

Table S1. Preserved PI variables and intervention values

Variable	Role	Frozen value/relationship
S _{o,ref}	Fixed manipulated reference	2.00; 1.90; 1.80 mg O ₂ /L
S _{o,4,m}	Measured DO in reactor 4	Preserved sensor dynamics/state
K _p	Proportional gain	25 m ³ g(COD) ⁻¹ d ⁻¹
T _i	Integral time	0.002 d
T _t	Anti-windup tracking time	0.001 d
KLa ₄	PI output	0–360 d ⁻¹
KLa ₃ ; KLa ₅	Coordinated aeration	KLa ₃ = KLa ₄ ; KLa ₅ = 0.5KLa ₄

Table S2. Production–rerun numerical comparison and audit controls

Control	Scope	Result
Tier A runs and trajectory sets	126 productions + 126 reruns = 252 runs / 252 trajectory sets	42 checkpoints × 3 references; one 48 h trajectory set per run; 126 production–rerun pairs
Tier A numeric variables	49 variables per run; 6174 variable-pair comparisons	42 time-indexed arrays + 7 scalar constants; all 49 compared in each of 126 pairs; exact numeric equality
Tier A final states	126 production + 126 rerun final states = 252 objects	126 native final-state pairs passed the archived MATLAB check. New selected replay comparisons inspect the serialized HDF5 representation separately.
Tier A horizon metrics	252 rows per phase; 504 across production and rerun	Each 48 h run supplies a 24 h and a 48 h assessment; these are not separate simulated trajectories
Historical metric reconstruction	Comparison under the historical post-processing formulas	Maximum discrepancies: metrics 4.91×10^{-11} ; paired deltas 4.83×10^{-13} ; rule components 4.97×10^{-14} . These do not validate the corrected BOD ₅ formula
Annual runs and trajectory sets	3 productions + 3 reruns = 6 runs / 6 trajectory sets	3 production–rerun pairs; each run has 34,945 timestamps and 34,944 intervals
Annual numeric variables	49 variables per run; 147 variable-pair comparisons	42 time-indexed arrays + 7 scalar constants in each of three archived production–rerun comparisons
Annual final states	3 production + 3 rerun final states = 6 objects	3 native final-state pairs checked separately from numeric arrays in the historical MATLAB procedure
Totals across tiers	129 productions + 129 reruns = 258 trajectory sets	258 final-state objects; 129 production–rerun pairs. These are computational counts, not independent biological replicates
Canonical model	Before and after executions	Recorded SHA-256 identity; the local SLX is also identical to the public BSM2_vR2019b release
Revision calculations	252 Tier-A metric rows; Tables 5 and 8; failures; sensitivity; KLa	8064 Tier-A numeric values checked; 48/48 Table 5 and 75/75 Table 8 entries at displayed precision; 21 NH ₄ + 2 BOD ₅ failures; all 720 NH ₄ outcomes and 756 KLa records checked
OCI	All branches	Not evaluated
Revision portable preflight	129 configurations; no simulation	126 Tier A + 3 annual configurations passed on the tested computational payload
Selected portable executions	N001_DO200 twice; N001_DO190 once; ANNUAL_DO200 once	Four executions on three existing cases; separate from 258 historical runs
Selected raw-output comparison	49 numeric fields per execution; final-state storage and RNG inspected separately	190,498 numeric values per short run; 34,490,722 for annual DO200; exact agreement with the stated references
Annual interval comparison	12/14 exact columns in the native comparison	Historical operation grouping gives 14/14 exact columns from the same raw output; both records retained

A trajectory set contains the stored signals from one run. The 49 numeric fields comprise 47 model outputs, tout and DO_ref_applied_mg_L: 42 time-indexed arrays and seven scalar constants (kla1in, kla2in and carbon1in–carbon5in). The DO-reference field is assigned from the case definition, not independently measured. ANNUAL_VARIABLES_49.csv lists names and shapes. Final states are stored separately. Historical MATLAB comparisons, revision Python recalculation and selected portable replays are distinct evidence. The latter comprise four executions on three cases and do not change the historical counts above. New final-state comparisons inspect stored HDF5 nodes, attributes and reference targets; they are not new native MATLAB isequaln tests. Replay logs report that initial solver state was discarded although block states were restored.

Table S3. Complete register of the 21 NH₄ excursion-rule failures

Anch or	Action	Horizon (h)	NH ₄ mean cand./PI (mg N/L)	Duration cand./PI (h)	Episodes cand./PI	Area cand./PI (mg N·h/L)	Peak excess cand./PI (mg N/L)	New	Class
N023	DO190	24	1.562/1.508	3.00/3.00	1/1	2.693/2.322	1.369/1.244	No	Material
N023	DO180	24	1.622/1.508	3.25/3.00	1/1	3.108/2.322	1.500/1.244	No	Material
N023	DO190	48	1.307/1.254	3.00/3.00	1/1	2.693/2.322	1.369/1.244	No	Material
N023	DO180	48	1.366/1.254	3.25/3.00	1/1	3.108/2.322	1.500/1.244	No	Material
N028	DO190	24	1.495/1.442	2.75/2.25	1/1	1.004/0.821	0.597/0.524	No	Material
N028	DO180	24	1.555/1.442	2.75/2.25	1/1	1.217/0.821	0.675/0.524	No	Material
N028	DO190	48	1.095/1.051	2.75/2.25	1/1	1.004/0.821	0.597/0.524	No	Material
N028	DO180	48	1.145/1.051	2.75/2.25	1/1	1.217/0.821	0.675/0.524	No	Material
N030	DO190	24	3.414/3.325	7.00/6.75	1/1	18.474/17.725	4.463/4.371	No	Small, rule-based
N030	DO180	24	3.511/3.325	7.25/6.75	1/1	19.279/17.725	4.562/4.371	No	Material
N030	DO190	48	2.480/2.404	7.00/6.75	1/1	18.474/17.725	4.463/4.371	No	Small, rule-based
N030	DO180	48	2.563/2.404	7.25/6.75	1/1	19.279/17.725	4.562/4.371	No	Material
N033	DO190	24	1.969/1.890	5.25/5.00	1/1	8.323/7.490	2.413/2.249	No	Material
N033	DO180	24	2.058/1.890	5.25/5.00	1/1	9.226/7.490	2.587/2.249	No	Material
N033	DO190	48	1.542/1.471	5.25/5.00	1/1	8.323/7.490	2.413/2.249	No	Material
N033	DO180	48	1.621/1.471	5.25/5.00	1/1	9.226/7.490	2.587/2.249	No	Material
N034	DO190	48	1.024/0.976	1.25/0.00	1/0	0.066/0.000	0.090/0.000	Yes	New exceedance
N034	DO180	48	1.079/0.976	1.75/0.00	1/0	0.299/0.000	0.246/0.000	Yes	New exceedance
N035	DO190	48	1.344/1.273	2.50/2.25	1/1	1.185/0.778	0.692/0.527	No	Material
N035	DO180	48	1.424/1.273	3.00/2.25	1/1	1.666/0.778	0.874/0.527	No	Material
N037	DO180	48	1.166/1.051	1.50/0.00	1/0	0.134/0.000	0.139/0.000	Yes	New exceedance

Values are candidate/matched PI2. Area is mg N·h/L; peak excess is mg N/L. The complete 1260-row five-pollutant matrix and the 21-row CSV are supplied in the data package. These 21 comparison rows involve seven distinct anchors. The Class column retains historical descriptive labels; it does not define pass/fail or establish operational severity. The five-pollutant matrix comprises 420 PI2 self-comparisons and 840 candidate comparisons; the NH₄ candidate subset has 168 rows. Two additional BOD₅ failures at N022 over 48 h are reported separately in Table S10. The complete candidate set therefore has 23 failed pairs across eight anchors.

Table S4. Mean absolute 48 h paired changes relative to matched PI2

Outcome	DO190 mean Δ	Direction	DO180 mean Δ	Direction
AE	−61.542 kWh/d	42/42 lower	−121.956 kWh/d	42/42 lower
NH ₄ mean	+0.027565 mg N/L	42/42 higher	+0.058909 mg N/L	42/42 higher
TN mean	−0.1559996 mg N/L	42/42 lower	−0.317919 mg N/L	42/42 lower
EQI (kg pollution units/d)	−18.924	38/42 lower	−37.276	38/42 lower
COD mean	+0.002294 mg/L	42/42 higher	+0.004859 mg/L	42/42 higher
BOD ₅ mean	+0.000592 mg/L	42/42 higher	+0.001254 mg/L	42/42 higher

Table S5. Annual relative changes from DO200

Indicator	DO190 vs DO200	DO180 vs DO200
AE	−1.4439%	−2.8568%
EQI	−0.6494%	−1.2840%
NH ₄ mean	+5.7157%	+12.2972%
NH ₄ maximum	+2.0821%	+4.3133%
Duration >4	+16.0839%	+36.3636%
Episodes >4	+18.1818%	+27.2727%
Excess area	+14.5041%	+32.6545%
TN mean	−1.7124%	−3.4580%

Table S6. Sensitivity of NH₄ excursion-rule failures to declared comparison margins

Scenario	Duration margin (h)	Episode margin	Area / peak margins	Failures /168	Unique anchors
Exact, no allowance	0.0	0	0.0 / 0.0	21	7
Study rule	0.25	1	0.0 / 0.0	21	7
Temporal-only relaxation	1.0	2	0.0 / 0.0	21	7
Moderate materiality	0.5	1	0.25 / 0.1	19	7
Broadest declared	1.0	2	1.0 / 0.25	10	6

Area and peak margins are in mg N·h/L and mg N/L, respectively. Numerical tolerances are 0 for the exact rule, 1.4×10^{-8} for the study rule, and 10^{-6} for the other listed scenarios. The complete 720-scenario grid is supplied as CSV. The primary rule is post hoc. The dated protocol in the authors' archive concerns additional analyses of existing trajectories, not prospective registration of the study. All 720 classifications were recalculated during this revision; the count ranges from 10 to 21 and is not a regulatory or universal risk measure.

Table S7. Short-horizon KLa₄ maxima and saturation counts on the 15 min reporting grid across 42 contexts

Setpoint / horizon	Global max (d ⁻¹)	Median branch max (d ⁻¹)	IQR (d ⁻¹)	At 360	≥356.4
DO200 / 24 h	233.119	181.844	174.930–200.459	0/42	0/42
DO200 / 48 h	236.920	184.254	176.549–202.297	0/42	0/42
DO190 / 24 h	228.607	178.370	171.705–196.220	0/42	0/42
DO190 / 48 h	232.414	180.765	172.947–198.411	0/42	0/42
DO180 / 24 h	224.176	174.877	168.431–192.012	0/42	0/42
DO180 / 48 h	227.978	177.050	169.370–194.546	0/42	0/42

The complete 756-row register is supplied as CSV. Maxima and saturation counts describe the stored reporting points; they do not exclude brief saturation between those points.

Table S8. Inventory of the corrected package prepared for this revision

Component	Exact files or directory	Purpose
Documentation	README.md; DATA_DICTIONARY.md; CHANGELOG.md; LICENSE.md	Scope, commands, units, correction history and licence
Corrected metrics	data/short_horizon/P3_INDEPENDENT_METRICS_252_V1_1.csv	252 branch-horizon rows; corrected BOD ₅ and EQI at N022 / 48 h
Corrected deltas	data/short_horizon/P3_INDEPENDENT_PAISED_DELTAS_252_V1_1.csv	252 differences from matched PI2
Corrected event matrix	data/short_horizon/P3_INDEPENDENT_NON_AGGRAVATION_1260_V1_1.csv	1260 pollutant rows; 23 candidate failures
Failure registers	data/short_horizon/TABLE_NH4_RULE_FAILURES_21.csv; TABLE_BOD5_RULE_FAILURES_2.csv; TABLE_ALL_RULE_FAILURES_23.csv	NH ₄ -only, BOD ₅ -only and combined failed subsets; all in data/short_horizon/
Historical Tier A audits	data/short_horizon/historical_audits/	Unchanged execution, identity, time-grid and reconstruction registers; historical formula scope
Annual results and audits	data/annual/	ANNUAL_FIXED_DO_METRICS.csv, annual relative changes, DO200/190/180 metrics, identities and event masks
Sensitivity and oxygen transfer	data/sensitivity/TABLE_NH4_SENSITIVITY_720.csv; TABLE_KLA_BRANCH_SATURATION_756.csv	Full grids with summaries in the same directory; unchanged
Plotting data	data/figure_sources/	FIGURE4_PAISED_48H_AE_NH4.csv; FIGURE8_ANNUAL_FIXED_DO_DOSE_RESPONSE.csv
Analysis and integrity	scripts/analyze_bsm2.py; scripts/validate_deposit.py	Recalculate 720 NH ₄ outcomes and 18 KLa summaries; validate hashes and row counts
Revision interval data	verification/TIER_A_INTERVALS.csv; verification/annual_intervals/	24,192 Tier-A intervals; three annual files with 34,944 intervals each
Revision verification	verification/verify_tier_a_intervals.py; verify_derived.py; verify_annual_intervals.py	Three standard-library verification scripts in verification/; expected tables, reports and provenance accompany them
Supplement and hashes	supplementary/BSM2_EEET_Supplement_REVISED.docx; manifests/DEPOSIT_SHA256_MANIFEST.txt	Tables S1–S10; manifest covers every other package file
Replay entry points	BSM2_ZENODO_REPLAY_20260915.zip: D03_PORTABLE_REPLAY_V1/	D03_PREFLIGHT_PORTABLE_V1.m; D03_REPLAY_CASE_V1.m; CASES_129.json
Runtime and checkpoints	payload/runtime/; payload/checkpoints/	Recorded model, scripts, inputs and 25 Windows MEX files; 42 Tier-A and one annual checkpoint
Selected replay evidence	evidence/; raw_comparisons/	Dated note, comparison records, native-log extracts, four output files and historical references for three cases
Updated provenance	provenance/PROVENANCE_UPDATE_20260915.json; SS_PROVENANCE_NOTE.md	State-generation chain, identities and limits; copies in verification/provenance/ of the derived-data archive

In Table S8, data and verification paths are relative to BSM2_DATA/; replay paths are relative to D03_PORTABLE_REPLAY_V1/. The data archive contains corrected metrics, separate NH₄ and BOD₅ failure registers, and interval checks. Historical audit files retain their original processing scope. The replay archive contains the recorded runtime files, 43 serialized checkpoints, entry points, saved outputs and historical references. Its computational files match those tested; the distribution requires a preflight against its own manifest before use.

Table S9. Source provenance and external reproduction requirements

Asset	Identification	Scope of verification
Public implementation	IWA task group; release BSM2_vR2019b; asset BSM2_R2019b.zip	https://github.com/wwtmodels/Benchmark-Simulation-Models/releases/tag/BSM2_vR2019b
bsm2_cl.slx	48,354 bytes; SHA-256 in SOURCE_PROVENANCE.json	Local file is byte-identical to the public release file; recorded Simulink model version 1.472
init_bsm2.m	917 bytes; SHA-256 in SOURCE_PROVENANCE.json	Recorded local identity equals the public release file
Local states_bsm2.mat	3,089 bytes; actual SHA-256 matches the recorded run identity	Matches the reference saved on 14 August 2026; native history and log trace bsm2_ss → stateset_bsm2 → timestamped copy. Release comparison: 14/17 variables and 275 scalars differ.
Influent and sensor noise	DYNINFLUENT_BSM2 and sensornoise_bsm2.mat; actual noise identity verified	Local SENSORNOISEFULL: 876,961 × 26 values, days 0–609. Its 25 noise channels repeat the public first 20,160 samples exactly; the day-14 endpoint is excluded from each cycle
Tier A checkpoints	42 identifiers and SHA-256 values in CHECKPOINT_IDENTITIES_42.csv	42 serialized operating points supplied in replay payload/checkpoints/tier_a/; identities retained for the three branches at each anchor
Annual checkpoint	Day 245; SHA-256 in annual_raw_recalculation.json	Serialized day-245 operating point supplied in replay payload/checkpoints/annual/; common checkpoint for the three annual branches
Runtime dependencies	MATLAB R2024a, Simulink, Windows MEX S-functions and model inputs	Recorded Windows MEX binaries supplied. Of 45 asset identities, 18 match the release and two MAT inputs differ; 25 local MEX binaries have no binary counterpart in the release. Historical compiler build remains unrecorded.

SOURCE_PROVENANCE.json and MAT_INPUT_REPRODUCIBLE_CHECK.json retain the release and local-input comparisons. The dated PROVENANCE_UPDATE_20260915.json and SS_PROVENANCE_NOTE.md add the recovered state-generation record without rewriting earlier audits. Native MATLAB history, the state-value log and the saved MAT identity connect the local initialization file to bsm2_ss and stateset_bsm2. This process-variable file differs from the 43 complete dynamic operating points supplied for replay. The historical compiler build and actual SS integration horizon are not established by those records; no value is inferred for either.

Table S10. Corrected BOD₅ comparison at N022 over 48 h

Reference	Mean (mg/L)	Maximum (mg/L)	Duration >10 (h)	Episodes	Excess area (mg·h/L)	Peak excess (mg/L)	Rule result
PI2 / DO200	4.105240	10.229631	0.25	1	0.057407762	0.229631	Reference
DO190	4.105853	10.229968	0.25	1	0.057492064	0.229968	Fail
DO180	4.106532	10.230335	0.25	1	0.057583697	0.230335	Fail

With the corrected BOD₅ calculation, each reference exceeds 10 mg/L for one interval. Relative to corrected PI2, DO190 and DO180 increase peak excess by 0.000337208 and 0.000703741 mg/L and integrated excess by 0.000084302 and 0.000175935 mg·h/L, respectively. Both fail the area/peak conditions beyond the BOD₅ numerical tolerance of 2.0 × 10⁻⁸. These small paired increases do not establish operational significance. Correcting the bypass contribution also raises the three EQI values by 3.679094 kg pollution units/d; Table 5 summaries are unchanged at displayed precision.