfiltered” scores every valid target record. The best of the three models in each row is bolded; rows without values are not applicable.

Protocol	Fence regime	Records	Random Forest	Deep backbone	PACE-A2G
W433 $\rightarrow$ S433	manuscript	863	20.74$\pm$0.04	21.18 $\pm$ 0.47	20.79 $\pm$ 0.11
W433 $\rightarrow$ S433	train, both	163	14.25$\pm$0.03	15.37 $\pm$ 0.73	14.49 $\pm$ 0.21
W433 $\rightarrow$ S433	train, test unfiltered	876	20.62$\pm$0.04	21.05 $\pm$ 0.47	20.68 $\pm$ 0.11
S433 $\rightarrow$ W433	manuscript	979	25.57 $\pm$ 0.07	22.62$\pm$0.10	24.62 $\pm$ 0.07
S433 $\rightarrow$ W433	train, both	82	7.67 $\pm$ 0.05	8.65 $\pm$ 1.35	7.42$\pm$0.12
S433 $\rightarrow$ W433	train, test unfiltered	1059	24.62 $\pm$ 0.07	21.93$\pm$0.07	23.73 $\pm$ 0.07
W433 $\rightarrow$ W740	manuscript	938	15.19 $\pm$ 0.05	13.94$\pm$0.70	14.69 $\pm$ 0.22
W433 $\rightarrow$ W740	train, both	875	14.85 $\pm$ 0.04	13.40$\pm$0.86	14.28 $\pm$ 0.26
W433 $\rightarrow$ W740	train, test unfiltered	988	15.91 $\pm$ 0.06	14.86$\pm$0.51	15.47 $\pm$ 0.19
W740 $\rightarrow$ W433	manuscript	979	12.25$\pm$0.04	15.75 $\pm$ 1.72	13.17 $\pm$ 0.52
W740 $\rightarrow$ W433	train, both	300	6.39$\pm$0.05	9.97 $\pm$ 2.28	7.19 $\pm$ 0.76
W740 $\rightarrow$ W433	train, test unfiltered	1059	12.66$\pm$0.03	15.82 $\pm$ 1.22	13.44 $\pm$ 0.34
M1500 $\rightarrow$ M3300	manuscript	527	28.15 $\pm$ 0.03	27.84$\pm$0.32	28.01 $\pm$ 0.11
M1500 $\rightarrow$ M3300	train, both	6	13.85 $\pm$ 0.09	12.48$\pm$0.47	13.40 $\pm$ 0.12
M1500 $\rightarrow$ M3300	train, test unfiltered	545	27.84 $\pm$ 0.03	27.51$\pm$0.31	27.70 $\pm$ 0.11
M3300 $\rightarrow$ M1500	manuscript	741	26.66$\pm$0.02	27.24 $\pm$ 0.37	26.81 $\pm$ 0.10
M3300 $\rightarrow$ M1500	train, both	0	—	—	—
M3300 $\rightarrow$ M1500	train, test unfiltered	745	26.60$\pm$0.02	27.18 $\pm$ 0.38	26.75 $\pm$ 0.10
water+suburb $\rightarrow$ mtn	manuscript	1290	13.89 $\pm$ 0.03	15.35 $\pm$ 0.26	13.85$\pm$0.05
water+suburb $\rightarrow$ mtn	train, both	1290	13.89 $\pm$ 0.03	15.35 $\pm$ 0.26	13.85$\pm$0.05
water+suburb $\rightarrow$ mtn	train, test unfiltered	1290	13.89 $\pm$ 0.03	15.35 $\pm$ 0.26	13.85$\pm$0.05
water+mtn $\rightarrow$ suburban	manuscript	863	21.10 $\pm$ 0.09	17.53$\pm$4.28	19.65 $\pm$ 1.38
water+mtn $\rightarrow$ suburban	train, both	876	20.98 $\pm$ 0.09	17.45$\pm$4.25	19.53 $\pm$ 1.37
water+mtn $\rightarrow$ suburban	train, test unfiltered	876	20.98 $\pm$ 0.09	17.45$\pm$4.25	19.53 $\pm$ 1.37
suburb+mtn $\rightarrow$ overwater	manuscript	2041	19.03 $\pm$ 0.04	17.85$\pm$0.97	17.94 $\pm$ 0.59
suburb+mtn $\rightarrow$ overwater	train, both	2047	19.01 $\pm$ 0.04	17.86$\pm$0.98	17.92 $\pm$ 0.58
suburb+mtn $\rightarrow$ overwater	train, test unfiltered	2047	19.01 $\pm$ 0.04	17.86$\pm$0.98	17.92 $\pm$ 0.58

S7 Out-of-Distribution Gate and Adaptive Recalibration: Per-Pair Results

This section accompanies Section 4.6, Table 7, and Figure 8 of the main paper. For each of the 20 ordered source–target pairs of the five datasets, the source model is trained and calibrated in distribution exactly as in Table 6 of the main paper for seeds 42 to 44 and transferred to the target without retraining. Table S6 lists, per pair and averaged over seeds, the point RMSE on the target, the detection AUROC of the coordinate-free feature distance (kNN), the TabM member standard deviation (std), and the raw interval width, the fraction of target records flagged by the coordinate-free distance at the 95th percentile of the in-distribution calibration slice (TPR), the target PICP@90 with the in-distribution correction E and after recalibration from $k = 10$ and $k = 50$ labelled target records (20 random draws each), the mean interval width at $k = 50$, and the overall PICP@90 of the gated pipeline on the mixed in-distribution plus target stream, with the value obtained when the in-distribution interval is used everywhere in parentheses. The AUROC of the 14-D feature distance is 1.00 on every pair and is not listed.

Table S5: Effect of removing the four absolute coordinates from the input, per protocol. Test RMSE in dB, mean $\pm$ std over seeds 42 to 44, for the canonical 14-D vector (14-D) and the coordinate-free 10-D variant (10-D). Group D rows are leave-one-dataset-out; the last block is the in-distribution random 70/30 split of each dataset.

Group	Protocol	Random Forest		Deep backbone		PACE-A2G	
		14-D	10-D	14-D	10-D	14-D	10-D
A x-scen	W433 $\rightarrow$ S433	20.74 $\pm$ 0.04	21.51 $\pm$ 0.10	21.18 $\pm$ 0.47	20.98 $\pm$ 0.43	20.79 $\pm$ 0.11	21.29 $\pm$ 0.19
A x-scen	S433 $\rightarrow$ W433	25.57 $\pm$ 0.07	23.93 $\pm$ 0.16	22.62 $\pm$ 0.10	20.99 $\pm$ 1.09	24.62 $\pm$ 0.07	22.98 $\pm$ 0.29
B x-freq	W433 $\rightarrow$ W740	15.19 $\pm$ 0.05	14.40 $\pm$ 0.07	13.94 $\pm$ 0.70	11.58 $\pm$ 0.81	14.69 $\pm$ 0.22	13.32 $\pm$ 0.29
B x-freq	W740 $\rightarrow$ W433	12.25 $\pm$ 0.04	13.64 $\pm$ 0.08	15.75 $\pm$ 1.72	16.24 $\pm$ 1.53	13.17 $\pm$ 0.52	14.33 $\pm$ 0.41
B x-freq	M1500 $\rightarrow$ M3300	28.15 $\pm$ 0.03	28.58 $\pm$ 0.07	27.84 $\pm$ 0.32	27.53 $\pm$ 1.50	28.01 $\pm$ 0.11	28.20 $\pm$ 0.44
B x-freq	M3300 $\rightarrow$ M1500	26.66 $\pm$ 0.02	26.20 $\pm$ 0.10	27.24 $\pm$ 0.37	26.95 $\pm$ 0.45	26.81 $\pm$ 0.10	26.40 $\pm$ 0.13
C LoSO	water+suburb $\rightarrow$ mountainous	13.89 $\pm$ 0.03	14.54 $\pm$ 0.04	15.35 $\pm$ 0.26	16.75 $\pm$ 0.27	13.85 $\pm$ 0.05	14.50 $\pm$ 0.13
C LoSO	water+mountainous $\rightarrow$ suburban	21.10 $\pm$ 0.09	20.36 $\pm$ 0.08	17.53 $\pm$ 4.28	19.86 $\pm$ 0.74	19.65 $\pm$ 1.38	20.16 $\pm$ 0.24
C LoSO	suburban+mountainous $\rightarrow$ overwater	19.03 $\pm$ 0.04	18.45 $\pm$ 0.09	17.85 $\pm$ 0.97	17.94 $\pm$ 0.51	17.94 $\pm$ 0.59	18.20 $\pm$ 0.20
D LoDO	others $\rightarrow$ W433	25.31 $\pm$ 0.08	24.44 $\pm$ 0.13	21.83 $\pm$ 0.48	20.33 $\pm$ 1.05	24.14 $\pm$ 0.19	23.02 $\pm$ 0.21
D LoDO	others $\rightarrow$ S433	21.10 $\pm$ 0.09	20.36 $\pm$ 0.08	17.53 $\pm$ 4.28	19.86 $\pm$ 0.74	19.65 $\pm$ 1.38	20.16 $\pm$ 0.24
D LoDO	others $\rightarrow$ W740	9.63 $\pm$ 0.05	10.69 $\pm$ 0.08	12.95 $\pm$ 1.70	12.90 $\pm$ 0.67	7.01 $\pm$ 1.94	8.72 $\pm$ 1.22
D LoDO	others $\rightarrow$ M1500	25.29 $\pm$ 0.07	25.50 $\pm$ 0.06	24.95 $\pm$ 0.43	23.97 $\pm$ 0.55	25.07 $\pm$ 0.19	24.73 $\pm$ 0.10
D LoDO	others $\rightarrow$ M3300	28.04 $\pm$ 0.01	26.19 $\pm$ 0.10	27.83 $\pm$ 1.05	26.81 $\pm$ 0.70	27.91 $\pm$ 0.34	26.23 $\pm$ 0.32
ID	W433 random 70/30	2.01 $\pm$ 0.08	2.06 $\pm$ 0.09	2.37 $\pm$ 0.19	2.81 $\pm$ 0.08	1.96 $\pm$ 0.09	2.08 $\pm$ 0.08
ID	S433 random 70/30	2.35 $\pm$ 0.11	2.47 $\pm$ 0.15	2.36 $\pm$ 0.08	2.93 $\pm$ 0.37	2.19 $\pm$ 0.09	2.39 $\pm$ 0.20
ID	W740 random 70/30	1.03 $\pm$ 0.09	1.04 $\pm$ 0.09	1.13 $\pm$ 0.04	1.22 $\pm$ 0.14	1.00 $\pm$ 0.08	1.02 $\pm$ 0.09
ID	M1500 random 70/30	2.07 $\pm$ 0.10	2.09 $\pm$ 0.10	2.18 $\pm$ 0.12	2.59 $\pm$ 0.05	1.95 $\pm$ 0.07	2.13 $\pm$ 0.09
ID	M3300 random 70/30	1.75 $\pm$ 0.09	1.76 $\pm$ 0.08	2.06 $\pm$ 0.19	2.34 $\pm$ 0.32	1.77 $\pm$ 0.12	1.77 $\pm$ 0.13

Table S6: Out-of-distribution detection, adaptive recalibration, and the gated interval pipeline for the 20 ordered source–target pairs, mean over seeds 42 to 44. Columns are defined in the text of this section.

Pair	RMSE (dB)	AUROC			TPR kNN	Target PICP@90			MPIW (dB)	Gated PICP@90
		kNN	std	width		ID	E	$k = 10$ $k = 50$		
W433 $\rightarrow$ S433	22.3	0.99	0.40	0.43	0.98	0.00	0.91	0.92	58	0.91 (0.24)
W433 $\rightarrow$ W740	13.9	1.00	0.37	0.40	1.00	0.01	0.93	0.92	39	0.92 (0.23)
W433 $\rightarrow$ M1500	21.5	1.00	0.40	0.40	1.00	0.00	0.89	0.92	57	0.92 (0.27)
W433 $\rightarrow$ M3300	9.4	1.00	0.46	0.35	1.00	0.15	0.90	0.91	29	0.92 (0.44)
S433 $\rightarrow$ W433	22.7	0.99	0.40	0.46	0.93	0.00	0.92	0.91	59	0.87 (0.19)
S433 $\rightarrow$ W740	9.3	1.00	0.41	0.51	1.00	0.18	0.92	0.92	31	0.92 (0.34)
S433 $\rightarrow$ M1500	5.8	1.00	0.38	0.44	1.00	0.39	0.93	0.92	20	0.91 (0.52)
S433 $\rightarrow$ M3300	26.6	1.00	0.39	0.37	1.00	0.00	0.91	0.92	65	0.91 (0.31)
W740 $\rightarrow$ W433	14.8	1.00	0.45	0.61	1.00	0.01	0.92	0.92	42	0.92 (0.21)
W740 $\rightarrow$ S433	9.4	1.00	0.42	0.59	1.00	0.09	0.89	0.93	31	0.92 (0.29)
W740 $\rightarrow$ M1500	11.5	1.00	0.53	0.51	1.00	0.06	0.91	0.93	37	0.91 (0.30)
W740 $\rightarrow$ M3300	19.9	1.00	0.45	0.59	1.00	0.00	0.91	0.93	52	0.92 (0.32)
M1500 $\rightarrow$ W433	23.0	1.00	0.34	0.36	1.00	0.00	0.93	0.92	61	0.93 (0.18)
M1500 $\rightarrow$ S433	5.3	1.00	0.33	0.40	1.00	0.42	0.91	0.92	19	0.92 (0.53)
M1500 $\rightarrow$ W740	10.1	1.00	0.37	0.44	1.00	0.11	0.92	0.92	32	0.93 (0.27)
M1500 $\rightarrow$ M3300	28.1	1.00	0.39	0.39	1.00	0.00	0.90	0.92	67	0.92 (0.29)
M3300 $\rightarrow$ W433	7.8	1.00	0.46	0.56	1.00	0.31	0.90	0.92	29	0.92 (0.40)
M3300 $\rightarrow$ S433	27.9	1.00	0.42	0.57	1.00	0.00	0.89	0.93	69	0.92 (0.15)
M3300 $\rightarrow$ W740	18.7	1.00	0.52	0.59	1.00	0.00	0.90	0.92	49	0.92 (0.14)
M3300 $\rightarrow$ M1500	27.5	1.00	0.47	0.52	1.00	0.00	0.92	0.92	68	0.92 (0.17)