Reduced Order Models for WRRF Operation - DARROW Case Study

Presented by: Cesar Arenas

PhD. Candidate

MEMBER OF
BASQUE RESEARCH
& TECHNOLOGY ALLIANCE

Welcome & Agenda

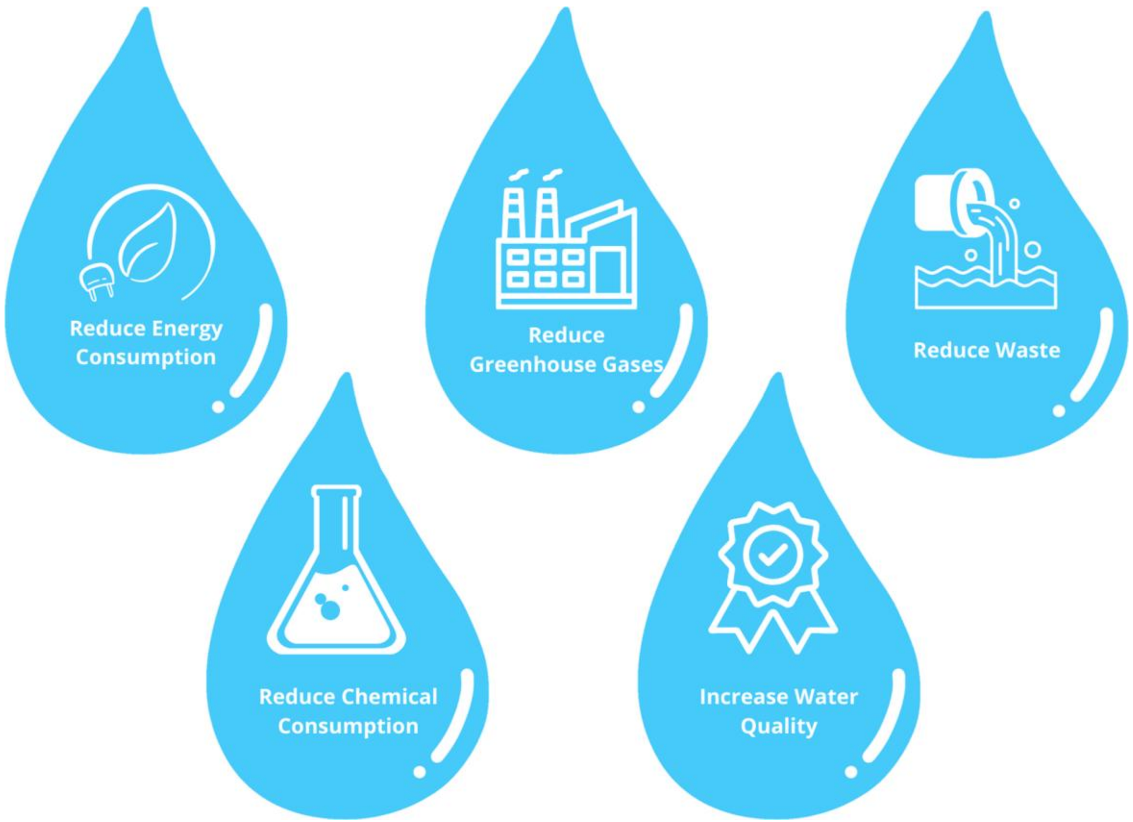
1. **Operational challenge in WRRFs**
2. **Motivation and Background**
3. **Proposed ROM-based framework**
4. **Water-line ROM**
    1. Data generation and dataset design
    2. Architecture selection
    3. Forecasting performance
5. **Sludge-line ROM**
    1. Real SCADA data
    2. Methane forecasting
6. **Decision Support System**
    1. Multi-objective optimization
    2. Day-ahead control policies
7. **Contributions**
8. **References**

Aerial view of DARROW WWTP (Tilburg, Netherlands)

DARROW

ceit
MEMBER OF
BASQUE RESEARCH
& TECHNOLOGY ALLIANCE

Motivation & Background



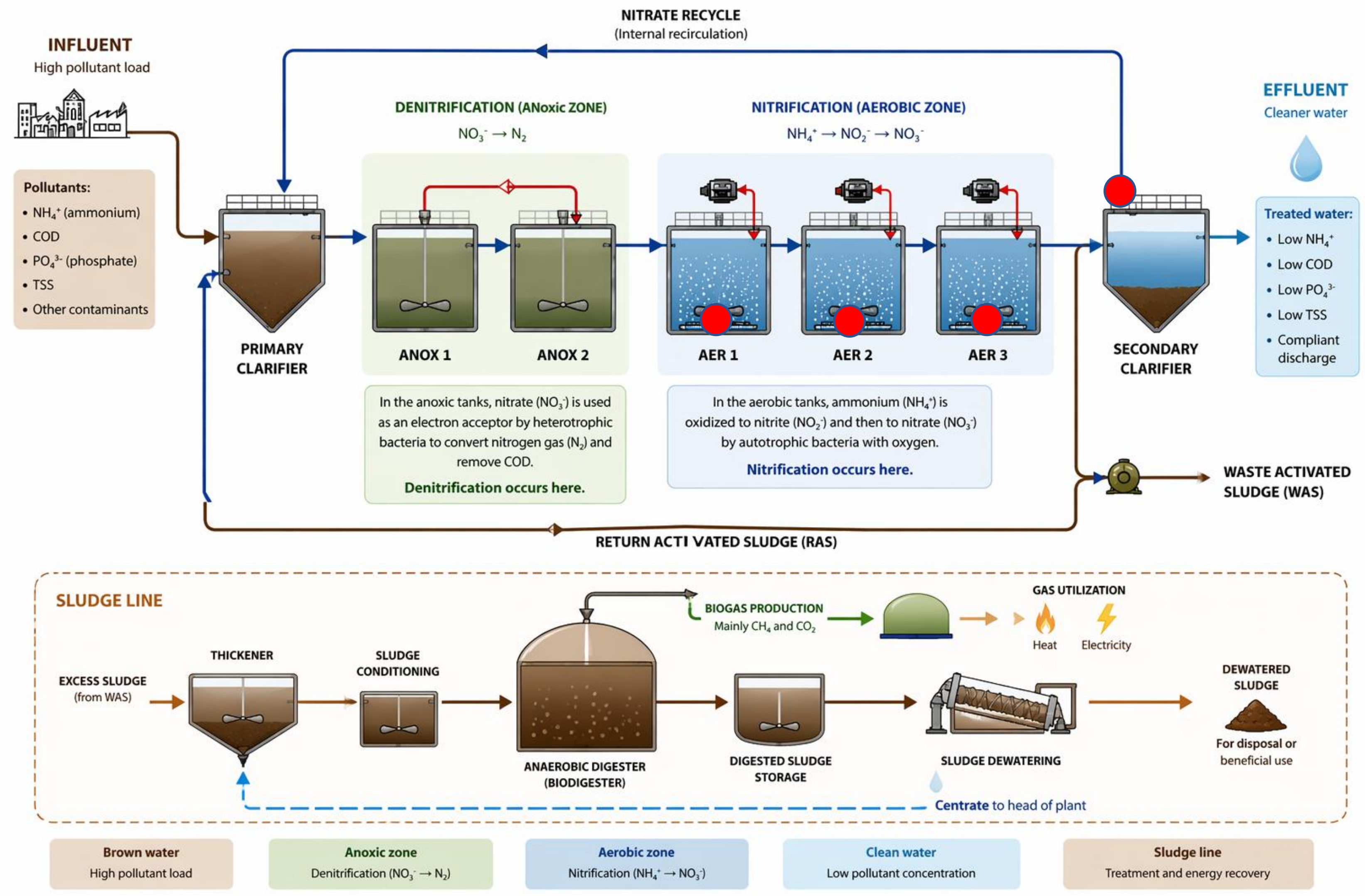
This work has been Funded by the European Union's Horizon Europe research and innovation programme within the project DARROW (grant agreement Nº 101070080).

DARROW

ceit

MEMBER OF
BASQUE RESEARCH
& TECHNOLOGY ALLIANCE

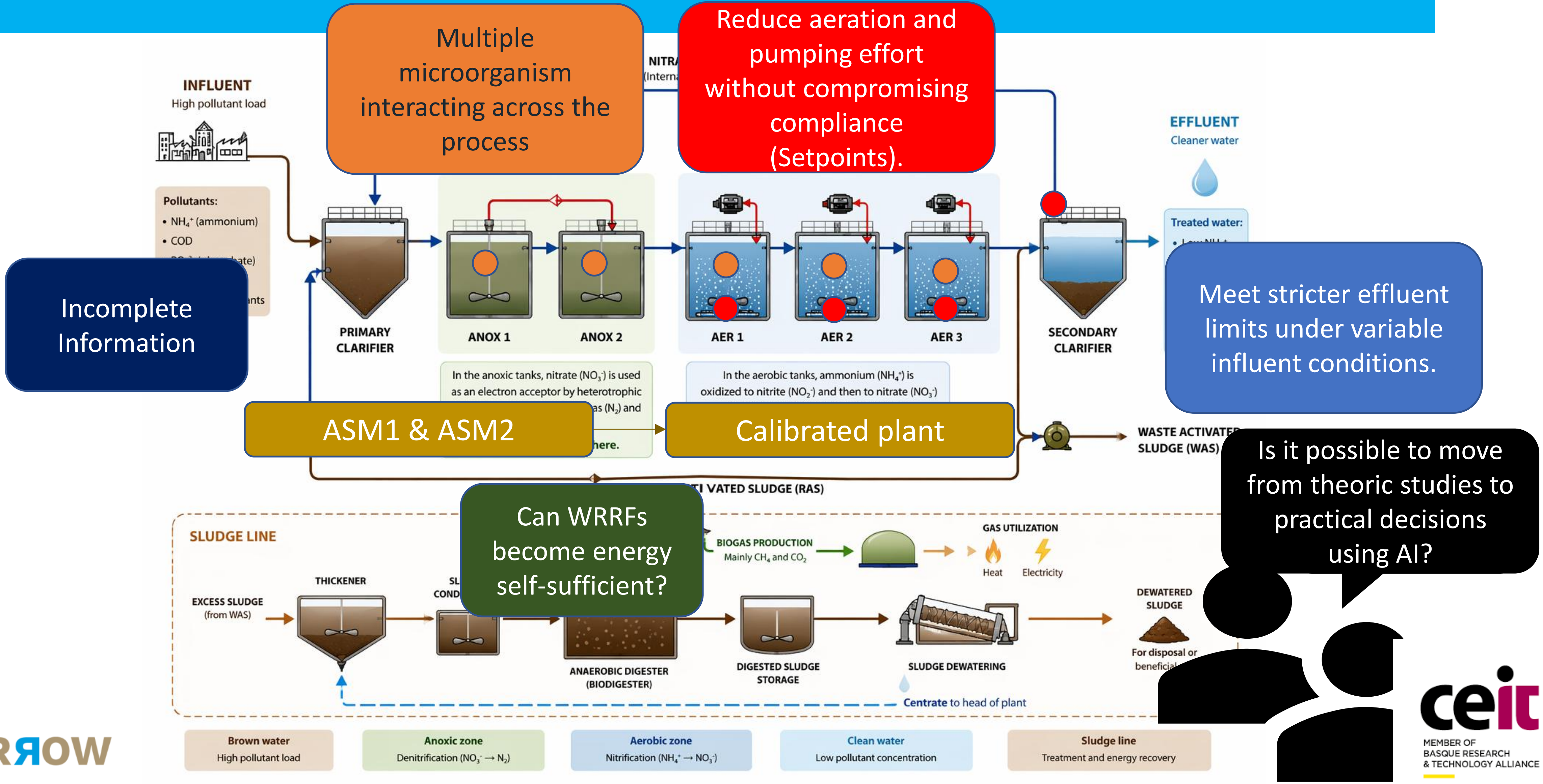
1.- Operational Challenge in WRRFs



DARROW

ceit
MEMBER OF
BASQUE RESEARCH
& TECHNOLOGY ALLIANCE

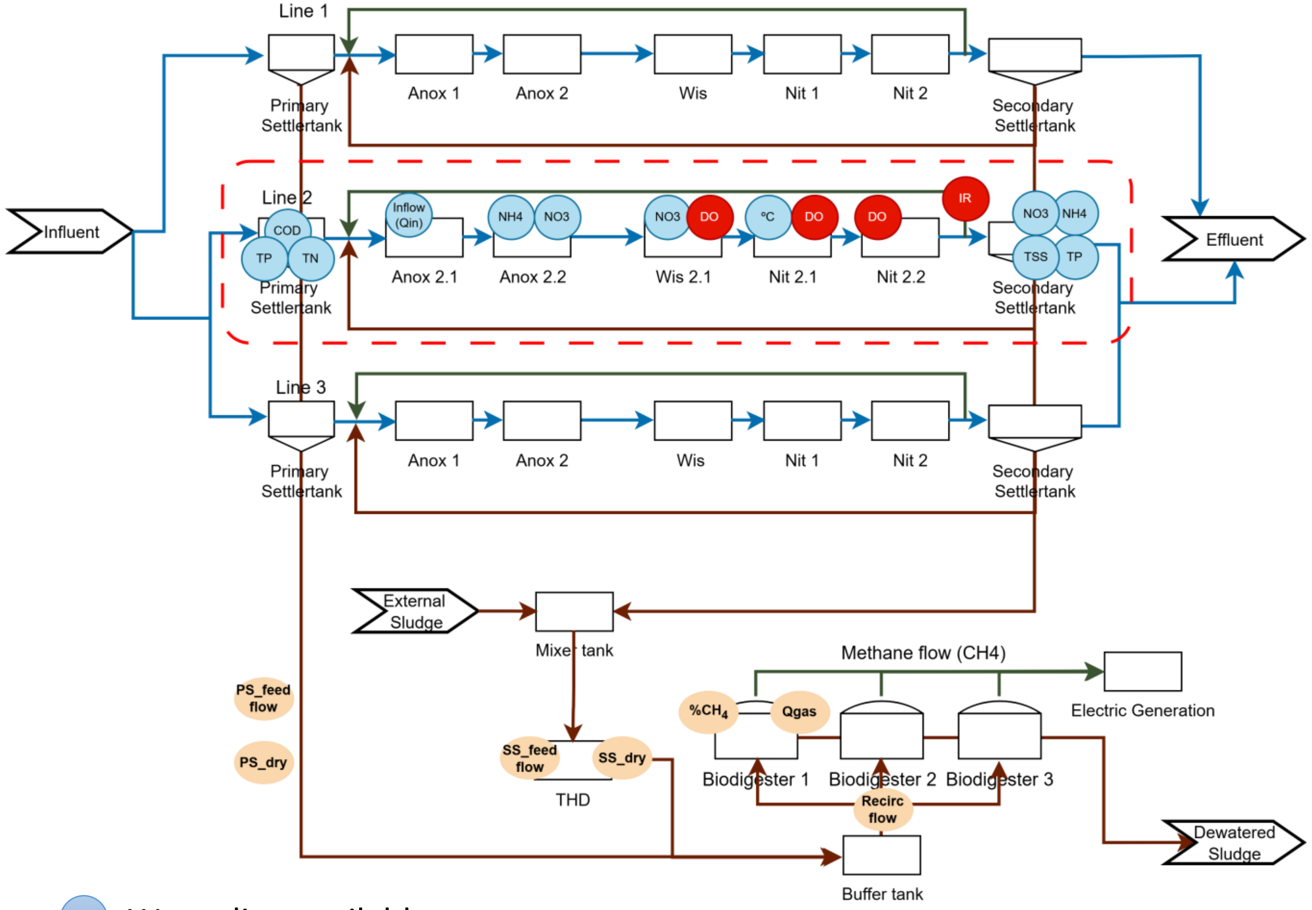
1.- Operational Challenge in WRRFs



DARROW

ceit
MEMBER OF BASQUE RESEARCH & TECHNOLOGY ALLIANCE

1.- Operational Challenge in WRRFs



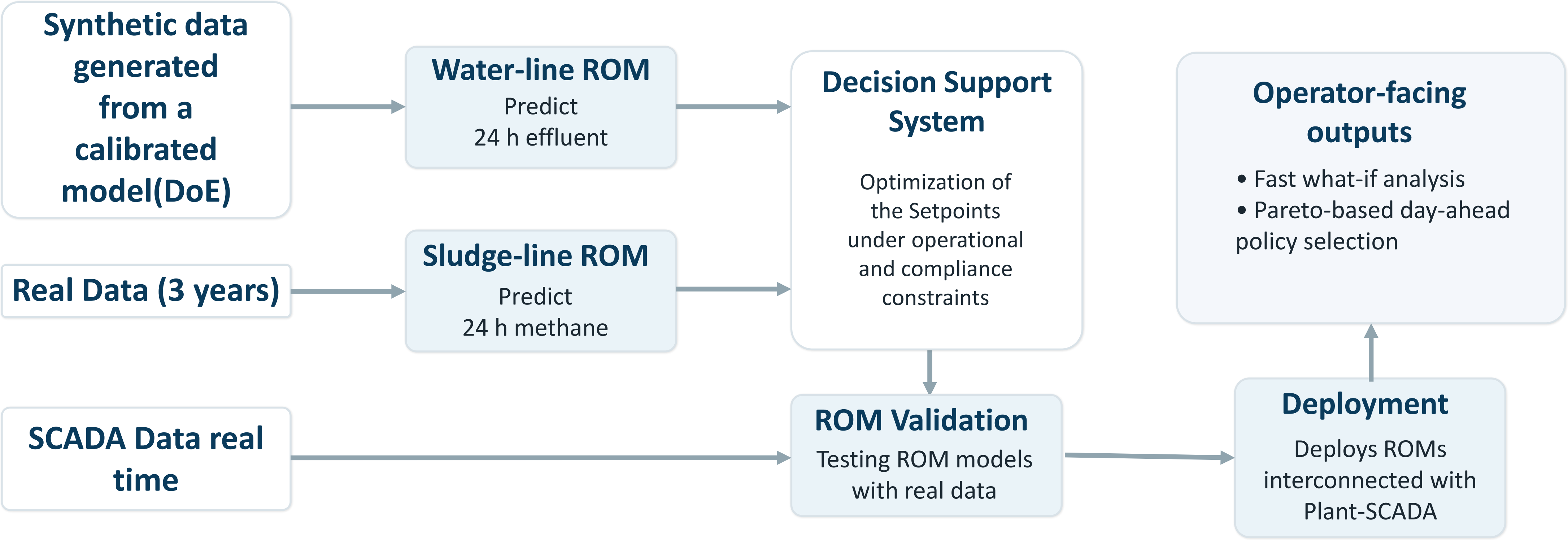
- Water-line available sensors
- Water-line manipulated variables
- sludge-line variables

Block	Water-line Variables	Unit
Input	Influent – Inflow	m^3/h
Input	Influent – COD	$gCOD/m^3$
Input	Influent – TN	gN/m^3
Input	Influent – TP	gP/m^3
Input	Anoxic Tank 2-NH_4	gN/m^3
Input	Anoxic Tank 2-NO_3	gN/m^3
Input	Wis Tank -NO_3	gN/m^3
Input	Nitrification Tank 2 -Temperature	ºC
Input	Wis Tank - Oxygen Setpoint	gO_2/m^3
Input	Nitrification Tank 1 - Oxygen Setpoint	gO_2/m^3
Input	Nitrification Tank 2 - Oxygen Setpoint	gO_2/m^3
Input	Internal Recirculation rate	%
Input	Biomass-Heterotrophs	g/ m^3
Input	Biomass-Autotrophs	g/ m^3
Input	Biomass-PAO	g/ m^3
Output	Aeration cost	M€
Output	Sludge production	Kg/h
Output	Effluent – COD	$gCOD/ m^3$
Output	Effluent - NH_4	gN/ m^3
Output	Effluent - NO_3	gN/ m^3
Output	Effluent - TN	gN/ m^3
Output	Effluent - PO_4	gP/ m^3
Output	Effluent – TP	gP/ m^3
Output	Effluent - TSS	g/ m^3

Block	Sludge-line Variables	Unit
Input	Primary sludge dry solids (PS_dry)	%
Input	Primary sludge feed flow to THD (PS_feed_Flow)	m^3/h
Input	Hydrolyzed sludge dry solids (SS_dry)	%
Input	Hydraulic sludge feed pump flow (SS_feed_Flow)	m^3/h
Input	Recirculation pump flow (Recirc_Flow)	m^3/h
Input	Biogas flow rate (Biogas_Flow)	m^3/h
Input	Methane content (Biogas_quality)	% CH_4
Output	Content of Methane (Methane Flow)	$m^3.CH_4/h$

MEMBER OF BASQUE RESEARCH & TECHNOLOGY ALLIANCE

2.- Proposed ROM-based framework

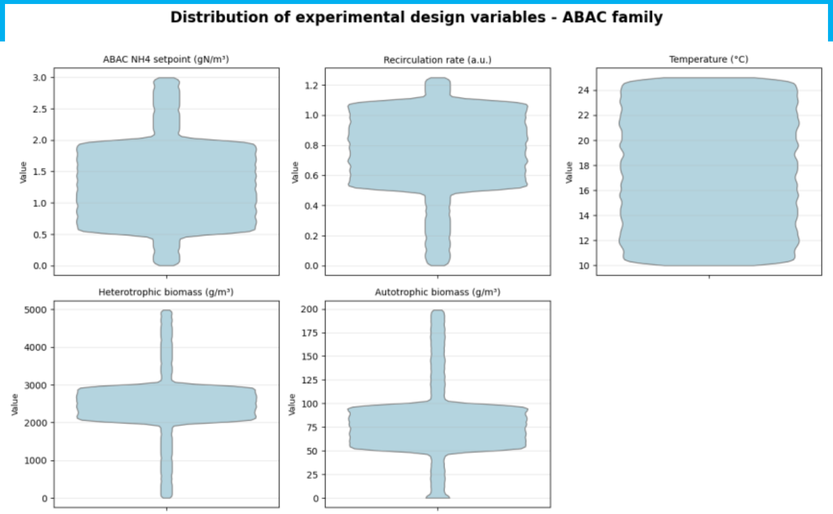


DARЯOW

3. Water-line ROM: Data generation and dataset design

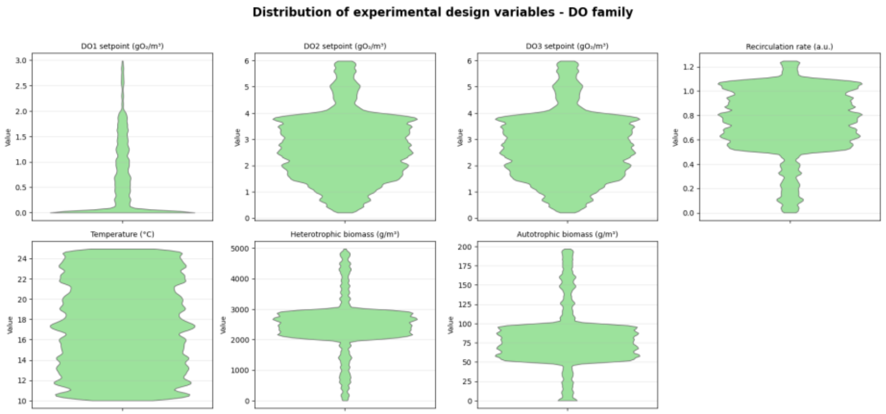
1000 three-month scenarios - sampling ranges for the two-regime ABAC design (75% typical, 25% extreme).

Variable	Typical (75%)	Extreme (25%)
NH4 setpoint	[0.5, 2]	[0, 3]
Recirculation setpoint (–)	[5, 10]	[0, 15]
Temperature (◦C)	[10, 25]	[10, 25]
XH	[2000, 3000]	[10^{-6}, 5000]
$XAUT$	[50, 100]	[10^{-6}, 200]



1000 three-month scenarios - sampling ranges for the two-regime fixed DO setpoint design (75% typical, 25% extreme).

Variable	Typical (75%)	Extreme (25%)
DO setpoint WIS	[0, 2]	[0, 3]
DO setpoint NIT1 & NIT2	[0.2, 4]	[0, 5]
Recirculation setpoint (–)	[5, 10]	[0, 15]
Temperature (◦C)	[10, 25]	[10, 25]
XH	[2000, 3000]	[10^{-6}, 5000]
$XAUT$	[50, 100]	[10^{-6}, 200]

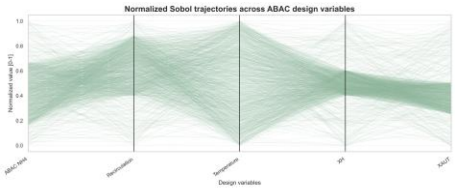


DARROW

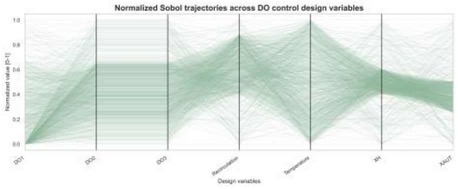
ceit
MEMBER OF
BASQUE RESEARCH
& TECHNOLOGY ALLIANCE

3. Water-line ROM: Data generation and dataset design

DoE Generation



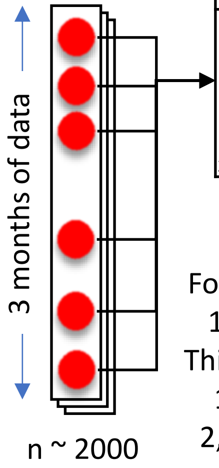
1000 three-month ABAC scenarios



1000 three-month DO scenarios

2000 scenarios of 3 months
each with an hourly frequency

Block Extraction



T	Qin	COD	NH4	NO3	…	XPAO
1	9	16	20	32	…	5
2	2	10	25	41	…	8
3	5	11	20	30	…	6
4	9	18	22	33	…	6
5	9	20	25	43	…	5
….	8	11	25	45	…	8
2400	6	12	22	33	…	5

…

For each simulation, six random 10-day blocks were selected. This resulted in a final dataset of 12,000 blocks, computed as 2,000 scenarios × 6 blocks per scenario, equivalent to **approximately 32 years of hourly data.**

Dataset Modelling

Time	Block	W_Id	Qin	NH4	…	EFL_TSS
1	a	1	9	20	…	2
2	a	1	2	25	…	8
3	a	1	5	20	…	9
…	a		4	30	…	7
240	a	1	6	25	…	2
1	a	2	1	29	…	4
2		2	6	27	…	9
3	a	2	6	20	…	2
…	a	2	5	23	…	9
240	a	2	8	28	…	4
			…			
1	n	1	11	26	…	6
2	n	9	20	21	…	1
3	n	8	10	22	…	3
…	n	7	14	30	…	2
240	n	8	18	23	…	4

The extracted **blocks were concatenated into a single dataset**. A cleaning step was then applied to remove potential outliers, infeasible values, and negative values that could arise from unrealistic operating combinations

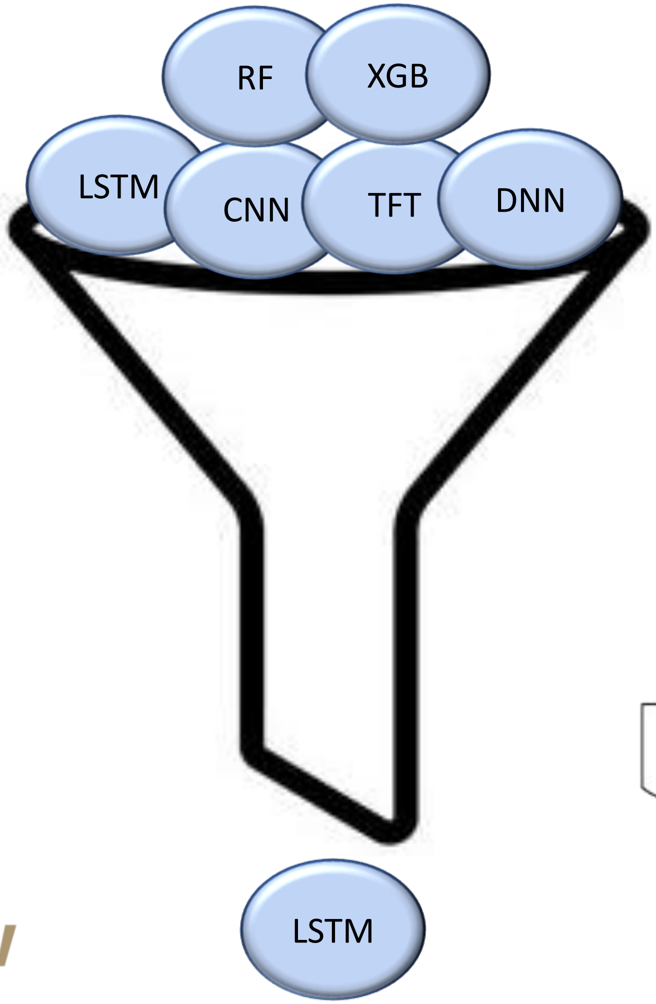
Noise Addition

Sensor acquisition **uncertainty was incorporated** through relative measurement errors derived from technical datasheets of representative sensor providers. In parallel**, increasing lead-time uncertainty was added to the future flow forecast**, enabling the ROM to be trained under realistic noisy operating conditions.

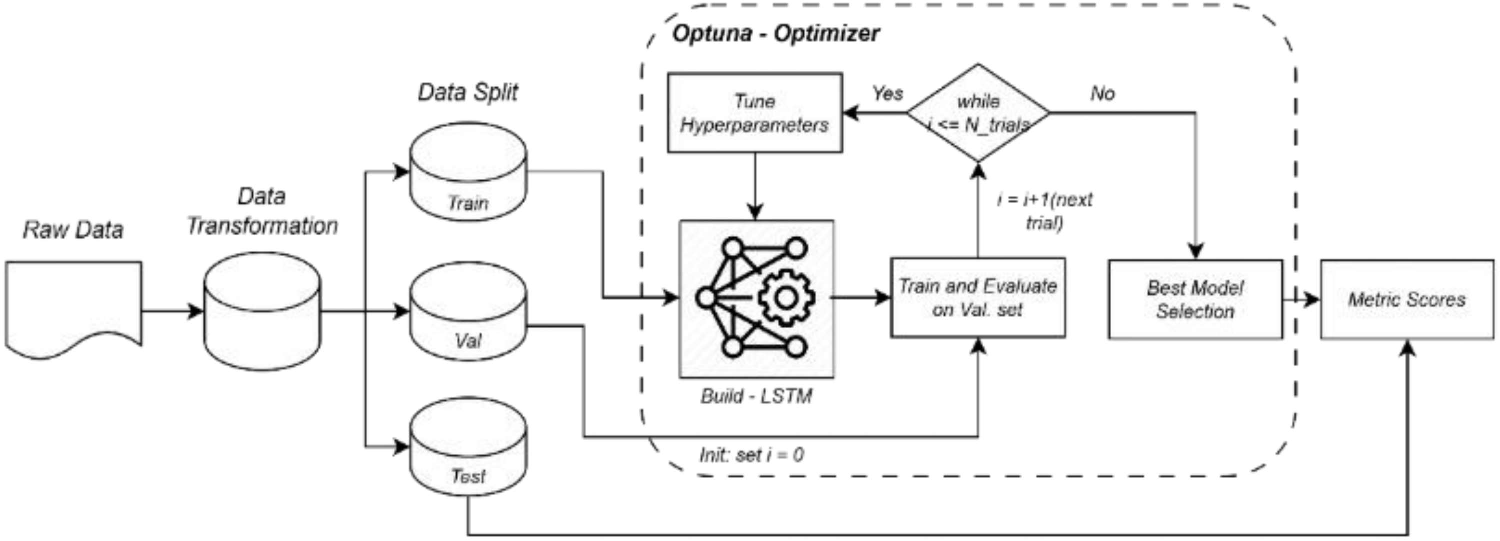
DARROW

ceit
MEMBER OF
BASQUE RESEARCH
& TECHNOLOGY ALLIANCE

3. Water-line ROM: Architecture selection



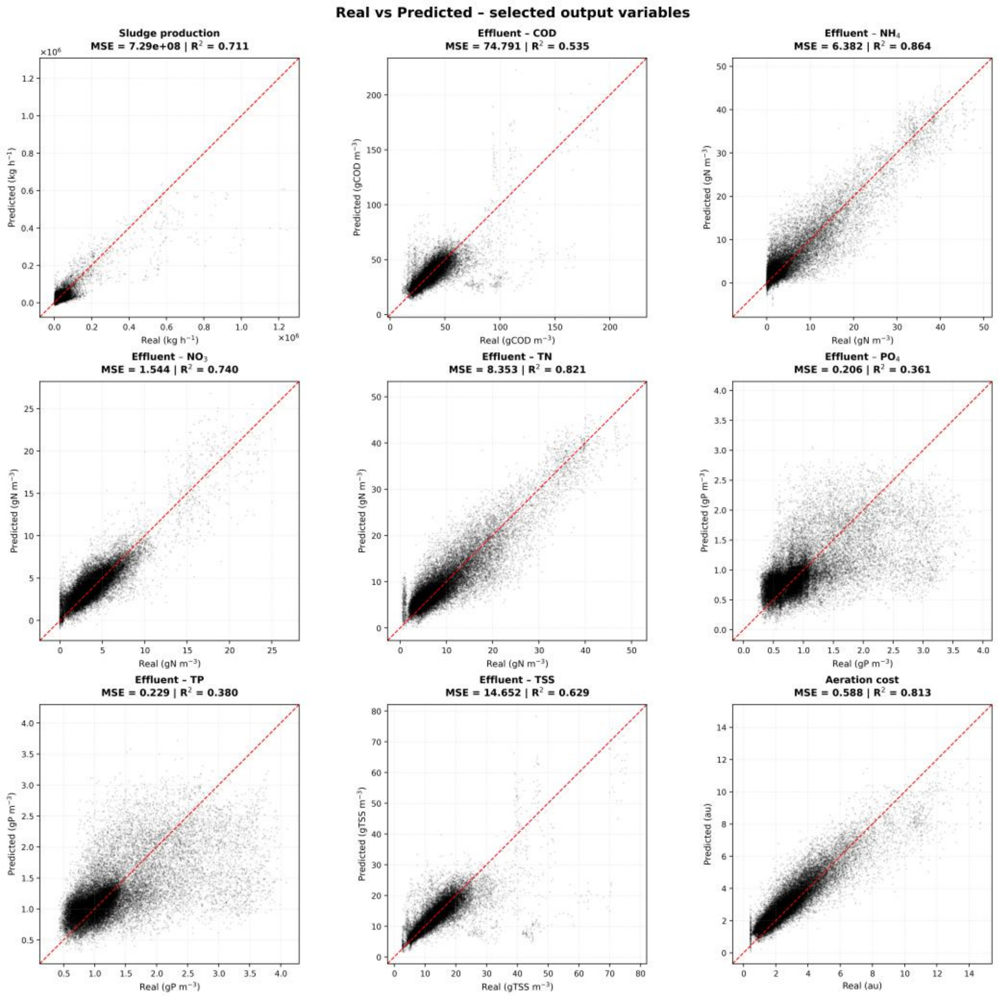
After evaluating several architectures we choose the LSTM because it offered the **best balance between accuracy, temporal consistency, and implementation feasibility for WRRF** operation. It captured the delayed effects of aeration, recirculation, and influent variations on effluent quality, while **providing a practical and deployable ROM backbone through TensorFlow/Keras.**



DARROW

3. Water-line ROM: Forecasting performance

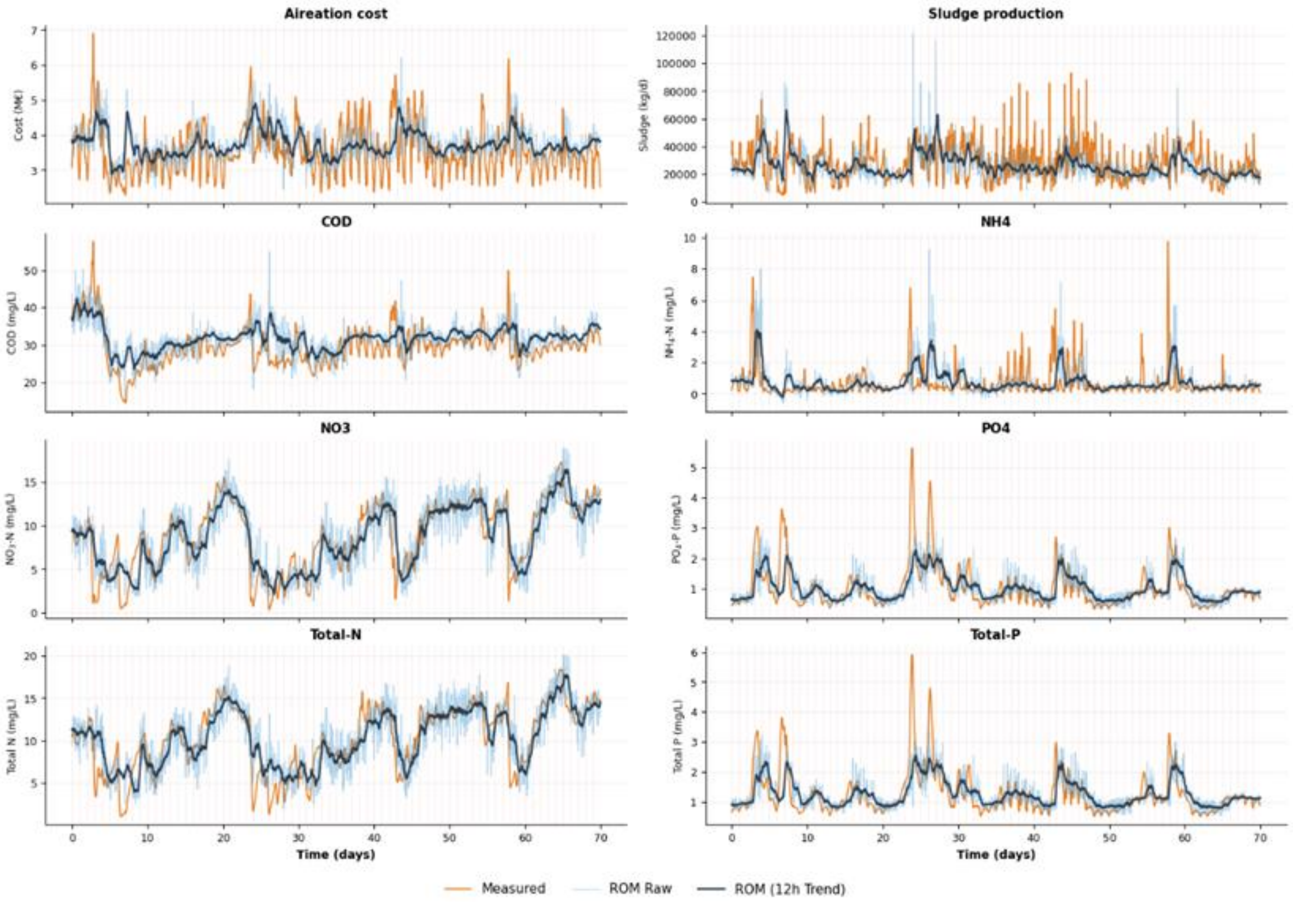
	Test			Train			Validation		
Variable	R2	MAE	RMSE	R2	MAE	RMSE	R2	MAE	RMSE
Effluent NH4 (gN/m3)	0.8641	1.5312	2.5263	0.8485	1.5441	2.5301	0.8662	1.4927	2.4824
Effluent TN (gN/m3)	0.8208	2.0446	2.8901	0.8083	2.0987	2.9327	0.8290	2.0736	2.9452
Aireation_cost_2.M	0.8134	0.5421	0.7669	0.8162	0.5291	0.7528	0.8184	0.5034	0.7053
Effluent NO3 (gN/m3)	0.7397	0.8871	1.2426	0.7693	0.9491	1.3440	0.8120	0.9364	1.3677
Prod_sludge_2	0.7114	12771.5	26994.8	0.6554	13095.3	24167.6	0.7375	13904.2	29923.7
Effluent TSS (gTSS/m3)	0.6292	1.9602	3.8278	0.6719	1.8970	3.4958	0.8334	1.8258	3.3516
Effluent COD (gCOD/m3)	0.5352	4.6891	8.6482	0.5566	4.5756	8.2777	0.8144	4.8049	10.2849
Effluent TP (gP/m3)	0.3802	0.3117	0.4786	0.3703	0.3106	0.4804	0.4901	0.3087	0.4752
Effluent PO4 (gP/m3)	0.3611	0.2954	0.4544	0.3504	0.2961	0.4559	0.3666	0.2904	0.4440



DARROW

ceit
MEMBER OF
BASQUE RESEARCH
& TECHNOLOGY ALLIANCE

3. Water-line ROM: Forecasting performance



Predictive performance of the LSTM-based ROM over a 70-day testing period. Each panel illustrates sequential 24-hour-ahead forecasts for operational costs, sludge production, and effluent quality metrics. Data obtained from mechanistic simulations (orange) are compared against raw ROM predictions (light blue) and a 12-hour moving average trend (dark blue). Red vertical lines delineate the daily simulation intervals across the evaluation timeline.

DARROW

MEMBER OF
BASQUE RESEARCH
& TECHNOLOGY ALLIANCE

4. Sludge-line ROM: Real SCADA data

Plant data foundation

3 years of SCADA data
286 raw variables at 15-min resolution

Selected core variables

Primary sludge dry solids
Feed flows
Recirculation
Biogas flow
Biogas quality

Hourly resampling

Direct multi-horizon

Modeling strategy

XGBoost predicts the next 24 hourly methane values in one shot.

Important modelling choice

Only information available up to time t is used.
No “known future” covariates are injected.

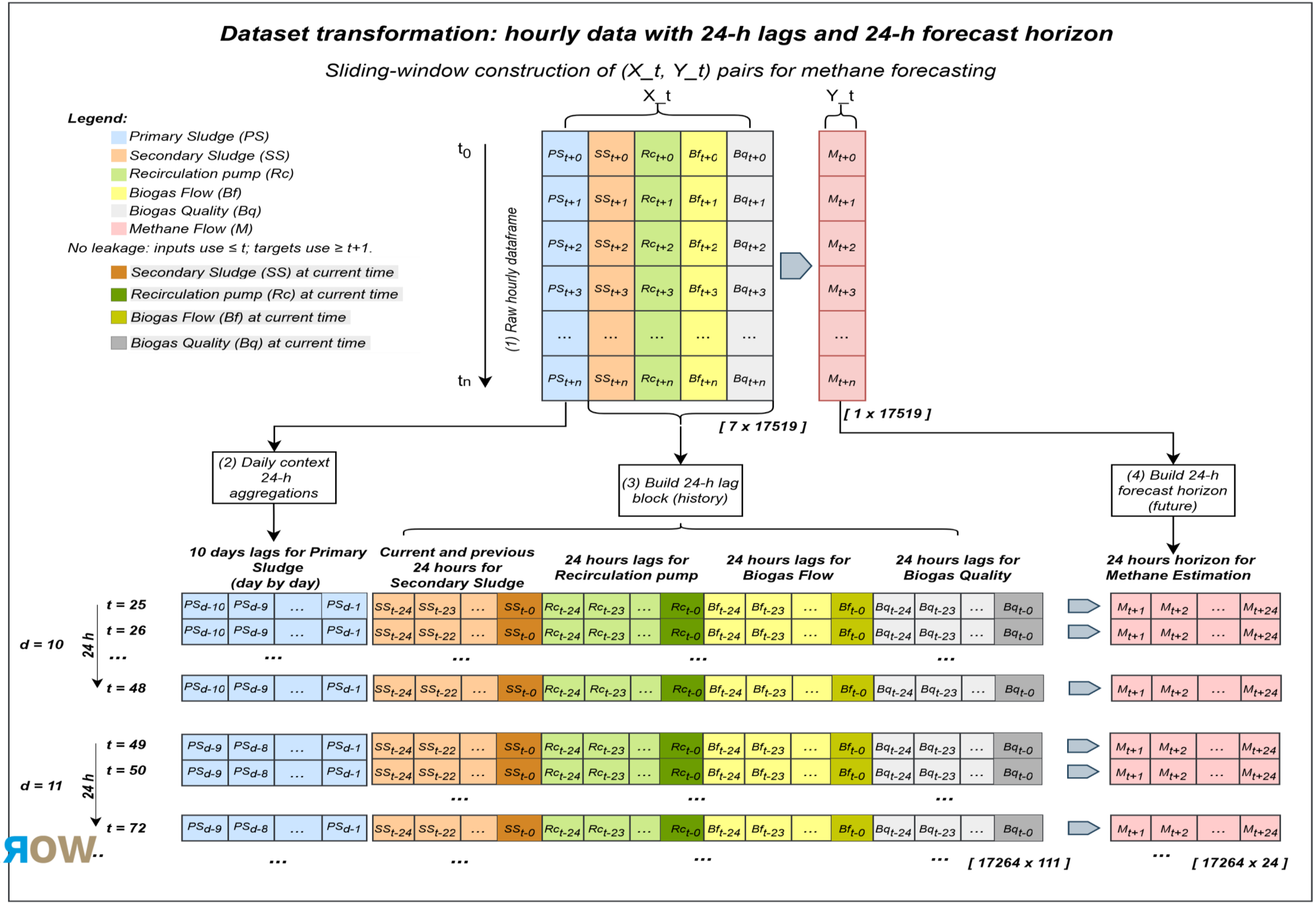
Result: a conservative but realistic day-ahead energy forecast.

Why include the sludge line?

- Extends the framework beyond effluent prediction.
- Connects process operation with energy recovery.
- Makes the DSS story more relevant for whole-plant decision support.

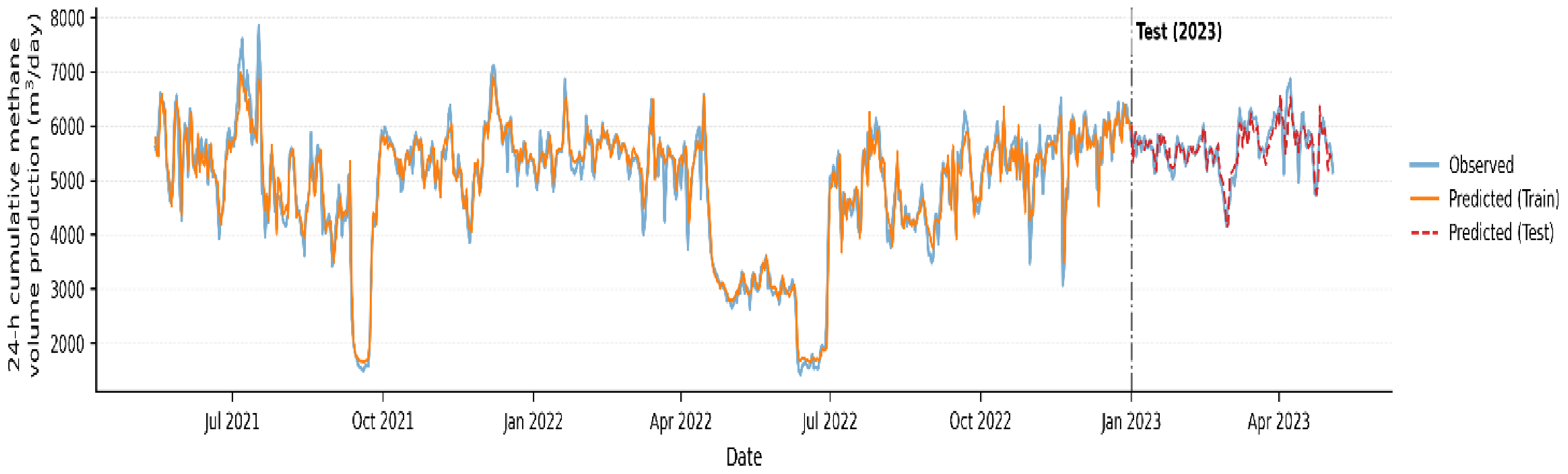
WRRF instead of WWTP

DARROW



Dataset transformation: hourly data with 24-h lags and 24-h forecast horizon
Sliding-window construction of (X_t, Y_t) pairs for methane forecasting
Legend:
Primary Sludge (PS)
Secondary Sludge (SS)
Recirculation pump (Rc)
Biogas Flow (Bf)
Biogas Quality (Bq)
Methane Flow (M)
No leakage: inputs use ≤ t; targets use ≥ t+1.
Secondary Sludge (SS) at current time
Recirculation pump (Rc) at current time
Biogas Flow (Bf) at current time
Biogas Quality (Bq) at current time
X_t
Y_t
t0
tn
(1) Raw hourly dataframe
PSt+0 SSt+0 Rct+0 Bft+0 Bqt+0
PSt+1 SSt+1 Rct+1 Bft+1 Bqt+1
PSt+2 SSt+2 Rct+2 Bft+2 Bqt+2
PSt+3 SSt+3 Rct+3 Bft+3 Bqt+3
PSt+n SSt+n Rct+n Bft+n Bqt+n
Mt+0 Mt+1 Mt+2 Mt+3 … Mt+n
[7 x 17519]
[1 x 17519]
(2) Daily context 24-h aggregations
(3) Build 24-h lag block (history)
(4) Build 24-h forecast horizon (future)
10 days lags for Primary Sludge (day by day)
Current and previous 24 hours for Secondary Sludge
24 hours lags for Recirculation pump
24 hours lags for Biogas Flow
24 hours lags for Biogas Quality
24 hours horizon for Methane Estimation
d = 10
d = 11
24 h
t = 25
t = 26
t = 48
t = 49
t = 50
t = 72
PSd-10 PSd-9 … PSd-1
PSd-9 PSd-8 … PSd-1
SSt-24 SSt-23 … SSt-0
SSt-24 SSt-22 … SSt-0
Rct-24 Rct-23 … Rct-0
Bft-24 Bft-23 … Bft-0
Bqt-24 Bqt-23 … Bqt-0
Mt+1 Mt+2 … Mt+24
[17264 x 111]
[17264 x 24]

4. Sludge-line ROM: Methane forecasting



Interpretation

- The framework is already useful for ranking control options.
- It is not replacing the full mechanistic model — it is complementing it.
- Errors reveal where future refinement should focus.

Sludge-line methane forecast

R^2	0.715	MAPE	3.28%
MAE	185 m³/day	RMSE	245 m³/day

DARROW

ceit
MEMBER OF BASQUE RESEARCH & TECHNOLOGY ALLIANCE

6. Decision Support System: Multi-objective optimization

Optimization setup

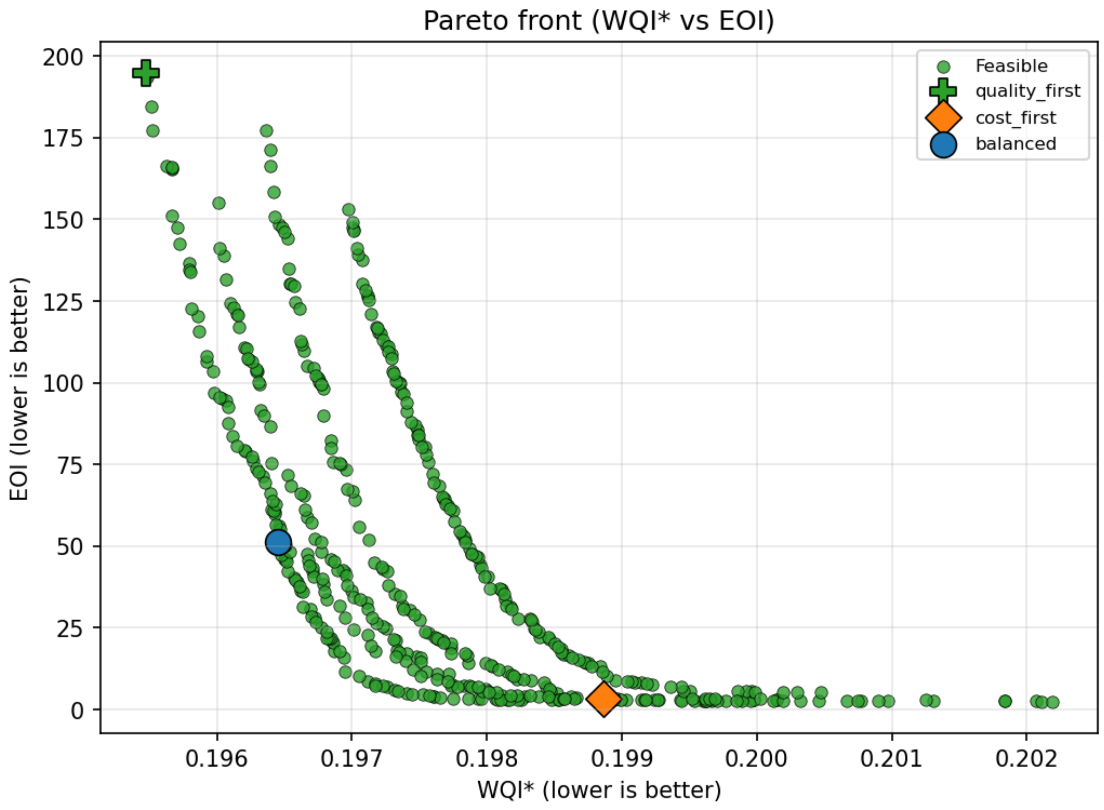
- Decision vector = 24 hourly control actions.
- Variables: DO setpoints and internal recirculation.
- Algorithm: NSGA-II multi-objective evolutionary search.
- Objectives: effluent quality (WQI*) and operating effort / energy (EOI).

Hard constraints

Equipment limits and feasible setpoints

Soft constraints

Effluent compliance violations penalized

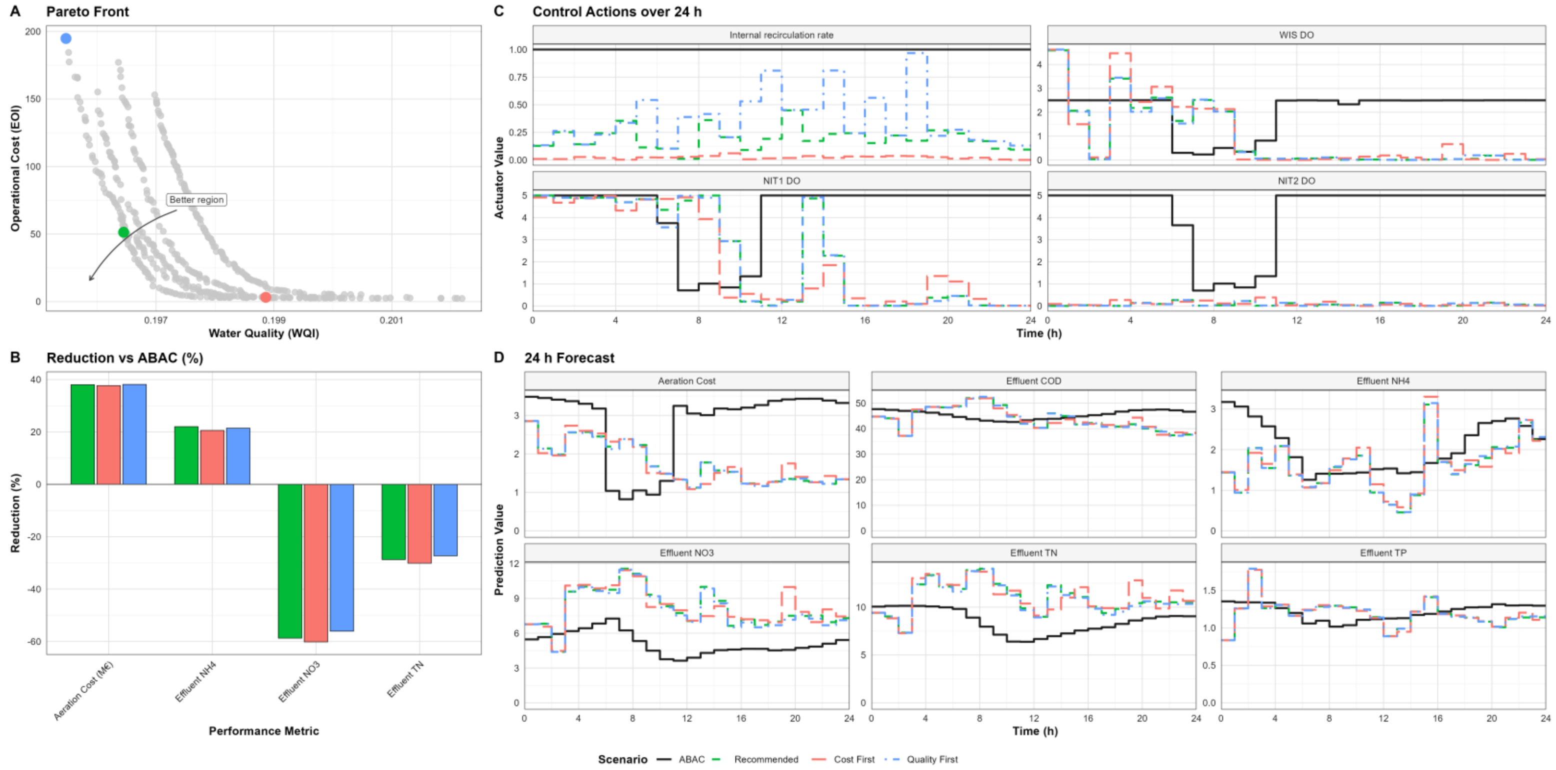


DARROW

ceit
MEMBER OF BASQUE RESEARCH & TECHNOLOGY ALLIANCE

6. Decision Support System: Day-ahead control policies

Compared with ABAC, the three representative policies improve both cost and effluent quality — but through different control philosophies.



Cost-first

- Most conservative actuation effort
- Strong cost reduction
- Moderate quality gains

Best when cost pressure dominates

Balanced

- Stable compromise across KPIs
- Good NH_4 reduction
- Easy operator narrative

Best default story for a broad audience

Quality-first

- Most aggressive aeration / recirculation shaping
- Largest NO_3 and TN reduction
- Best compliance-focused option

Best when compliance risk is the priority

Main operational lever: dynamic internal recirculation

7.- Contributions

This work proposes a ROM-based framework to support practical WRRF operation.

- A LSTM-based reduced-order model was developed to predict 24-hour-ahead effluent quality and operational variables under dynamic influent and control conditions.
- A data-driven methane forecasting model was built from real SCADA data, extending the framework from effluent prediction to whole-plant energy awareness.
- The ROMs were connected to a multi-objective optimization layer to generate day-ahead control policies balancing effluent quality and operating effort.
- The framework provides a fast, deployable layer for what-if analysis, policy screening and operator-oriented decision support.

8.-References

- A. J. Wang *et al.*, "Digital Twins for Wastewater Treatment: A Technical Review," *Engineering*, vol. 36, pp. 21–35, May 2024, doi: 10.1016/J.ENG.2024.04.012.
- J. Sun, Y. Xu, S. Nairat, J. Zhou, and Z. He, "Prediction of biogas production in anaerobic digestion of a full-scale wastewater treatment plant using ensembled machine learning models," *Water Environment Research*, vol. 95, no. 6, p. e10893, Jun. 2023, doi: 10.1002/WER.10893.
- C. Sweetapple, G. Fu, and D. Butler, "Multi-objective optimisation of wastewater treatment plant control to reduce greenhouse gas emissions," *Water Res.*, vol. 55, 2014, doi: 10.1016/j.watres.2014.02.018.
- S. K. Heo, K. J. Nam, S. Tariq, J. Y. Lim, J. Park, and C. K. Yoo, "A hybrid machine learning–based multi-objective supervisory control strategy of a full-scale wastewater treatment for cost-effective and sustainable operation under varying influent conditions," *J. Clean. Prod.*, vol. 291, 2021, doi: 10.1016/j.jclepro.2021.125853.
- A. J. Wang *et al.*, "Digital Twins for Wastewater Treatment: A Technical Review," 2024. doi: 10.1016/j.eng.2024.04.012.
- I. M. Sobol', "On the distribution of points in a cube and the approximate evaluation of integrals," *USSR Computational Mathematics and Mathematical Physics*, vol. 7, no. 4, pp. 86–112, Jan. 1967, doi: 10.1016/0041-5553(67)90144-9.
- T. Akiba, S. Sano, T. Yanase, T. Ohta, and M. Koyama, "Optuna: A Next-generation Hyperparameter Optimization Framework," *Proceedings of the ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, pp. 2623–2631, Jul. 2019, doi: 10.1145/3292500.3330701.
- M. Maktabifard, E. Zaborowska, and J. Makinia, "Energy neutrality versus carbon footprint minimization in municipal wastewater treatment plants," *Bioresour. Technol.*, vol. 300, 2020, doi: 10.1016/j.biortech.2019.122647.
- M. M. Uddin and M. M. Wright, "Anaerobic digestion fundamentals, challenges, and technological advances," *Physical Sciences Reviews*, vol. 8, no. 9, 2023, doi: 10.1515/psr-2021-0068.