SUPPLEMENTAL MATERIALS

ASCE *Journal of Environmental Engineering*

Unit Operation and Process Modeling with Physics-Informed Machine Learning

Haochen Li, David Spelman, and John Sansalone

DOI: 10.1061/JOEEDU.EEENG-7467

MATERIALS AND METHODS

Case 1. continuous stirred-tank reactor (CSTR)

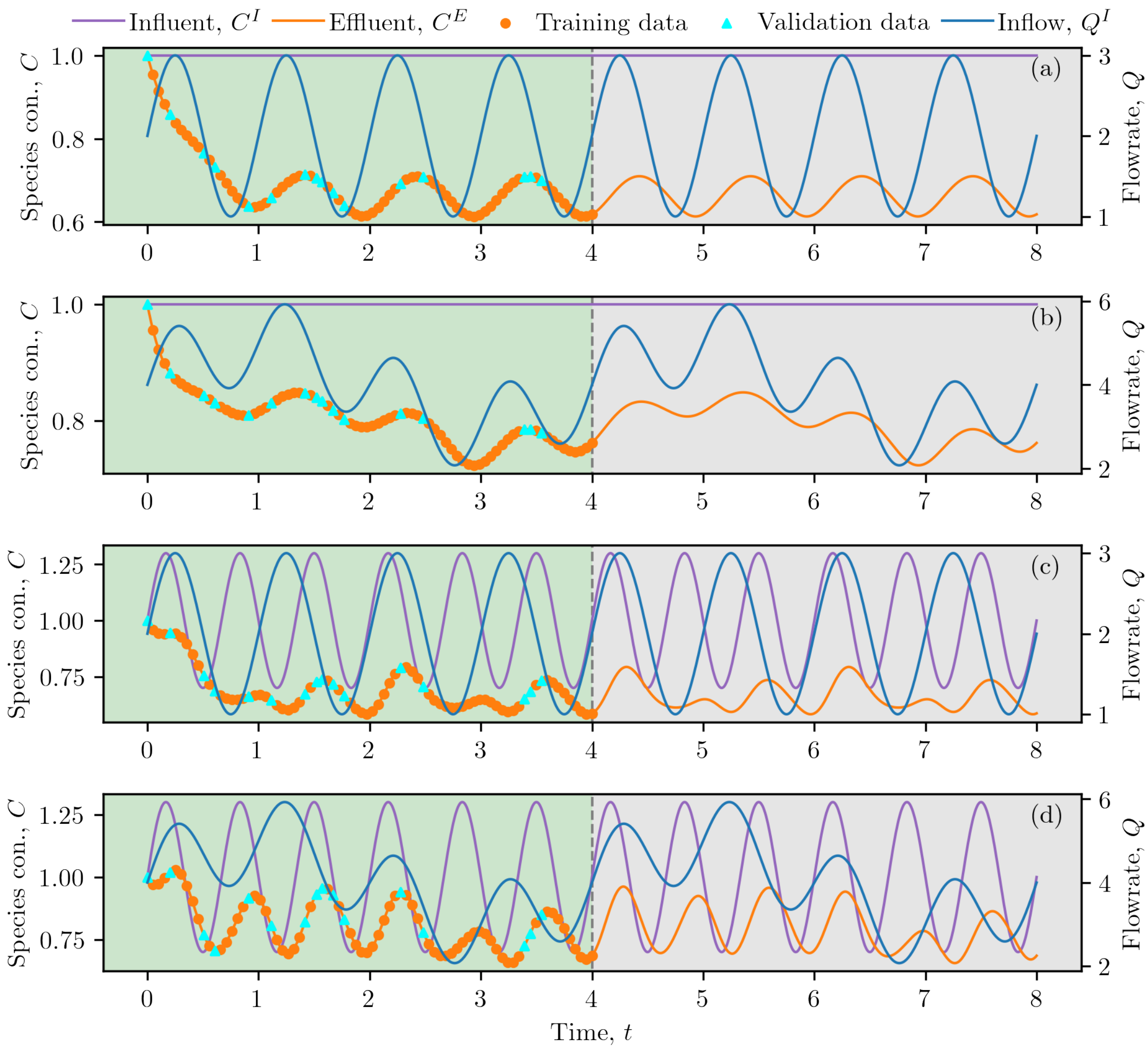

Fig. S1. Input loading and training data in CSTR case.

The coefficient of determination is defined in Eq. S1.

$$R^2(y, \hat{y}) = 1 - \frac{\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}{\sum_{i=1}^{n}(y_i - \bar{y})^2}. \tag{S1}$$

In this equation, $\hat{y}_i$ is the predicated value of the i-th sample and y_i is is the corresponding true value, $\bar{y}$ is the mean value. A negative R^2 indicates that the model poorly match the data.

Case 2. activated sludge reactor

Full expression of activated sludge model No.1 (ASM1)

$$\frac{dS_I}{dt} = 0 \tag{S2}$$

$$\begin{aligned}\frac{dS_S}{dt} = &-\frac{1}{Y_H}\mu_H\left(\frac{S_S}{K_S+S_S}\right)\left(\frac{S_O}{K_{O,H}+S_O}\right)X_{B,H} - \frac{1}{Y_H}\mu_H\left(\frac{S_S}{K_S+S_S}\right)\left(\frac{K_{O,H}}{K_{O,H}+S_O}\right)\left(\frac{S_{NO}}{K_{NO}+S_{NO}}\right)\eta_g X_{B,H} \\ &+ k_h\frac{X_S/X_{BH}}{K_X+(X_S/X_{BH})}\left[\left(\frac{S_O}{K_{O,H}+S_O}\right)+\eta_b\left(\frac{K_{O,H}}{K_{O,H}+S_O}\right)\left(\frac{S_{NO}}{K_{NO}+S_{NO}}\right)\right]X_{BH}\end{aligned} \tag{S3}$$

$$\frac{dX_I}{dt} = 0 \tag{S4}$$

$$\begin{aligned}\frac{dX_S}{dt} = &(1-f_p)\,b_H X_{BH} + (1-f_p)\,b_H X_{BA} \\ &- k_h\frac{X_S/X_{BH}}{K_X+(X_S/X_{BH})}\left[\left(\frac{S_O}{K_{OH}+S_O}\right)+\eta_b\left(\frac{K_{OH}}{K_{OH}+S_O}\right)\left(\frac{S_{NO}}{K_{NO}+S_{NO}}\right)\right]X_{BH}\end{aligned} \tag{S5}$$

$$\frac{dX_{BH}}{dt} = \mu_H\left(\frac{S_S}{K_S+S_S}\right)\left(\frac{S_O}{K_{OH}+S_O}\right)X_{BH} + \mu_H\left(\frac{S_S}{K_S+S_S}\right)\left(\frac{K_{OH}}{K_{OH}+S_O}\right)\left(\frac{S_{NO}}{K_{NO}+S_{NO}}\right)\eta_g X_{BH} - b_H X_{BH} \tag{S6}$$

$$\frac{dX_{BA}}{dt} = \mu_A\left(\frac{S_{NH}}{K_{NH}+S_{NH}}\right)\left(\frac{S_O}{K_{OA}+S_O}\right)X_{BA} - b_A X_{BA} \tag{S7}$$

$$\frac{dX_p}{dt} = f_P b_H X_{BH} + f_P b_A X_{BH} \tag{S8}$$

$$\frac{dS_O}{dt} = -\frac{1-Y_H}{Y_H}\mu_H\left(\frac{S_S}{K_S+S_S}\right)\left(\frac{S_O}{K_{OH}+S_O}\right)X_{BH} - \frac{4.57-Y_A}{Y_A}\mu_A\left(\frac{S_{NH}}{K_{NH}+S_{NH}}\right)\left(\frac{S_O}{K_{OA}+S_O}\right)X_{BA} \tag{S9}$$

$$\frac{dS_{NO}}{dt} = -\frac{1-Y_H}{2.86Y_H}\mu_H\left(\frac{S_S}{K_S+S_S}\right)\left(\frac{K_{OH}}{K_{OH}+S_O}\right)\left(\frac{S_{NO}}{K_{NO}+S_{NO}}\right)\eta_g X_{B,H} + \frac{1}{Y_A}\mu_A\left(\frac{S_{NH}}{K_{NH}+S_{NH}}\right)\left(\frac{S_O}{K_{OA}+S_O}\right)X_{BA} \tag{S10}$$

$$\begin{aligned}\frac{dS_{NH}}{dt} = &-i_{XB}\mu_H\left(\frac{S_S}{K_S+S_S}\right)\left(\frac{S_O}{K_{OH}+S_O}\right)X_{B,H} - i_{XB}\mu_H\left(\frac{S_S}{K_S+S_S}\right)\left(\frac{K_{OH}}{K_{OH}+S_O}\right)\left(\frac{S_{NO}}{K_{NO}+S_{NO}}\right)\eta_g X_{BH} \\ &-\left(i_{XB}+\frac{1}{Y_A}\right)\mu_A\left(\frac{S_{NH}}{K_{NH}+S_{NH}}\right)\left(\frac{S_O}{K_{OA}+S_O}\right)X_{BA}\end{aligned} \tag{S11}$$

$$\frac{dS_{ND}}{dt} = -k_a S_{ND} X_{BH} + k_h\frac{X_S/X_{BH}}{K_X+(X_S/X_{BH})}\left[\left(\frac{S_O}{K_{OH}+S_O}\right)+\eta_b\left(\frac{K_{OH}}{K_{OH}+S_O}\right)\left(\frac{S_{NO}}{K_{NO}+S_{NO}}\right)\right]X_{BH}\,(X_{ND}/X_S) \tag{S12}$$

$$
\begin{aligned}
\frac{dX_{ND}}{dt} = &-\frac{i_{XB}}{14}\mu_H\left(\frac{S_S}{K_S+S_S}\right)\left(\frac{S_O}{K_{OH}+S_O}\right)X_{BH} \\
&+\frac{1}{14}\left(\frac{1-Y_H}{2.86Y_H}-i_{XB}\right)\mu_H\left(\frac{S_S}{K_S+S_S}\right)\left(\frac{K_{OH}}{K_{OH}+S_O}\right)\left(\frac{S_{NO}}{K_{NO}+S_{NO}}\right)\eta_g X_{BH} \\
&+\left(\frac{i_{XB}}{14}-\frac{1}{7Y_A}\right)\mu_A\left(\frac{S_{NH}}{K_{NH}+S_{NH}}\right)\left(\frac{S_O}{K_{OA}+S_O}\right)X_{BA}-\frac{1}{14}k_a S_{ND} X_{BH}
\end{aligned}
\tag{S13}
$$

In these equations, $S_I, S_S, X_I, X_S, X_{BH}, X_{BA}, X_P, S_O, S_{NO}, S_{NH}, S_{ND}, X_{ND}, S_{ALK}$ are the 13 state variables: soluble inert organic matter, readily biodegradable substrate, particulate inert organic matter, slowly biodegradable substrate, active heterotrophic biomass, active autotrophic biomass, particulate products arising from biomass decay, oxygen, nitrate and nitrite nitrogen, $NH_4^+ + NH_3$ nitrogen, soluble biodegradable organic nitrogen, particulate biodegradable organic nitrogen, and alkalinity.

TABLE S1. Stoichiometric parameters in activated sludge model 1 (ASM1).

Name	Symbol	Value	Unit
Autotrophic yield	Y_A	0.24	g cell COD formed (g N oxidized)$^{-1}$
Heterotrophic yield	Y_H	0.67	g cell COD formed (g COD oxidized)$^{-1}$
Fraction of biomass yielding particulate products	f_p	0.08	dimensionless
Mass N/Mass COD in biomass	i_{XB}	0.08	g N (g COD)$^{-1}$ in biomass
Mass N/Mass COD in products from biomass	i_{XP}	0.06	g N (g COD)$^{-1}$ in particulate products

TABLE S2. kinetic parameters in activated sludge model 1 (ASM1).

Name	Symbol	Value	Unit
Heterotrophic growth and decay	μ_H	4.0	d^{-1}
	K_s	10.0	g COD.m^{-3}
	K_{OH}	0.2	g (-COD).m^{-3}
	K_{NO}	0.5	g NO_3-N.m^{-3}
	b_H	0.3	d^{-1}
Anoxic growth of heterotrophs	η_g	0.8	dimensionless
Anoxic hydrolysis	η_h	0.8	dimensionless
Hydrolysis	k_h	3.0	g slowly biodegradable COD.(g cell COD.d)$^{-1}$
	K_X	0.1	g slowly biodegradable COD.(g cell COD)$^{-1}$
Autotrophic growth and decay	μ_A	0.5	d^{-1}
	K_{NH}	1.0	g NH_3-N.m^{-3}
	b_A	0.05	d^{-1}
	K_{OA}	0.4	g (-COD).m^{-3}
Ammonification	k_a	0.05	m^3 (g COD.d)$^{-1}$

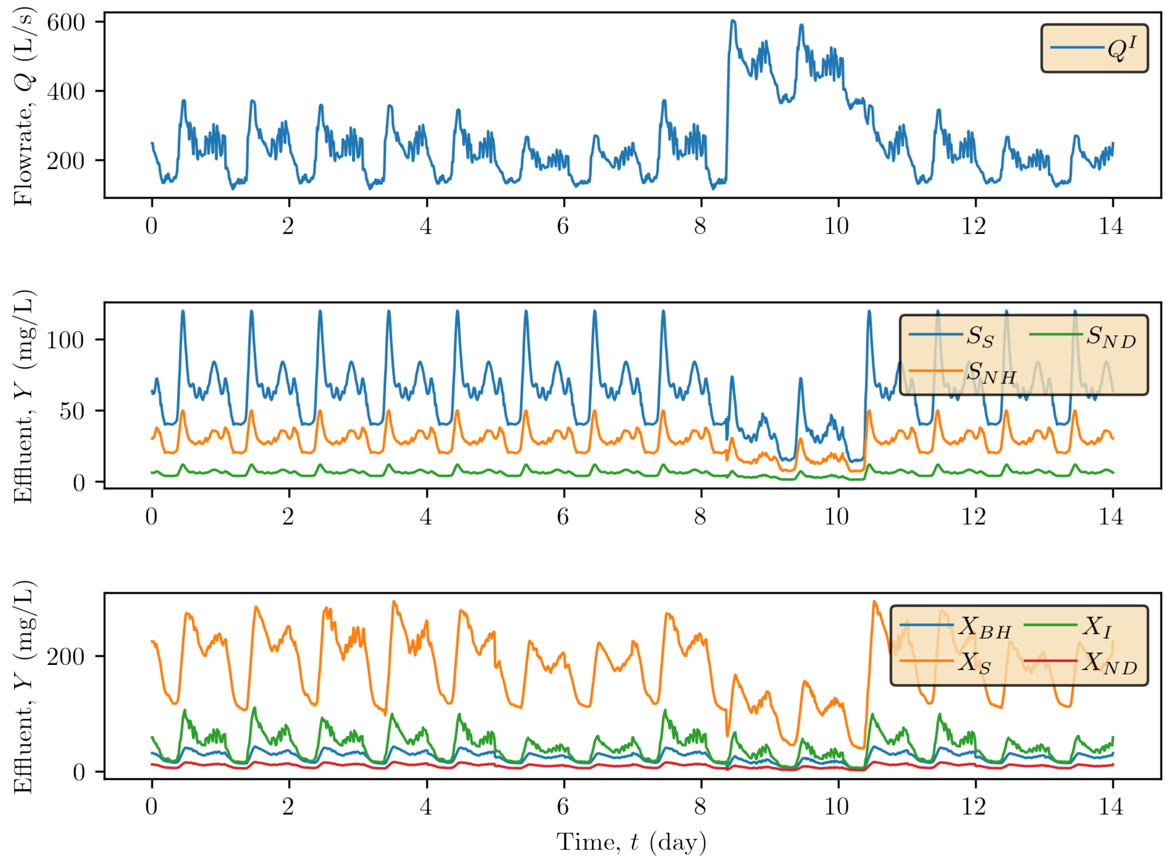

Fig. S2. Inflow hydrograph and influent loading graph in activated sludge reactor. The input file is publicly available at IWA website (http://iwa-mia.org/wp-content/uploads/2019/04/Inf_rain_2006.txt).

The implementation of ASM1 in this study is using Python 3.9.12. The ordinary differential equations (ODE) system of ASM1 is solved with SciPy.integrate.solve_ivp. The numerical implementation of ASM1 is verified with the solution of the well-established Benchmark Simulation Model No. 1 (BSM1) (Alex et al. 2018) (http://iwa-mia.org/benchmarking/). As summarized in Table S3, the implementation of ASM1 agrees well with the results reported by Alex et al. (2018) for the open-loop steady-state condition, with the relative differences are generally smaller than 1% except for S_O.

TABLE S3. Relative differences between the numerical implementation of this study and BSM1 reports for the open-loop steady-state condition. (S_I, S_S, X_I, X_S, X_{BH}, X_{BA}, X_P, S_O, S_{NO}, S_{NH}, S_{ND}, X_{ND}, S_{ALK}: soluble inert organic matter, readily biodegradable substrate, particulate inert organic matter, slowly biodegradable substrate, active heterotrophic biomass, active autotrophic biomass, particulate products arising from biomass decay, oxygen, nitrate and nitrite nitrogen, $NH_4^+ + NH_3$ nitrogen, soluble biodegradable organic nitrogen, particulate biodegradable organic nitrogen, and alkalinity, BSM1 reports is available at http://iwa-mia.org/benchmarking/)

Reactor #	S_I	S_S	X_I	X_S	X_{BH}	X_{BA}	X_P	S_O	S_{NO}	S_{NH}	S_{ND}	X_{ND}	S_{ALK}	TSS
1	0.000	-0.001	0.000	0.000	0.000	0.003	0.000	0.000	0.000	0.000	-0.003	0.001	0.000	0.000
2	0.000	-0.001	0.000	0.000	0.000	0.002	-0.001	-0.052	0.001	0.001	0.000	0.000	0.000	0.000
3	0.000	0.000	0.000	-0.001	0.000	0.000	0.001	-0.001	0.000	0.000	0.000	0.001	0.001	0.000
4	0.000	0.000	0.000	0.000	0.000	-0.003	0.001	0.000	0.000	-0.001	0.000	0.000	0.001	0.000
5	0.000	0.001	0.000	0.000	0.000	-0.001	0.000	0.000	0.001	0.002	0.000	-0.001	-0.001	0.000

Case 3. fixed-bed granular adsorption reactor

The 1D version of fixed-bed granular adsorption reactor is defined as following

$$\phi \frac{\partial Y}{\partial t} + \frac{\partial uY}{\partial x} = \frac{\partial}{\partial x}\left(\phi D_h \frac{\partial Y}{\partial x}\right) - \phi \beta R, \quad \text{(S14)}$$

$$\frac{\partial q}{\partial t} = R, \quad \text{(S15)}$$

$$R = \frac{1}{\beta} k_f Y \left(1 - \frac{q}{q_{max}}\right) - k_r q. \quad \text{(S16)}$$

In this equation, Y is the dissolved phase species concentration, u is the superficial velocity for the granular media bed, $u_p = u/\phi$ is the pore velocity, ϕ is the reactor porosity, D_h is the hydrodynamic dispersion coefficient for isotropic granular media (Bear 1972; Istok 1989; Horgue et al. 2015), $\beta = \rho_p (1-\phi)/\phi$ represents the specific mass of the granular media (divided by the porosity ϕ), ρ_p is the granular media density, R is the total/net reaction rate, q is the mass fraction of the adsorbed phase and adsorbent (kg adsorbed/kg adsorbent), k_f is the forward (adsorption) constant and k_r is the reverse (desorption) constant, q_{max} is the adsorption capacity (kg adsorbed/kg adsorbent). The mass of adsorbent is on a dry basis.

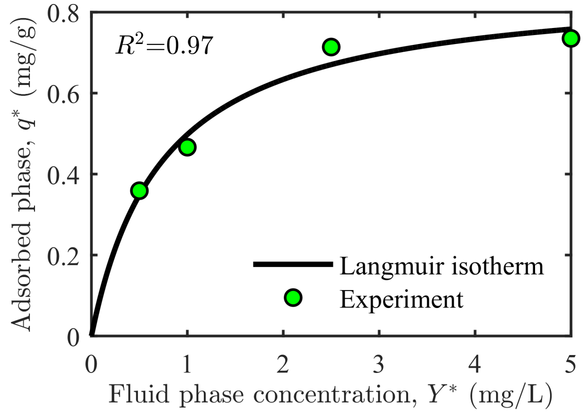

Fig. S3. Adsorption isotherm for AOCM-Clay media. q^* is mass fraction of adsorbed phase at adsorption equilibrium, Y^* dissolved phase species concentration at adsorption equilibrium. Experiment data from Sansalone and Ma (2011).

TABLE S4. Summary of input features, output feature, and the number of data pairs used for training and testing in each case. CSTR: continuous stirred tank reactor, ASR: activated sludge reactor, FBR: fixed-bed reactor, N_{train} is the number of paired data for training, N_{test} is the number of data paired for testing. In the FBR case, N_{test} is the total paired data for all loading conditions.).

Case	Input feature	Output feature	N_{train}	N_{test}
CSTR	Q^I, Y^I, t	Y^E	160	160
ASR	$Q^I, \mathbf{Y}^I, t$	$\mathbf{Y}^E$	480	864
FBR	Y^I, t	Y^E	40	120

RESULTS

Case 1. continuous stirred-tank reactor (CSTR)

TABLE S5. Coefficient of determination (R^2) and root mean square error (RMSE) of common ML models and PINNs in CSTR. (R^2/RMSE are calculated for the test period. Floats are presented to two decimal points. Loading A, B, C, and D are shown in Fig. 3a, 3b, 3c, and 3d.).

Model	Loading A	Loading B	Loading C	Loading D
Vanilla ANN	-45.08/0.23	-7.82/0.10	-37.91/0.37	-3.41/0.19
Lagged ANN	1.00/0.00	0.30/0.03	-0.13/0.06	-1.85/0.15
LSTM	1.00/0.00	0.37/0.03	-3.23/0.12	-3.09/0.18
PINNs	1.00/0.00	1.00/0.00	1.00/0.00	1.00/0.00

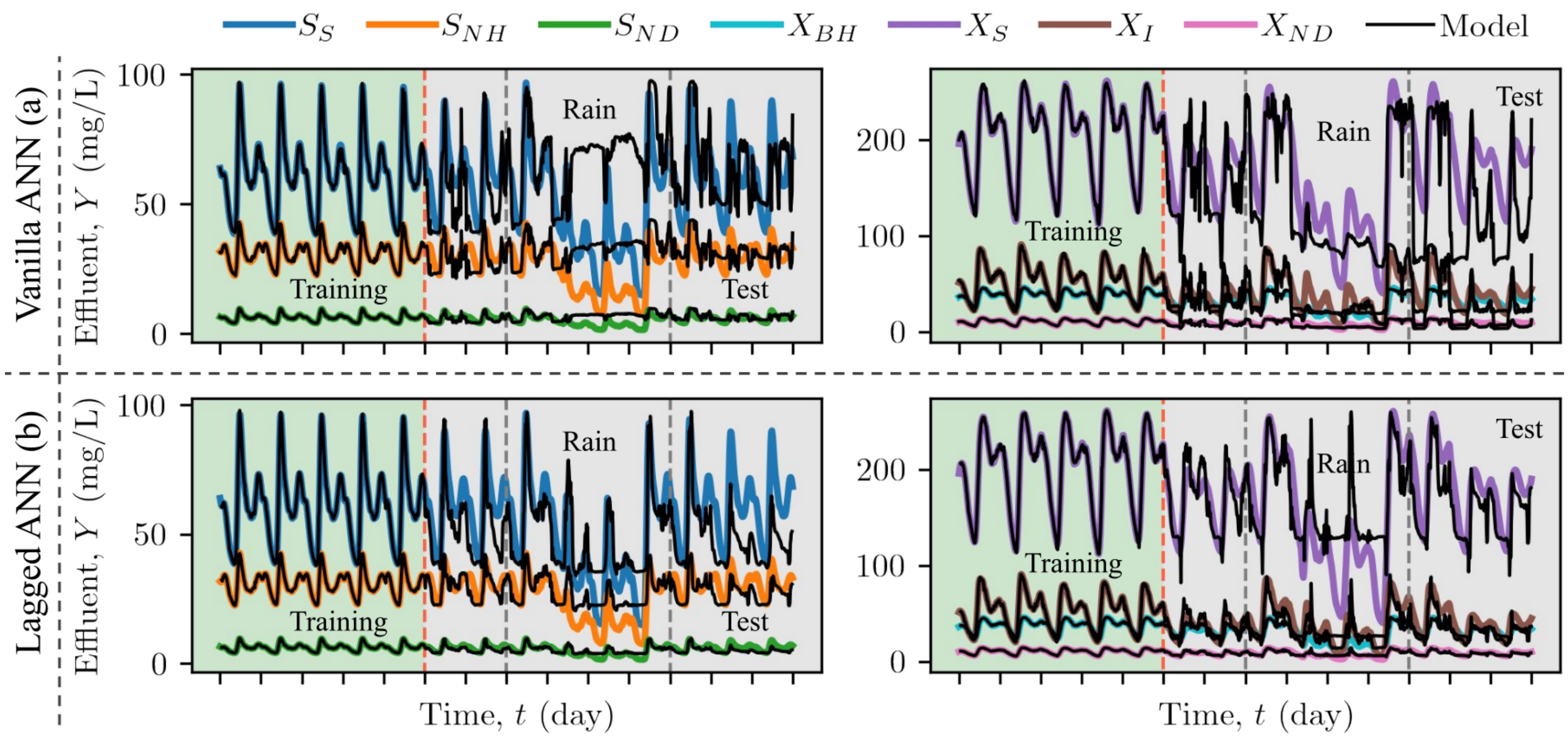

Fig. S4. Prediction of vanilla ANN and lagged ANN model for activated sludge reactor under wet weather loading conditions. S_S, S_{NH}, S_{ND}, X_{BH}, X_S, X_I, X_{ND} are readily biodegradable substrate, $NH_4^+ + NH_3$ nitrogen, soluble biodegradable organic nitrogen, active heterotrophic biomass, slowly biodegradable substrate, particulate inert organic matter, particulate biodegradable organic nitrogen. The activated sludge reactor dynamics is defined in Eq. 7 and Eqs. S2-S13. The test period includes a dry weather condition (5 days $< t \leq$ 7 days), a rain event (7 days $< t \leq$ 11 days) and a post event ($t >$ 11 days). Coefficients of determination (R^2) are summarized in Table S5.

Case 2. activated sludge reactor

TABLE S6. Coefficient of determination (R^2) and root mean square error (RMSE) of common ML models and PINNs in activated sludge reactor. (Dry weather: 5 days $< t \leq 7$ days; Rain event: 7 days $< t \leq 11$ days; Post event: $t > 11$ days. R^2/RMSE are averaged for S_S, S_{NH}, S_{ND}, X_{BH}, X_S, X_I, and X_{ND}. Floats are presented to two decimal points.).

Model	Dry weather	Rain event	Post event
Vanilla ANN	-0.22/9.46	-0.20/13.85	-0.94/16.31
Lagged ANN	0.46/5.94	0.59/11.26	0.15/8.91
LSTM	0.83/3.15	0.58/11.38	0.81/3.94
PINNs	1.00/0.08	1.00/0.09	1.00/0.08

Case 3. fixed-bed granular adsorption reactor

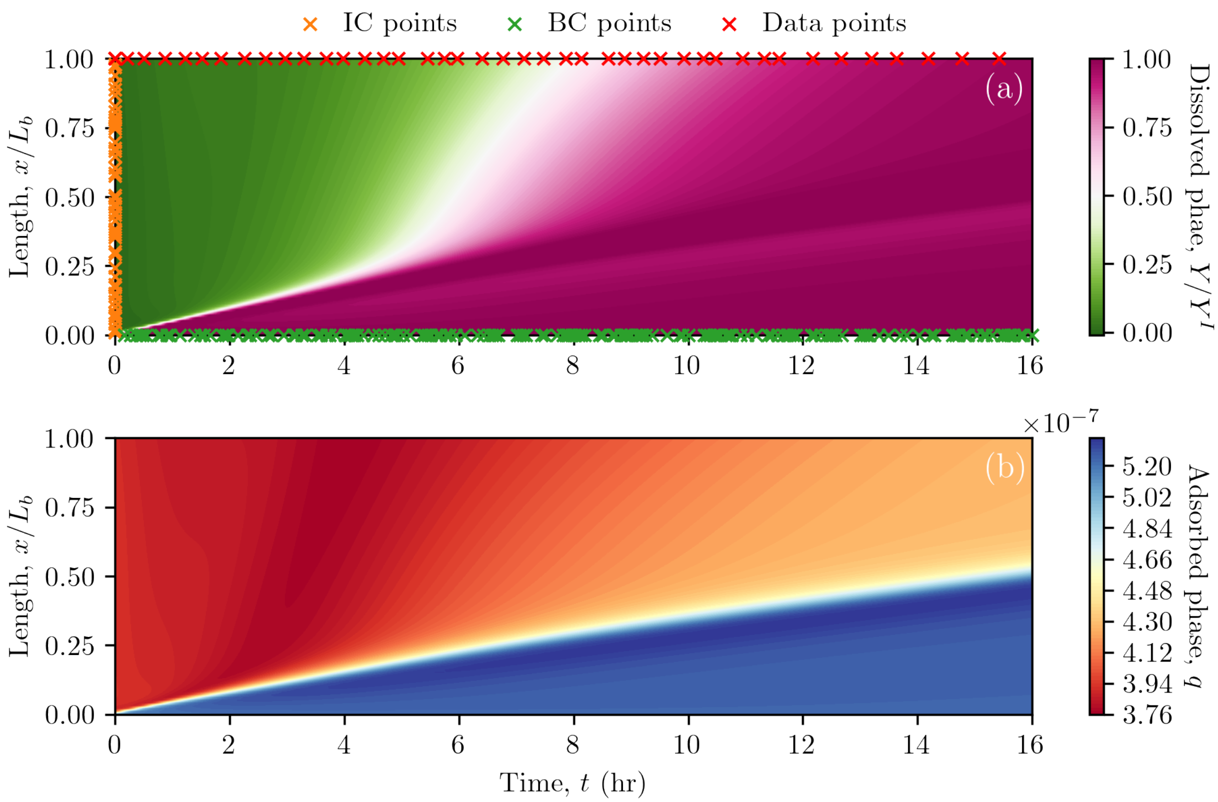

Fig. S5. Prediction of pure data-driven ANN in fixed-bed granular adsorption reactor. The fixed-bed granular adsorption reactor dynamics is defined in Eqs. 8-10.

TABLE S7. Coefficient of determination (R^2) and root mean square error (RMSE) of common ML models and PINNs in fixed-bed granular adsorption reactor. (R^2/RMSE are presented to two decimal points. Influent TDP loading of 0.5 mg/L is for training. Influent TDP loading of 1.0, 2.5, and 5.0 mg/L are for test).

Model	0.5 mg/L	1.0 mg/L	2.5 mg/L	5.0 mg/L
Vanilla ANN	1.00/0.02	-0.24/0.38	-1.79/0.54	-4.29/0.66
Lagged ANN	1.00/0.02	-1.40/0.53	-3.12/0.66	-5.49/0.74
LSTM	1.0/0.02	-1.05/0.49	-3.17/0.66	-5.57/0.74
PINNs	1.0/0.02	0.93/0.09	0.74/0.16	0.58/0.19

DISCUSSION

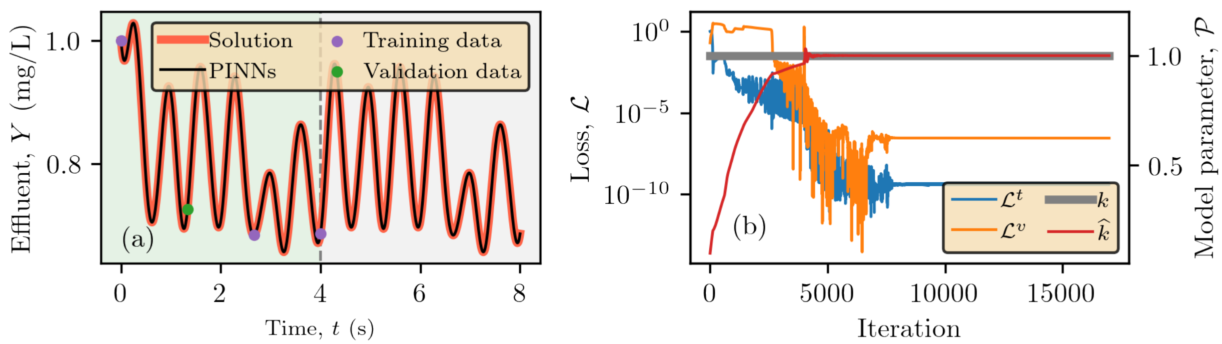

Fig. S6. PINNs model prediction in a continuous stirred-tank reactor (CSTR) with limited training data. The CSTR configuration and loading condition is identical to Fig. 3d.

REFERENCES

Alex, J., Benedetti, L., Copp, J., Gernaey, K., Jeppsson, U., Nopens, I., Pons, M., Steyer, J., and Vanrolleghem, P., 2018. *Benchmark Simulation Model no. 1 (BSM1)*. Technical Report. International Water Association. London, United Kingdom.

Bear, J., 1972. *Dynamics of Fluids in Porous Media*. Dover Publications, New York, NY.

Horgue, P., Soulaine, C., Franc, J., Guibert, R., and Debenest, G., 2015. "An open-source toolbox for multiphase flow in porous media". *Computer Physics Communications* 187: 217–226. https://doi.org/10.1016/j.cpc.2014.10.005.

Istok, J., 1989. "Derivation of Equations of Solute Transport", in: Groundwater Modeling by the Finite Element Method. American Geophysical Union, Washington, D.C., pp. 469–477. https://agupubs.onlinelibrary.wiley.com/doi/abs/10.1002/9781118665541.app3, https://doi.org/https://doi.org/10.1002/9781118665541.app3.

Sansalone, J., and Ma, J., 2011. "Parametric Analysis and Breakthrough Modeling of Phosphorus from Al-Oxide Filter Media". *Journal of Environmental Engineering* 137 (2): 108–118. https://doi.org/10.1061/(ASCE)EE.1943-7870.0000304.