A Solar-Powered Multiparameter IoT Buoy for Dissolved-Oxygen Monitoring and Short-Term Water-Quality Forecasting in a High-Altitude Trout Raceway

Supplementary Materials

Sections are numbered in the order of the corresponding topics of the main text. All values refer to the single 89-day campaign (22 April–19 July 2025) in one raceway at 2100 m a.s.l.; no additional field campaign was conducted. Units follow the main text: DO in mg L^{-1}, conductivity in µS cm^{-1} and turbidity in NTU-equivalent units (NTU-eq.) on the scale of the calibration references.

S1. Installation and Buoy Geometry

The raceway (Section 2.1; Figure S1) is 20 m long and 3 m wide, with a structural height of 1.20 m and an operating depth of about 1.00 m; three 6 in gravity inlets discharge from about 0.60 m above the water surface. The buoy is moored near the center of the raceway by four monofilament branches to a 15 kg bottom deadweight. The probe position, about 0.30 m below the float-center datum C, is measured from that datum to the cartridge reference; the offset between the water surface and datum C was not measured, so the record describes one vertical position whose depth below the surface is unknown.

The CAD envelope is 587 × 326 × 420.5 mm (length × width × height) and the electronics enclosure 230 × 180 × 88 mm. In Figure 3a of the main text the exploded distances are illustrative and the monofilament in Figure 3b is emphasized for visibility. Photograph (a) of Figure 2 shows the installation arrangement; photographs (b) and (c) were taken on 6 November 2024, before the 2025 campaign.

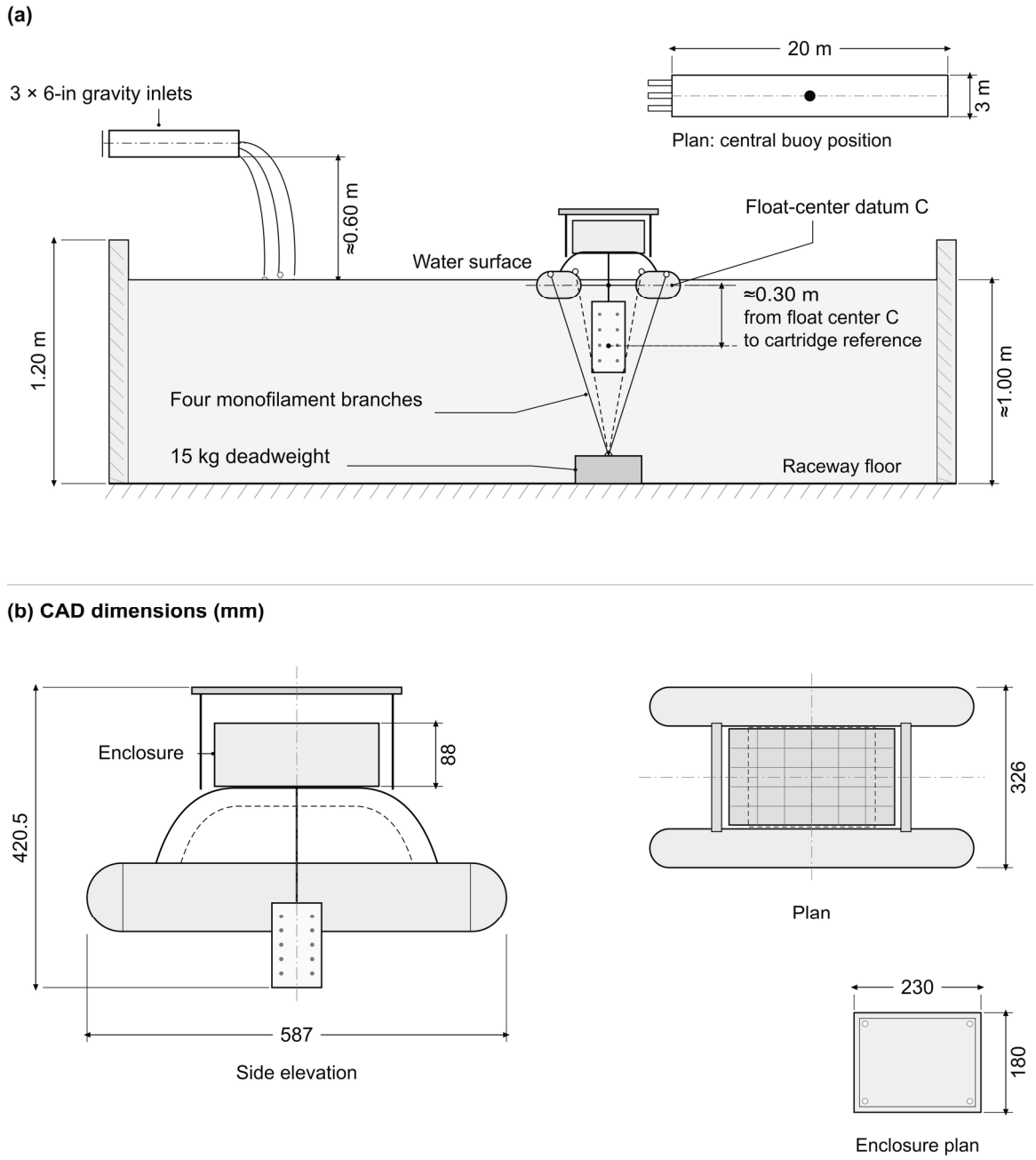


Figure S1. Dimensioned installation and buoy geometry. (a) Longitudinal section of the raceway (20 m × 3 m; 1.20 m structural height; ≈1.00 m operating depth; horizontal distances not to scale) with the three 6 in gravity inlets (≈0.60 m drop), the passive mooring of four monofilament branches to a 15 kg deadweight, and the probe cartridge ≈0.30 m below the float-center datum C; the plan inset shows the buoy position near the center. (b) CAD dimensions (mm) of the platform in side elevation and plan, and of the electronics enclosure.

S2. Power Architecture, Energy Monitor and Energy Scenarios

S2.1. Analog Front End and Power-Path Details

The charger stage (Figure 4a of the main text) uses a BQ24650 controller [1] with L = 22 µH, R_{SR} = 50 mΩ and a 600 kHz switching frequency (ripple ΔI_L = 0.239 A and peak current 0.919 A; Equation (9) of the main text), input protection by a 2 A fuse, a reverse-polarity

MOSFET and a transient-voltage suppressor, and a 3S1P pack of three Panasonic NCR18650B cells (3.35 Ah typical) with protection and balancing. The shore gateway is powered independently.

pH (SEN0161-V2), DO (SEN0237-A) and conductivity (SEN0244) enter the ADS1115 inputs AIN0, AIN1 and AIN2 through 10 kΩ/100 nF filters (pH, conductivity) or a 4.7 kΩ/47 kΩ divider with 100 nF (DO; V_{module} = 1.1 V_{ADC}) [2]. Turbidity (SEN0189) enters the ESP32-S3 ADC through a 12 kΩ/15 kΩ divider with 100 nF (V_{module} = 1.8 V_{ADC}), and its conversion uses the vendor voltage calibration [3]. With worst-case resistor tolerances, the turbidity divider gives 2.522 V at the ESP32-S3 input (nominal 2.500 V), and the ±50 mV ADC error corresponds to ±90 mV at the sensor output. Temperature (DS18B20) is read on a dedicated 1-Wire bus on GPIO16 with a 4.7 kΩ pull-up [4]; GPIO8/9 carry I^2C, and GPIO4–7 drive the four TPS22917 load switches, whose enable inputs have 100 kΩ pull-down resistors. For the SEN0237-A, range and response time refer to the current product page and accuracy to the 2019 DFRobot selection guide [5,6].

S2.2. Energy-Monitor Input Ranges

The INA228 was external test instrumentation, connected at the battery terminal of the deployed node during the 72 h test of 15–17 June 2025 and not part of the node's acquisition electronics. It senses the shunt differential voltage and the bus voltage with a 20-bit converter. With the 0.020 Ω shunt, its ±40.96 mV and ±163.84 mV shunt input ranges correspond to ±2.048 A and ±8.192 A [7], well above the −0.776 to +0.463 A observed in the 72 h test. The exported series has 1 mV and 0.1 mA resolution at 1 s. Duration summaries of ON, OFF and BOOT runs used only complete runs bounded on both sides by the opposite state.

S2.3. Manufacturer Operating-State Currents

Table S1 gives the published operating-state currents of the components in Table 2 of the main text. Radio peaks, conversion, standby and switch quiescence are separate conditions; currents at different supply voltages cannot be added as battery current.

Table S1. Manufacturer operating-state currents and reference power for the listed device quantities.

Component and state	Datasheet supply (V)	Current per device	Devices	Reference power	Basis
ESP32-S3, Wi-Fi TX reference	3.3	340 mA	1	1.12 W	peak reference
ESP32-S3, Wi-Fi RX reference	3.3	88 mA	1	290 mW	peak reference
ESP32-S3, light sleep	3.3	240 µA	1	792 µW	typical
ESP32-S3, deep sleep (RTC memory only)	3.3	7 µA	1	23.1 µW	typical
ESP32-S3, deep sleep (RTC peripherals retained)	3.3	8 µA	1	26.4 µW	typical
ADS1115, operating	3.3	150 µA	1	495 µW	typical
ADS1115, power-down	3.3	0.5 µA	1	1.65 µW	typical

Component and state	Datasheet supply (V)	Current per device	Devices	Reference power	Basis
TPS22917, enabled (quiescent only)	5.0	0.5 µA	4	10 µW	typical
TPS22917, disabled (quiescent only)	5.0	10 nA	4	0.2 µW	typical
BQ24650, sleep state	12.6	15 µA	1	189 µW	maximum
DS18B20, active conversion/EEPROM write	5.0	1–1.5 mA	1	5–7.5 mW	typical–maximum
DS18B20, standby	5.0	0.75–1 µA	1	3.75–5 µW	typical–maximum
SEN0244 (conductivity/TDS), working range	5.0	3–6 mA	1	15–30 mW	minimum–maximum
SEN0189 (turbidity), operating maximum	5.0	40 mA	1	200 mW	maximum
INA228, active monitoring	3.3	640 µA	1	2.11 mW	typical

Sources: [1,2,4,7–11]. Power = supply voltage × current × number of devices. ADS1115 values refer to the 8 SPS, ±2.048 V datasheet condition and DS18B20 values to a 5 V supply. The pH (SEN0161-V2) and DO (SEN0237-A) interface boards have no published supply current, and board-level regulation losses are not listed.

S2.4. Night-Time ON Profile and Details of the 72 h Trace

During the 72 h test (15–17 June 2025), battery voltage ranged from 10.578 to 12.595 V and net current from −0.776 to +0.463 A; daily discharge was 23.30, 25.15 and 24.42 Wh. Complete ON runs lasted 84–96 s and complete OFF runs 112–129 s. At night, ON runs accounted for 41.20 Wh and OFF runs for 8.43 Wh. The mean current of the 561 complete night-time ON runs fell from 342 mA in the first second to a plateau of 249 mA; an exponential fit to the mean profile gives a time constant of 6.9 s (Figure S2). The within-run SD of the ON current was 20.4 mA and that of the OFF current 9.0 mA.

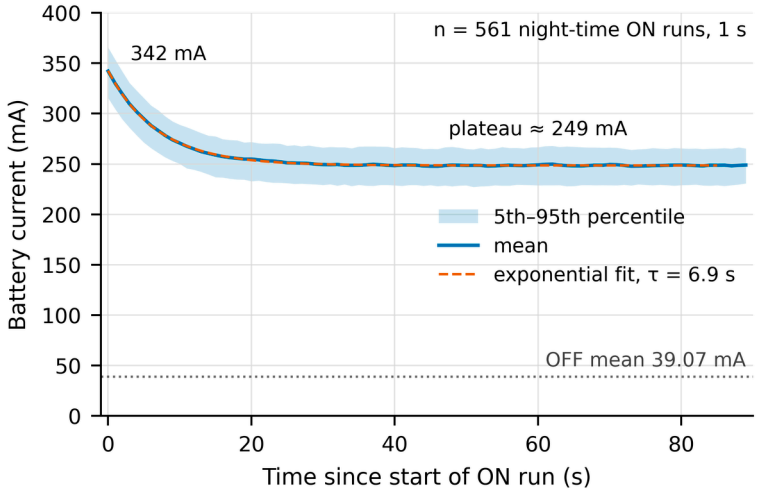


Figure S2. Battery-terminal current during night-time ON runs of the 72 h INA228 test (1 s resolution; n = 561 complete runs). Mean current and 5th–95th percentile band against time since the start of the run, with the least-squares exponential fit to the mean profile (time constant 6.9 s); the dotted line marks the night-time OFF mean (39.07 mA).

Within the ON increment of 217.27 mA, the manufacturer currents allocate at most 40 mA to turbidity, 3–6 mA to conductivity and less than 2 mA to temperature and the ADS1115 (Table 2 of the main text). Charger output, estimated as gross load minus net battery current, was 29.4, 26.0 and 35.2 Wh on 15, 16 and 17 June, at NASA POWER irradiation of 3.03, 3.07 and 3.13 kWh m^{-2} day^{-1} (campaign range 1.93–6.27 kWh m^{-2} day^{-1}). The implied charger output at the most negative currents was 0.806 A in ON and 0.815 A in OFF, matching the 0.80 A setpoint; the output was at least 0.76 A for 387 s and at least 0.70 A for 2,255 s in 72 h. Charging began between 06:33 and 07:13 and ended between 17:03 and 17:20 local time. Because the charge current is limited to 0.80 A, charger output cannot exceed about 9.6 W (0.80 A at 12 V), about two-thirds of the 15 W rating, so clipping can occur only near clear-sky irradiance; the 89-day balance applies this ceiling (Equation (18)). NASA POWER air temperature over the campaign averaged 15.5 °C, exceeded 25 °C during 19 h, fell below 5 °C during 15 h and ranged from 11.0 to 21.3 °C during the electrical test. The archived campaign export does not include the battery-status field, so the 89-day energy balance is calculated (Section S2.6).

S2.5. Load Scenarios S0–S5

In the six load scenarios of Section 2.7 and Figure 8c, mean currents follow Equation (17) of the main text, daily energy uses the 11.42 V mean pack voltage, and autonomy without sun uses the 28.94 Wh available above a 20% reserve of the installed 36.18 Wh pack, whose cells the revised design keeps (Equation (20); Section 4.3 of the main text).

S0 (test sequence): 90 s ON/120 s OFF with the measured night-time state currents (256.34 and 39.07 mA); mean current 132.18 mA, 3,172 mAh day^{-1} and 36.2 Wh day^{-1}; autonomy 0.80 days (0.72 days with 10% less capacity on cold mornings).

S1 (five-minute field cycle): 90 s ON/210 s OFF with the same currents; 104.25 mA, 2,502 mAh day^{-1} and 28.6 Wh day^{-1} (21.1% less than S0); autonomy 1.01 days.

S2: S1 with the OFF state replaced by a 1 mA sleep floor built from manufacturer quiescent currents: ESP32-S3 deep sleep (7–8 µA), INA228 (0.64 mA, kept as a conservative allowance for an on-board battery monitor), BQ24650 sleep (15 µA) and ADS1115 power-down (0.5 µA), plus regulator and protection-board quiescent currents [1,2,7,8]; 77.60 mA, 1,862 mAh day^{-1} (41.3% less); autonomy 1.36 days.

S3: S2 with a 30 s powered window; 26.53 mA, 637 mAh day^{-1} and 7.27 Wh day^{-1} (79.9% less); autonomy 3.98 days (3.58 days with 10% less capacity); night-time discharge

falls to about 0.35 Ah and battery throughput, scaled by daily demand as in Section 3.4 of the main text, to about 0.13 equivalent full cycles per day.

S4: S3 with acquisition at radio-off and batched uploads every 30 min (six records per batch, 8 s at the ON current per batch); 453 mAh day^{-1} (85.7% less; autonomy 5.60 days) under linear regulation from the pack (H1) and 592 mAh day^{-1} (81.3% less; 4.28 days) with buck regulators of 85% efficiency (H2), the relevant hypothesis for the revised design.

S5: S4 with a 20 s powered window and a 0.2 mA floor; 301 mAh day^{-1} (90.5% less; 8.42 days) under H1 and 394 mAh day^{-1} (87.6% less; 6.43 days) under H2.

S2–S5 are design scenarios built from manufacturer currents, not implemented on the buoy; they omit the time and current of Wi-Fi re-association after each deep-sleep wake-up. H1 and H2 are regulator hypotheses used to allocate the measured current; they are not design options. Autonomy values use $\bar{P}_{bat} = E_{day}/24$ h, including conversion losses.

S2.6. Calculated 89-Day Energy Balance

Table S2 gives the hourly energy balance of Section 2.7 (Equations (18) and (19) of the main text) for the installed 15 W module, the 10.3 W and 20 W brackets and the 13 and 14 W effective ratings calibrated against the 72 h test, under the test sequence (S0) and the five-minute field cycle (S1). Hourly NASA POWER all-sky irradiance at the site (daily totals 1,928–6,267 Wh m^{-2}; mean 3,922 Wh m^{-2}) [12] was multiplied by the module rating and a constant system factor of 0.69 (central performance ratio 0.75 × charger efficiency 0.92). Demand was the cycle-mean battery-terminal current extended to a full day at the 11.42 V mean pack voltage (36.2 Wh day^{-1} for S0 and 28.6 Wh day^{-1} for S1), and hourly PV power was capped at the charger ceiling of about 9.6 W, which was not binding for ratings up to 15 W in the hours that set the daily minima. Stored energy started full on 22 April and was bounded by the 36.18 Wh nominal pack energy and a 20% reserve, below which demand was counted as unserved. With the 15 W module, the charge returned during the 72 h test (charger output 26.0–35.2 Wh day^{-1}) agrees with the 31–32 Wh day^{-1} expected from the same model. Run for 15–17 June with the test sequence, the measured initial SOC of 94.3% and no reserve, the model with the 15 W rating gave a minimum SOC of 27.8% and a final SOC of 48.1%, against the measured 12.0% and 41.1%, so it is optimistic for those days. Effective ratings of 13 and 14 W gave minima of 14.1% and 21.5% and final values of 22.8% and 35.9%; Table S2 lists them as test-calibrated effective ratings.

Table S2. Calculated 89-day energy balance for five photovoltaic ratings and two operating cycles (hourly balance, NASA POWER irradiance, system factor 0.69, 9.6 W charger ceiling, 36.18 Wh storage with a 20% reserve).

Module rating	Cycle (ON/OFF)	Demand (Wh day^{-1})	Mean PV available (Wh day^{-1})	Days PV < demand	Days with unserved energy	Unserved energy (% of demand)	Minimum daily storage (%)
10.3 W (CAD outline)	S0, 90 s/120 s	36.2	27.9	74	69	22.6	20
10.3 W (CAD outline)	S1, 90 s/210 s	28.6	27.9	49	36	7.9	20
13 W (test-calibrated effective rating)	S0, 90 s/120 s	36.2	35.2	50	43	9.7	20
13 W (test-calibrated effective rating)	S1, 90 s/210 s	28.6	35.2	25	4	0.49	20

Module rating	Cycle (ON/OFF)	Demand (Wh day^{-1})	Mean PV available (Wh day^{-1})	Days PV < demand	Days with unserved energy	Unserved energy (% of demand)	Minimum daily storage (%)
14 W (test-calibrated effective rating)	S0, 90 s/120 s	36.2	37.9	41	30	6.3	20
14 W (test-calibrated effective rating)	S1, 90 s/210 s	28.6	37.9	17	0	0	24
15 W (installed)	**S0, 90 s/120 s**	**36.2**	**40.6**	**33**	**23**	**3.9**	**20**
15 W (installed)	**S1, 90 s/210 s**	**28.6**	**40.6**	**13**	**0**	**0**	**31**
20 W	S0, 90 s/120 s	36.2	54.1	11	1	0.04	20
20 W	S1, 90 s/210 s	28.6	54.1	2	0	0	50

S0, test sequence (90 s ON/120 s OFF); S1, five-minute field cycle (90 s ON/210 s OFF). Days PV < demand counts days on which the available photovoltaic energy was below the daily demand; unserved energy is demand that could not be met without discharging below the 20% reserve. 10.3 W = 295 mm × 175 mm × 20% × 1000 W m^{-2} (CAD outline). Demand is the cycle-mean battery-terminal current at the 11.42 V mean pack voltage, as in Table 2 of the main text (36.23 Wh day^{-1} to two decimals). Hourly PV power is capped at the charger ceiling (≈9.6 W). The 13 and 14 W rows are effective ratings that bring the model close to the 72 h test (see text above). The S1 balance assumes the five-minute cycle on every one of the 89 days and therefore does not represent the three test days (15–17 June), when the node ran the 90 s ON/120 s OFF sequence and the measured SOC fell to 12.0% before recovering to 41.1% (Section S2.4).

S3. Calibration by Reference Level, Session and Triplicate

Reference agreement was evaluated in six sessions at 18-day intervals: 18 April, 6 May, 24 May, 11 June, 29 June and 17 July 2025. Reference values of temperature, pH, DO and conductivity were read with a Hanna Instruments HI9829 multiparameter meter; the turbidity references were the standard suspensions themselves. Each session comprised six reference levels in triplicate: 8, 12, 16, 20, 24 and 28 °C for temperature; pH 6.50, 6.86, 7.00, 7.40, 8.00 and 8.50; 2, 4, 6, 8, 10 and 12 mg L^{-1} for DO; 0, 5, 10, 20, 40 and 80 NTU for turbidity; and 50, 100, 150, 250, 500 and 900 µS cm^{-1} for conductivity. Table S3 and Figure S3 give the three agreement metrics of Section 2.6 by reference level (18 pairs per level) and by session (18 pairs per session).

Pooled over the 108 pairs, raw RMSE was 0.371 °C, 0.098 pH units, 0.312 mg L^{-1} DO, 1.651 NTU-eq. and 16.00 µS cm^{-1}; post-calibration RMSE was 0.046 °C, 0.012, 0.049 mg L^{-1}, 0.290 NTU-eq. and 5.227 µS cm^{-1} (pooled biases −0.006, +0.0003, −0.003, +0.003 and −0.167 in channel units); and LOSO RMSE was 0.211 °C, 0.057, 0.100 mg L^{-1}, 0.984 NTU-eq. and 16.74 µS cm^{-1} (Table 4 of the main text).

DO post-calibration RMSE was 0.041–0.060 mg L^{-1} in individual sessions, and the raw DO bias varied between sessions from −0.389 mg L^{-1} (18 April) to −0.217 mg L^{-1} (17 July), with intermediate values of −0.270, −0.363, −0.233 and −0.319 mg L^{-1}; the bias did not change monotonically. Temperature LOSO error was 0.046–0.22 °C in the sessions other than 24 May (0.40 °C). Turbidity LOSO RMSE was 0.51–0.62 NTU-eq. at 10–20 NTU and rose to 1.90 NTU-eq. at 80 NTU. Conductivity LOSO RMSE was 3.30–3.42 µS cm^{-1} at the 50 and 100 µS cm^{-1} levels, 7.22 µS cm^{-1} at 150 µS cm^{-1} and 34.8 µS cm^{-1} at 900 µS cm^{-1}; the field range (101.9–125.5 µS cm^{-1}) lies between the 100 and 150 µS cm^{-1} levels.

Table S3. Calibration agreement by reference level and by session for raw readings, archived post-calibration readings and leave-one-session-out (LOSO) affine transfer.

Channel (unit)	Subset	Raw RMSE	Raw bias	Post-calibration RMSE	Post-calibration bias	LOSO RMSE	LOSO bias
Temperature (°C)	Level 8 °C	0.267	+0.257	0.045	−0.009	0.090	+0.001
	Level 12 °C	0.302	+0.281	0.029	−0.016	0.130	−0.001
	Level 16 °C	0.365	+0.324	0.042	+0.000	0.196	+0.015
	Level 20 °C	0.384	+0.337	0.052	−0.001	0.218	+0.002
	Level 24 °C	0.417	+0.359	0.051	−0.007	0.252	−0.003
	Level 28 °C	0.456	+0.376	0.055	−0.003	0.305	−0.013
	Session 18 Apr	0.393	+0.383	0.046	−0.003	0.092	+0.073
	Session 6 May	0.509	+0.496	0.051	−0.005	0.224	+0.208
	Session 24 May	0.077	+0.007	0.034	−0.001	0.399	−0.375
	Session 11 Jun	0.246	+0.242	0.052	−0.009	0.117	−0.096
	Session 29 Jun	0.473	+0.461	0.044	−0.024	0.182	+0.166
	Session 17 Jul	0.348	+0.343	0.048	+0.005	0.046	+0.025
pH (pH units)	Level pH 6.50	0.097	−0.086	0.009	+0.000	0.052	+0.002
	Level pH 6.86	0.093	−0.079	0.008	+0.003	0.058	+0.007
	Level pH 7.00	0.093	−0.081	0.013	+0.001	0.054	+0.005
	Level pH 7.40	0.104	−0.093	0.011	−0.001	0.055	−0.009
	Level pH 8.00	0.095	−0.083	0.016	−0.000	0.056	+0.000
	Level pH 8.50	0.103	−0.087	0.015	−0.001	0.065	−0.006
	Session 18 Apr	0.156	−0.155	0.013	−0.002	0.087	−0.084
	Session 6 May	0.061	−0.056	0.013	+0.000	0.043	+0.035
	Session 24 May	0.037	−0.030	0.013	−0.002	0.068	+0.065
	Session 11 Jun	0.052	−0.048	0.013	+0.003	0.049	+0.044

Channel (unit)	Subset	Raw RMSE	Raw bias	Post-calibration RMSE	Post-calibration bias	LOSO RMSE	LOSO bias
Session 29 Jun	0.117	−0.115	0.009	+0.002	0.041	−0.036	
Session 17 Jul	0.108	−0.105	0.011	−0.000	0.034	−0.024	
DO (mg L^{-1})	Level 2 mg L^{-1}	0.289	−0.274	0.042	+0.004	0.101	+0.006
Level 4 mg L^{-1}	0.290	−0.281	0.055	−0.006	0.077	+0.006	
Level 6 mg L^{-1}	0.300	−0.291	0.047	−0.003	0.083	+0.004	
Level 8 mg L^{-1}	0.347	−0.332	0.040	+0.000	0.114	−0.029	
Level 10 mg L^{-1}	0.308	−0.292	0.052	−0.003	0.113	+0.018	
Level 12 mg L^{-1}	0.334	−0.322	0.056	−0.007	0.106	−0.004	
Session 18 Apr	0.393	−0.389	0.051	+0.025	0.121	−0.109	
Session 6 May	0.277	−0.270	0.041	−0.018	0.068	+0.034	
Session 24 May	0.369	−0.363	0.042	−0.014	0.098	−0.078	
Session 11 Jun	0.239	−0.233	0.046	−0.014	0.098	+0.079	
Session 29 Jun	0.332	−0.319	0.053	−0.008	0.094	−0.024	
Session 17 Jul	0.224	−0.217	0.060	+0.013	0.113	+0.098	
Turbidity (NTU-eq.)	Level 0 NTU	1.56	+1.52	0.19	+0.03	0.39	+0.08
Level 5 NTU	1.39	+1.34	0.26	−0.05	0.38	−0.09	
Level 10 NTU	1.62	+1.55	0.21	+0.03	0.51	+0.13	
Level 20 NTU	1.49	+1.38	0.21	−0.03	0.62	−0.03	
Level 40 NTU	1.67	+1.35	0.37	+0.12	1.11	−0.03	
Level 80 NTU	2.09	+1.30	0.43	−0.09	1.90	−0.03	
Session 18 Apr	1.45	+1.38	0.38	−0.03	0.43	−0.03	
Session 6 May	2.63	+2.41	0.34	−0.01	1.78	+1.22	

Channel (unit)	Subset	Raw RMSE	Raw bias	Post-calibration RMSE	Post-calibration bias	LOSO RMSE	LOSO bias
Session 24 May	1.57	+1.46	0.34	+0.00	0.59	+0.07	
Session 11 Jun	1.10	+0.89	0.19	−0.01	0.92	−0.61	
Session 29 Jun	1.34	+1.14	0.21	+0.03	0.78	−0.32	
Session 17 Jul	1.36	+1.15	0.22	+0.03	0.82	−0.31	
Conductivity (µS cm^{-1})	Level 50 µS cm^{-1}	5.67	+4.59	1.45	+0.09	3.30	+0.16
Level 100 µS cm^{-1}	4.99	+3.72	1.60	+0.50	3.42	−0.68	
Level 150 µS cm^{-1}	9.03	+5.88	1.78	+0.25	7.22	+1.52	
Level 250 µS cm^{-1}	7.90	+2.76	2.87	−0.80	7.82	−1.54	
Level 500 µS cm^{-1}	18.52	+8.43	4.37	+0.16	18.38	+4.33	
Level 900 µS cm^{-1}	31.50	+0.25	11.35	−1.21	34.77	−3.55	
Session 18 Apr	13.27	−0.61	5.08	+0.73	14.79	−5.85	
Session 6 May	23.57	+15.20	6.40	+1.07	24.18	+13.24	
Session 24 May	13.12	+4.66	5.82	−1.81	12.32	+0.46	
Session 11 Jun	12.69	+4.60	3.64	+0.90	11.87	+0.39	
Session 29 Jun	15.59	−4.89	4.90	−2.43	19.72	−10.92	
Session 17 Jul	15.13	+6.68	5.11	+0.54	14.06	+2.90	

Errors are reading minus reference in channel units; turbidity references in NTU, readings in NTU-eq. Level rows pool the six sessions (n = 18 per level); session rows pool the six levels (n = 18 per session). Raw: readings in engineering units before the session coefficients; post-calibration: archived readings after the coefficients of that session had been applied; LOSO: affine raw-to-reference relation fitted on the other five sessions (90 pairs) and applied to the held-out session.

Triplicate repeatability was summarized as the within-session SD of the three replicates at each level (36 level–session groups per channel; Figure S3b). Pooled raw and post-calibration triplicate SDs were 0.041 and 0.048 °C for temperature, 0.022 and 0.014 for pH, 0.066 and 0.049 mg L^{-1} for DO, 0.618 and 0.296 NTU-eq. for turbidity and 13.8 and 5.2 µS cm^{-1} for conductivity; the median post-calibration-to-raw ratio was 1.08, 0.63, 0.89, 0.56 and 0.34, respectively.

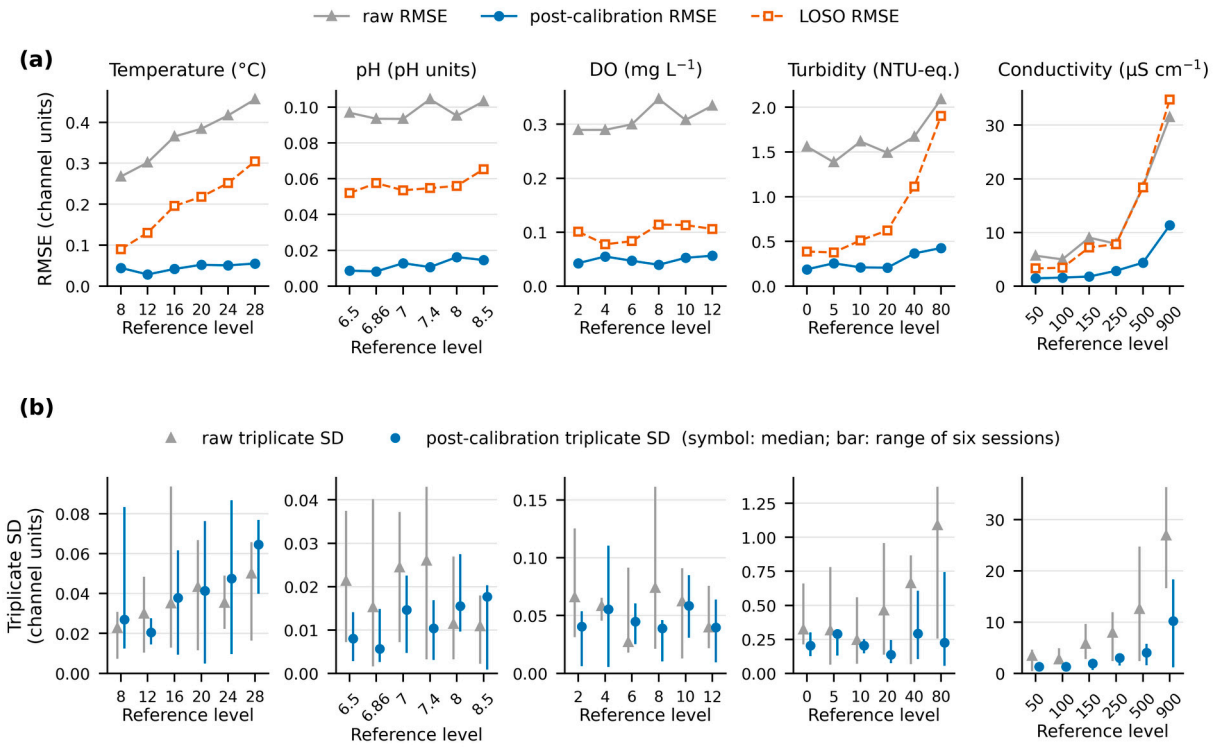


Figure S3. Calibration agreement by reference level and triplicate repeatability. (a) RMSE at each of the six reference levels for raw readings, archived post-calibration readings and LOSO affine transfer (n = 18 pairs per level: six sessions × three replicates). (b) Within-session triplicate SD at each level for raw and post-calibration readings (symbol, median of the six sessions; bar, range). Channel units; turbidity in NTU-eq.; levels on a categorical axis.

Expressed in the resolution scales of Sections 2.6 and 3.3 of the main text, the campaign SD of temperature, DO and turbidity spans 3.5–4.2 LOSO transfer-error units; the median DO daily amplitude (0.95 mg L^{-1}) is 9.5 times the DO LOSO RMSE, and the DO error at the only reference level inside the field range (6 mg L^{-1}) is 0.083 mg L^{-1} under transfer. The raw DO bias (about −0.30 mg L^{-1}) is of the order of the campaign SD (0.35 mg L^{-1}), so session coefficients are required. The diel cycle carries 91% of the temperature variance. The pH daily cycle (0.18 units) is 3.25 times its transfer error, while the SD of daily pH means (0.019) is below it. The turbidity P5–P95 span equals 10.8 transfer-error units. The conductivity successive-difference scale (0.254 µS cm^{-1}) is 0.015 times its transfer error, and the first-to-last-week increase of 4.75 µS cm^{-1} lies inside the range of session raw biases (−4.89 to +15.20 µS cm^{-1}). Successive-difference scales were 0.021 °C, 0.004 pH units, 0.021 mg L^{-1} DO and 0.185 NTU-eq.

S4. Telemetry

Of the 25,522 records (Section 3.1; Figure S4), 25,355 were delivered directly, 154 by backfill and 13 not at all; direct delivery was at least 99% on 72 of 89 days, with a minimum of 85.8% (25 April, when 38 records were backfilled). The 13 undelivered records fell on nine dates, with at most three on one date. Backfilled packets kept their original sequence identifiers and measurement timestamps and arrived 77–1,202 s after storage (median 633 s). The local and cloud logs agreed in RSSI, retry count and payload size for every packet, and the cloud never received a packet before its local-save time.

The 22 acquisition gaps removed 110 five-minute records (550 min in total); they lasted 10–60 min (median 22.5 min; 90th percentile 40 min) and began at various times of day, with six of them between 00:00 and 06:00 local time.

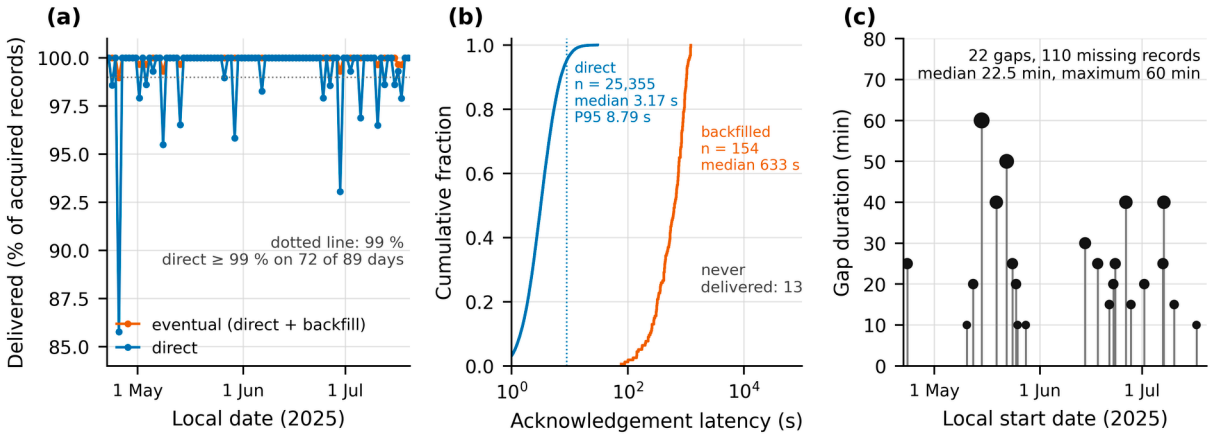


Figure S4. Daily telemetry delivery, acknowledgement latency and acquisition gaps. (a) Daily share of the acquired (locally stored) records delivered directly and eventually (direct + backfill) over the 89 days; the dotted line marks 99%. (b) Cumulative distribution of acknowledgement latency for directly delivered (n = 25,355) and backfilled (n = 154) packets (logarithmic axis), with the median and 95th percentile of direct latency. (c) The 22 acquisition gaps by local start date; stem height gives the duration and marker size the number of missing records.

S5. Turbidity Events and Daily Low-DO Exposure

S5.1. Turbidity Runs and Responses

Turbidity runs were defined above strict cutoffs of 10, 15 and 20 NTU-eq. A rising edge was counted as confirmed only when the preceding value was valid and at or below the cutoff; a run beginning after a gap was classified as a restart. The 6 h windows before and after each edge are centered on the edge, include future values and are used descriptively.

Turbidity averaged 5.62 NTU-eq. With cutoffs of 10, 15 and 20 NTU-eq., it formed 54, 41 and 15 runs, with 48, 39 and 15 confirmed rising edges and six, two and zero restarts after gaps. Around the 48 confirmed edges above 10 NTU-eq., the median change from the preceding to the following 6 h was −0.260 mg L^{-1} for DO, −0.011 for pH, +0.623 µS cm^{-1} for conductivity and +0.068 °C for temperature (Figure S5). The 1,042 samples below 6.0 mg L^{-1} formed 45 runs, and 912 of them (87.5%) coincided with turbidity above 10 NTU-eq. On the 21 event days (maximum turbidity above 15 NTU-eq.), the median daily mean DO was 6.42 versus 6.70 mg L^{-1} on the other 68 days and the median daily DO amplitude 1.43 versus 0.91 mg L^{-1}; across days, maximum turbidity and mean DO had a Spearman correlation of −0.63 (n = 89). Over the campaign, the median daily minimum DO fell from 6.422 mg L^{-1} in the first seven days to 6.066 mg L^{-1} in the last seven, and daily median conductivity increased by 0.0421 µS cm^{-1} day^{-1} (Theil–Sen slope).

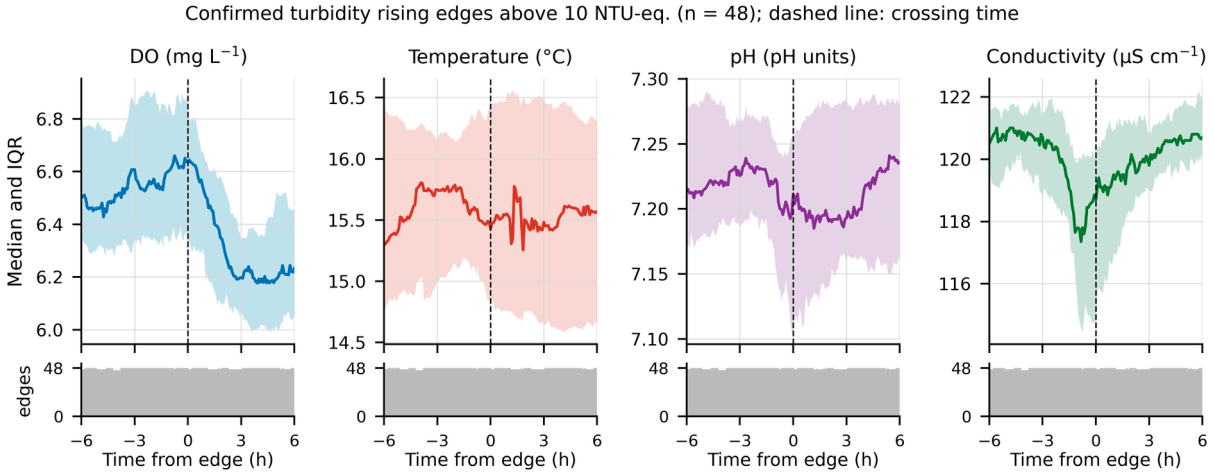


Figure S5. Water-quality profiles around turbidity rising edges. Median (line) and interquartile range (band) of DO, temperature, pH and conductivity from 6 h before to 6 h after each of the 48 confirmed rising edges above 10 NTU-eq.; the dashed line marks the crossing time, and the lower strips give the number of edges contributing at each 5 min offset.

S5.2. Daily Low-DO Exposure and Diel Detail

Of the 25,486 QC-OK observations, 8,724 (34.23%) were below 6.5 mg L^{-1}, 1,042 (4.09%) below 6.0 mg L^{-1} and 9,171 (35.98%) above 16 °C; the DO median was 6.655 mg L^{-1} and water temperature ranged from 12.951 to 17.239 °C. Nineteen days contained values below 6.0 mg L^{-1}. The largest daily share below 6.5 mg L^{-1} was 91.61% on 27 May (262 of 286 samples), followed by 17 July (85.6%), 18 July (77.1%) and 3 June (75.3%), and the lowest 7-day mean of daily minima (5.78 mg L^{-1}) corresponded to the week ending 31 May. The median diel DO curve ranged by 0.730 mg L^{-1}, from 7.036 mg L^{-1} at 12:30 to 6.306 mg L^{-1} at 19:20, and the hourly share below 6.5 mg L^{-1} peaked at 92.0% at 19:00. Of all samples, 35.5% were below 84.4% saturation, the saturation equivalent of the 6.5 mg L^{-1} endpoint at 2100 m; the median saturation curve rose after 07:00. The median oxygen deficit relative to air equilibrium was 1.07 mg L^{-1} (P5–P95 0.54–1.65 mg L^{-1}). DO autocorrelation was 0.753 at 2 h and 0.390 at 48 h.

S6. Feeding and Cleaning Comparisons

The feeding log contains 163 entries totaling 303.21 kg over 89 dates, all between 07:00 and 08:59 (84 entries) or between 15:00 and 16:59 (79 entries) local time; two meals were recorded on 74 days and one on 15 days (10 without the afternoon and five without the morning meal). The cleaning log contains five dates without execution times (9 May, 24 May, 10 June, 23 June and 7 July 2025). Biometry and feeding entries were recorded at each farm visit and in the producer's field notebook and were entered remotely through an AppSheet application.

For each feeding entry, the associated change was the median of t + 5 to t + 60 min minus the median of t − 30 to t − 5 min, requiring at least nine of twelve and five of six QC-OK samples, and was compared with clock-matched controls: the same local clock time within ±7 days, a complete ±90 min record and no recorded feeding. Each entry's control change was the median across its eligible controls, and control reuse was retained. Sensor coverage was adequate for 162 entries, of which 119 had matched controls; their 179 links used 106 distinct control centers, 48 of them reused, with up to four entries sharing one center. Median DO changed by −0.121 mg L^{-1} between the preceding 30 min and the following 60 min, and the event-minus-control changes were +0.03 °C, −0.017 pH units, −0.109 mg L^{-1} DO, +0.51 NTU-eq. and approximately 0 µS cm^{-1}. Cleaning was evaluated by date as the median of the following day minus that of the preceding day (at least 216 of 288 samples per day): −0.08 °C, +0.024 pH units, +0.061 mg L^{-1} DO, +0.23 NTU-eq. and −0.20 µS cm^{-1}.

Relative to the local ±7-day median for the same clock slot, the DO anomaly aligned on feeding entries stayed within ±0.022 mg L^{-1} of zero from 2 h before to 6 h after the entry (Figure S6a), consistent with feeding being part of the regular diel cycle. On the 10 days without an afternoon meal, the median evening (17:00–22:59) minimum DO was 6.31 versus 6.22 mg L^{-1} on the other days; on the 5 days without a morning meal, the median DO at 09:00–13:59 was 7.09 versus 6.99 mg L^{-1}. Across days, the afternoon ration was weakly correlated with the evening minimum (Spearman $\rho = -0.12$; $n = 89$; Figure S6b) and the daily ration with the daily minimum ($\rho = -0.18$). These differences (≈0.1 mg L^{-1}) are small compared with the 0.58 mg L^{-1} lower daily minimum on turbidity-event days.

The + Feeding configurations (Section 2.9 of the main text) added past feeding flags and quantities lagged by one sample and summed over 1, 6 and 24 h, each window requiring at least 75% observed values, together with their available fractions (12 columns; 47 predictors with

DO history and 107 with Five channels). The time at which entries became available to an online service was not recorded, so the comparison is retrospective (Table S4).

Table S4. DO forecast error with and without past feeding records (+ Feeding configurations), test period.

h	Inputs	p	Test n	RMSE (95% interval)	MAE	Bias	ΔRMSE vs. reference (95% interval)
1 h	DO history	35	3,559	0.150 (0.113–0.188)	0.100	+0.011	Reference
1 h	Five channels	95	3,559	0.134 (0.109–0.161)	0.097	+0.033	Reference
1 h	DO history + Feeding	47	3,559	0.149 (0.114–0.186)	0.101	+0.013	−0.0004 (−0.0024 to +0.0018)
1 h	Five channels + Feeding	107	3,559	0.135 (0.109–0.163)	0.098	+0.033	+0.0015 (+0.0002 to +0.0028)
2 h	DO history	35	3,542	0.210 (0.153–0.267)	0.143	+0.025	Reference
2 h	Five channels	95	3,542	0.174 (0.145–0.205)	0.138	+0.063	Reference
2 h	DO history + Feeding	47	3,542	0.209 (0.153–0.265)	0.142	+0.026	−0.0010 (−0.0030 to +0.0014)
2 h	Five channels + Feeding	107	3,542	0.175 (0.144–0.208)	0.139	+0.061	+0.0007 (−0.0020 to +0.0032)
3 h	DO history	35	3,529	0.260 (0.173–0.345)	0.171	+0.080	Reference
3 h	Five channels	95	3,529	0.180 (0.150–0.212)	0.139	+0.045	Reference
3 h	DO history + Feeding	47	3,529	0.256 (0.177–0.335)	0.171	+0.070	−0.0036 (−0.0110 to +0.0059)
3 h	Five channels + Feeding	107	3,529	0.180 (0.151–0.209)	0.141	+0.050	−0.0007 (−0.0077 to +0.0054)
6 h	DO history	35	3,497	0.305 (0.208–0.390)	0.202	+0.065	Reference
6 h	Five channels	95	3,497	0.359 (0.290–0.421)	0.246	+0.007	Reference

h	Inputs	p	Test n	RMSE (95% interval)	MAE	Bias	ΔRMSE vs. reference (95% interval)
6 h	DO history + Feeding	47	3,497	0.314 (0.217–0.401)	0.208	+0.063	+0.0095 (+0.0010 to +0.0182)
6 h	Five channels + Feeding	107	3,497	0.332 (0.260–0.398)	0.230	+0.025	−0.0265 (−0.0427 to −0.0120)

p, number of predictors. Errors in mg L^{-1}; bias is prediction minus observation. Each feeding variant is compared with its own sensor-only reference on common origins and with the same estimator family and settings (Random Forest at 1, 2 and 6 h; gradient boosting at 3 h). Intervals are percentiles of 10,000 paired resamples of the 14 test dates and are pointwise.

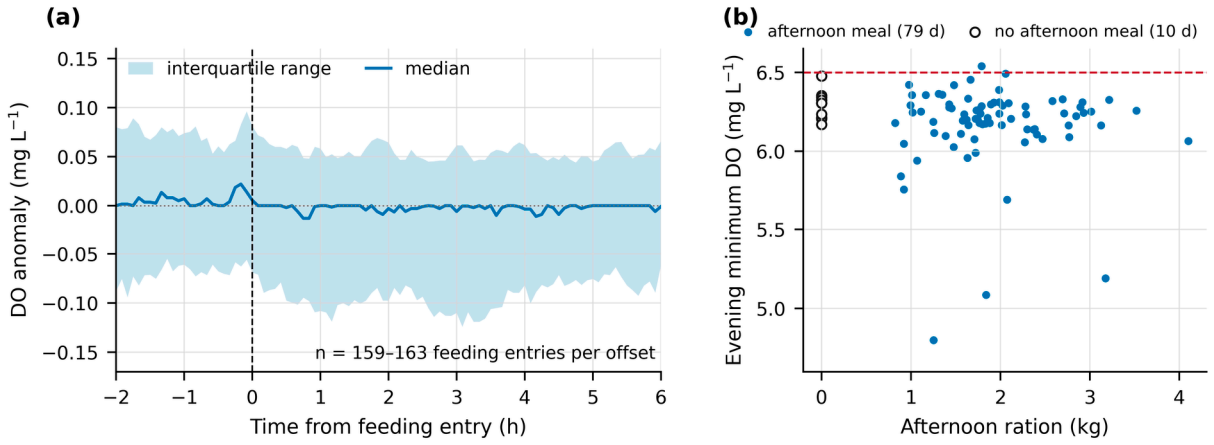


Figure S6. Feeding comparisons. (a) DO anomaly (observed minus the local ±7-day median for the same 5 min clock slot) aligned on the feeding entries, from 2 h before to 6 h after the entry: median and interquartile range. (b) Evening (17:00–22:59) minimum DO against the afternoon ration for each of the 89 days; open symbols mark days without an afternoon meal and the dashed line the 6.5 mg L^{-1} endpoint.

S7. Retrospective NASA POWER Comparisons

Hourly NASA POWER estimates (community AG) of air temperature, relative humidity, wind speed, all-sky shortwave irradiance and precipitation were retrieved for −6.35444°, −77.89917° and 20 April–20 July 2025 in UTC [12]. Each source hour h was assigned the end time h + 1 h and matched to the following twelve 5 min forecast origins, with a maximum tolerance of 55 min, so that no value entered before the end of its source hour. Because NASA reports latencies of 2–3 days for meteorological and 5–7 days for solar variables [13], these inputs were not available at the forecast origin in real time and are used only retrospectively. The + NASA configurations added 66 NASA-derived columns to DO history (101 predictors) or Five channels (161).

NASA inputs lowered 3 h DO-history RMSE from 0.260 to 0.250 mg L^{-1} and changed Five channels by −0.0008 mg L^{-1} at 3 h; at 6 h both NASA variants were less accurate than their sensor-only references (0.312 versus 0.305 and 0.383 versus 0.359 mg L^{-1}; Table S5), so the effect of the weather inputs was not consistent across horizons.

Table S5. DO forecast error with NASA POWER inputs, test period.

h	Test n	Inputs (p)	RMSE (95% interval)	MAE	Bias	R^2	Skill vs. persistence (%)	ΔRMSE vs. Hour26 (95% interval)
1 h	3,559	DO history (35)	0.150 (0.113–0.188)	0.100	+0.011	0.855	15.7	−0.210 (−0.316 to −0.083)
	DO history + NASA (101)	0.148 (0.114–0.183)	0.099	+0.008	0.858	16.6	−0.211 (−0.321 to −0.080)	
	Five channels (95)	0.134 (0.109–0.161)	0.097	+0.033	0.885	24.7	−0.226 (−0.345 to −0.086)	
	Five channels + NASA (161)	0.134 (0.109–0.161)	0.097	+0.026	0.883	24.4	−0.225 (−0.345 to −0.085)	
2 h	3,542	DO history (35)	0.210 (0.153–0.267)	0.143	+0.025	0.719	25.7	−0.152 (−0.240 to −0.042)
	DO history + NASA (101)	0.206 (0.152–0.260)	0.140	+0.014	0.729	27.0	−0.156 (−0.247 to −0.042)	
	Five channels (95)	0.174 (0.145–0.205)	0.138	+0.063	0.805	38.1	−0.187 (−0.303 to −0.048)	
	Five channels + NASA (161)	0.167 (0.142–0.194)	0.132	+0.048	0.821	40.7	−0.194 (−0.313 to −0.052)	
3 h	3,529	DO history (35)	0.260 (0.173–0.345)	0.171	+0.080	0.573	28.5	−0.103 (−0.168 to −0.018)
	DO history + NASA (101)	0.250 (0.175–0.325)	0.169	+0.032	0.605	31.2	−0.112 (−0.182 to −0.020)	
	Five channels (95)	0.180 (0.150–0.212)	0.139	+0.045	0.794	50.3	−0.182 (−0.297 to −0.045)	
	Five channels + NASA (161)	0.180 (0.150–0.210)	0.139	+0.052	0.796	50.6	−0.183 (−0.302 to −0.043)	

h	Test n	Inputs (p)	RMSE (95% interval)	MAE	Bias	R^2	Skill vs. persistence (%)	ΔRMSE vs. Hour26 (95% interval)
6 h	3,497	DO history (35)	0.305 (0.208–0.390)	0.202	+0.065	0.385	38.3	−0.047 (−0.092 to +0.009)
	DO history + NASA (101)	0.312 (0.211–0.403)	0.204	+0.030	0.356	36.8	−0.040 (−0.089 to +0.014)	
	Five channels (95)	0.359 (0.290–0.421)	0.246	+0.007	0.148	27.4	+0.007 (−0.075 to +0.112)	
	Five channels + NASA (161)	0.383 (0.304–0.452)	0.265	−0.045	0.029	22.5	+0.031 (−0.064 to +0.147)	

p, number of predictors. Errors in mg L^{-1}; bias is prediction minus observation. Rows within a horizon share the same test origins and estimator family. Intervals are percentiles of 10,000 paired resamples of the 14 test dates. Skill = $100 \times (1 - RMSE_{model}/RMSE_{persistence})$.

Over hourly pairs (n ≈ 2,100 per lag), the Spearman correlation between NASA precipitation and measured turbidity h hours later ranged from −0.09 (5 h) to 0.13 (19 h) for lags of 0–24 h, and from −0.06 to 0.08 when turbidity led precipitation; at this spatial resolution, NASA POWER precipitation did not anticipate the turbidity events of Section S5.

S8. Model Selection and Forecasting Design

S8.1. Predictors

For the target channel, predictors comprised the current observation; lags of 5, 10, 15, 30, 60, 120, 180, 360, 720 and 1440 min; and rolling means, SDs, minima and maxima over 1, 3, 6, 12 and 24 h windows ending at the preceding sample (31 columns). Each other channel contributed its current value, lags of 1, 3, 6, 12 and 24 h, and rolling means and SDs over 1, 6 and 24 h (12 columns); rolling minima and maxima over the same windows were added for DO, temperature and turbidity (18 columns). Every rolling statistic required at least 75% of its samples and no fewer than three. Four sine–cosine terms encoded local time of day and day of year. Hence DO history and Own history have 35 predictors; DO + T and DO + Turb 53; DO + pH and DO + Cond 47; and Five channels 95 for DO, temperature and turbidity but 101 for pH and conductivity, because DO, temperature and turbidity then all enter as non-target channels with rolling extrema. Calendar4 uses only the four sine–cosine terms, and Hour26 uses 24 hour indicators and two daily harmonics. Shared-canal discharge, aeration-related flags, cleaning dates and fouling indices were not used as predictors (Table S7).

S8.2. Chronological Boundaries

The eligible span of origins (23 April 05:00 UTC to 20 July 03:55 UTC 2025 at 1 h) was divided in elapsed time into training (to 23 June 18:38 UTC), selection (to 30 June 08:58 UTC), calibration (to 6 July 23:17 UTC) and test periods; at 6 h the boundaries were 23 June 15:09, 30 June 05:05 and 6 July 19:02 UTC. Origin counts per period were 16,721/1,745/1,826/3,559 at 1 h and 16,642/1,626/1,710/3,497 at 6 h. Final estimators were fitted on training and selection

data (18,486 origins at 1 h and 18,381 at 6 h; these totals exceed the sum of the two periods because the one-horizon purge at the training–selection boundary is not applied when the periods are merged for fitting); calibration supplied interval quantiles and probability thresholds; the test period (14 local dates, 6–19 July 2025) was used only for evaluation. The percentages locate boundaries in elapsed time rather than exact fractions of retained rows, and one forecast horizon was purged on each side of every boundary.

S8.3. Hyperparameter Grids and Selection Scores

The grids were Ridge regression with $\alpha \in \{0.1, 1, 10, 100\}$ and scaling fitted on training data; Random Forest [14] with 200 trees, bootstrap fraction 0.85, feature fractions 0.4 or 0.8 and minimum leaf size 5 or 15; and histogram gradient boosting [15] with 200 iterations, learning rates 0.03 or 0.08, 15 or 31 terminal nodes, minimum leaf size 25 and no early stopping, all in scikit-learn [16]. Settings were chosen with four expanding-window folds separated by a horizon-length gap and then frozen; the final calculation refitted the estimators without a new search. For each target and horizon, the family with the lowest selection-period RMSE for Five channels served all input configurations (Table S6). The low-DO classifiers compared logistic regression ($C \in \{0.01, 0.1, 1, 10\}$) and histogram gradient boosting (the regression grid with 200 iterations and minimum leaf size 25) by selection-period average precision; gradient boosting was selected at both endpoints (learning rate 0.08 and 15 terminal nodes at 6.5 mg L^{-1}; 0.03 and 15 at 7.0 mg L^{-1}).

Table S6. Selection-period RMSE of the three candidate regression families for Five channels in each of the 20 target–horizon tasks (60 scores), with the frozen setting of each family.

Target	h	Ridge RMSE (setting)	Random Forest RMSE (setting)	Gradient boosting RMSE (setting)	Selected
DO (mg L^{-1})	1 h	0.1075 (α = 100)	**0.1071 (features 0.4; leaf 15)**	0.1107 (rate 0.03; nodes 15)	Random Forest
2 h	0.1468 (α = 100)	**0.1453 (features 0.4; leaf 15)**	0.1527 (rate 0.03; nodes 15)	Random Forest	
3 h	0.1607 (α = 100)	0.1560 (features 0.8; leaf 5)	**0.1547 (rate 0.03; nodes 15)**	Gradient boosting	
6 h	0.1552 (α = 100)	**0.1539 (features 0.4; leaf 15)**	0.1648 (rate 0.03; nodes 15)	Random Forest	
Temperature (°C)	1 h	**0.0618 (α = 100)**	0.0653 (features 0.4; leaf 15)	0.0648 (rate 0.03; nodes 15)	Ridge
2 h	**0.0769 (α = 100)**	0.0951 (features 0.4; leaf 15)	0.0883 (rate 0.03; nodes 15)	Ridge	
3 h	**0.0866 (α = 100)**	0.1229 (features 0.4; leaf 5)	0.1108 (rate 0.03; nodes 31)	Ridge	
6 h	**0.1039 (α = 100)**	0.1627 (features 0.4; leaf 15)	0.1429 (rate 0.03; nodes 15)	Ridge	
pH (pH units)	1 h	0.0140 (α = 100)	**0.0133 (features 0.4; leaf 15)**	0.0133 (rate 0.03; nodes 15)	Random Forest
2 h	**0.0173 (α = 100)**	0.0183 (features 0.4; leaf 15)	0.0177 (rate 0.03; nodes 31)	Ridge	
3 h	**0.0190 (α = 100)**	0.0200 (features 0.4; leaf 15)	0.0202 (rate 0.03; nodes 15)	Ridge	

Target	h	Ridge RMSE (setting)	Random Forest RMSE (setting)	Gradient boosting RMSE (setting)	Selected
6 h	**0.0217 (α = 100)**	0.0322 (features 0.4; leaf 5)	0.0469 (rate 0.03; nodes 15)	Ridge	
Turbidity (NTU-eq.)	1 h	0.459 (α = 100)	0.366 (features 0.4; leaf 5)	**0.348 (rate 0.03; nodes 15)**	Gradient boosting
2 h	0.647 (α = 100)	0.677 (features 0.4; leaf 15)	**0.462 (rate 0.03; nodes 15)**	Gradient boosting	
3 h	**0.866 (α = 100)**	1.639 (features 0.4; leaf 15)	1.487 (rate 0.03; nodes 31)	Ridge	
6 h	**1.433 (α = 100)**	5.017 (features 0.4; leaf 5)	7.308 (rate 0.08; nodes 15)	Ridge	
Conductivity (µS cm^{-1})	1 h	**0.620 (α = 100)**	0.702 (features 0.4; leaf 15)	1.041 (rate 0.03; nodes 15)	Ridge
2 h	**0.700 (α = 100)**	1.105 (features 0.4; leaf 15)	1.648 (rate 0.08; nodes 31)	Ridge	
3 h	**0.731 (α = 100)**	1.627 (features 0.4; leaf 15)	2.830 (rate 0.03; nodes 15)	Ridge	
6 h	**0.994 (α = 100)**	4.050 (features 0.4; leaf 5)	4.000 (rate 0.03; nodes 31)	Ridge	

RMSE in the unit of each target. Bold, selected family (lowest selection-period RMSE), which was then used for every input configuration of that target and horizon. Settings: α, Ridge penalty; features, Random Forest feature fraction; leaf, minimum leaf size; rate, gradient-boosting learning rate; nodes, maximum terminal nodes. Ridge was selected in 13 of the 20 tasks, always with α = 100, the largest value in the grid.

S8.4. Changes from the Preliminary Analysis

The revised design was evaluated on the same July test period as the preliminary analysis of the same record (Section 4.7 of the main text discusses this reuse); Table S7 lists every change. Test observations did not enter estimator fitting, threshold selection or interval calibration, and Section S12 compares the water-only results of the two designs.

Table S7. Changes to the forecasting design between the preliminary analysis and the final analysis.

Design component	Preliminary analysis	Final analysis	Reason
Forecast targets	DO at 1, 2, 3 and 6 h	All five water channels at 1, 2, 3 and 6 h	Show whether each channel is itself forecastable and what it adds to DO forecasts
Linear regressor	Elastic Net; α 0.0001–0.1; L1 ratios 0.1, 0.5 or 0.9	Ridge; α 0.1, 1, 10 or 100	Single-penalty linear reference for many correlated lag predictors; smaller grid
Random Forest	350 trees; feature fractions 0.4/0.8; minimum leaf size 1/5/15	200 trees; feature fractions 0.4/0.8; minimum leaf size 5/15	Lower model complexity across 20 target–horizon tasks; no single-sample leaves

Design component	Preliminary analysis	Final analysis	Reason
Histogram gradient boosting	250 iterations; learning rates 0.03/0.08; 15/31 terminal nodes; minimum leaf size 15/35; L2 penalty 0/0.1	200 iterations; learning rates 0.03/0.08; 15/31 terminal nodes; minimum leaf size 25	Lower model complexity; same grid for regression and classification
Classifier candidates	Logistic regression, Random Forest, Extra Trees and histogram boosting (400 trees or 250 iterations)	Logistic regression and histogram boosting (200 iterations; minimum leaf size 25)	One linear and one boosting candidate, as for regression
Time boundaries	70% and 85%; training, validation and test	70%, 77.5% and 85%; training, selection, calibration and test	Separate period for interval quantiles and probability thresholds, so that neither is set on selection or test data
Final fitting	All eligible origins before the 85% boundary minus the horizon	All eligible origins before the 77.5% boundary minus the horizon	Keep the calibration week out of estimator fitting
Family comparison	Full 251-predictor configuration; family shared by input groups	Five-channel configuration; family shared by input groups	The full configuration contained inputs withdrawn from the analysis
DO family at 1/2/3/6 h	Gradient boosting / gradient boosting / Random Forest / Random Forest	Random Forest / Random Forest / gradient boosting / Random Forest	Consequence of the new grids, partitions and comparison configuration
Additional inputs	Meteorology, shared-channel flow and operational records in the full configuration	NASA POWER and past feeding examined separately and retrospectively; no flow, aeration, cleaning or fouling predictors	Flow describes the canal that feeds four raceways, not the monitored raceway; the raceway has no aerator; the fouling index was not defined by a measurement and cleaning dates have no execution times; the original meteorological series was replaced by a documented hourly NASA POWER retrieval aligned to the last completed hour (Section S7)

Both designs purge one forecast horizon on each side of every boundary. Table S9 compares the results of the two designs.

S8.5. Training-History Curves

Six trailing fractions (25–100%) of each DO Five-channel training set were evaluated at a common training endpoint and against the same selection period, giving 24 additional fits with the frozen settings (Figure S7b). Training counts at 25% and 100% were 4,180/16,721 at 1 h, 4,175/16,702 at 2 h, 4,171/16,686 at 3 h and 4,160/16,642 at 6 h, and selection counts were 1,745, 1,721, 1,695 and 1,626. Increasing the fraction extends the history backward and therefore also changes the seasons and conditions represented in training, so the curves do not isolate a sample-size effect. At 3 h, RMSE was not monotonic in the training fraction: the 70% window

gave 0.150 mg L^{-1} versus 0.155 mg L^{-1} for the complete training period. Neither calibration nor test observations were used, and the curves did not select a model or a training fraction.

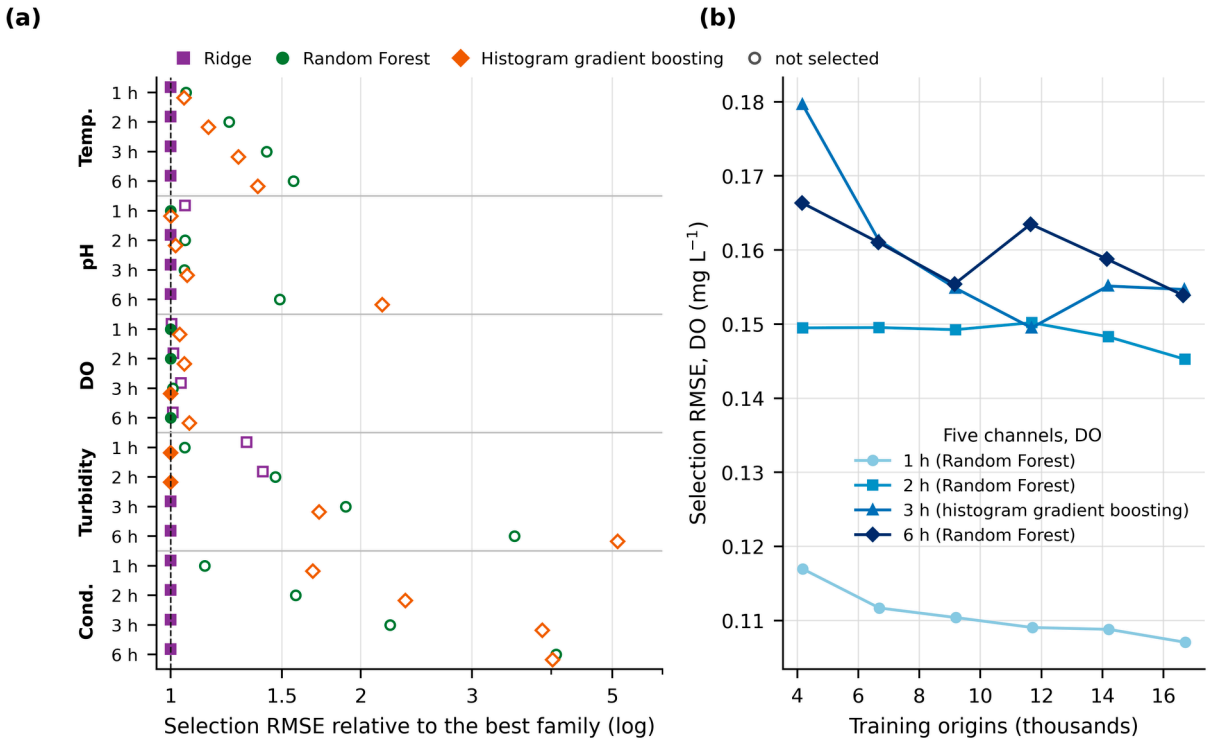


Figure S7. Model-selection scores and training-history curves. (a) Selection-period RMSE of Ridge, Random Forest and histogram gradient boosting for Five channels in each target × horizon task, relative to the best family of the task (logarithmic axis; filled symbols, selected family; Table S6). (b) Selection-period RMSE of the DO Five-channel forecasts at 1, 2, 3 and 6 h against the number of training origins when training is truncated to its most recent 25–100% (same selection period and frozen settings).

S8.6. Additional Forecast Details

Other channels. Own-history temperature RMSE was 0.11–0.20 times the test SD. pH RMSE was 0.32–0.49 test SD at 1–2 h; the 3 h pH skill relative to Hour26 was 5.2% (Own history) and 8.1% (Five channels), and at 6 h pH RMSE was 0.041 for Hour26, 0.047 for Own history and 0.049 for Five channels. Five-channel turbidity RMSE was 0.15 and 0.29 test SD at 1 and 2 h, and the 6 h skill relative to Hour26 was 3% and 2% with intervals including zero. Conductivity forecast RMSE was 1.48–2.19 µS cm^{-1}, with R^2 of −0.08 and −0.22 at 6 h.

DO input configurations. At 1 h DO + T reached 0.152 mg L^{-1}; 3 h skill relative to persistence ranged from 28.5% (DO history) to 52.0% (DO + Cond); at 6 h DO + pH had lower daily error than Hour26 on 9 of 14 days, DO + Cond had R^2 = −0.05, and DO + Turb, DO + T and DO history reached 0.300, 0.304 and 0.305 mg L^{-1}. At 6 h DO history was 0.054 mg L^{-1} more accurate than Five channels (Diebold–Mariano p = 0.049). At 3 h Five channels reached 50.3% skill relative to persistence and 50.2% relative to Hour26, whose RMSE was about 0.36 mg L^{-1} at every horizon.

Daily baselines. DO-history RMSE of 0.150, 0.210, 0.260 and 0.305 mg L^{-1} at 1, 2, 3 and 6 h compares with 0.468, 0.469, 0.470 and 0.461 mg L^{-1} for Seasonal-24 h and 0.228, 0.359, 0.452 and 0.564 mg L^{-1} for Diurnal anomaly; 15 of the 16 model–baseline paired intervals lay below zero, the exception being Five channels against Seasonal-24 h at 6 h (−0.102 mg L^{-1}; −0.232 to +0.028).

Six-hour diagnostics. The mean change for each clock slot explained 36% of the variance of the 3 h DO change, and Hour26 RMSE was 0.90–0.91 test SD. Six-hour DO-history RMSE by origin hour was 0.36–0.44 mg L^{-1} for 21:00–23:00 and 0.39–0.43 mg L^{-1} for 03:00–07:00, but only 0.14–0.20 mg L^{-1} for 11:00–19:00 (129–155 origins per hour); Five-channel bias was

−0.25 to −0.37 mg L^{-1} for origins at 12:00–16:00. Training contained 3.7% DO below 6.0 mg L^{-1} and 9.1% turbidity above 10 NTU-eq., whereas the selection and calibration periods contained no DO below 6.0 mg L^{-1}. Between the two test blocks (n = 1,619 for 6–12 July and 1,878 for 13–19 July), Five-channel RMSE rose from 0.296 to 0.406 mg L^{-1} and Calendar4 RMSE from 0.231 to 0.453 mg L^{-1}. On the 326 low-DO targets, Five channels overpredicted DO by 0.568 mg L^{-1}. The DO-history mean squared error comprised 73.7% within-day error, 21.7% between-day level shifts and 4.6% overall bias, and the Five-channel predicted-to-observed variance ratio was 1.16. Six-hour test error was about twice that of the calibration week, and test-period low-DO prevalence (46.9%) exceeded that of calibration (27.6%).

S9. Coverage of Prediction Intervals

Split-conformal 90% intervals [17] were calibrated on the calibration period and evaluated on the test period, with 95% day-bootstrap intervals and Christoffersen tests of unconditional coverage and independence on non-overlapping origins [18] (Figure S8). Test coverage ranged from 0.67 to 0.85 for DO, 0.86 to 0.93 for temperature, 0.70 to 0.84 for pH, 0.54 to 0.83 for turbidity and 0.77 to 0.88 for conductivity. For DO, six-hour half-widths were 0.242 (DO history) and 0.253 mg L^{-1} (Five channels), and the upper limits of the 3 h coverage intervals reached 90.1% and 91.7%. Misses were serially clustered for DO history at 1–2 h (p_{ind} = 0.003 and 0.024) but not at 3–6 h (0.28–0.52); at 6 h p_{ind} was 0.37 for DO history and 0.51 for Five channels.

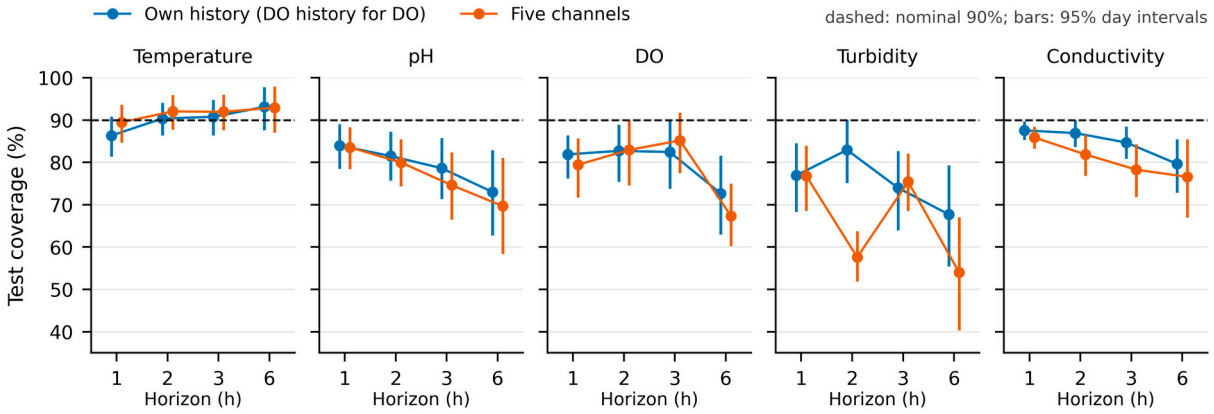


Figure S8. Coverage of nominal 90% residual intervals for the five channels. Test coverage of the split-conformal intervals by horizon for Own history (DO history for DO) and Five channels, with 95% day-bootstrap intervals (10,000 resamples of the 14 test dates); the dashed line marks the nominal 90%.

S10. Threshold Sensitivity

S10.1. Elevation Assumed for Oxygen Saturation

Repeating the saturation calculation of Section 2.8 at 2000 m (79.58 kPa) and 2200 m (77.63 kPa) instead of 2100 m (78.60 kPa) [19–21] gave a median saturation of the record of 85.1% and 87.3%, against 86.2%; a solubility at the median water temperature (15.505 °C) of 7.80 and 7.60 mg L^{-1}, against 7.70 mg L^{-1}; and saturation equivalents of the 6.5 mg L^{-1} endpoint of 83.4% and 85.5%, against 84.4%. The 6.0 and 7.0 mg L^{-1} endpoints corresponded to 76.9–78.9% and 89.8–92.1% saturation over the same elevation range. These values bracket the elevation assumption; they are not confidence limits for the weather-driven variation in pressure. The 6.0 and 7.0 mg L^{-1} endpoints follow the former EU salmonid-water values, under which all samples should exceed 7 mg L^{-1} and concentrations below 6 mg L^{-1} required evidence of no harm [22].

S10.2. Probability Thresholds and DO Endpoints

Table S8 reports the operating points of the 3 h low-DO classifier under calibration-sensitivity constraints of 0.70, 0.80 and 0.90, the classifier fitted at the 7.0 mg L^{-1} endpoint, and the concentration rule at the 6.0 mg L^{-1} endpoint, for which no classifier was fitted because the selection and calibration periods contained no positive cases. Low-DO prevalence (DO below 6.5 mg L^{-1} three hours ahead) was 33.7% in training, 32.6% in selection, 27.6% in calibration and 46.9% in the test period, consistent with the underestimation of test-period risk by the predicted probabilities. Changing the constraint moved the operating point in the intended direction for every input configuration.

Table S8. Threshold-sensitivity analyses for low-DO warnings on the test period: calibration-sensitivity constraint at 6.5 mg L^{-1}, the 7.0 mg L^{-1} endpoint and the 6.0 mg L^{-1} concentration rule.

Endpoint and analysis	Inputs	Setting	Probability threshold	Sensitivity (%)	Specificity (%)	Precision (%)	Balanced accuracy (%)
6.5 mg L^{-1} at 3 h, classifier: calibration-sensitivity constraint (test prevalence 46.9%)	DO history	≥0.70	0.766	74.8	95.1	93.2	85.0
		≥0.80 (primary)	0.306	85.6	87.7	86.0	86.7
		≥0.90	0.093	91.2	78.9	79.3	85.0
	Five channels	≥0.70	0.571	72.6	94.3	91.8	83.5
		≥0.80 (primary)	0.093	92.6	80.4	80.7	86.5
		≥0.90	0.022	96.5	73.8	76.5	85.1
	Calendar4	≥0.70	0.922	56.8	95.8	92.2	76.3
		≥0.80 (primary)	0.193	71.9	87.8	83.9	79.8
		≥0.90	0.064	81.4	69.1	70.0	75.3
	Hour26	≥0.70	0.711	57.3	95.6	92.0	76.4
		≥0.80 (primary)	0.411	73.3	87.7	84.1	80.5
		≥0.90	0.340	78.6	76.3	74.6	77.4
7.0 mg L^{-1} at 3 h, classifier (test prevalence 87.8%)	DO history	≥0.80; ROC-AUC 0.877; AP 0.981; Brier 0.088	0.549	91.2	62.0	94.5	76.6
	Five channels	≥0.80; ROC-AUC 0.886; AP 0.983; Brier 0.100	0.562	85.1	73.8	95.9	79.5

Endpoint and analysis	Inputs	Setting	Probability threshold	Sensitivity (%)	Specificity (%)	Precision (%)	Balanced accuracy (%)
Calendar4	≥0.80; ROC-AUC 0.851; AP 0.976; Brier 0.213	0.553	74.3	78.0	96.0	76.1	
Hour26	≥0.80; ROC-AUC 0.831; AP 0.973; Brier 0.182	0.428	72.6	78.9	96.1	75.8	
6.0 mg L^{-1} at 3 h, concentration rule (test prevalence 10.3%; n = 3,529)	Persistence of state	Forecast < 6.0 mg L^{-1}	—	55.1	95.3	57.3	75.2
DO history	Forecast < 6.0 mg L^{-1}	—	31.0	98.3	67.3	64.6	
Five channels	Forecast < 6.0 mg L^{-1}	—	61.4	99.1	88.2	80.2	
Calendar4	Forecast < 6.0 mg L^{-1}	—	0.0	100.0	—	50.0	
Hour26	Forecast < 6.0 mg L^{-1}	—	0.0	100.0	—	50.0	
6.0 mg L^{-1} at 6 h, concentration rule (test prevalence 9.3%; n = 3,497)	Persistence of state	Forecast < 6.0 mg L^{-1}	—	41.1	93.0	37.6	67.1
DO history	Forecast < 6.0 mg L^{-1}	—	4.6	99.2	37.5	51.9	
Five channels	Forecast < 6.0 mg L^{-1}	—	7.1	88.4	5.9	47.7	
Calendar4	Forecast < 6.0 mg L^{-1}	—	0.0	100.0	—	50.0	
Hour26	Forecast < 6.0 mg L^{-1}	—	0.0	100.0	—	50.0	

Each probability threshold maximized calibration-period specificity subject to the stated minimum calibration sensitivity; ≥0.80 is the primary setting (Table 7 of the main text). Precision is the positive predictive value. The concentration rule classifies a target as low when the regression forecast at that horizon is below 6.0 mg L^{-1}; for persistence of state, the forecast is the current DO. Calendar4 and Hour26 forecasts never fell below 6.0 mg L^{-1}. AP, average precision. n = 3,529 origins at 3 h and 3,497 at 6 h.

At the primary constraint and without confirmation, balanced accuracy was 86.7% for DO history and 86.5% for Five channels, precision 86.0% and 80.7%, average precision 0.916 and 0.935, and Brier score 0.103 and 0.113 (Table 7 of the main text); the ROC-AUC intervals of the controls were 0.760–0.937 (Calendar4) and 0.770–0.941 (Hour26). With one, two and three confirmations, DO-history sensitivity was 85.6, 83.6 and 81.8% at specificities of 87.7, 89.2 and 90.2%, and Five-channel sensitivity 92.6, 90.6 and 88.6% at 80.4, 81.6 and 82.5%. Completely false warning runs per support-day were 1.39, 0.82 and 0.49 for DO history and 0.90, 0.49 and 0.33 for Five channels. With two-sample confirmation, Five channels detected all 24 complete episodes, whereas DO history detected 15 and the calendar controls 16; the episodes missed by DO history were short (median five samples). The median retrospective lead equals the 180 min bound, i.e., at least half of the detections occurred at the earliest possible origin.

S11. Negative Turbidity Forecasts

The six-hour turbidity model selected for Five channels is an unconstrained Ridge regression (Table S6). It produced 13 negative forecasts among the 3,497 test origins (0.37%; minimum −0.295 NTU-eq.), although observed test turbidity never fell below 3.27 NTU-eq.; the Own-history Ridge model produced none (minimum 1.83 NTU-eq.). The negative forecasts were kept in the error evaluation. Physical constraints, such as truncation at zero or a log-transformed target, would have to be specified and validated in advance rather than applied after the test results; truncating 13 small negative values would not change the reported turbidity RMSE at the precision given in the main text.

S12. Water-Only Results under the Preliminary and Final Analysis Designs

Table S9 compares the DO results of the preliminary analysis with those of the final analysis for the input configurations that both designs contain. The preliminary design used the 70/85% partitions, the candidate grids of Table S7, final fitting before the 85% boundary and 3,595 test origins at 3 h; the final analysis uses the design of Section 2.9 of the main text (3,529 origins at 3 h). Both evaluate the same July test period. The water-only configurations reproduced the preliminary results closely: the 3 h DO-history RMSE was 0.253 versus 0.260 mg L^{-1}, the 3 h low-DO ROC-AUC of DO history 0.929 versus 0.932, and the ranking of the six configurations by skill agreed with Spearman $\rho = 0.89$–1.00 across horizons (median absolute skill difference 2.0 points; maximum 6.4). Configurations of the preliminary analysis that included shared-canal flow, the original meteorological series or operational records (119, 179 and 251 predictors) are not carried forward, because they contain the withdrawn inputs (Table S7).

Table S9. Water-only DO results under the two analysis designs: preliminary analysis / final analysis.

Quantity	Inputs	1 h	2 h	3 h	6 h
RMSE (mg L^{-1})	DO history	0.145 / 0.150	0.216 / 0.210	0.253 / 0.260	0.312 / 0.305
	Hour26	0.363 / 0.359	0.365 / 0.361	0.365 / 0.362	0.358 / 0.352
Skill vs. persistence (%)	DO history	18.6 / 15.7	23.7 / 25.7	30.6 / 28.5	36.7 / 38.3
	DO + T	16.8 / 14.3	26.6 / 25.4	30.3 / 32.7	38.6 / 38.5
	DO + pH	26.0 / 22.7	37.0 / 37.3	45.2 / 46.7	41.6 / 42.1
	DO + Turb	29.8 / 25.6	37.8 / 39.2	42.9 / 42.3	39.5 / 39.3
	DO + Cond	28.9 / 23.0	42.4 / 40.7	48.0 / 52.0	18.1 / 19.4
	Five channels	31.1 / 24.7	40.1 / 38.1	45.4 / 50.3	25.4 / 27.4

Quantity	Inputs	1 h	2 h	3 h	6 h
Rank agreement of the six configurations	Spearman ρ, preliminary vs. final skill	0.94	0.89	0.94	1.00
Median (maximum) absolute skill difference, points	3.8 (6.4)	1.5 (2.0)	2.2 (4.9)	0.9 (2.0)	
Three-hour low-DO ROC-AUC (DO < 6.5 mg L^{-1})	DO history	—	—	0.929 / 0.932	—

Each cell gives preliminary / final. Skill = 100 × (1 − $RMSE_{model}/RMSE_{persistence}$). DO history includes the four calendar terms in both designs ("DO with calendar features" in the preliminary analysis), where Hour26 was a tuned time-of-day control. Preliminary test origins: 3,595 at 3 h (prevalence 46.6%); final: 3,559, 3,542, 3,529 and 3,497 at 1, 2, 3 and 6 h (3 h prevalence 46.9%). Differences reflect the joint change in models, grids, partitions and fitting periods (Table S7), not a controlled comparison of single factors.