

Online Resource 1

Supplementary Material

Leakage-Controlled Machine Learning for European Cd, Hg, and Pb Emission Inventories

Water, Air, & Soil Pollution

Tianyi Guan and Jennifer Uyen-Vi Nguyen

Department of Physical and Environmental Sciences, University of Toronto, Toronto, Ontario, Canada

Department of Computer and Mathematical Sciences, University of Toronto, Toronto, Ontario, Canada

Department of Biological Sciences, University of Toronto, Toronto, Ontario, Canada

Corresponding author: tianyi.guan@mail.utoronto.ca

Overview

This document accompanies the main manuscript and reports model configurations, complete candidate-model comparisons, residual and target diagnostics, feature summaries, and supplementary figures. Numerical tables and figures were generated from the Dataset v3 products and model-training outputs.

Abbreviations used in tables and figures: BC, black carbon; BaP, benzo[a]pyrene; BbF, benzo[b]fluoranthene; BkF, benzo[k]fluoranthene; DIOX, dioxins; HCB, hexachlorobenzene; MAE, mean absolute error; MedAE, median absolute error; NMVOC, non-methane volatile organic compounds; NO_x, nitrogen oxides; PM, particulate matter; RMSE, root mean squared error; and SO_x, sulfur oxides.

S1 Candidate models and tuning

Table S1 Candidate algorithms and principal training or tuning settings. HGB denotes histogram gradient boosting; MLP denotes multilayer perceptron

Algorithm	Model family		Principal configuration
ElasticNetCV	Regularized	linear	Seven l_1 ratios (0.05–0.98), 80 α values, five-fold internal cross-validation.
HGB	Boosted trees		1,200 iterations, learning rate 0.025, 63 leaf nodes, $L_2 = 0.05$; four validation candidates and early stopping.
Gradient boosting	Boosted trees		1,200 trees, learning rate 0.025, depth 4, subsample 0.85, minimum leaf size 3.
Random forest	Bagged trees		800 trees, $\sqrt{p}$ features per split, minimum leaf size 2, unrestricted depth.
Extra Trees	Randomized trees		1,000 trees, $\sqrt{p}$ features per split, minimum leaf size 1, unrestricted depth.
Scikit-learn MLP	Neural network		Hidden widths 512–256–128–64, 600 iterations, learning rate 0.0007, $L_2 = 10^{-4}$, early stopping.
PyTorch MLP	Neural network		Hidden widths 512–512–256–128–64, dropout 0.10, 350 epochs, batch 2,048, learning rate 0.0007, weight decay 10^{-4} .
XGBoost	Boosted trees		3,500 trees; four candidates spanning learning rate 0.012–0.025, depth 6–9, row/column sampling 0.82–0.90, and minimum child weight 2–8.
LightGBM	Boosted trees		5,000 trees; five candidates spanning learning rate 0.009–0.018, 63–255 leaves, row sampling 0.82–0.90, column sampling 0.85–0.95, and minimum child size 8–30.
CatBoost	Boosted trees		3,500 iterations, learning rate 0.018, depth 8, L_2 leaf penalty 4, RMSE loss.

Table S2 Validation-selected boosting configurations and recorded iteration endpoints

Metal Task	Model	Learning rate	Complexity	Row/column sampling	L_2	Minimum child	Best iteration	Val. RMSE
Cd Absolute	LightGBM	0.010	191 leaves	0.90/0.90	2.5	16 samples	5,000	0.117
Cd Change	LightGBM	0.012	255 leaves	0.82/0.90	3.0	10 samples	4,996	0.110
Hg Absolute	LightGBM	0.012	127 leaves	0.85/0.85	2.0	20 samples	5,000	0.151
Hg Change	LightGBM	0.012	255 leaves	0.82/0.90	3.0	10 samples	5,000	0.147
Pb Absolute	LightGBM	0.012	255 leaves	0.82/0.90	3.0	10 samples	5,000	0.122
Pb Change	XGBoost	0.012	depth 9	0.82/0.82	3.0	weight 8	3,499	0.126

S2 Complete candidate-model comparisons

The six tables in this section report all ten candidate algorithms for each metal and task. Bold model names identify the single model selected by validation RMSE; held-out spatial-block test results were not used for model selection.

Table S3 Complete ten-model comparison for Cd absolute updating. Bold indicates the model selected by validation RMSE

Model	RMSE	MAE	MedAE	R^2	Pearson r	Spearman ρ	Bias	Training (s)
<i>Panel A: validation metrics</i>								
LightGBM	0.117	0.055	0.023	0.988	0.994	0.995	0.002	389.3
XGBoost	0.119	0.060	0.029	0.988	0.994	0.994	0.001	445.4
CatBoost	0.124	0.069	0.035	0.987	0.993	0.994	-0.002	142.6
HistGradientBoosting	0.125	0.064	0.030	0.986	0.993	0.994	0.000	180.1
RandomForest	0.165	0.091	0.045	0.976	0.988	0.989	-0.004	40.3
GradientBoosting	0.168	0.111	0.072	0.975	0.988	0.988	-0.008	4575.6
ExtraTrees	0.170	0.088	0.038	0.975	0.988	0.989	-0.006	34.7
SklearnDeepMLP	0.170	0.086	0.045	0.975	0.987	0.990	-0.008	111.2
TorchDeepMLP	0.178	0.097	0.054	0.972	0.987	0.989	0.021	1019.9
ElasticNetCV	0.321	0.200	0.120	0.910	0.954	0.966	-0.019	371.2
<i>Panel B: held-out spatial-block test metrics</i>								
LightGBM	0.150	0.076	0.031	0.985	0.993	0.991	-0.013	389.3
XGBoost	0.150	0.078	0.036	0.985	0.993	0.991	-0.010	445.4
CatBoost	0.157	0.088	0.045	0.984	0.992	0.990	-0.013	142.6
HistGradientBoosting	0.160	0.085	0.038	0.983	0.992	0.989	-0.015	180.1
RandomForest	0.194	0.112	0.054	0.975	0.988	0.984	-0.027	40.3
GradientBoosting	0.196	0.125	0.079	0.975	0.988	0.984	-0.023	4575.6
ExtraTrees	0.199	0.114	0.053	0.974	0.987	0.983	-0.031	34.7
SklearnDeepMLP	0.269	0.110	0.052	0.952	0.977	0.987	-0.024	111.2
TorchDeepMLP	0.216	0.120	0.062	0.969	0.985	0.985	0.003	1019.9
ElasticNetCV	0.368	0.249	0.155	0.911	0.955	0.959	-0.022	371.2

Table S4 Complete ten-model comparison for Cd change updating. Bold indicates the model selected by validation RMSE

Model	RMSE	MAE	MedAE	R^2	Pearson r	Spearman ρ	Bias	Training (s)
<i>Panel A: validation metrics</i>								
LightGBM	0.110	0.048	0.015	0.896	0.947	0.918	0.003	386.4
XGBoost	0.111	0.054	0.021	0.893	0.945	0.907	0.002	524.6
HistGradientBoosting	0.119	0.061	0.027	0.877	0.937	0.887	0.000	188.3
CatBoost	0.122	0.064	0.029	0.872	0.935	0.886	-0.002	144.6
RandomForest	0.140	0.065	0.021	0.831	0.917	0.888	0.000	47.9
ExtraTrees	0.142	0.065	0.018	0.824	0.911	0.885	-0.000	30.9
SklearnDeepMLP	0.145	0.071	0.031	0.818	0.905	0.864	-0.006	135.2
TorchDeepMLP	0.148	0.070	0.026	0.810	0.902	0.866	0.020	716.5
GradientBoosting	0.157	0.099	0.061	0.785	0.890	0.779	-0.006	4853.0
ElasticNetCV	0.319	0.200	0.118	0.119	0.353	0.378	-0.014	347.1
<i>Panel B: held-out spatial-block test metrics</i>								
LightGBM	0.144	0.069	0.025	0.861	0.928	0.918	-0.007	386.4
XGBoost	0.144	0.073	0.031	0.862	0.929	0.918	-0.007	524.6
HistGradientBoosting	0.153	0.082	0.037	0.843	0.919	0.901	-0.011	188.3
CatBoost	0.155	0.085	0.039	0.839	0.917	0.901	-0.011	144.6
RandomForest	0.173	0.089	0.035	0.800	0.899	0.900	-0.013	47.9
ExtraTrees	0.187	0.097	0.040	0.766	0.879	0.885	-0.019	30.9
SklearnDeepMLP	0.182	0.093	0.042	0.779	0.883	0.876	-0.010	135.2
TorchDeepMLP	0.183	0.093	0.038	0.775	0.882	0.878	0.015	716.5
GradientBoosting	0.184	0.113	0.065	0.773	0.881	0.849	-0.016	4853.0
ElasticNetCV	0.368	0.250	0.153	0.093	0.324	0.356	-0.019	347.1

Table S5 Complete ten-model comparison for Hg absolute updating. Bold indicates the model selected by validation RMSE

Model	RMSE	MAE	MedAE	R^2	Pearson r	Spearman ρ	Bias	Training (s)
<i>Panel A: validation metrics</i>								
LightGBM	0.151	0.074	0.032	0.979	0.990	0.991	-0.010	667.2
XGBoost	0.153	0.078	0.036	0.979	0.989	0.991	-0.009	180.7
HistGradientBoosting	0.156	0.081	0.039	0.978	0.989	0.990	-0.009	293.5
CatBoost	0.157	0.083	0.042	0.978	0.989	0.990	-0.012	166.3
GradientBoosting	0.181	0.111	0.067	0.970	0.985	0.986	-0.012	5733.4
SklearnDeepMLP	0.198	0.104	0.050	0.965	0.982	0.986	-0.007	296.1
TorchDeepMLP	0.203	0.120	0.071	0.963	0.981	0.986	0.014	176.1
ExtraTrees	0.205	0.118	0.058	0.962	0.981	0.983	-0.021	72.8
RandomForest	0.207	0.118	0.061	0.961	0.981	0.983	-0.017	83.3
ElasticNetCV	0.317	0.214	0.150	0.909	0.955	0.953	-0.022	4.3
<i>Panel B: held-out spatial-block test metrics</i>								
LightGBM	0.170	0.084	0.031	0.975	0.987	0.984	0.001	667.2
XGBoost	0.175	0.090	0.036	0.973	0.987	0.983	0.003	180.7
HistGradientBoosting	0.173	0.090	0.037	0.974	0.987	0.983	0.001	293.5
CatBoost	0.171	0.091	0.039	0.974	0.987	0.983	-0.003	166.3
GradientBoosting	0.189	0.115	0.066	0.969	0.984	0.980	-0.004	5733.4
SklearnDeepMLP	0.205	0.109	0.048	0.963	0.981	0.980	0.002	296.1
TorchDeepMLP	0.208	0.119	0.063	0.962	0.981	0.980	0.026	176.1
ExtraTrees	0.199	0.111	0.052	0.965	0.983	0.977	-0.007	72.8
RandomForest	0.201	0.114	0.054	0.965	0.982	0.978	-0.001	83.3
ElasticNetCV	0.323	0.230	0.151	0.909	0.956	0.951	-0.054	4.3

Table S6 Complete ten-model comparison for Hg change updating. Bold indicates the model selected by validation RMSE

Model	RMSE	MAE	MedAE	R^2	Pearson r	Spearman ρ	Bias	Training (s)
<i>Panel A: validation metrics</i>								
LightGBM	0.147	0.066	0.020	0.788	0.888	0.862	-0.005	658.9
XGBoost	0.148	0.069	0.025	0.788	0.888	0.852	-0.005	179.3
HistGradientBoosting	0.154	0.075	0.030	0.768	0.877	0.840	-0.006	310.4
CatBoost	0.154	0.077	0.032	0.767	0.876	0.828	-0.009	168.5
RandomForest	0.167	0.078	0.026	0.729	0.855	0.832	-0.005	85.3
TorchDeepMLP	0.167	0.078	0.027	0.727	0.853	0.827	0.004	289.6
ExtraTrees	0.168	0.077	0.023	0.726	0.853	0.837	-0.006	66.9
GradientBoosting	0.176	0.101	0.054	0.699	0.837	0.756	-0.010	5662.4
SklearnDeepMLP	0.179	0.087	0.032	0.688	0.833	0.796	-0.005	374.5
ElasticNetCV	0.316	0.209	0.136	0.023	0.207	0.161	-0.028	15.6
<i>Panel B: held-out spatial-block test metrics</i>								
LightGBM	0.168	0.078	0.020	0.712	0.844	0.854	0.002	658.9
XGBoost	0.169	0.081	0.024	0.711	0.843	0.848	0.002	179.3
HistGradientBoosting	0.172	0.085	0.029	0.700	0.837	0.841	0.001	310.4
CatBoost	0.174	0.088	0.031	0.693	0.833	0.828	-0.000	168.5
RandomForest	0.193	0.092	0.024	0.620	0.792	0.826	-0.001	85.3
TorchDeepMLP	0.194	0.095	0.029	0.619	0.788	0.809	0.014	289.6
ExtraTrees	0.196	0.092	0.024	0.611	0.785	0.832	-0.005	66.9
GradientBoosting	0.187	0.108	0.055	0.645	0.805	0.773	-0.000	5662.4
SklearnDeepMLP	0.200	0.104	0.038	0.595	0.776	0.796	0.015	374.5
ElasticNetCV	0.319	0.223	0.139	-0.034	0.116	0.009	-0.068	15.6

Table S7 Complete ten-model comparison for Pb absolute updating. Bold indicates the model selected by validation RMSE

Model	RMSE	MAE	MedAE	R^2	Pearson r	Spearman ρ	Bias	Training (s)
<i>Panel A: validation metrics</i>								
LightGBM	0.122	0.056	0.023	0.988	0.994	0.996	-0.000	399.2
XGBoost	0.125	0.061	0.027	0.987	0.994	0.995	-0.001	453.4
HistGradientBoosting	0.133	0.068	0.033	0.986	0.993	0.995	-0.002	192.1
CatBoost	0.134	0.071	0.038	0.986	0.993	0.994	-0.002	145.1
ExtraTrees	0.164	0.089	0.044	0.978	0.989	0.991	-0.007	37.1
RandomForest	0.173	0.097	0.051	0.976	0.988	0.990	-0.006	41.2
GradientBoosting	0.176	0.109	0.070	0.975	0.987	0.990	-0.003	5136.5
SklearnDeepMLP	0.189	0.093	0.045	0.971	0.986	0.993	0.006	72.6
TorchDeepMLP	0.227	0.137	0.085	0.958	0.980	0.990	0.038	377.5
ElasticNetCV	0.373	0.211	0.116	0.888	0.942	0.967	0.024	463.0
<i>Panel B: held-out spatial-block test metrics</i>								
LightGBM	0.181	0.086	0.029	0.973	0.987	0.982	-0.021	399.2
XGBoost	0.182	0.089	0.032	0.973	0.987	0.982	-0.022	453.4
HistGradientBoosting	0.185	0.094	0.039	0.972	0.986	0.982	-0.018	192.1
CatBoost	0.180	0.097	0.046	0.974	0.987	0.983	-0.023	145.1
ExtraTrees	0.219	0.118	0.051	0.961	0.981	0.976	-0.034	37.1
RandomForest	0.228	0.126	0.058	0.957	0.979	0.973	-0.033	41.2
GradientBoosting	0.224	0.133	0.076	0.959	0.980	0.973	-0.029	5136.5
SklearnDeepMLP	0.244	0.123	0.053	0.951	0.976	0.978	-0.012	72.6
TorchDeepMLP	0.278	0.162	0.092	0.936	0.968	0.970	0.017	377.5
ElasticNetCV	0.416	0.256	0.153	0.858	0.927	0.925	-0.033	463.0

Table S8 Complete ten-model comparison for Pb change updating. Bold indicates the model selected by validation RMSE

Model	RMSE	MAE	MedAE	R^2	Pearson r	Spearman ρ	Bias	Training (s)
<i>Panel A: validation metrics</i>								
XGBoost	0.126	0.057	0.020	0.892	0.945	0.902	-0.000	516.6
LightGBM	0.127	0.057	0.020	0.891	0.944	0.900	0.001	392.3
CatBoost	0.135	0.067	0.030	0.877	0.938	0.884	-0.004	146.2
HistGradientBoosting	0.136	0.067	0.029	0.874	0.936	0.882	-0.001	193.4
RandomForest	0.166	0.071	0.021	0.814	0.907	0.879	-0.003	46.5
ExtraTrees	0.168	0.071	0.018	0.809	0.904	0.885	-0.003	33.8
GradientBoosting	0.168	0.098	0.055	0.808	0.901	0.819	-0.003	4812.4
TorchDeepMLP	0.183	0.075	0.024	0.774	0.885	0.880	0.034	746.1
SklearnDeepMLP	0.185	0.082	0.025	0.769	0.878	0.847	0.001	124.9
ElasticNetCV	0.367	0.216	0.118	0.091	0.384	0.551	0.018	352.0
<i>Panel B: held-out spatial-block test metrics</i>								
XGBoost	0.185	0.087	0.026	0.832	0.915	0.886	-0.018	516.6
LightGBM	0.192	0.088	0.026	0.818	0.908	0.884	-0.023	392.3
CatBoost	0.186	0.096	0.037	0.829	0.915	0.876	-0.023	146.2
HistGradientBoosting	0.196	0.097	0.035	0.812	0.904	0.880	-0.021	193.4
RandomForest	0.245	0.112	0.030	0.706	0.850	0.842	-0.027	46.5
ExtraTrees	0.255	0.117	0.026	0.680	0.834	0.831	-0.027	33.8
GradientBoosting	0.236	0.131	0.062	0.726	0.859	0.812	-0.029	4812.4
TorchDeepMLP	0.237	0.102	0.030	0.725	0.854	0.877	0.024	746.1
SklearnDeepMLP	0.243	0.113	0.035	0.711	0.848	0.847	-0.027	124.9
ElasticNetCV	0.416	0.254	0.144	0.152	0.410	0.568	-0.008	352.0

S3 Residual, target, and feature diagnostics

Table S9 Held-out test residual summaries for the selected models. Residuals are prediction minus observation in log10 units

Metal	Task	Model	n	Mean	Median	SD	Mean absolute	95th pct. absolute
Cd	Absolute	LightGBM	20,669	-0.013	-0.003	0.150	0.076	0.300
Cd	Change	LightGBM	20,669	-0.007	-0.001	0.144	0.069	0.280
Hg	Absolute	LightGBM	19,454	0.001	0.001	0.170	0.084	0.365
Hg	Change	LightGBM	19,454	0.002	0.000	0.168	0.078	0.369
Pb	Absolute	LightGBM	18,123	-0.021	-0.001	0.180	0.086	0.379
Pb	Change	XGBoost	18,123	-0.018	-0.000	0.184	0.087	0.391

Table S10 Positive-target availability for the compatible 2018–2019 **nationaltotal** demonstration targets

Metal	Year	Non-missing cells	Positive cells retained	Zero cells excluded	Positive (%)
Cd	2018	76,979	76,613	366	99.52
Cd	2019	76,979	75,813	1,166	98.49
Hg	2018	77,312	77,286	26	99.97
Hg	2019	77,312	76,050	1,262	98.37
Pb	2018	77,980	77,967	13	99.98
Pb	2019	77,980	76,883	1,097	98.59

Table S11 Target-family, sparsity, and pooling audit across all available metal-year columns

Metal	Year	Family label	Role	Positive cells	Zero cells	Positive (%)	Pooled
Cd	2017	nationaltotal	Baseline	76,979	0	100.00	Yes
Cd	2018	nationaltotal	Primary	76,613	366	99.52	Yes
Cd	2019	nationaltotal	Primary	75,813	1,166	98.49	Yes
Cd	2020	sumallsectors	Exploratory	450	76,529	0.58	No
Cd	2021	sumallsectors	Exploratory	535	76,444	0.69	No
Cd	2022	sumallsectors	Exploratory	542	76,437	0.70	No
Cd	2023	sumallsectors	Exploratory	574	76,405	0.75	No
Hg	2017	nationaltotal	Baseline	77,312	0	100.00	Yes
Hg	2018	nationaltotal	Primary	77,286	26	99.97	Yes
Hg	2019	nationaltotal	Primary	76,050	1,262	98.37	Yes
Hg	2020	sumallsectors	Exploratory	388	76,924	0.50	No
Hg	2021	sumallsectors	Exploratory	418	76,894	0.54	No
Hg	2022	sumallsectors	Exploratory	378	76,934	0.49	No
Hg	2023	sumallsectors	Exploratory	404	76,908	0.52	No
Pb	2017	nationaltotal	Baseline	77,980	0	100.00	Yes
Pb	2018	nationaltotal	Primary	77,967	13	99.98	Yes
Pb	2019	nationaltotal	Primary	76,883	1,097	98.59	Yes
Pb	2020	sumallsectors	Exploratory	20,338	57,642	26.08	No
Pb	2021	sumallsectors	Exploratory	23,672	54,308	30.36	No
Pb	2022	sumallsectors	Exploratory	24,785	53,195	31.78	No
Pb	2023	sumallsectors	Exploratory	24,922	53,058	31.96	No

All rows labelled “Exploratory” in Supplementary Table S11 belonged to the **sumallsectors** family and were treated as incompatible with the pooled **nationaltotal** demonstration targets.

Table S12 Permutation importance of baseline and target-year variables and the three leading additional predictors for each selected model. Values are increases in validation RMSE after permutation

Metal	Task	Baseline	Target year	Leading additional predictors (importance)
Cd	Absolute	0.895	0.198	2016 NO _x -solvents (0.060); 1990 PM _{coarse} -other stationary combustion (0.037); 2010 BC-all sectors (0.031).
Cd	Change	0.447	0.205	2014 BC-waste (0.070); 2017 HCB-other agriculture (0.051); 2010 NH ₃ -other stationary combustion (0.050).
Hg	Absolute	0.844	0.139	2017 BbF-road transport (0.016); 2013 BC-solvents (0.012); 1995 NH ₃ -other (0.008).
Hg	Change	0.315	0.165	1995 NH ₃ -other (0.075); 2013 BC-solvents (0.072); 2017 BbF-road transport (0.046).
Pb	Absolute	0.724	0.151	2010 BC-all sectors (0.058); 2017 HCB-national total (0.030); 2017 BkF-national total (0.020).
Pb	Change	0.437	0.131	2017 BbF-road transport (0.053); 2017 BbF-off-road (0.050); 1990 NH ₃ -other (0.028).

Table S13 Selected predictor counts by feature-family label in the three Dataset v3 builds

Feature family	Cd	Hg	Pb
BC	11	15	11
BaP	4	3	3
BbF	3	4	4
BkF	5	3	4
CO	17	19	21
DIOX	2	2	1
HCB	2	2	3
NH ₃	31	47	41
NMVOC	41	45	28
NO _x	19	33	36
PM ₁₀	42	31	20
PM _{2.5}	20	13	15
PM _{coarse}	52	58	70
SO _x	51	24	43
Annual precipitation	0	1	0
Total	300	300	300

Table S14 Programmed strict-machine-learning checks recorded for each Dataset v3 build

Check		Cd	Hg	Pb	Recorded implementation
Spatial control	leakage	Pass	Pass	Pass	Spatial block splitting was used; the maximum number of split labels within any block was one.
Split before selection	leakage	Pass	Pass	Pass	Feature screening, scoring, correlation pruning, imputation, and scaling were fitted after split creation using training data only.
Temporal control	leakage	Pass	Pass	Pass	Predictors containing years after the 2017 target year were removed before feature selection.
Target control	leakage	Pass	Pass	Pass	Metal target and target-derived columns were stored outside the model predictor table.
Train-only preprocessing	preprocessing	Pass	Pass	Pass	Imputation medians and scaling parameters were fitted on the training split and applied unchanged to validation and test data.

Table S15 Recorded PyTorch multilayer-perceptron training-history summaries

Metal	Task	Epochs recorded	Best epoch	Minimum val. RMSE	Final val. RMSE	Final train loss
Cd	Absolute	206	161	0.178	0.272	0.095
Cd	Change	153	108	0.148	0.212	0.025
Hg	Absolute	114	69	0.203	0.227	0.074
Hg	Change	191	146	0.167	0.172	0.015
Pb	Absolute	71	26	0.227	0.284	0.068
Pb	Change	138	93	0.183	0.218	0.026

S4 Supplementary figures

Unless otherwise specified, panels follow the same order: Cd absolute-emission updating, Cd relative-change updating, Hg absolute-emission updating, Hg relative-change updating, Pb absolute-emission updating, and Pb relative-change updating.

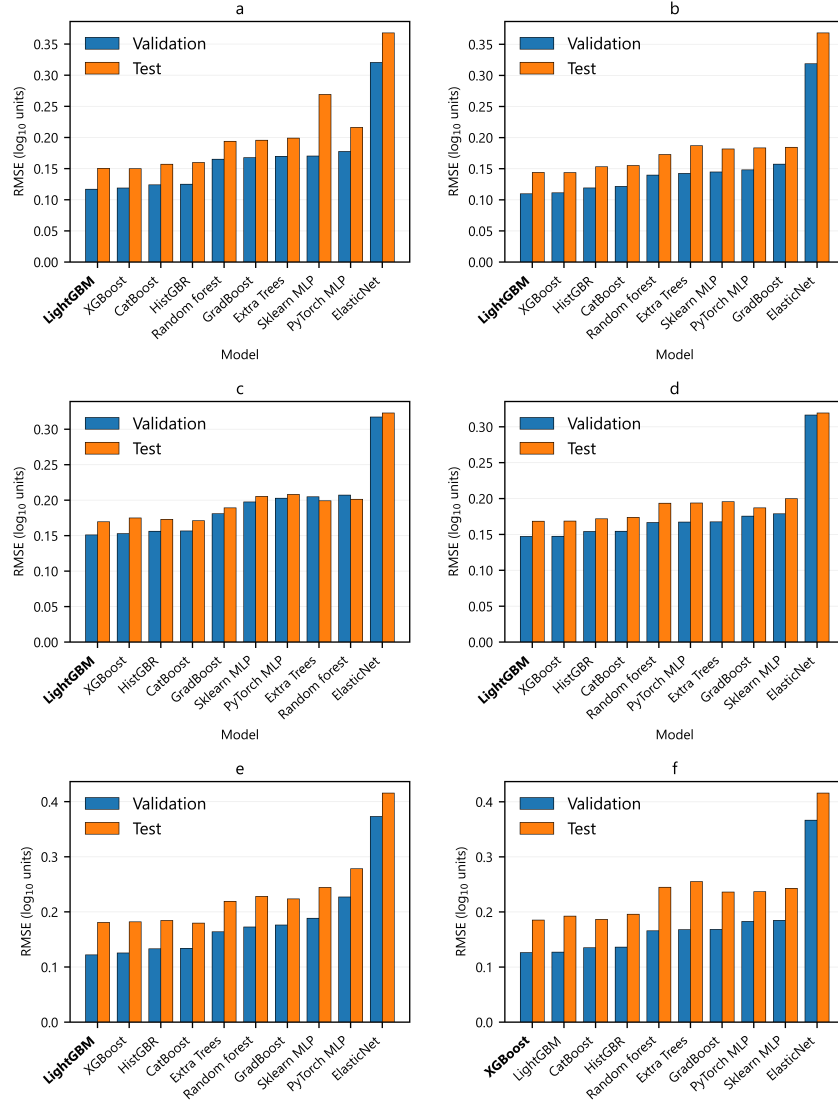


Fig. S1 Complete candidate-model RMSE comparisons for the validation and held-out spatial-block test sets. Panels a and b show Cd absolute-emission and relative-change updating, panels c and d show the corresponding Hg tasks, and panels e and f show the corresponding Pb tasks. Models are ordered by validation RMSE within each panel; bold model labels identify the validation-selected models

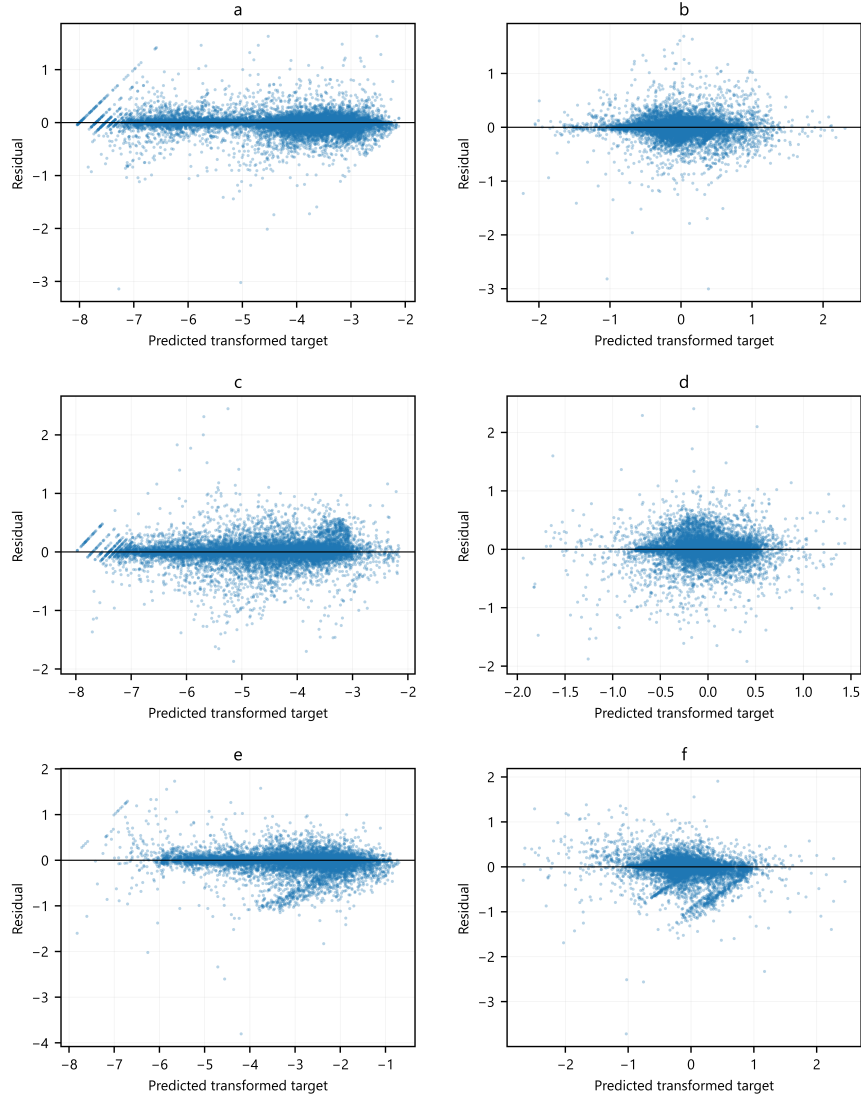


Fig. S2 Residual-versus-prediction diagnostics for the six selected metal-task models. Panels a–f follow the common task order stated above. Residuals are prediction minus observation on the transformed target scale, and the horizontal reference line denotes zero residual

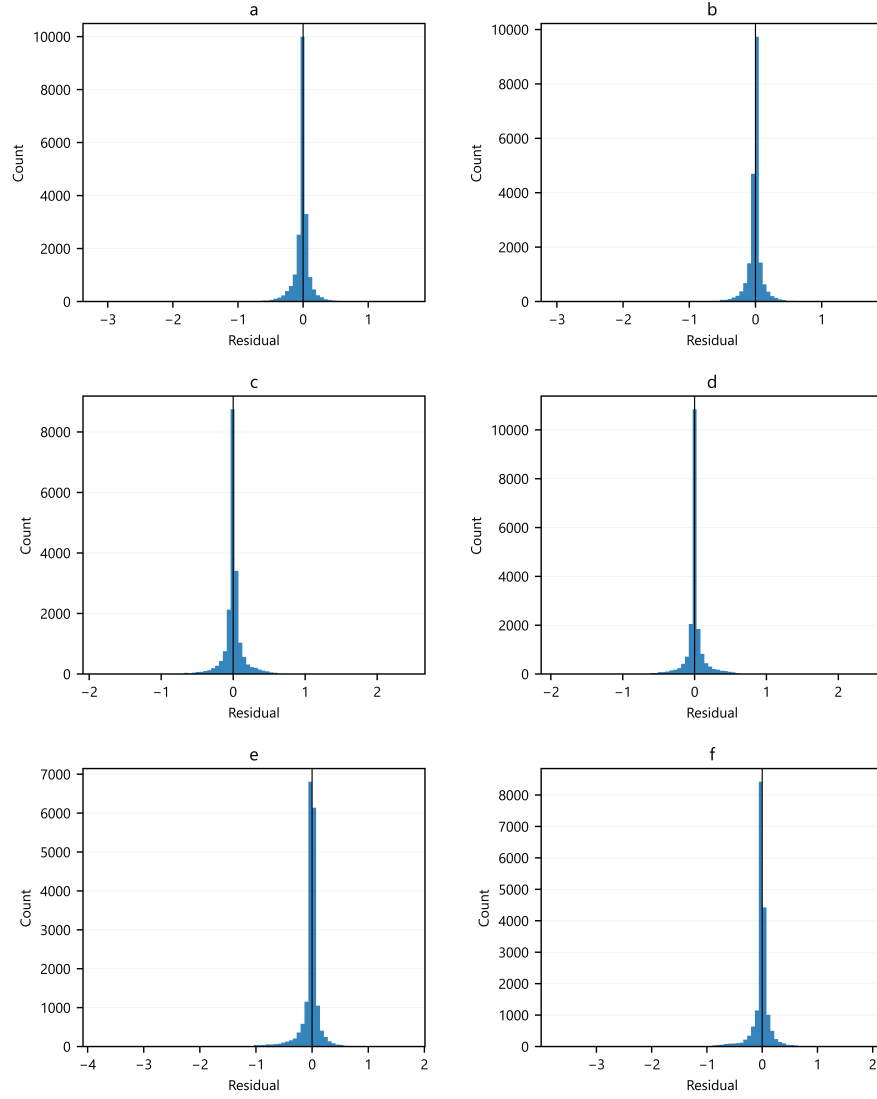


Fig. S3 Held-out spatial-block test residual distributions for the six selected metal-task models. Panels a–f follow the common task order stated above. Residuals are prediction minus observation on the transformed target scale, and the vertical reference line denotes zero residual

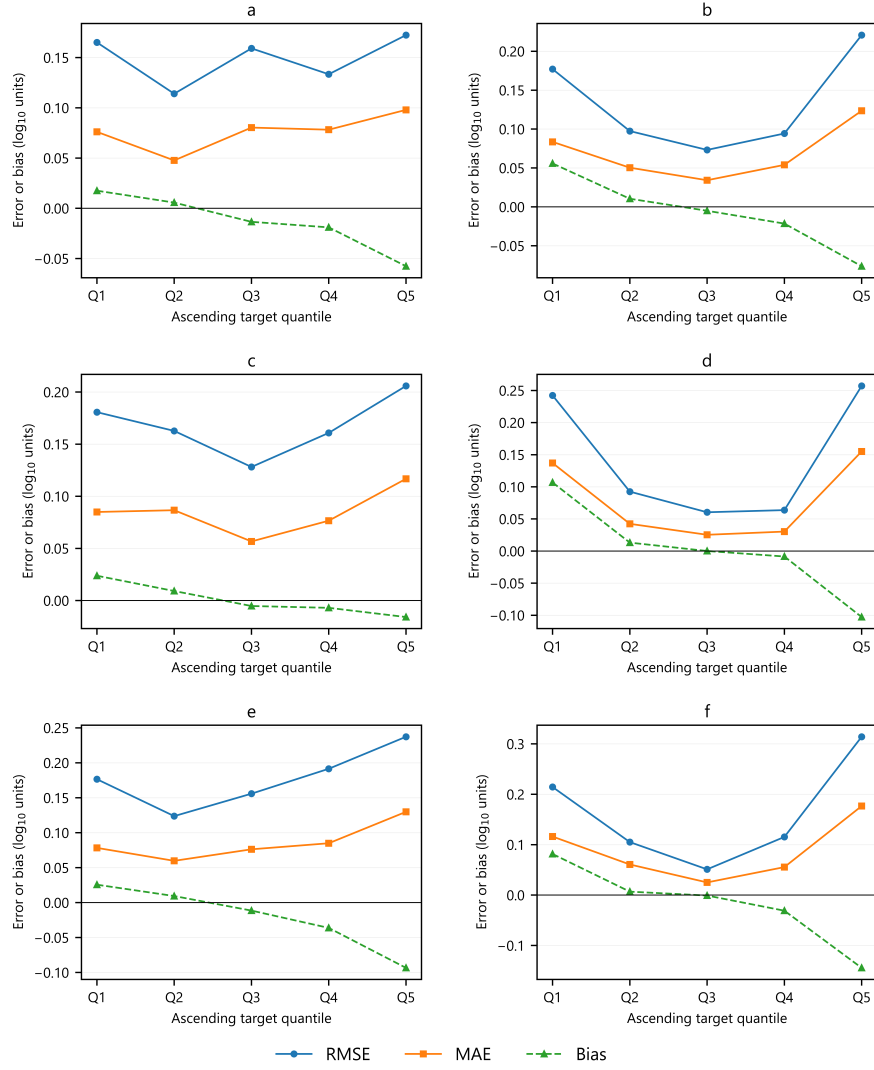


Fig. S4 Target-quantile error profiles for the six selected metal-task models. Panels a–f follow the common task order stated above. Q1–Q5 denote ascending quantiles of the observed held-out target; curves report RMSE, MAE, and bias in $\log_{10}$ units

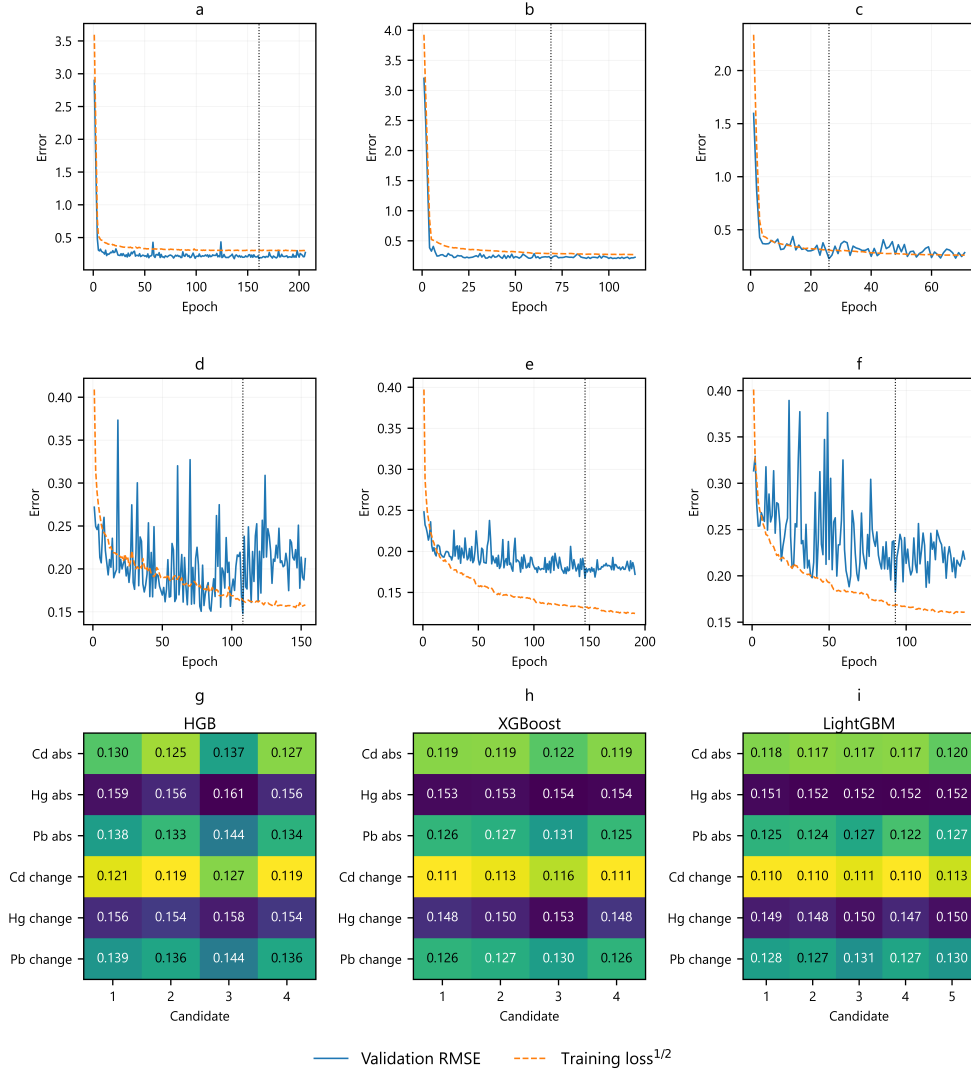


Fig. S5 Recorded neural-network training histories and validation-based tuning diagnostics. Panels a–c show the PyTorch multilayer-perceptron histories for Cd, Hg, and Pb absolute-emission updating, respectively; panels d–f show the corresponding relative-change tasks. Solid curves denote validation RMSE, dashed curves denote the square root of training loss, and vertical dotted lines mark the minimum validation-RMSE epochs. Panels g–i show validation RMSE across the configured candidates for histogram gradient boosting, XGBoost, and LightGBM, respectively; rows correspond to the six metal-task combinations

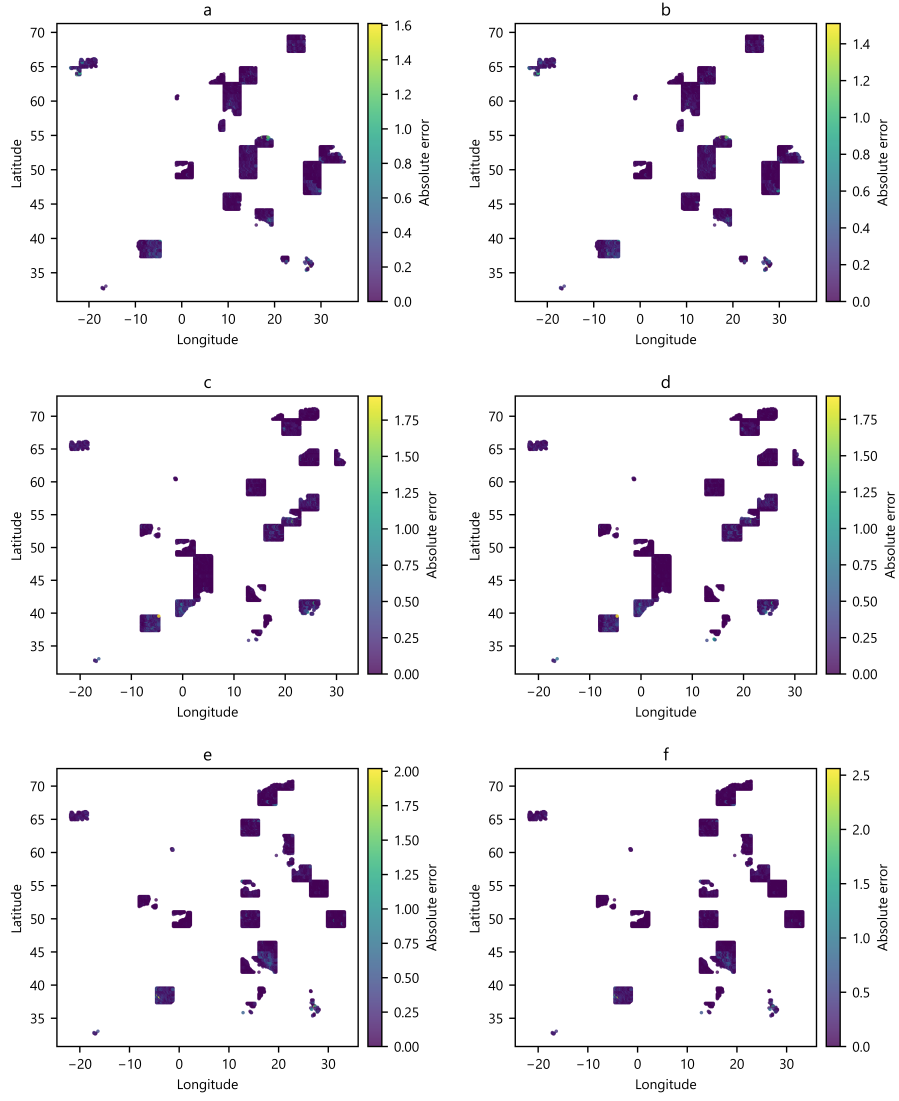


Fig. S6 Spatial distributions of absolute held-out test errors for the six selected metal-task models. Panels a–f follow the common task order stated above. Only cells assigned to the held-out spatial blocks are displayed. Color scales are panel-specific and should not be compared directly across metals or tasks

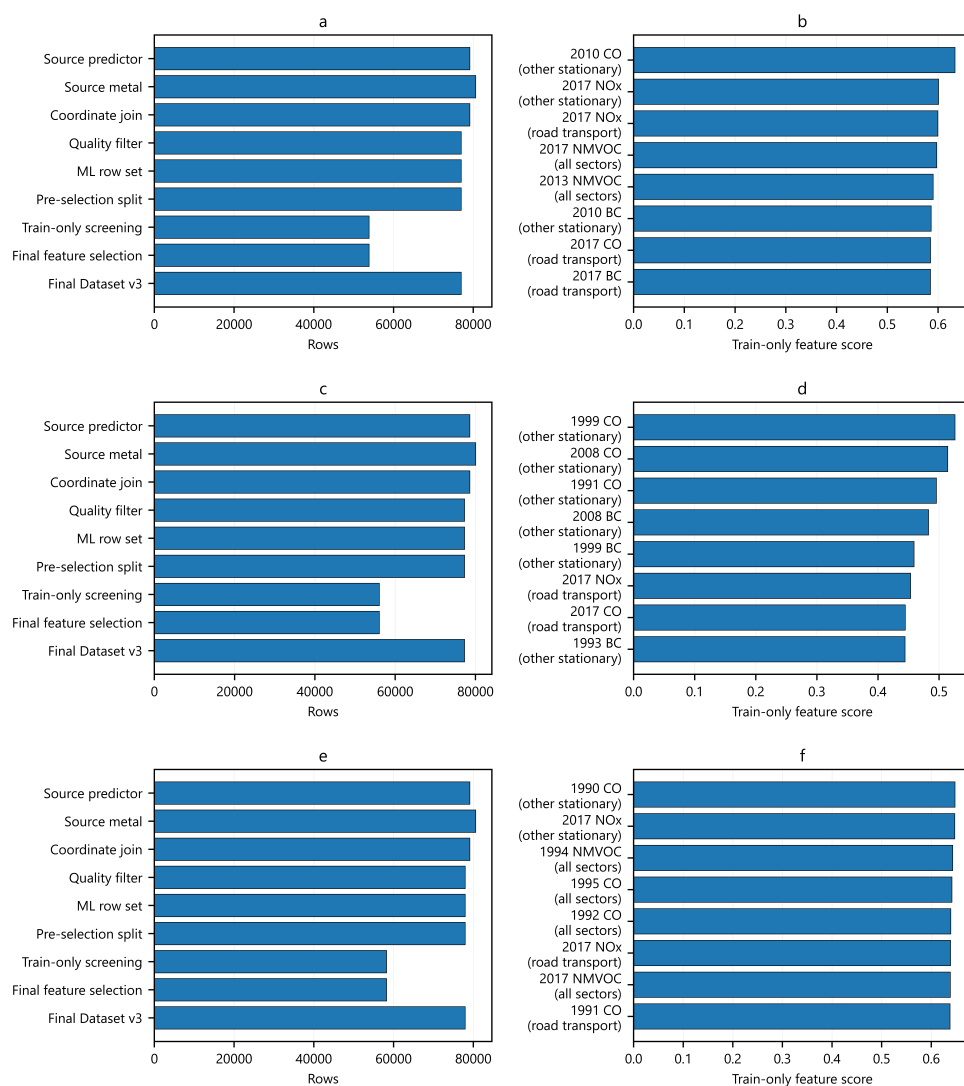


Fig. S7 Dataset v3 quality-control and train-only feature-screening diagnostics. Panels a and b show the Cd row-retention audit and eight highest-scoring selected predictors, panels c and d show the corresponding Hg diagnostics, and panels e and f show the corresponding Pb diagnostics. Feature scores were calculated using the training partition only