Supplementary Information

Turbidity-associated and warm-season manganese patterns at river and canal drinking-water intakes in Hai Phong City, Vietnam: statistical evidence and same-day alert rules from routine monitoring data

Contents: Tables S1–S13; Supplementary Document S1 (scope of accreditation of the analysing laboratory); Supplementary Note S2 (data and code). All tables are generated by the scripts in the code and data repository (see the Data availability statement of the article); each table lists the file it was built from. Table S8 and Table S11 contain the full model-skill and alert-rule tables that were Tables 3 and 5 of the first submission.

Table S1. Number of determinations and of missing values by intake, year and variable, 2021–2024

Intake	Year	Days	Mn	Fe	Turb.	Temp.	PI	NH_4^+-N	NO_2^--N	NO_3^--N	pH	EC
QV	2021	261	261 (0)	104 (157)	255 (6)	255 (6)	259 (2)	261 (0)	105 (156)	104 (157)	255 (6)	255 (6)
QV	2022	260	260 (0)	104 (156)	254 (6)	254 (6)	258 (2)	260 (0)	104 (156)	104 (156)	254 (6)	253 (7)
QV	2023	258	258 (0)	104 (154)	252 (6)	252 (6)	256 (2)	258 (0)	110 (148)	104 (154)	252 (6)	252 (6)
QV	2024	267	263 (4)	106 (161)	258 (9)	257 (10)	261 (6)	263 (4)	114 (153)	0 (267)	258 (9)	260 (7)
QV	All	1046	1042 (4)	418 (628)	1019 (27)	1018 (28)	1034 (12)	1042 (4)	433 (613)	312 (734)	1019 (27)	1020 (26)
HP3	2021	260	260 (0)	157 (103)	236 (24)	236 (24)	231 (29)	240 (20)	157 (103)	157 (103)	236 (24)	236 (24)
HP3	2022	260	260 (0)	156 (104)	160 (100)	160 (100)	158 (102)	160 (100)	156 (104)	156 (104)	160 (100)	160 (100)
HP3	2023	228	228 (0)	155 (73)	157 (71)	157 (71)	153 (75)	159 (69)	158 (70)	155 (73)	157 (71)	157 (71)
HP3	2024	265	262 (3)	158 (107)	163 (102)	163 (102)	160 (105)	163 (102)	162 (103)	0 (265)	163 (102)	165 (100)
HP3	All	1013	1010 (3)	626 (387)	716 (297)	716 (297)	702 (311)	722 (291)	633 (380)	468 (545)	716 (297)	718 (295)
CN	2021	127	75 (52)	52 (75)	127 (0)	127 (0)	125 (2)	80 (47)	55 (72)	52 (75)	127 (0)	127 (0)
CN	2022	115	105 (10)	78 (37)	114 (1)	115 (0)	114 (1)	105 (10)	79 (36)	78 (37)	113 (2)	111 (4)
CN	2023	108	107 (1)	63 (45)	105 (3)	107 (1)	106 (2)	107 (1)	106 (2)	63 (45)	105 (3)	105 (3)
CN	2024	111	109 (2)	53 (58)	110 (1)	110 (1)	108 (3)	109 (2)	107 (4)	0 (111)	110 (1)	111 (0)
CN	All	461	396 (65)	246 (215)	456 (5)	459 (2)	453 (8)	401 (60)	347 (114)	193 (268)	455 (6)	454 (7)
HD	2021	124	122 (2)	104 (20)	124 (0)	124 (0)	121 (3)	123 (1)	104 (20)	104 (20)	124 (0)	124 (0)
HD	2022	111	110 (1)	78 (33)	92 (19)	111 (0)	109 (2)	110 (1)	103 (8)	78 (33)	91 (20)	91 (20)
HD	2023	67	67 (0)	53 (14)	63 (4)	67 (0)	65 (2)	67 (0)	67 (0)	53 (14)	63 (4)	63 (4)
HD	2024	71	68 (3)	52 (19)	69 (2)	69 (2)	67 (4)	68 (3)	67 (4)	0 (71)	69 (2)	71 (0)
HD	All	373	367 (6)	287 (86)	348 (25)	371 (2)	362 (11)	368 (5)	341 (32)	235 (138)	347 (26)	349 (24)
VB1	2021	121	114 (7)	52 (69)	113 (8)	12 (109)	112 (9)	114 (7)	97 (24)	54 (67)	113 (8)	113 (8)
VB1	2022	58	58 (0)	51 (7)	58 (0)	11 (47)	57 (1)	58 (0)	56 (2)	52 (6)	58 (0)	58 (0)
VB1	2023	53	53 (0)	51 (2)	53 (0)	12 (41)	52 (1)	53 (0)	53 (0)	51 (2)	53 (0)	53 (0)
VB1	2024	58	56 (2)	51 (7)	56 (2)	14 (44)	56 (2)	56 (2)	56 (2)	6 (52)	56 (2)	58 (0)
VB1	All	290	281 (9)	205 (85)	280 (10)	49 (241)	277 (13)	281 (9)	262 (28)	163 (127)	280 (10)	282 (8)
VB2	2021	52	28 (24)	28 (24)	52 (0)	12 (40)	52 (0)	28 (24)	28 (24)	26 (26)	52 (0)	52 (0)
VB2	2022	53	35 (18)	34 (19)	53 (0)	11 (42)	52 (1)	35 (18)	35 (18)	33 (20)	53 (0)	53 (0)
VB2	2023	53	50 (3)	46 (7)	53 (0)	12 (41)	52 (1)	50 (3)	49 (4)	44 (9)	53 (0)	53 (0)
VB2	2024	58	57 (1)	52 (6)	57 (1)	13 (45)	57 (1)	57 (1)	57 (1)	0 (58)	57 (1)	58 (0)
VB2	All	216	170 (46)	160 (56)	215 (1)	48 (168)	213 (3)	170 (46)	169 (47)	103 (113)	215 (1)	216 (0)
NL	2021	146	35 (111)	15 (131)	146 (0)	14 (132)	143 (3)	35 (111)	29 (117)	28 (118)	146 (0)	146 (0)
NL	2022	66	45 (21)	15 (51)	66 (0)	30 (36)	64 (2)	43 (23)	39 (27)	24 (42)	66 (0)	66 (0)
NL	2023	51	51 (0)	22 (29)	51 (0)	50 (1)	50 (1)	51 (0)	51 (0)	31 (20)	51 (0)	51 (0)
NL	2024	55	54 (1)	28 (27)	55 (0)	55 (0)	53 (2)	54 (1)	54 (1)	0 (55)	55 (0)	55 (0)
NL	All	318	185 (133)	80 (238)	318 (0)	149 (169)	310 (8)	183 (135)	173 (145)	83 (235)	318 (0)	318 (0)
All	2021	1091	895 (196)	512 (579)	1053 (38)	780 (311)	1043 (48)	881 (210)	575 (516)	525 (566)	1053 (38)	1053 (38)
All	2022	923	873 (50)	516 (407)	797 (126)	692 (231)	812 (111)	771 (152)	572 (351)	525 (398)	795 (128)	792 (131)
All	2023	818	814 (4)	494 (324)	734 (84)	657 (161)	734 (84)	745 (73)	594 (224)	501 (317)	734 (84)	734 (84)
All	2024	885	869 (16)	500 (385)	768 (117)	681 (204)	762 (123)	770 (115)	617 (268)	6 (879)	768 (117)	778 (107)
All	All	3717	3451 (266)	2022 (1695)	3352 (365)	2810 (907)	3351 (366)	3167 (550)	2358 (1359)	1557 (2160)	3350 (367)	3357 (360)

Note: Days, intake-days with at least one record. Each cell gives the number of determinations and, in parentheses, the number of intake-days on which the variable was not determined (missing values). Temp., water temperature measured on site (direct readings only, before construction of the regional series). Nitrate was not reported in 2024. Rows "All" give totals over years or over intakes. Source file: t07_counts_missing.csv.

Table S2. Number of total Mn determinations by intake, year and month

Intake	Year	1	2	3	4	5	6	7	8	9	10	11	12	Total
QV	2021	21	20	23	22	21	22	22	22	22	21	22	23	261
QV	2022	21	20	23	21	22	22	21	23	22	21	22	22	260
QV	2023	22	20	23	20	23	22	21	21	21	22	22	21	258
QV	2024	23	21	21	22	24	20	23	22	21	23	21	22	263
HP3	2021	21	20	23	21	21	22	22	22	22	21	22	23	260
HP3	2022	21	20	23	21	22	22	21	23	22	21	22	22	260
HP3	2023	22	12	23	20	23	14	13	15	21	22	22	21	228
HP3	2024	23	21	21	22	23	20	23	22	21	23	21	22	262
CN	2021	4	6	6	4	4	8	5	11	7	7	5	8	75
CN	2022	5	6	10	8	9	12	10	10	8	10	8	9	105
CN	2023	9	8	9	8	8	10	11	8	8	9	8	11	107
CN	2024	10	7	12	8	12	8	11	9	9	7	7	9	109
HD	2021	11	11	10	11	9	10	10	10	9	9	11	11	122
HD	2022	10	7	12	10	9	11	8	9	8	8	9	9	110
HD	2023	6	8	8	4	5	5	7	4	5	5	5	5	67
HD	2024	5	6	5	5	5	5	8	8	5	7	5	4	68
VB1	2021	9	7	16	9	9	7	12	10	9	8	9	9	114
VB1	2022	6	6	5	4	4	5	6	5	4	4	4	5	58
VB1	2023	5	4	5	5	5	4	4	4	4	4	5	4	53
VB1	2024	4	5	4	4	5	4	7	5	5	5	4	4	56
VB2	2021	2	1	3	4	1	2	5	2	1	2	2	3	28
VB2	2022	1	3	3	2	2	3	2	4	3	4	3	5	35
VB2	2023	5	4	4	5	3	4	4	4	4	4	5	4	50
VB2	2024	4	5	4	4	5	4	7	5	4	5	5	5	57
NL	2021	2	2	3	3	4	7	2	3	2	2	3	2	35
NL	2022	2	3	4	2	4	4	4	5	4	4	5	4	45
NL	2023	5	4	4	4	4	4	4	5	4	5	4	4	51
NL	2024	5	4	4	4	5	4	5	4	4	6	4	5	54

Note: Columns 1–12, calendar months. Source file: t06_counts_intake_year_month.csv.

Table S3. Direct water-temperature readings at each intake and error of the regional series rebuilt without that intake (leave-one-intake-out), by season

Intake	Direct readings	Mn samples with reading, %	Season	n	Bias, °C	RMSE, °C	5th pct., °C	95th pct., °C
QV	1018	97.5	all	1018	−0.04	0.21	−0.30	0.16
			rainy (May-Nov)	596	−0.04	0.19	−0.30	0.20
			dry (Dec-Apr)	422	−0.05	0.25	−0.30	0.15
HP3	716	70.9	all	716	0.01	0.15	−0.20	0.20
			rainy (May-Nov)	426	0.01	0.14	−0.20	0.20
			dry (Dec-Apr)	290	0.01	0.15	−0.15	0.20
CN	459	100.0	all	459	0.05	0.21	−0.10	0.30
			rainy (May-Nov)	268	0.04	0.17	−0.13	0.25
			dry (Dec-Apr)	191	0.07	0.26	−0.10	0.35
HD	371	100.0	all	371	0.05	0.17	−0.12	0.30
			rainy (May-Nov)	211	0.05	0.18	−0.15	0.30
			dry (Dec-Apr)	160	0.05	0.17	−0.10	0.25
VB1	49	17.4	all	49	−0.04	0.34	−0.90	0.37
			rainy (May-Nov)	30	−0.07	0.41	−0.95	0.37
			dry (Dec-Apr)	19	0.01	0.17	−0.21	0.27
VB2	48	21.8	all	48	−0.04	0.31	−0.69	0.47
			rainy (May-Nov)	29	−0.03	0.37	−0.80	0.67
			dry (Dec-Apr)	19	−0.05	0.17	−0.29	0.17
NL	149	72.4	all	149	0.01	0.24	−0.20	0.30
			rainy (May-Nov)	95	0.03	0.19	−0.20	0.30
			dry (Dec-Apr)	54	−0.02	0.31	−0.34	0.23

Note: Error = reading at the intake minus the regional series built from the other intakes (mean of same-day readings, gaps of up to 7 days carried forward). Rainy season May–November; dry season December–April. Source file: t06_temperature_by_intake.csv.

Table S4. Automated screening of the monitoring records (2021–2024, seven intakes)

Check on the records as received	Result
Rows (intake-days) in 2021-2024	3717
Duplicate intake-date rows	0
Mn: non-numeric entries	0
Mn: values <= 0	0
Mn: minimum	0.029
Mn: maximum	0.898
Mn: reported decimal places (mode)	3
Mn: robust |z| > 4 within intake (log scale), flagged not removed	1
Fe: non-numeric entries	0
Fe: values <= 0	0
Fe: minimum	0.1
Fe: maximum	2.92
Fe: reported decimal places (mode)	2
Fe: robust |z| > 4 within intake (log scale), flagged not removed	2
Turb: non-numeric entries	0
Turb: values <= 0	0
Turb: minimum	2.87
Turb: maximum	251
Turb: reported decimal places (mode)	1
Turb: robust |z| > 4 within intake (log scale), flagged not removed	2
T: non-numeric entries	0
T: values <= 0	0
T: minimum	12
T: maximum	32.2
T: reported decimal places (mode)	0
pH: non-numeric entries	0
pH: values <= 0	0
pH: minimum	6.86
pH: maximum	9.13
pH: reported decimal places (mode)	2
Mn/Fe mass ratio > 1 (Mn exceeding Fe)	1
pH outside 5-10	0

Note: Flagged values were inspected and retained, because none was a non-numeric entry, a duplicate or a physically implausible value. Robust z = ($\log_{10}$ x − intake median)/(1.4826 × MAD). Source file: t06_data_screening.csv.

Table S5. Sensitivity of the lithogenic Mn bound to the reference Mn/Fe ratio

Reference composition	Mn/Fe (mass)	Above bound, dry, %	Median excess, dry, %	Above bound, all, %	Median excess, all, %
Taylor (1964), continental crust	0.0169	100.0	91.8	100.0	90.2
Wedepohl (1995), upper crust	0.0171	100.0	91.7	100.0	90.0
Rudnick & Gao (2003), upper crust	0.0198	100.0	90.4	100.0	88.5
Roberts et al. (2022), Red River SPM 2018	0.0216	100.0	89.4	100.0	87.4
Roberts et al. (2022), Red River SPM 2019	0.0258	100.0	87.4	100.0	84.9
Twice Taylor (1964)	0.0337	100.0	83.5	100.0	80.3
Three times Taylor (1964)	0.0506	100.0	75.3	100.0	70.5

Note: Bound = reference Mn/Fe × total Fe of the sample; excess share = 1 − bound/total Mn. Dry season: 836 samples with Mn and Fe (December–April); all: 2,022 samples. Taylor (1964): Mn 950 mg kg^{-1}, Fe 5.63%; Wedepohl (1995) upper crust: Mn 527 mg kg^{-1}, Fe 30,890 mg kg^{-1}; Rudnick and Gao (2003) upper crust: MnO 0.10%, FeOT 5.04%; Roberts et al. (2022): ratio of annual Mn and Fe fluxes on suspended particulate matter of the Red River (near-total HNO_3–HF digestion). Source file: t06_crustal_ratio.csv.

Table S6. Sensitivity of the Mn breakpoint on the 30-day water-temperature index to sampling imbalance and to the control variables

Data set	n	ψ, °C	95% CI, °C	SSE within 1%, °C	Slope below, % °C^{-1}	Slope above, % °C^{-1}
All intakes, unweighted (main analysis)	3,034	25.4	23.2–26.4	23.5–26.6	0.9	14.2
All intakes, each intake weighted equally	3,034	26.0	22.8–27.5	23.7–27.4	1.1	12.9
Without the canal intakes (VB1, VB2)	2,596	25.4	23.2–26.4	23.7–26.4	1.0	15.3
Without QV	2,038	25.3		23.4–26.9	0.1	12.9
Without HP3	2,338	25.5		23.4–26.8	1.6	12.9
Without CN	2,646	25.5		23.8–26.7	1.0	14.9
Without HD	2,701	25.2		23.4–26.5	0.6	14.1
Without VB1	2,764	25.5		23.7–26.6	1.0	15.0
Without VB2	2,866	25.2		23.6–26.5	0.7	14.2
Without NL	2,851	25.2		23.4–26.4	0.5	14.4
QV and HP3 thinned to <= 5 samples per month (200 random draws): median		25.5	25.2–26.0			
Mn; controls: intake, turbidity (as in Table 2)	3,034	25.4	23.2–26.4		0.9	14.2
Mn; controls: intake, turbidity, PI, NH_4^+-N, pH	2,979	25.2	22.1–28.8		1.9	6.1
Fe; controls: intake, turbidity (as in Table 2)	1,977	22.4	21.8–27.7		−4.6	4.5
Fe; controls: intake, turbidity, PI, NH_4^+-N, pH	1,942	21.9	19.4–30.5		−2.6	2.0

Note: Segmented regression of $\log_{10}$ Mn on T_{30} with intake fixed effects and $\log_{10}$ turbidity (Eq. 3). Equal weighting: each observation weighted by 1/(number of samples at its intake), so that every intake contributes equally. Thinning: QV and HP3 reduced at random to at most five samples per calendar month (200 draws; median ψ and 2.5th–97.5th percentiles of the draws). CIs from 1,000 bootstrap resamples of year-months; SSE within 1%: range of ψ whose residual sum of squares lies within 1% of the minimum. The last four rows give the segmented regressions of $\log_{10}$ Mn and $\log_{10}$ Fe with the two sets of control variables (ΔAIC versus linear/quadratic fits: Mn 276/32 and 36/5; Fe 65/5 and 9/7). Source files: t06_breakpoint_sensitivity.csv, t02_breakpoints.csv.

Table S7. Model settings, hyperparameters, random seed and software environment

Item	Setting
Random seed (all analyses)	20260924 (NumPy default_rng; random_state of scikit-learn and XGBoost)
Bootstrap	1,000 resamples of whole year-months (48 clusters); percentile 95% CIs; same scheme for pooled validation metrics and rule scores
Window L of the temperature index	candidates 1, 7, 14, 30, 45, 60 d; chosen within each outer fold by grouped inner cross-validation of the linear model (maximum NSE)
Linear regression	ordinary least squares on T_L, $\log_{10}$ turbidity, PI, $\log_{10}$ NH_4^+-N, pH (scikit-learn)
Random forest	500 trees, min_samples_leaf = 5, max_features = 0.6 (not tuned)
XGBoost, fixed	learning_rate 0.05, min_child_weight 10, subsample 0.8, colsample_bytree 0.8, reg_lambda 1.0
XGBoost, tuned in each outer fold	max_depth ∈ {2, 3, 4}; n_estimators ∈ {300, 600}; grouped inner CV
Back-transformation	Duan smearing factor from out-of-fold residuals of the calibration set
SHAP	TreeSHAP, interventional, 500 background samples; final model L = 45 d, depth 2, 300 trees
GAM	pyGAM LinearGAM, 12 basis functions per spline, common λ chosen by GCV from 10^{-2}–10^{3}
Segmented regression	grid search of ψ from 17.5 to 30.5 °C in 0.1 °C steps
Trigger model	intake-specific logistic regression logit $P(Mn > 0.1) = a_i + b_i T_{30}$; $\theta_p = (\text{logit } p - a_i)/b_i$
Software	Python 3.11.15; pandas 3.0.2, NumPy 2.4.4, SciPy 1.16.3, statsmodels 0.15.0, scikit-learn 1.8.0, XGBoost 3.2.0, shap 0.51.0, pyGAM 0.12.0, Matplotlib 3.10.9, openpyxl 3.1.5

Note: All settings are defined at the top of the corresponding scripts (02_statistical_analysis.py, 03_ML_models_SHAP.py, 04_operational_triggers.py); the selected L and XGBoost hyperparameters of every outer fold are listed in Table S9.

Table S8. Calibration and validation skill of the five Mn estimation models

Scheme	Model	R^2	NSE	KGE	RMSE, mg L^{-1}	PBIAS, %
Intakes withheld	Climatology	0.40 / 0.34	0.39 / 0.33	0.39 / 0.34	0.070 / 0.074	+1.3 / +3.7
	Season + turbidity	0.61 / 0.57	0.61 / 0.57	0.58 / 0.55	0.056 / 0.059	−0.6 / +0.2
	Linear regression	0.77 / 0.75	0.77 / 0.75	0.81 / 0.79	0.050 / 0.054	+0.5 / +1.2
	Random forest	0.90 / 0.74	0.90 / 0.74	0.83 / 0.74	0.032 / 0.049	−1.9 / −0.2
	XGBoost	0.84 / 0.75	0.84 / 0.75	0.85 / 0.78	0.037 / 0.048	−0.4 / +0.3
Sources withheld	Climatology	0.39 / 0.27	0.38 / 0.24	0.36 / 0.27	0.069 / 0.075	+1.0 / −0.7
	Season + turbidity	0.60 / 0.51	0.60 / 0.51	0.57 / 0.54	0.056 / 0.059	−0.6 / +0.5
	Linear regression	0.77 / 0.73	0.77 / 0.73	0.81 / 0.77	0.049 / 0.055	+0.5 / −0.9
	Random forest	0.90 / 0.71	0.89 / 0.71	0.82 / 0.76	0.032 / 0.048	−1.9 / +3.1
	XGBoost	0.83 / 0.72	0.83 / 0.72	0.84 / 0.79	0.037 / 0.047	−0.5 / +1.8
Validated on 2023	Climatology	0.50 / 0.30	0.49 / 0.24	0.47 / 0.29	0.060 / 0.080	+0.6 / −5.9
	Season + turbidity	0.70 / 0.69	0.70 / 0.68	0.65 / 0.68	0.049 / 0.054	−0.7 / +1.8
	Linear regression	0.79 / 0.75	0.79 / 0.74	0.80 / 0.77	0.048 / 0.051	+0.5 / −4.9
	Random forest	0.91 / 0.68	0.90 / 0.67	0.81 / 0.61	0.031 / 0.058	−1.9 / −7.6
	XGBoost	0.86 / 0.68	0.86 / 0.67	0.82 / 0.67	0.035 / 0.057	−0.8 / −6.3
Validated on 2024	Climatology	0.42 / 0.35	0.41 / 0.28	0.38 / 0.33	0.067 / 0.077	+0.9 / −9.6
	Season + turbidity	0.70 / 0.62	0.70 / 0.60	0.67 / 0.61	0.049 / 0.062	−0.8 / −8.4
	Linear regression	0.78 / 0.79	0.78 / 0.79	0.81 / 0.77	0.048 / 0.060	+0.4 / +0.8
	Random forest	0.90 / 0.81	0.90 / 0.80	0.81 / 0.80	0.031 / 0.044	−2.0 / 0.0
	XGBoost	0.87 / 0.81	0.87 / 0.81	0.85 / 0.84	0.033 / 0.043	−0.9 / +0.3

Note: Each cell: calibration (mean over folds) / validation (withheld data, pooled). Calibration years: 2021–2022 for validation on 2023, 2021–2023 for validation on 2024. R^2 and NSE on the $\log_{10}$ scale; KGE, RMSE and PBIAS on the concentration scale. Random-forest calibration values are in-sample; its out-of-bag NSE (mean over folds) was 0.80–0.82.

Table S9. Validation skill of each model in each of the 13 outer folds

Withheld	Model	n	NSE log	R^2 log	KGE conc.	RMSE, mg L^{-1}	PBIAS, %	L, d	XGB depth/trees
Intake QV	Climatology	968	0.60	0.60	0.60	0.050	17.9	60	2/300

	Season + turbidity		0.64	0.66	0.57	0.043	−3.9		
	Linear		0.77	0.79	0.78	0.035	7.8		
	Random forest		0.77	0.79	0.81	0.038	12.9		
	XGBoost		0.79	0.80	0.81	0.037	12.0		
Intake HP3	Climatology	664	0.26	0.49	0.26	0.090	−24.9	45	2/300
	Season + turbidity		0.70	0.79	0.52	0.063	−9.5		
	Linear		0.81	0.84	0.71	0.049	−5.7		
	Random forest		0.74	0.78	0.60	0.063	−11.8		
	XGBoost		0.78	0.81	0.66	0.059	−9.4		
Intake CN	Climatology	378	0.09	0.41	0.38	0.064	39.4	45	2/300
	Season + turbidity		0.13	0.56	0.50	0.055	38.4		
	Linear		0.70	0.71	0.62	0.029	−0.8		
	Random forest		0.68	0.68	0.64	0.029	1.6		
	XGBoost		0.69	0.70	0.66	0.028	1.7		
Intake HD	Climatology	315	0.05	0.57	0.31	0.061	58.2	45	2/300
	Season + turbidity		0.15	0.63	0.44	0.051	46.4		
	Linear		0.68	0.71	0.72	0.026	10.2		
	Random forest		0.71	0.75	0.71	0.025	9.8		
	XGBoost		0.69	0.72	0.72	0.026	9.4		
Intake VB1	Climatology	252	−0.27	0.33	0.23	0.101	−27.1	60	3/300
	Season + turbidity		0.24	0.40	0.48	0.085	−11.7		
	Linear		0.59	0.71	0.65	0.083	−1.4		
	Random forest		0.59	0.71	0.75	0.059	−10.2		
	XGBoost		0.61	0.71	0.79	0.059	−7.5		
Intake VB2	Climatology	165	−0.64	0.16	0.12	0.113	−28.1	60	2/300
	Season + turbidity		−0.18	0.25	0.33	0.100	−20.9		
	Linear		0.24	0.43	0.33	0.138	11.9		
	Random forest		0.36	0.45	0.58	0.090	−0.2		
	XGBoost		0.38	0.46	0.61	0.092	3.1		
Intake NL	Climatology	174	−1.18	0.21	−0.21	0.072	71.0	45	2/600
	Season + turbidity		0.17	0.36	0.51	0.027	−0.5		
	Linear		−0.27	0.48	0.41	0.030	−25.0		
	Random forest		0.10	0.31	0.40	0.027	−10.9		
	XGBoost		−0.36	0.28	0.24	0.032	−24.0		
Source Re	Climatology	1632	0.35	0.52	0.25	0.071	−14.0	45	2/300
	Season + turbidity		0.63	0.72	0.50	0.055	−8.1		
	Linear		0.78	0.83	0.69	0.042	−0.8		
	Random forest		0.75	0.76	0.74	0.049	7.3		
	XGBoost		0.77	0.79	0.79	0.046	6.0		
Source DaDo	Climatology	693	−0.02	0.47	0.33	0.065	50.3	30	2/300
	Season + turbidity		0.03	0.60	0.49	0.053	40.4		
	Linear		0.69	0.71	0.63	0.028	−1.3		
	Random forest		0.66	0.70	0.61	0.029	1.8		
	XGBoost		0.66	0.69	0.64	0.028	3.5		
Source ChanhDuong	Climatology	417	−0.48	0.25	0.24	0.101	−19.9	60	2/300
	Season + turbidity		0.03	0.33	0.47	0.089	−8.4		

	Linear		0.43	0.59	0.46	0.113	4.1		
	Random forest		0.43	0.58	0.71	0.072	−5.7		
	XGBoost		0.45	0.59	0.72	0.074	−7.7		
Source Gia	Climatology	174	−1.18	0.21	−0.29	0.077	77.2	30	2/300
	Season + turbidity		0.17	0.36	0.52	0.027	1.5		
	Linear		−0.33	0.47	0.38	0.031	−26.7		
	Random forest		0.06	0.29	0.38	0.028	−13.5		
	XGBoost		−0.24	0.29	0.28	0.031	−23.1		
Year 2023	Climatology	719	0.24	0.30	0.29	0.080	−5.9	60	2/300
	Season + turbidity		0.68	0.69	0.68	0.054	1.8		
	Linear		0.74	0.75	0.77	0.051	−4.9		
	Random forest		0.67	0.68	0.61	0.058	−7.6		
	XGBoost		0.67	0.68	0.67	0.057	−6.3		
Year 2024	Climatology	754	0.28	0.35	0.33	0.077	−9.6	14	3/300
	Season + turbidity		0.60	0.62	0.61	0.062	−8.4		
	Linear		0.79	0.79	0.77	0.060	0.8		
	Random forest		0.80	0.81	0.80	0.044	0.0		
	XGBoost		0.81	0.81	0.84	0.043	0.3		

Note: Seven intake folds (leave-one-intake-out), four source folds (leave-one-source-out) and two forward year folds (calibrated on earlier years). L and XGBoost hyperparameters were selected within the calibration data of each fold. Fold-level predictions (every withheld sample, observed and predicted) are in t03_predictions.csv. Source file: t06_fold_metrics.csv.

Table S10. Check for information leakage through the regional temperature series (leave-one-intake-out)

Withheld intake	Model	L, d	NSE, original series	NSE, series without the intake	Difference
QV	Linear	60	0.775	0.775	0.000
QV	XGBoost	60	0.785	0.785	−0.001
HP3	Linear	45	0.812	0.812	−0.000
HP3	XGBoost	45	0.779	0.783	0.004
CN	Linear	45	0.700	0.700	−0.000
CN	XGBoost	45	0.691	0.693	0.001
HD	Linear	45	0.675	0.675	−0.000
HD	XGBoost	45	0.693	0.694	0.001
VB1	Linear	60	0.592	0.592	−0.000
VB1	XGBoost	60	0.606	0.607	0.001
VB2	Linear	60	0.237	0.236	−0.000
VB2	XGBoost	60	0.382	0.382	−0.000
NL	Linear	45	−0.268	−0.268	−0.000
NL	XGBoost	45	−0.356	−0.335	0.021

Note: The regional series and T_L were rebuilt from the other six intakes only and the models refitted with the same L and hyperparameters. Source file: t06_leakage_T_series.csv.

Table S11. Out-of-sample comparison of six intake-specific daily alert rules (leave-one-year-out, 2021–2024)

Rule	POD	FAR	CSI [95% CI]	ΔCSI vs calendar	ΔCSI vs persistence	Intake-mean ΔCSI vs calendar
$p*$ = 0.5 (n = 3,034; exceedances = 1,445)						
Fixed calendar months	0.76	0.22	0.62 [0.52, 0.70]	–	−0.056 [−0.113, −0.011]	–
Water-temperature index T_{30}	0.8	0.23	0.64 [0.56, 0.71]	+0.020 [−0.009, 0.053]	−0.036 [−0.075, −0.002]	+0.003 [−0.034, 0.037]
Same-day turbidity only	0.72	0.26	0.58 [0.53, 0.61]	−0.048 [−0.114, 0.029]	−0.104 [−0.145, −0.060]	−0.047 [−0.100, 0.012]
Calendar harmonic + turbidity	0.84	0.17	0.72 [0.65, 0.77]	+0.092 [0.050, 0.141]	+0.036 [0.008, 0.068]	+0.075 [0.046, 0.113]
T_{30} + turbidity	0.83	0.18	0.71 [0.64, 0.75]	+0.081 [0.039, 0.133]	+0.025 [−0.003, 0.056]	+0.063 [0.029, 0.108]
Persistence (previous sample)	0.81	0.19	0.68 [0.61, 0.73]	+0.056 [0.011, 0.113]	–	+0.063 [0.020, 0.106]
$p*$ = 0.2 (n = 3,034; exceedances = 1,445)						
Fixed calendar months	0.92	0.37	0.60 [0.52, 0.67]	–	−0.080 [−0.119, −0.045]	–
Water-temperature index T_{30}	0.92	0.35	0.61 [0.53, 0.68]	+0.014 [−0.002, 0.028]	−0.066 [−0.106, −0.030]	+0.022 [0.002, 0.041]
Same-day turbidity only	0.94	0.41	0.57 [0.49, 0.64]	−0.029 [−0.058, 0.001]	−0.108 [−0.152, −0.067]	−0.022 [−0.049, 0.001]
Calendar harmonic + turbidity	0.95	0.28	0.69 [0.63, 0.74]	+0.091 [0.060, 0.130]	+0.011 [−0.018, 0.041]	+0.063 [0.038, 0.091]
T_{30} + turbidity	0.94	0.29	0.68 [0.61, 0.73]	+0.080 [0.050, 0.116]	+0.001 [−0.030, 0.033]	+0.058 [0.031, 0.087]
Persistence (previous sample)	0.81	0.19	0.68 [0.61, 0.73]	+0.080 [0.045, 0.119]	–	+0.044 [0.006, 0.082]

Note: Intake-specific rules fitted on three years and tested on the fourth, rotating over 2021–2024. Persistence: flag if the previous sample exceeded 0.1 mg L^{-1}, independent of $p*$. POD, probability of detection; FAR, false-alarm ratio; CSI, critical success index. 95% CIs from 1,000 bootstrap resamples of year-months.

Table S12. Sensitivity of the main estimates to random analytical error in Mn and Fe and to values near the limit of quantification

Data set	β T_{30}, log Mn	β turb., log Mn	β T_{30}, log Fe	β turb., log Fe	β T_{30}, log(Mn/Fe)	β turb., log(Mn/Fe)	ψ, °C	Slope above ψ, % °C^{-1}	Mn > 0.1, %
Original data	0.056	0.113	0.020	0.109	0.036	0.004	25.4	14.2	47.6
Random error CV 10 % (median of 200)	0.056	0.113	0.020	0.109	0.036	0.004	25.4	14.1	47.6
Random error CV 10 % (p2.5 of 200)	0.053	0.110	0.017	0.106	0.033	0.000	25.2	13.7	47.0
Random error CV 10 % (p97.5 of 200)	0.058	0.115	0.022	0.111	0.039	0.007	25.5	14.6	48.2
Random error CV 20 % (median of 200)	0.056	0.112	0.020	0.109	0.036	0.004	25.4	14.1	47.2
Random error CV 20 % (p2.5 of 200)	0.051	0.108	0.015	0.105	0.029	−0.002	25.1	13.5	46.4
Random error CV 20 % (p97.5 of 200)	0.060	0.118	0.025	0.113	0.043	0.011	25.7	15.1	47.9
Mn < 0.04 and Fe < 0.10 mg/L excluded	0.052	0.108	0.019	0.107	0.033	0.002	25.2	13.8	48.7

Note: β, standardized coefficients of the station fixed-effects regressions (Eq. 2, same samples as Table 2); ψ and slope above ψ from the segmented regression with intake and turbidity controls (Eq. 3). Random error: every Mn and Fe value was multiplied by an independent log-normal factor with a coefficient of variation (CV) of 10% or 20% (mean-preserving), 200 replicates; the median and the 2.5th and 97.5th percentiles of the replicates are given. Last row: values below twice the limit of quantification (Mn < 0.04 mg L^{-1}, 81 values; Fe < 0.10 mg L^{-1}, none) excluded. Source file: t07_analytical_error.csv.

Table S13. Audit of the separation between calibration and validation data in the 13 outer folds

Scheme	Withheld	Calibration rows	Validation rows	Intake-days in both	Intakes in both	Last calibration date	First validation date
Intake	QV	2409	1042	0	0	2024-12-31	2021-01-01
Intake	HP3	2441	1010	0	0	2024-12-31	2021-01-01
Intake	CN	3055	396	0	0	2024-12-31	2021-01-04
Intake	HD	3084	367	0	0	2024-12-31	2021-01-05
Intake	VB1	3170	281	0	0	2024-12-31	2021-01-04
Intake	VB2	3281	170	0	0	2024-12-31	2021-01-14
Intake	NL	3266	185	0	0	2024-12-31	2021-01-11
Source	Re	1399	2052	0	0	2024-12-31	2021-01-01
Source	DaDo	2688	763	0	0	2024-12-31	2021-01-04
Source	ChanhDuong	3000	451	0	0	2024-12-31	2021-01-04
Source	Gia	3266	185	0	0	2024-12-31	2021-01-11
Year	2023	1768	814	0	7	2022-12-30	2023-01-02
Year	2024	2582	869	0	7	2023-12-29	2024-01-01

Note: Rows are intake-days with a Mn determination. In the intake and source folds, no intake-day and no intake appears in both sets; the sets share calendar dates because the intakes were sampled on the same days, and the regional temperature index shared on those dates was tested separately (Table S10). In the year folds, every calibration date precedes every validation date. The window L, the XGBoost hyperparameters and the Duan smearing factor were selected or estimated within the calibration set of each fold only (Table S9). Source file: t07_leakage_audit.csv.

Supplementary Document S1. Scope of accreditation of the analysing laboratory (ISO/IEC 17025:2017, VILAS 449, Bureau of Accreditation of Vietnam; Decision 290/QĐ-VPCNCLQG of 31 January 2026, valid to 30 January 2031)

The four pages of the accreditation scope are reproduced below; the methods and limits of quantification used in the article (Mn, TCVN 6002:1995, 0.02 mg L^{-1}; Fe, TCVN 6177:1996, 0.05 mg L^{-1}; turbidity, SMEWW 2130B:2023, 0.12 NTU; NH_4^+-N, NO_2^--N, NO_3^--N, PI, pH) are items 5, 10, 8, 11, 9, 6, 4 and 7 of the list. The original PDF, digitally signed by the Bureau of Accreditation, is included in the Zenodo record (Supplementary Note S2).

DANH MỤC PHÉP THỬ ĐƯỢC CÔNG NHẬN

LIST OF ACCREDITED TESTS

(Kèm theo quyết định số/attachment with decision: 290 ***/QĐ - VPCNCLQG ngày*** 31 ***tháng 01 năm 2026 của giám đốc Văn phòng Công nhận chất lượng quốc gia/of BoA Director)***

Tên phòng thí nghiệm:	**Phòng Kiểm tra Chất lượng nước**
Laboratory:	***Water Quality Control Department***
Tổ chức /Cơ quan chủ quản:	**Công ty Cổ phần Cấp nước Hải Phòng**
Organization:	***Hai phong Water Supply Company***
Số hiệu/ *Code:*	**VILAS 449**
Chuẩn mực công nhận *Accreditation criteria*	**ISO/IEC 17025:2017**
Lĩnh vực:	**Hóa, Sinh**
Field:	***Chemical, Biological***
Người quản lý:	**Ngô Quỳnh Hoa**
Laboratory manager:	***Ngo Quynh Hoa***
Hiệu lực công nhận *Period of Validation:*	***Từ ngày/ from*** 31 ***/01/2026 đến/to*** 30 ***/01/2031***
Địa chỉ: *Address*:	**Số 54 Đinh Tiên Hoàng, phường Hồng Bàng, thành phố Hải Phòng** ***54, Đinh Tien Hoang street, Hong Bang ward, Hai Phong city***
Địa điểm: *Location:*	**Số 249 Tôn Đức Thắng, Phường An Biên, thành phố Hải Phòng** ***249, Ton Duc Thang street, An Bien ward, Hai Phong city***
Điện thoại/ *Tel*:	**02253.835602**
Email:	**hpwaterlab@gmail.com**
Website:	**www.capnuochaiphong.com.vn**

DANH MỤC PHÉP THỬ ĐƯỢC CÔNG NHẬN

LIST OF ACCREDITED TESTS

VILAS 449

Lĩnh vực thử nghiệm: Hóa

Field of testing: Chemical

TT	Tên sản phẩm, vật liệu được thử/ *Materials or product tested*	Tên phép thử cụ thể/ *The name of specific tests*	Giới hạn định lượng (nếu có)/ Phạm vi đo *Limit of quantitation (if any)/range of measurement*	Phương pháp thử/ *Test method*
1.	**Nước sạch, nước mặt** ***Domestic water, surface water***	Xác định độ mầu *Determination of colour*	6 mg/LPt	TCVN 6185:2015
2.		Xác định tổng Canxi và Magiê Phương pháp chuẩn độ EDTA *Determination of the sum of Calcium and Magnesium* *EDTA titrimetric method*	6,0 mg$CaCO_3$/L	TCVN 6224:1996
3.		Xác định hàm lượng Clorua Chuẩn độ bạc nitrat với chỉ thị Cromat (phương pháp MO) *Determination of Chloride content* *Silver nitrate titration with chromate indicator (Mohr's method)*	5,0 mg Cl^-/L	TCVN 6194:1996
4.		Xác định chỉ số Pemanganat Phương pháp chuẩn độ *Determination of Permanganate index* *Titration method*	0,6 mgO_2/L	TCVN 6186:1996
5.		Xác định hàm lượng Mangan Phương pháp trắc quang dùng Fomaldoxim *Determination of Manganese content* *Formaldoxime spectrometric method*	0,02 mg/L	TCVN 6002:1995
6.		Xác định hàm lượng Nitrat Phương pháp trắc phổ dùng Axit Sunfosalixylic *Determination of Nitrate content* *Spectrometric method using sulfosalicylic acid*	0,04 mg/L	TCVN 6180:1996
7.		Xác định pH *Determination of pH*	2 ~ 12	TCVN 6492:2011
8.		Xác định Độ đục *Determination of turbidity*	0,12 NTU	SMEWW 2130 B:2023

TT	Tên sản phẩm, vật liệu được thử/ *Materials or product tested*	Tên phép thử cụ thể/ *The name of specific tests*	Giới hạn định lượng (nếu có)/ Phạm vi đo *Limit of quantitation (if any)/range of measurement*	Phương pháp thử/ *Test method*
9.	**Nước sạch, nước mặt** ***Domestic water, surface water***	Xác định hàm lượng Nitrit Phương pháp trắc phổ hấp thụ phân tử *Determination of Nitrite content* *Molecular absorption Spectrometric method*	0,01 mg/L	TCVN 6178:1996
10.		Xác định hàm lượng Sắt Phương pháp trắc phổ dùng thuốc thử 1,10 - Phenantrolin *Determination of Iron content* *Spectrometric method using 1,10 Phenanthroline*	0,05 mg/L	TCVN 6177:1996
11.		Xác định hàm lượng Amoni Phương pháp trắc phổ thao tác bằng tay. *Determination of Amonium content* *Manual spectrometric method*	0,03 mg/L	TCVN 6179-1:1996
12.	**Nước sạch** ***Domestic water***	Xác định tổng chất rắn hòa tan Phương pháp sấy khô ở 180^0C *Determination of total dissolved solids* *Dried method at 180^0C*	10,0 mg/L	SMEWW 2540C:2023
13.		Xác định hàm lượng Sunfat Phương pháp độ đục *Determination of sunfate content* *Turbidimetric method*	5,0 mg/L	SMEWW 4500. SO_4^{2-} E:2023
14.	**Nước mặt** ***Surface water***	Xác định hàm lượng cặn lơ lửng Phương pháp sấy khô ở (103~105)^{0}C *Determination of total suspended solids* *Dried method at (103 ~ 105)^{0}C*	5,0 mg/L	SMEWW 2540D:2023
15.		Xác định nhu cầu ô xi hóa học (COD) Phương pháp chuẩn độ *Determination of Chemical oxygen Demand (COD).* *Closed reflux, Titrimetric method*	(40 ~ 400) mg/L	SMEWW 5220C:2023

DANH MỤC PHÉP THỬ ĐƯỢC CÔNG NHẬN

LIST OF ACCREDITED TESTS

VILAS 449

Lĩnh vực thử nghiệm: Sinh

Field of testing: Biological

TT	Tên sản phẩm, vật liệu được thử/ *Materials or product tested*	Tên phép thử cụ thể/ *The name of specific tests*	Giới hạn định lượng (nếu có)/ Phạm vi đo *Limit of quantitation (if any)/range of measurement*	Phương pháp thử/ *Test method*
1.	**Nước sạch, nước mặt** ***Domestic water, surface water***	Định lượng coliform tổng số Phương pháp màng lọc *Enumeration of total Coliform Membrane filter method*		SMEWW 9222B:2023
2.		Định lượng Fecal coliform Phương pháp màng lọc *Enumeration of Fecal Coliform Membrane filter method*		SMEWW 9222D:2023

Chú thích/*Note*:

- TCVN: Tiêu chuẩn Quốc gia Việt Nam/*Vietnamese National Standards*
- ISO: Tổ chức Tiêu chuẩn quốc tế/*International Organization for Standardization.*

Trường hợp Công ty Cổ phần Cấp nước Hải Phòng cung cấp dịch vụ thử nghiệm chất lượng sản phẩm, hàng hoá thì Công ty Cổ phần Cấp nước Hải Phòng phải đăng ký hoạt động và được cấp giấy chứng nhận đăng ký hoạt động theo quy định của pháp luật trước khi cung cấp dịch vụ này/ *It is mandatory for Hai phong Water supply Company that provides product quality testing services must register their activities and be granted a certificate of registration according to the law before providing the service.*

Supplementary Note S2. Data and code

The complete data and code package that regenerates every number, table and figure of the article and of this Supplementary Information is openly available in Zenodo at https://doi.org/10.5281/zenodo.23333057 (concept DOI, resolving to the latest version; version 1.3 accompanies the second revision). The record contains: the original laboratory workbook of the utility (unmodified, with SHA-256 checksum), the analysis input extracted from it by 00a_parse_original_records.py, the daily rainfall records, a data dictionary, all intermediate results, the fold-level predictions of every withheld sample (t03_predictions.csv), all scripts (Python 3.11; run_all.py regenerates every output with the fixed seed 20260924), the exact package versions (requirements.txt; Table S7), Figures

2–8 and all tables, and the scope of accreditation of the laboratory (Supplementary Document S1). The package is licensed CC BY 4.0 (data) and MIT (code).