National Alliance
for Water Innovation

Incorporating Detailed Water Chemistry into Process-Scale Cost Optimization with Machine Learning

Alex Dudchenko (SLAC)

Tim Bartholomew (NETL)

Chemistry governs operation of water treatment processes, but is difficult to model

WaterTAP typically uses property packages for non-electrolyte solutions

1. Non-electrolyte -> components are water and salt (i.e., NaCl, TDS, etc)
   - $Properties = f(salt\ concentration, temperature, pressure)$
   - Good for bulk properties like density, osmotic pressure, viscosity, specific enthalpy

2. Electrolyte -> components are water and all the potential species
   - Must track numerous electrolyte species and chemical reactions
   - Essential for ion activities, solubility/scaling tendencies, precipitation

Components
Na
K
Ca
Mg
Cl
SO4
HCO3
Si

Carbonation Process

H2O ←→ H+ + OH-

CaOH+ ←→ Ca2+ + OH-

CaHCO3+ ←→ Ca2+ + HCO3-

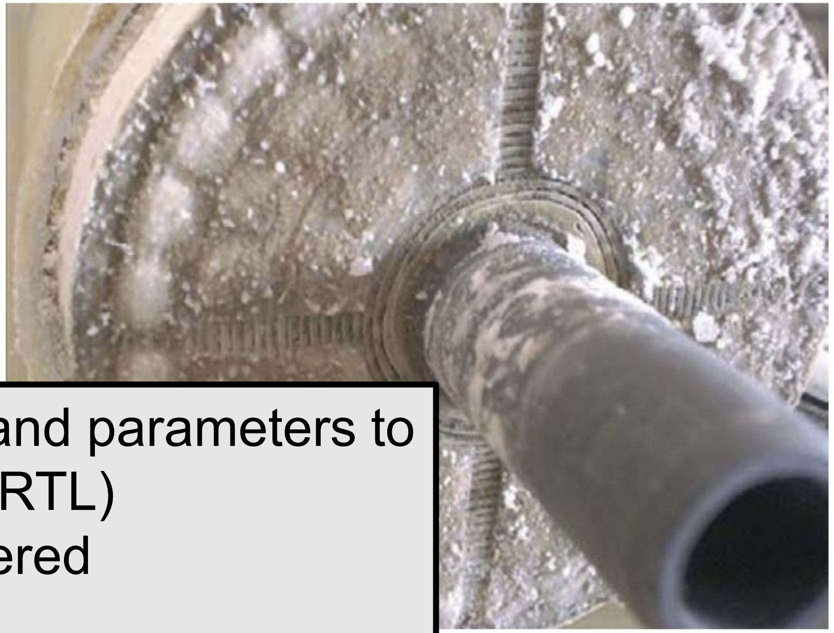
CaCO3 ←→ Ca2+ + CO32-

H2CO3 ←→ CO2 (aq) + H2O

H2CO3 ←→ H+ + HCO3-

HCO3- ←→ H+ + CO32-

Chemistry governs operation of water treatment processes, but is difficult to model



WaterTAP typically uses property packages for non-electrolyte solutions

1. Non-electrolyte -> components are water and salt (i.e., NaCl, TDS, etc)
    - G
      vi

- Electrolyte theoretical models have numerous terms and parameters to represent all of the interactions (e.g., MSE, Pitzer, eNRTL)
- Data availability limits the species that can be considered
- Inherently large models with many complications
    - Numerous species and reactions
    - Species can be at 0 concentration and increase by many orders of magnitude (round-off errors can be problematic)

2. Electroly
    - Must track numerous electrolyte species and chemical reactions
    - Essential for ion activities, solubility/scaling tendencies, precipitation

nation Process

	→ H+ + OH-
K	←→ Ca2+ + OH-
Ca	CaHCO3+ ←→ Ca2+ + HCO3-
Mg	CaCO3 ←→ Ca2+ + CO32-
Cl	H2CO3 ←→ CO2 (aq) + H2O
SO4	H2CO3 ←→ H+ + HCO3-
HCO3	HCO3- ←→ H+ + CO32-
Si	

NAWI

WaterTAP has 3 approaches for water chemistry

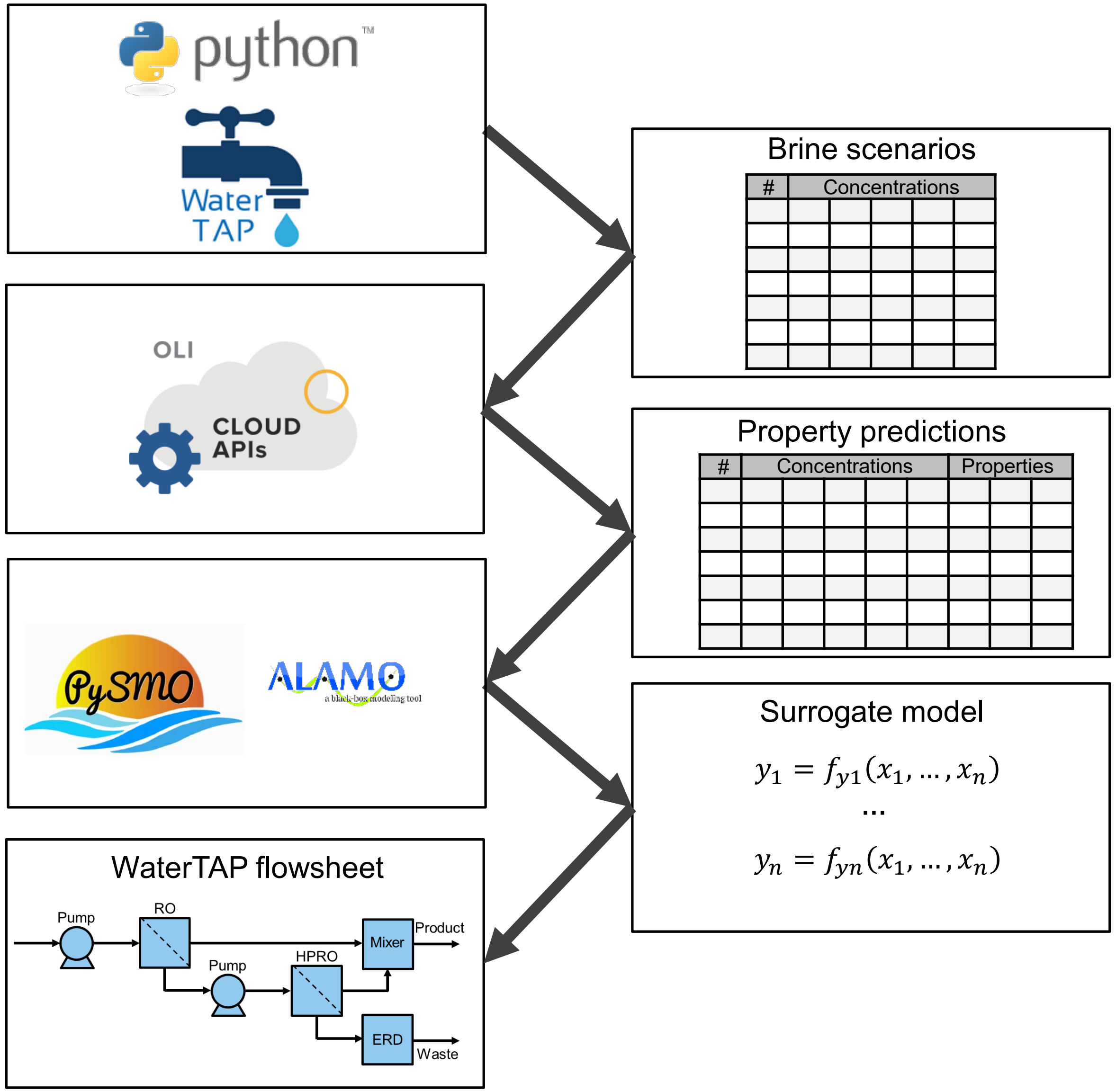
All approaches use external water chemistry software:

1. Narrow surrogate models
    - Inputs are the key decision variables of the flowsheet
    - Polynomial functions, Radial basis functions (interpolative model), etc.
2. Broad surrogate models
    - Inputs are apparent species concentrations, pH, pressure, temperature
    - Neural net (machine learning model)
3. Direct integration (Demo on Thursday at 5 PM)
    - Use pyomo External Grey Box Model
    - Requires the external water chemistry software to provide the Jacobian (and Hessian)
    - Possible with Reaktoro-pse (Repository on watertap-org)

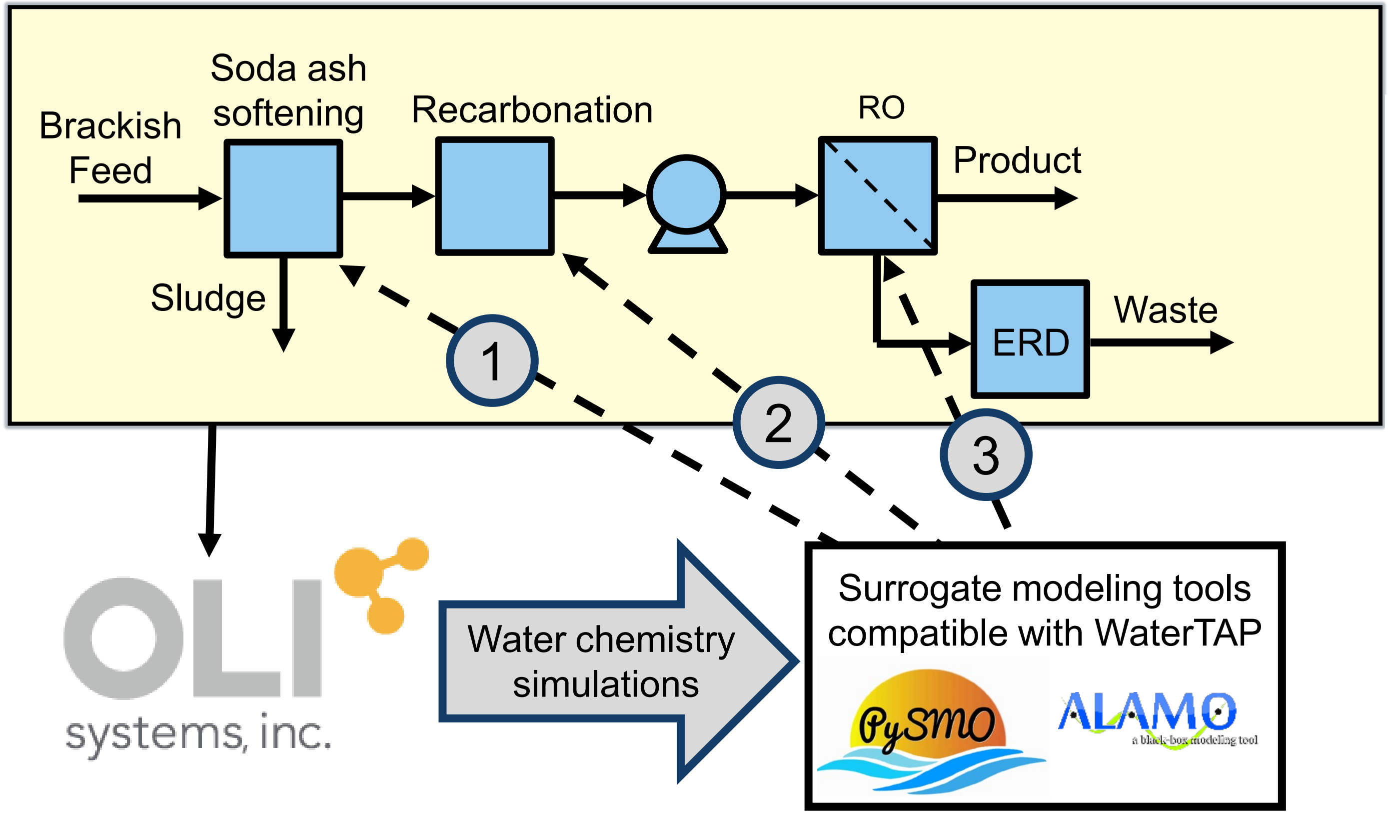
Reaktoro
for Python and C++

Building and integrating surrogate models

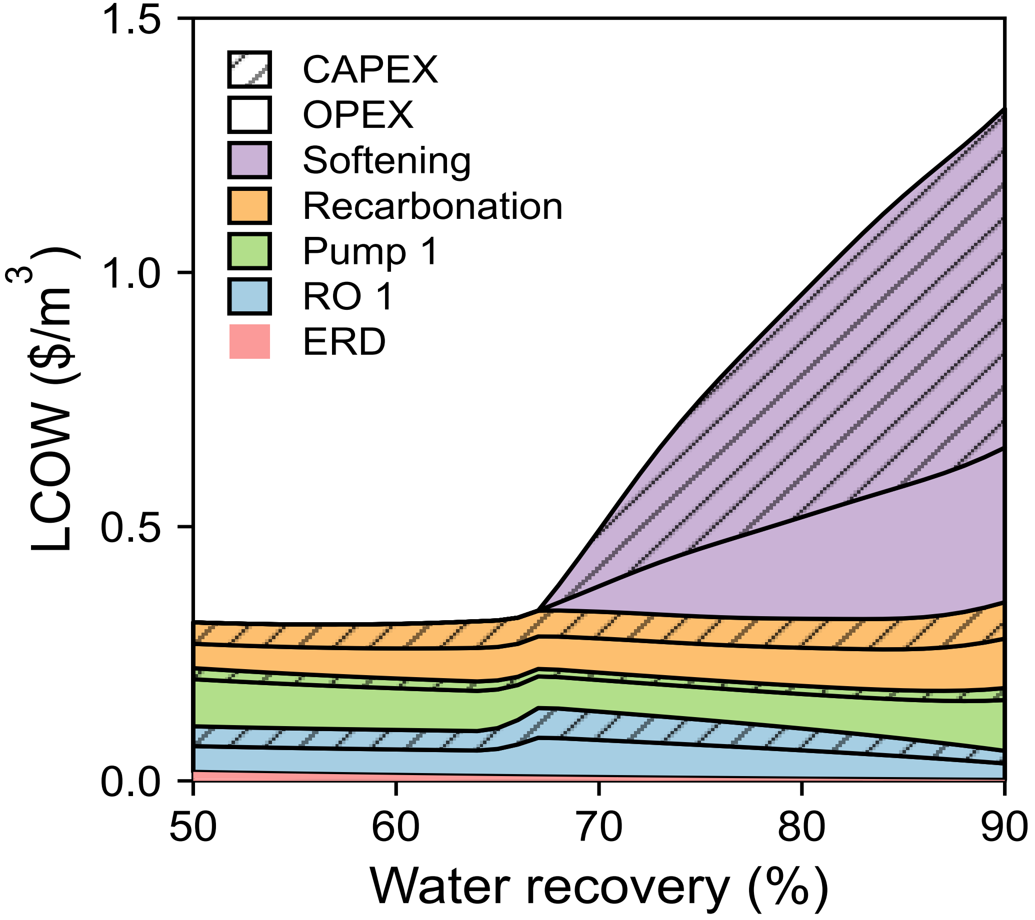
1. Generate relevant brine scenarios
2. Use OLI Cloud API to calculate properties for brine scenarios
3. Use IDAES tools to fit or integrate models into IDAES compatible models
4. Use WaterTAP flowsheet with the OLI surrogate model



Our first approach was narrow surrogate models

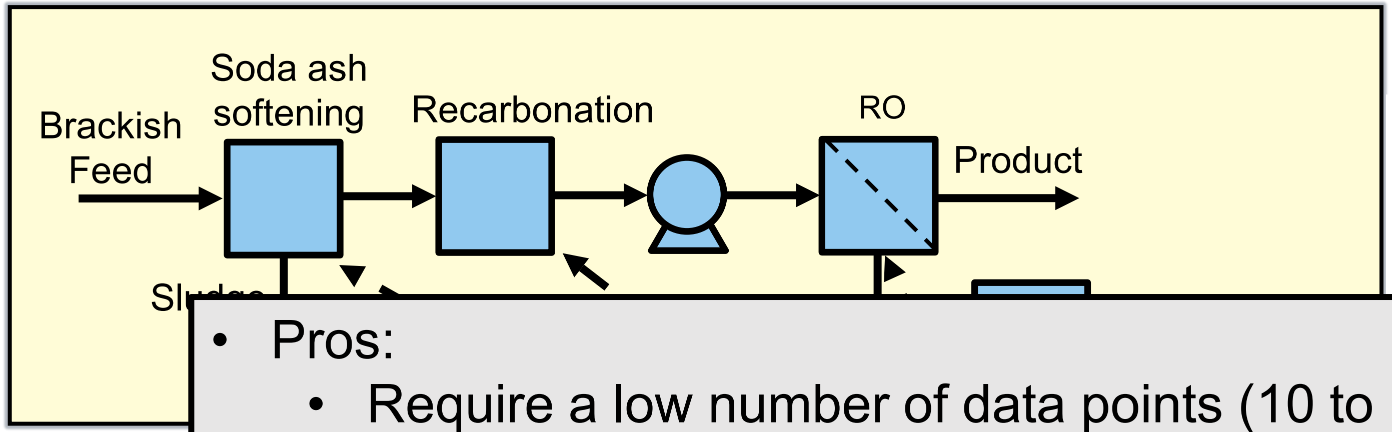


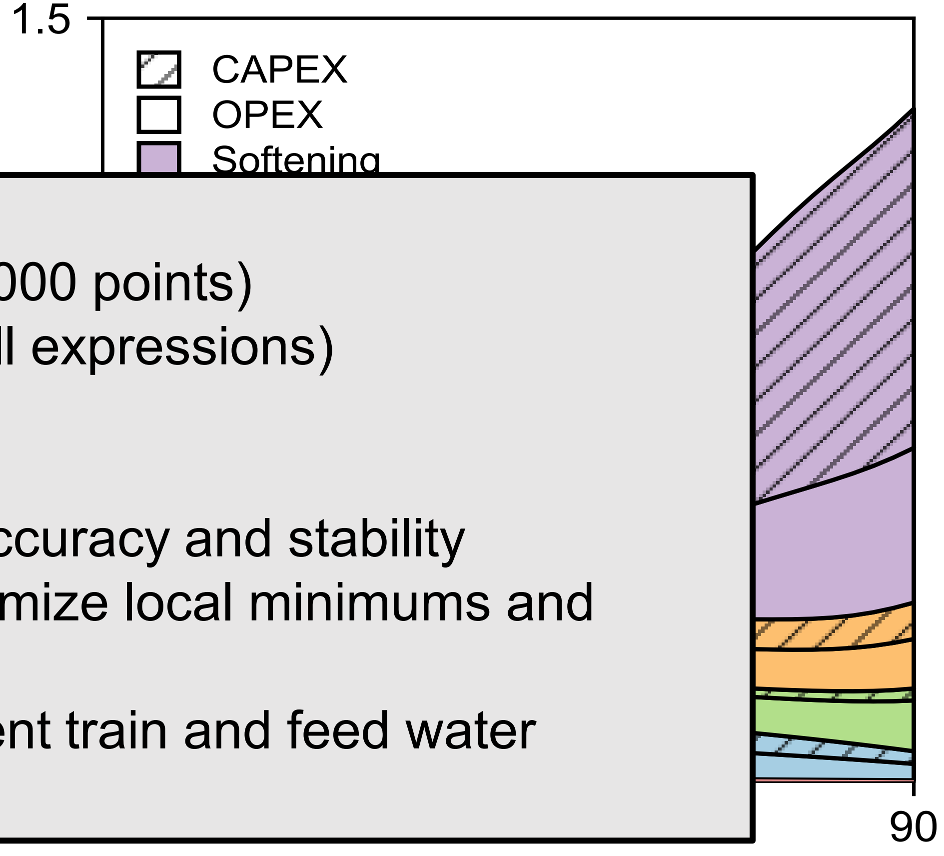
1. Chemical precipitation *f(soda ash)*
2. pH adjustment *f(soda ash, CO2)*
3. Mineral scaling prediction *f(soda ash, CO2 dose, pressure, water recovery)*



Amusat et al. *ACS ES&T Engineering.* 2024

Our first approach was narrow surrogate models





- Pros:
  - Require a low number of data points (10 to 100,000 points)
  - Low computational requirements (relatively small expressions)

- Cons:
  - Significant time requirements to achieve good accuracy and stability
    - Tailored data generation and training to minimize local minimums and optimize accuracy
  - Surrogates can be only used for specific treatment train and feed water composition

syste

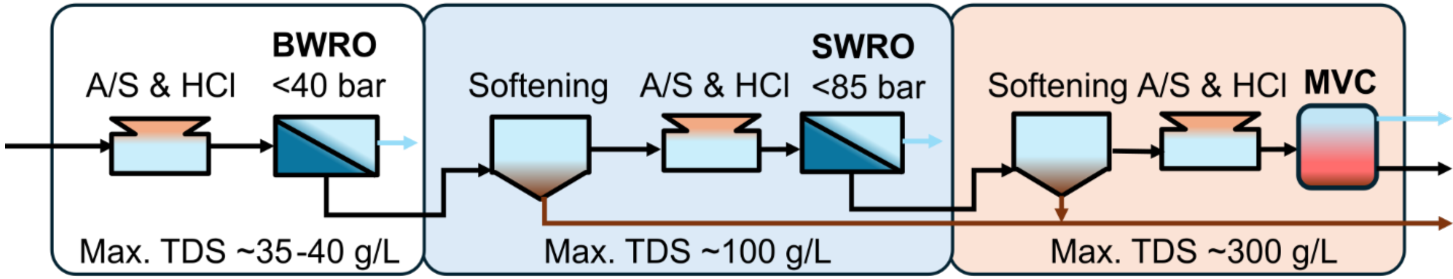
1. Chem
2. pH a
3. Mineral scaling prediction f(*soda ash, CO2 dose, pressure, water recovery)*

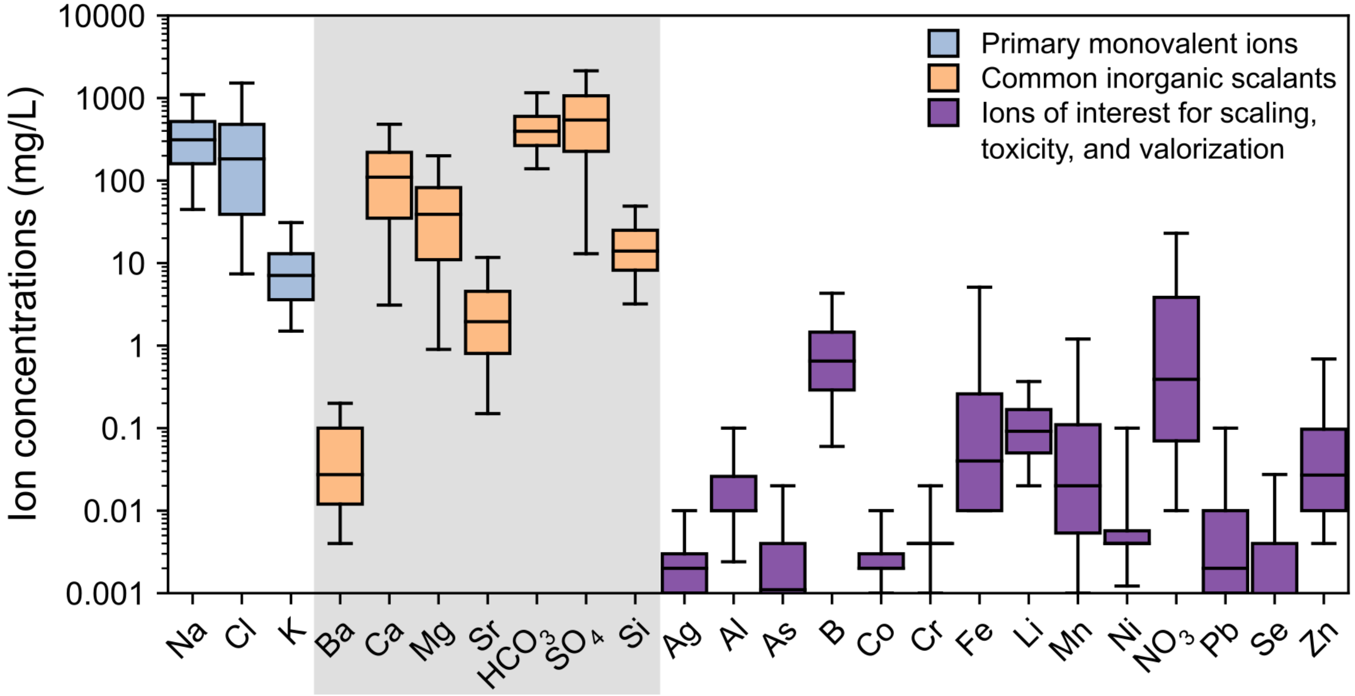
Amusat et al. *ACS ES&T Engineering.* 2024

Broad surrogate models are needed to assess different train configurations and feed compositions

- Large number of treatment train configurations
  - Use of recycle loops
  - Multiple stages
  - Different combinations of driving forces
- Water compositions vary dramatically across US

Proposed high recovery brackish water treatment train

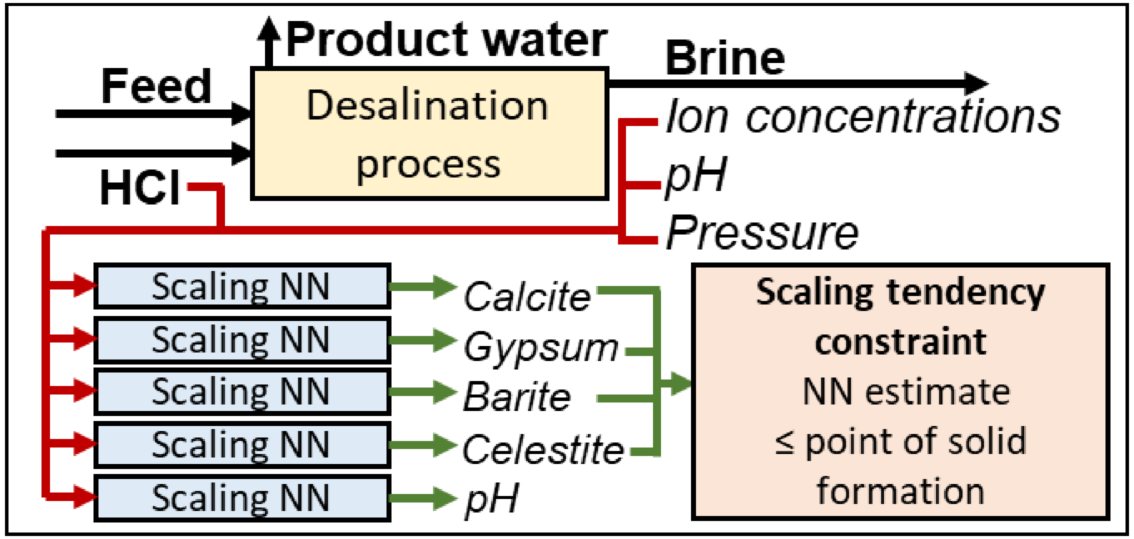




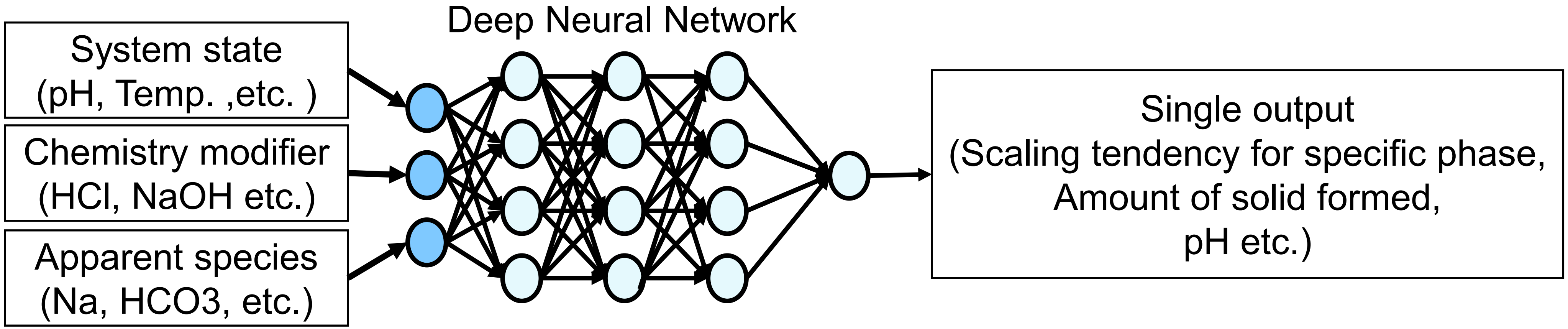
NAWI

Machine learning models can enable generation of broad surrogate models

Enables adding neural networks to Pyomo models:
- Supports dense and convolution layers
- Supports a number of activation functions
- Supports Keras and ONNX standards



Key Questions:
(1) Can deep neural networks provide broad range of chemistry estimates?
(2) How does NN architecture impact solver like IPOPT



Machine learning models can enable generation of broad surrogate models … *or suggest we add glue*

MLT

Enab
mode
- Su
- Su
- Su

Key Questions:
(1) Can deep neu
chemistry esti
(2) How does NN

of
OPT
eura

≡ Forbes

FORBES > LEADERSHIP > CAREERS

Google's AI Recommended Adding Glue To Pizza And Other Misinformation—What Caused The Viral Blunders?

Jack Kelly Senior Contributor

Jack Kelly covers career growth, job market and workplace trends.

Follow

0

May 31, 2024, 06:00am EDT

Sundar Pichai, chief executive officer of Alphabet Inc./Photographer: David Paul Morris/Bloomberg © 2024 BLOOMBERG FINANCE LP

nature

View all journals | Search | Log in

Explore content ˅ About the journal ˅ Publish with us ˅

nature > articles > article

Download PDF

Article | Open access | Published: 29 November 2023

An autonomous laboratory for the acc… materials

Nathan J. Szymanski, Bernardus Rendy, Yuxing Fei, Rishi E. Kumar, Tanjin He… Cubuk, Amil Merchant, Haegyeom Kim, Anubhav Jain, Christopher J. Bartel,…

Nature **624**, 86–91 (2023) | Cite this article

139k Accesses | **106** Citations | **1047** Altmetric | Metrics

Abstract

Challenges in high-throughput inorganic material prediction and autonomous synthesis

Josh Leeman,† Yuhan Liu,† Joseph Stiles,†,¶ Scott B. Lee,† Prajna Bhatt,† Leslie M. Schoop,*,†,¶ and Robert G. Palgrave*,‡

†Department of Chemistry, Princeton University, Princeton, NJ 08540, USA
‡Department of Chemistry, UCL, 20 Gordon St, London, WC1H0AJ, United Kingdom.
¶Princeton Materials Institute, Princeton University, Princeton, NJ 08540, USA

E-mail: lschoop@princeton.edu; r.palgrave@ucl.ac.uk

…t water
…ation
…ess

Brine
Ion concentrations
pH
…ressure

Scaling tendency constraint
NN estimate
≤ point of solid formation

…cific phase,
…ned,

The Verge / Tech / Reviews / Science / Entertainment / AI / More +

TESLA / ELECTRIC CARS / CARS

Tesla's Autopilot and Full Self-Driving linked to hundreds of crashes, dozens of deaths

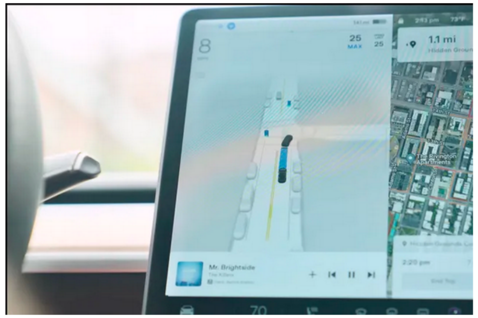


Image: Owen Grove / The Verge

/ NHTSA found that Tesla's driver-assist features are insufficient at keeping drivers engaged in the task of driving, which can often have fatal results.

By Andrew J. Hawkins, transportation editor with 10+ years of experience who covers EVs, public transportation, and aviation. His work has appeared in The New York Daily News and City & State.

Apr 26, 2024, 7:15 AM PDT

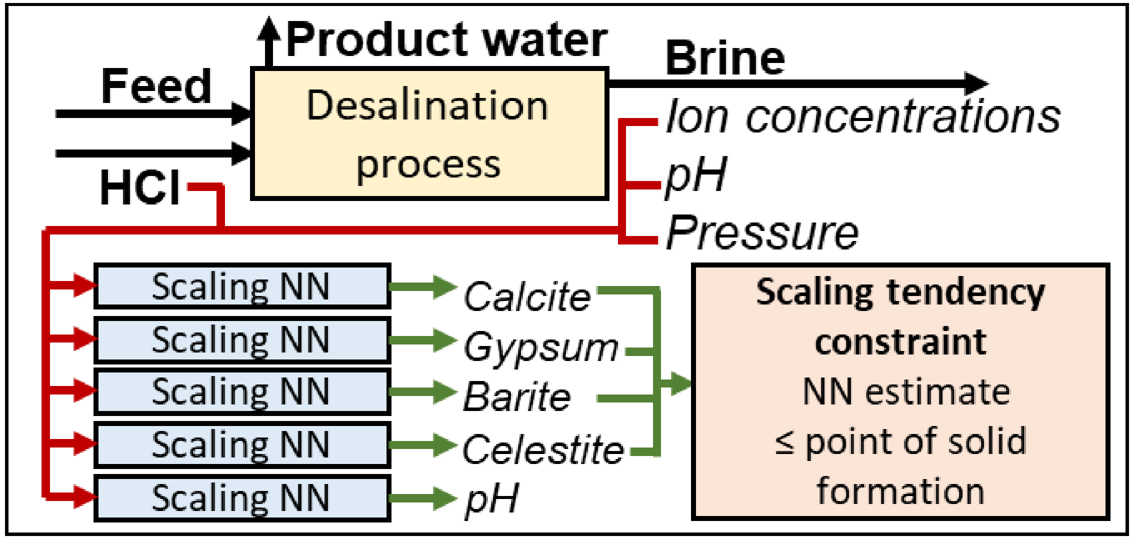
295 Comments (295 New)

NAWI

Machine learning models can enable generation of broad surrogate models … *or suggest we add glue*

OMLT

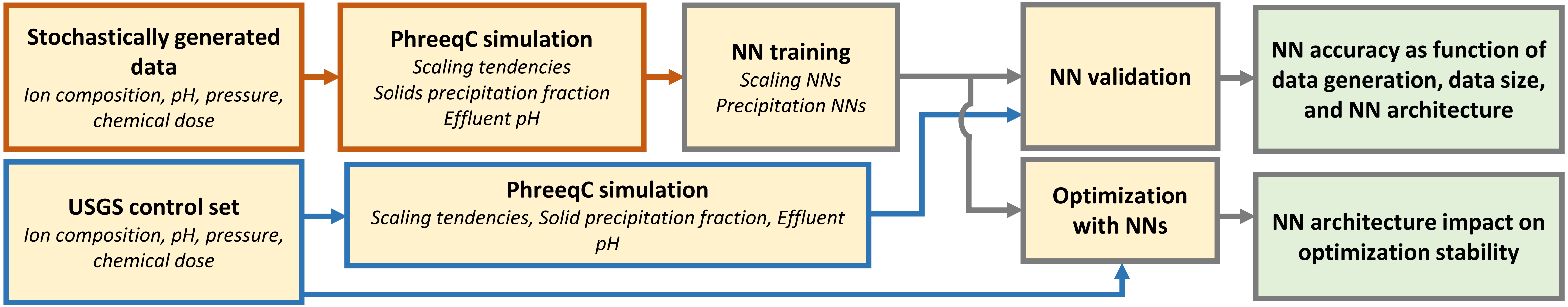
Enables adding neural networks to Pyomo models:
- Supports dense and convolution layers
- Supports a number of activation functions
- Supports Keras and ONNX standards



Key Questions:

(1) Can deep neural networks provide broad range of chemistry estimates?

(2) How does NN architecture impact solver like IPOPT

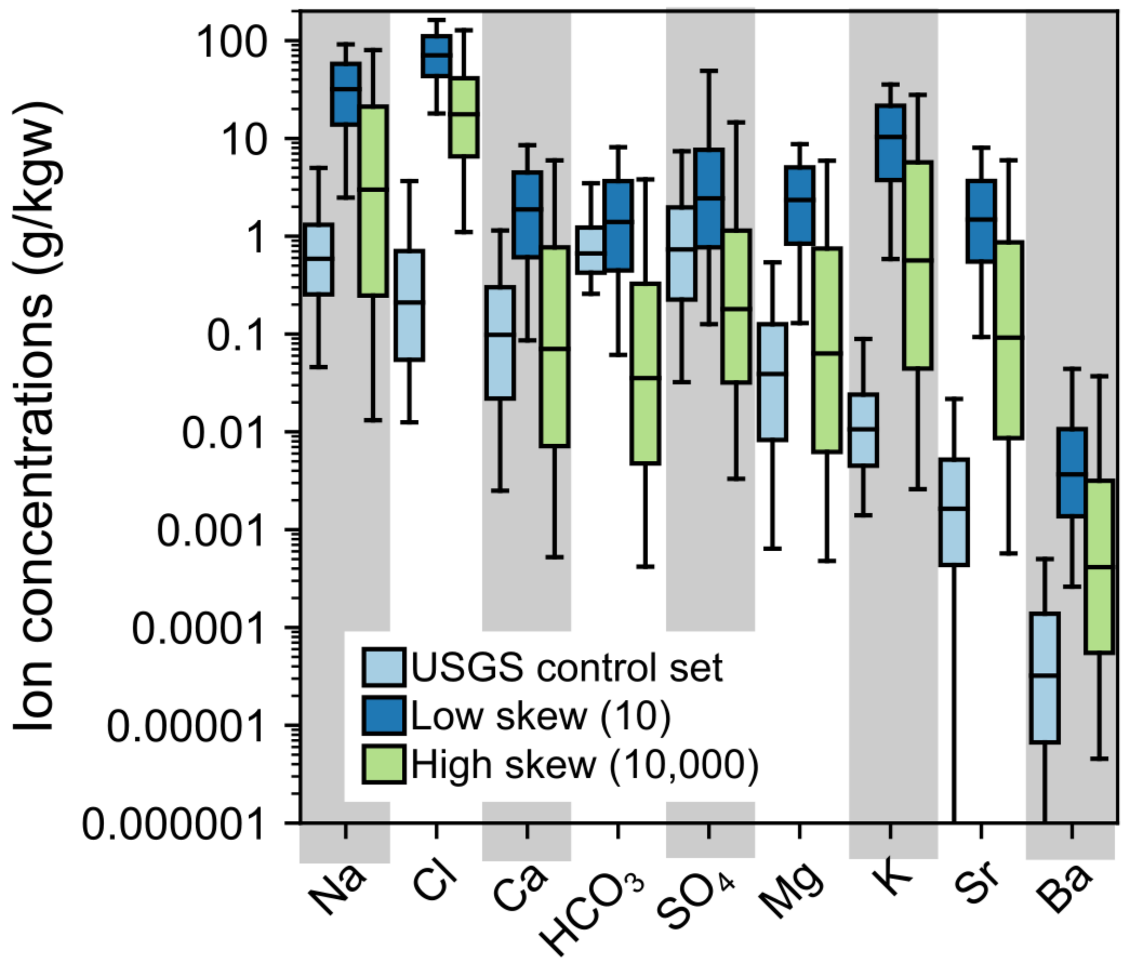


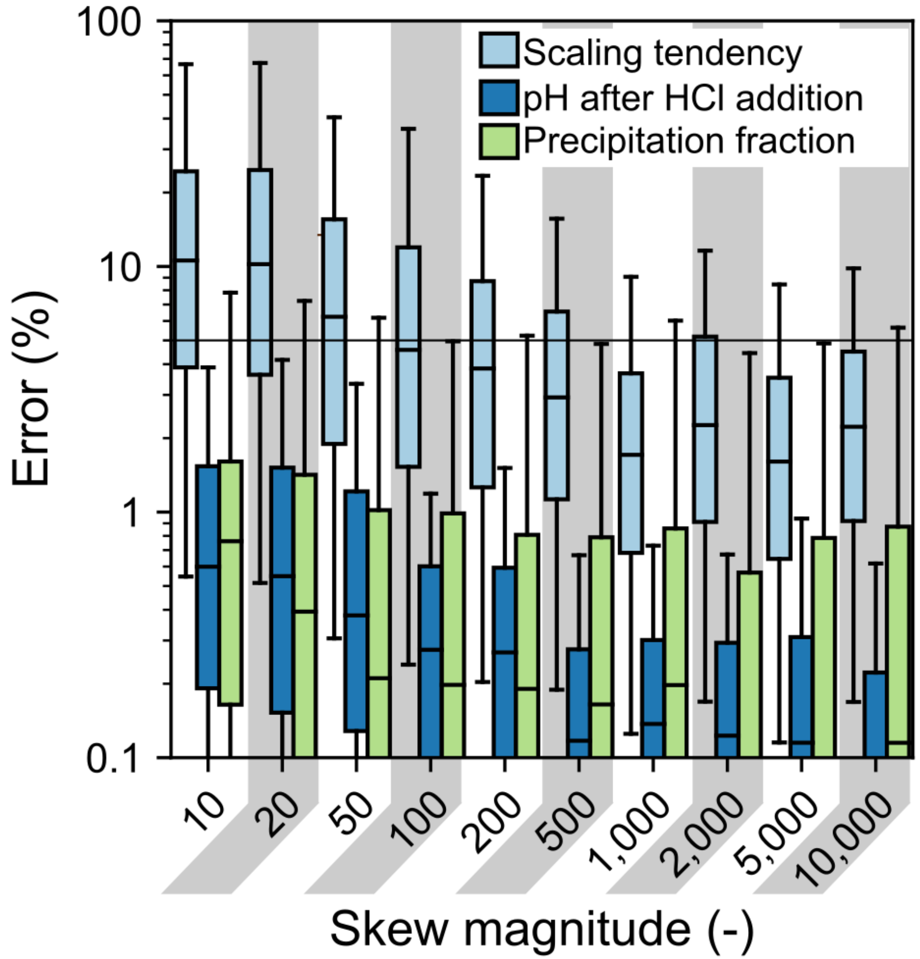
Dudchenko, A. V. & Amusat, O. *Systems & Control Transactions* (2024)

Tailoring data sampling is key for good accuracy

Simple exponential skewing of ion concentrations provides closer match to real waters

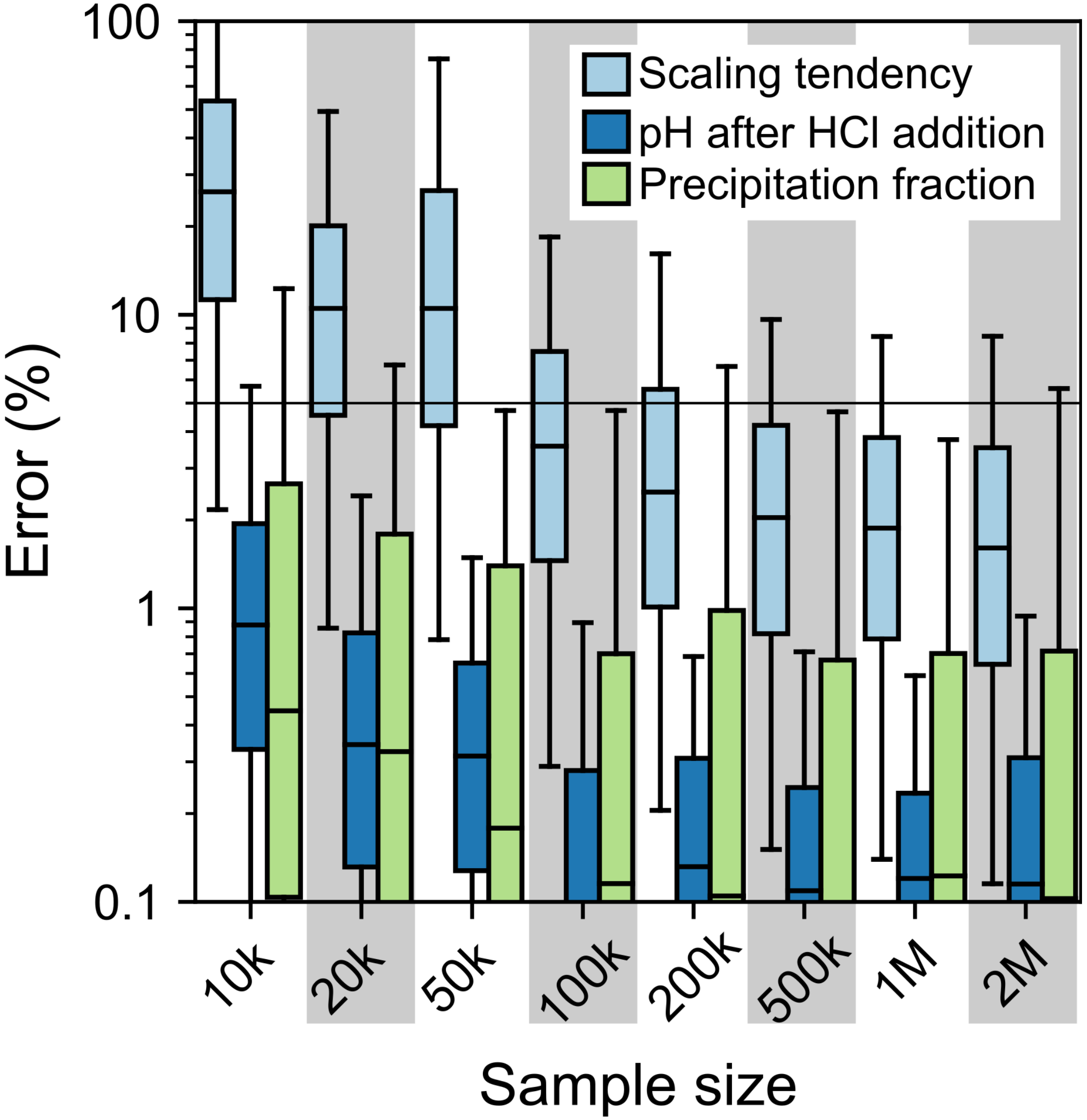
$$f_{exp-skew}(x) = (S^x - 1) * \frac{R_{high} - R_{low}}{S-1} + R_{low}$$

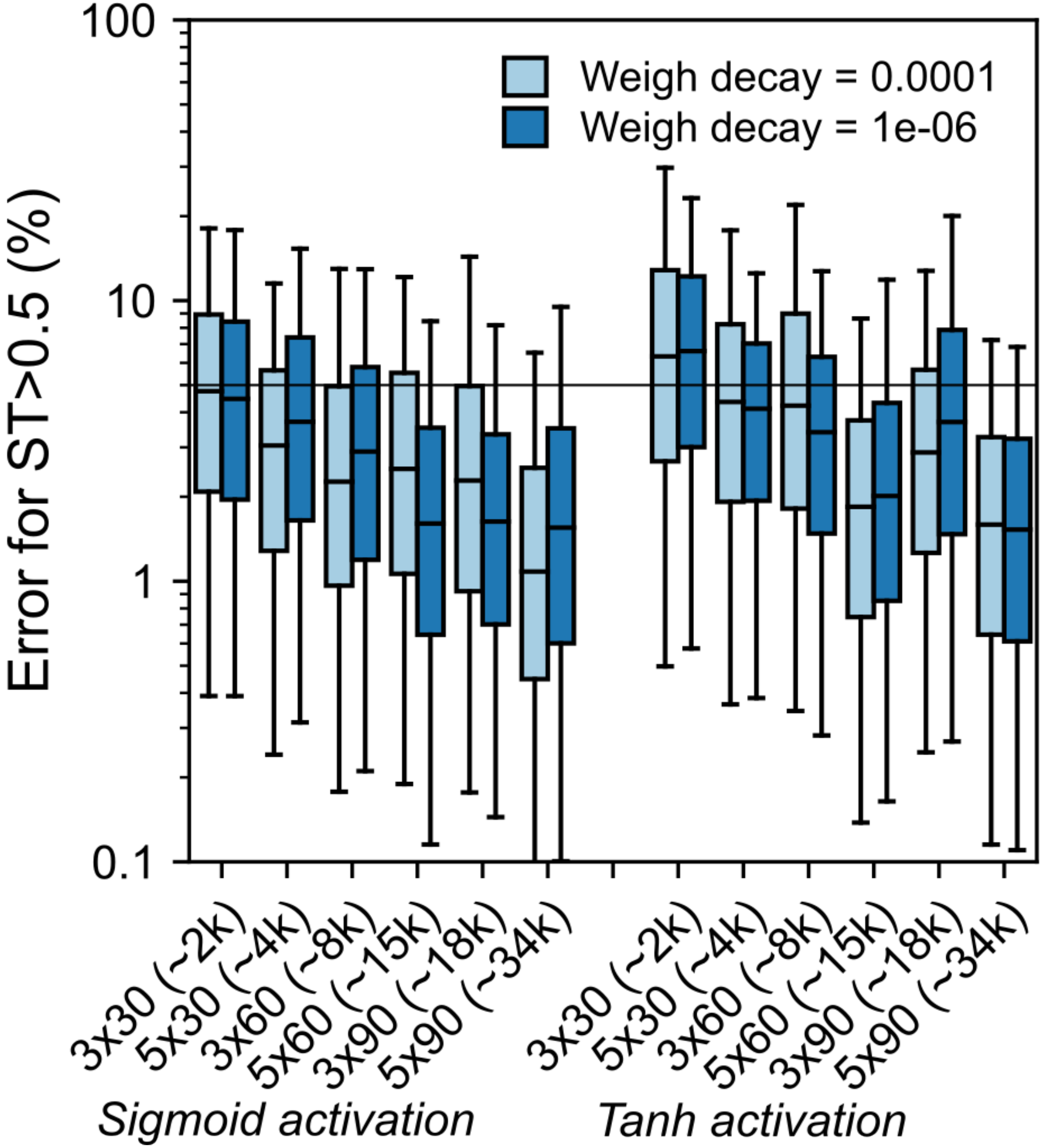




Dudchenko, A. V. & Amusat, O. *Systems & Control Transactions* (2024)

Data size and NN architecture play a "secondary" role in accuracy



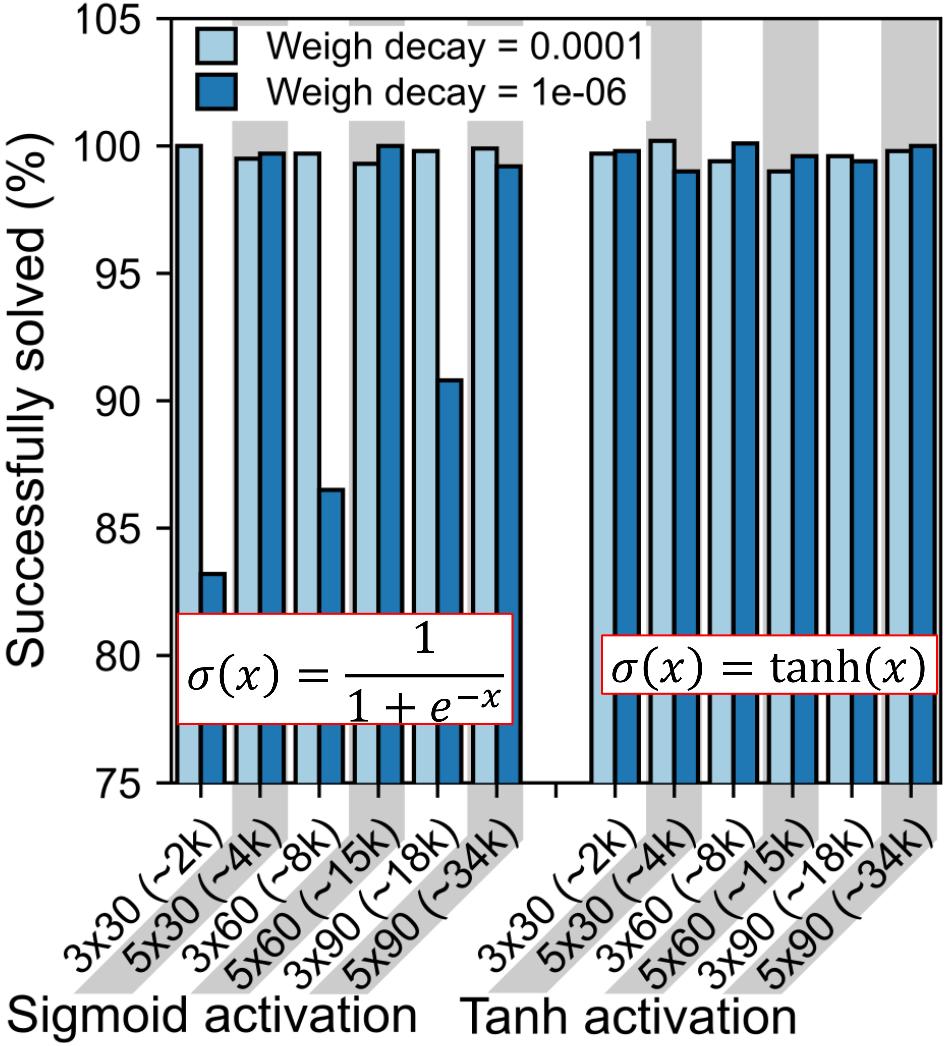


NAWI

Dudchenko, A. V. & Amusat, O. *Systems & Control Transactions* (2024)

Large networks and *tanh* are key for use in EO frameworks

NN accuracy tested against USGS brackish water data set (non-synthetic data)
Solvability tested using NNs in a black box desalination model using USGS brackish water data set
Solved using IPOPT with MA27 linear solver – tested 500 different feed compositions and 2 different guesses



Dudchenko, A. V. & Amusat, O. *Systems & Control Transactions* (2024)

NAWI

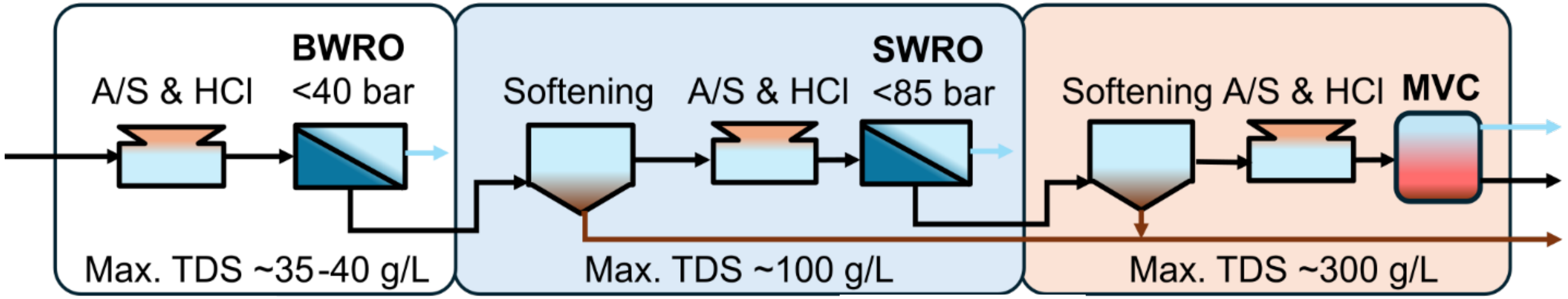
NNs enable assessment of complex treatment trains

PHREEQC ML models for:

- pH control
- precipitation
- mineral scaling

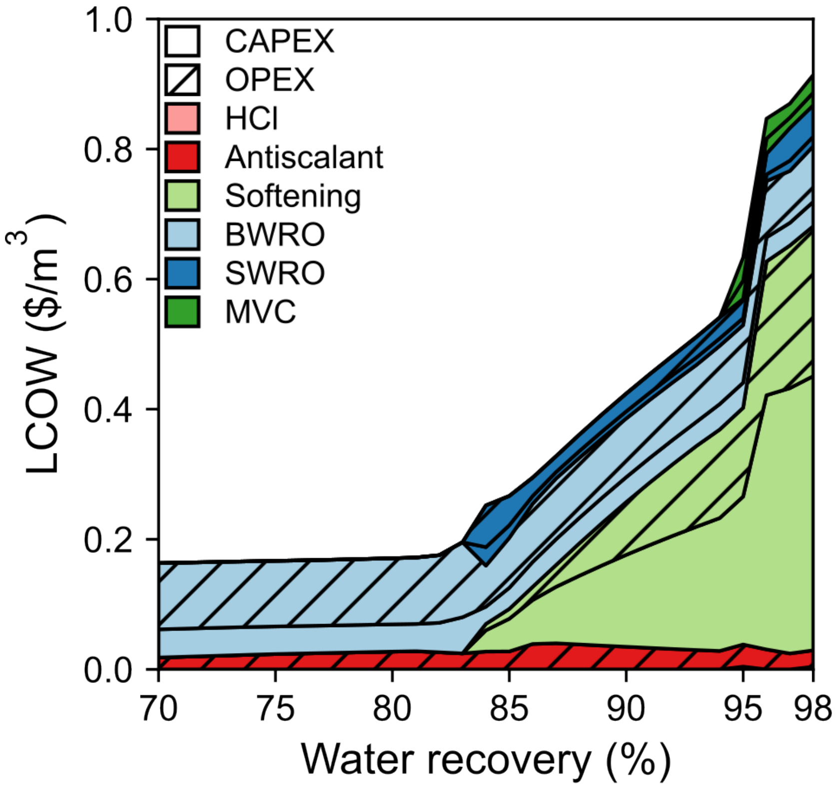
Flowsheets contained about 30 NN models with 30,000 parameters each

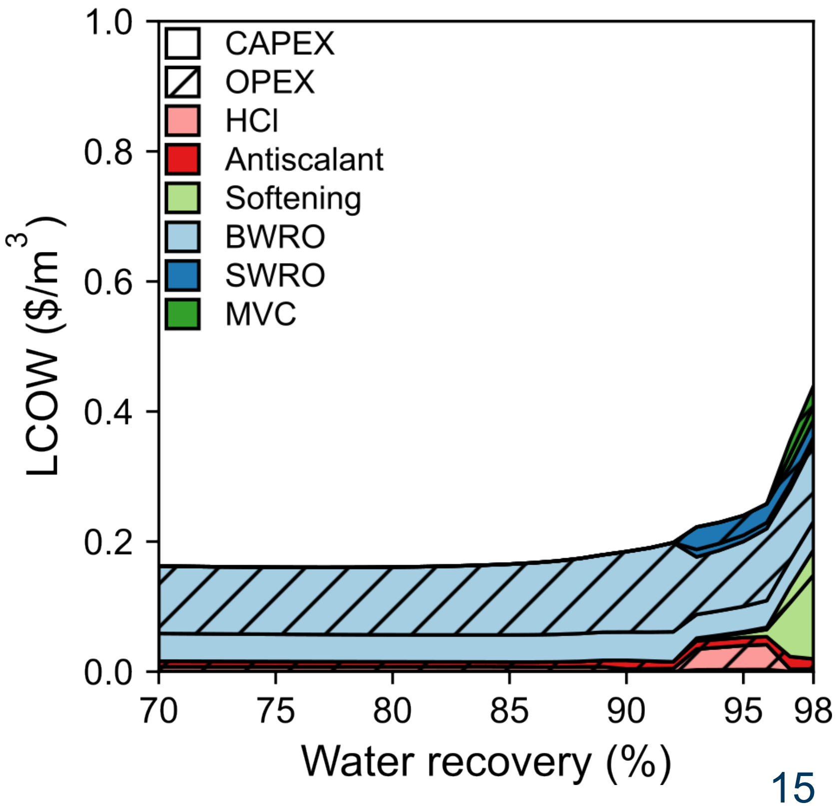
Solved in <5 min.



18 decision variables (degrees of freedom):

- 3 HCl acid doses, 3 antiscalant doses
- 2 lime doses, 2 soda ash doses
- 3 RO design and operating variables
- 2 MVC design and operating variables





Ion (mg/L)	Case 1	Case 2
Na	739	1120
Cl	870	1750
K	9	15
Ca	258	150
Mg	90	33
SO4	1011	260
HCO3	385	250
Sr	3	0.08
SiO2	25	30.5
TDS	3397	3609

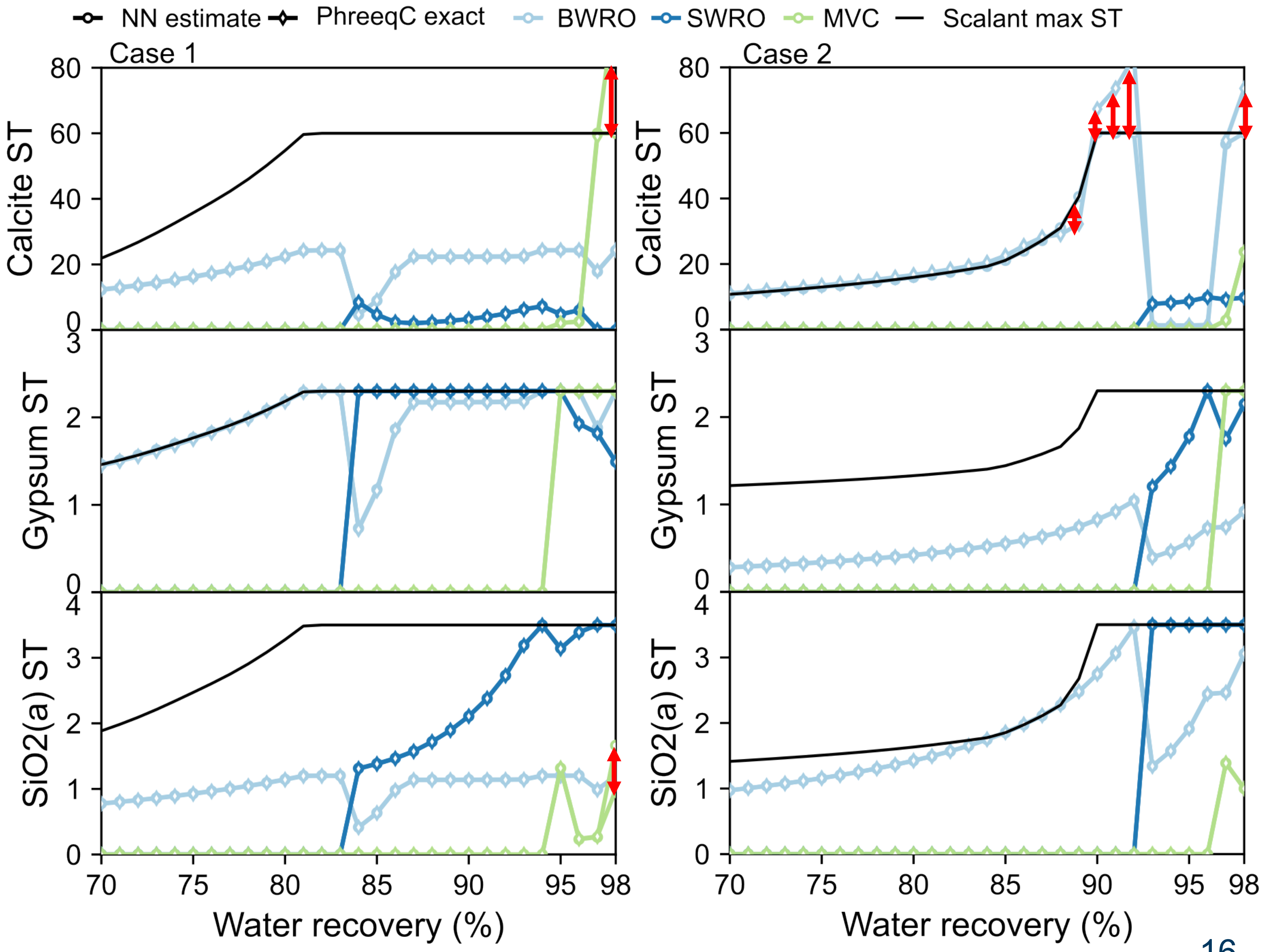
Are our networks also suggesting we add "glue" to fix water treatment?

NN accuracy for Scaling tendency *prediction*:

- Average error: 0.9%
- 95th percentile of error: ~5%
- 99th percentile of error: 18.5%

Out of 56 simulations, 6 points had poor estimates

***NNs provide great accuracy on "average" but can unpredictably and rapidly degrade in performance*.**



NAWI

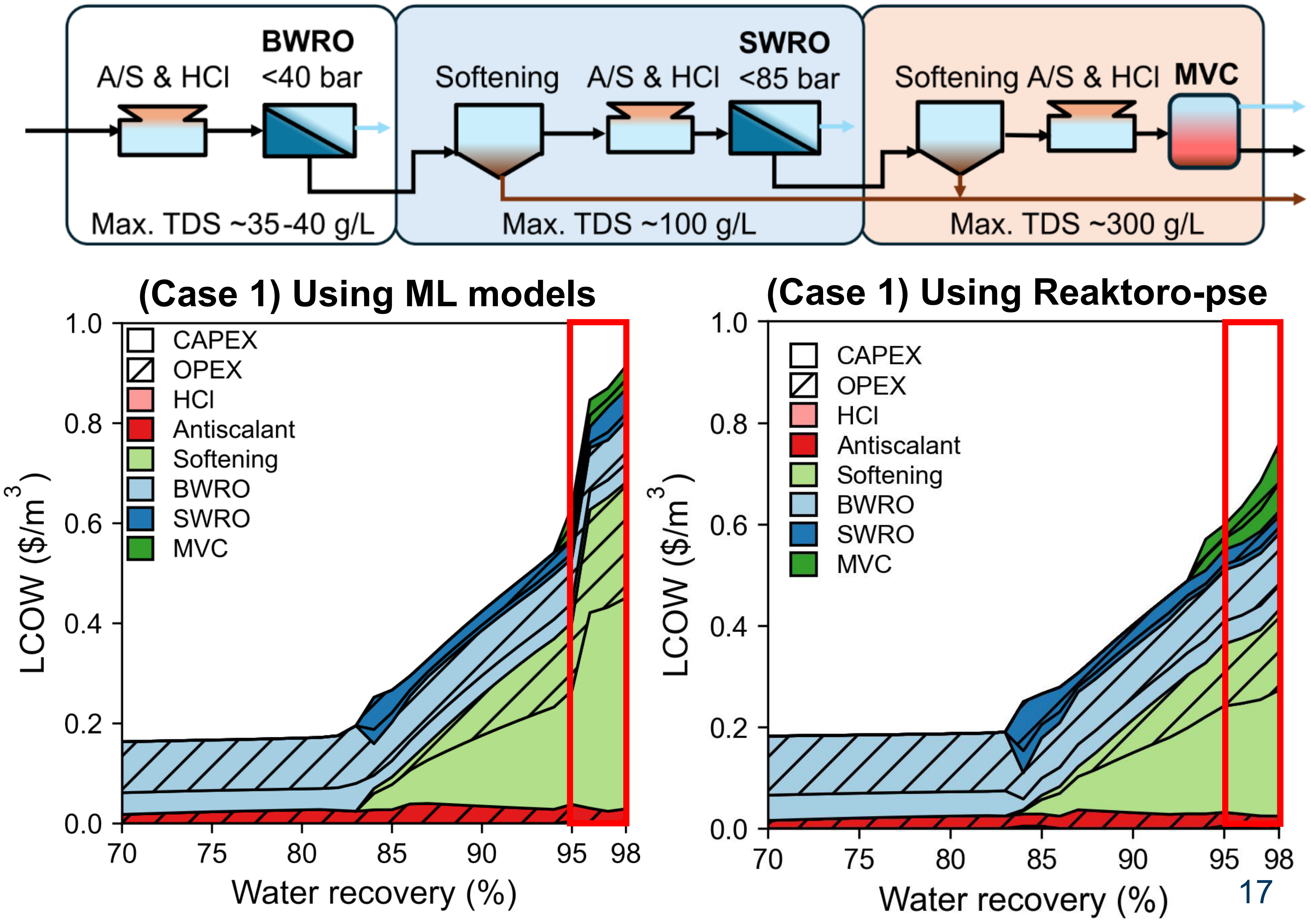
Reaktoro-pse enables exact calculation, enabling us for the first time to "fact check" the ML surrogates

Reaktoro-PSE integrates Reaktoro chemistry models directly into IDEAS and IDAES compatible libraries.

Reaktoro-PSE blocks are applied to estimate track changes in:

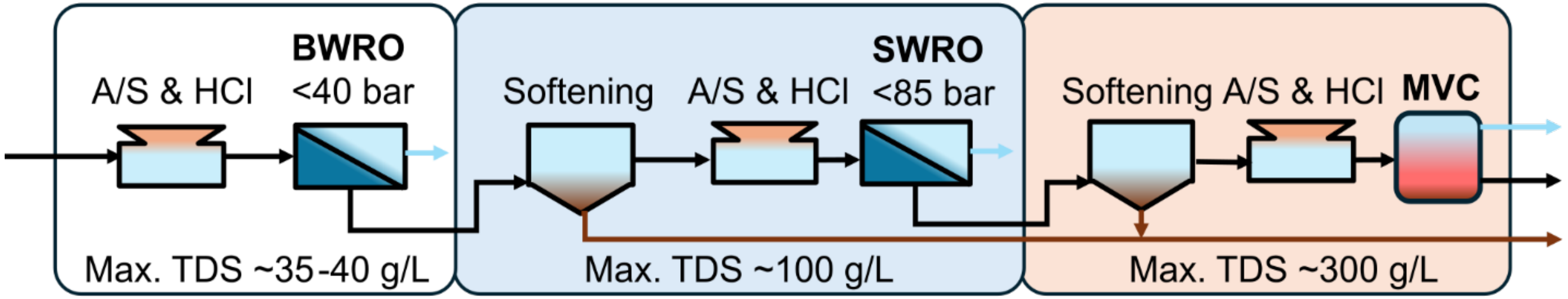
- pH
- Scaling tendencies
- Precipitation amount

Uses same database as ML models imitating them as closely as possible



Reaktoro-pse enables exact calculation, enabling us for the first time to "fact check" the ML surrogates

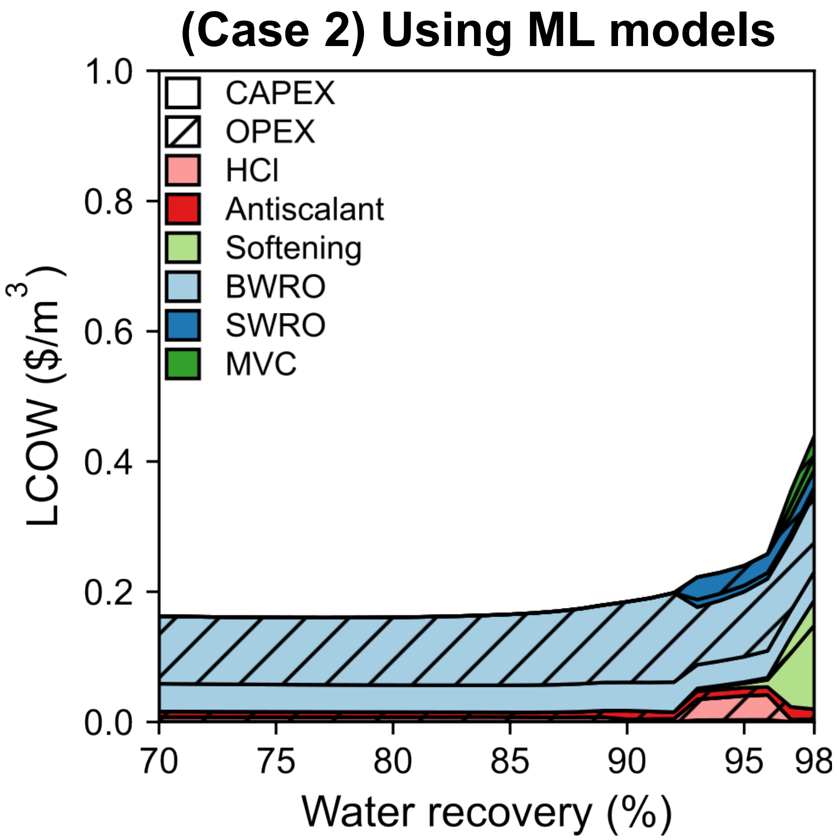
Reaktoro-PSE integrates Reaktoro chemistry models directly into IDEAS and IDAES compatible libraries.

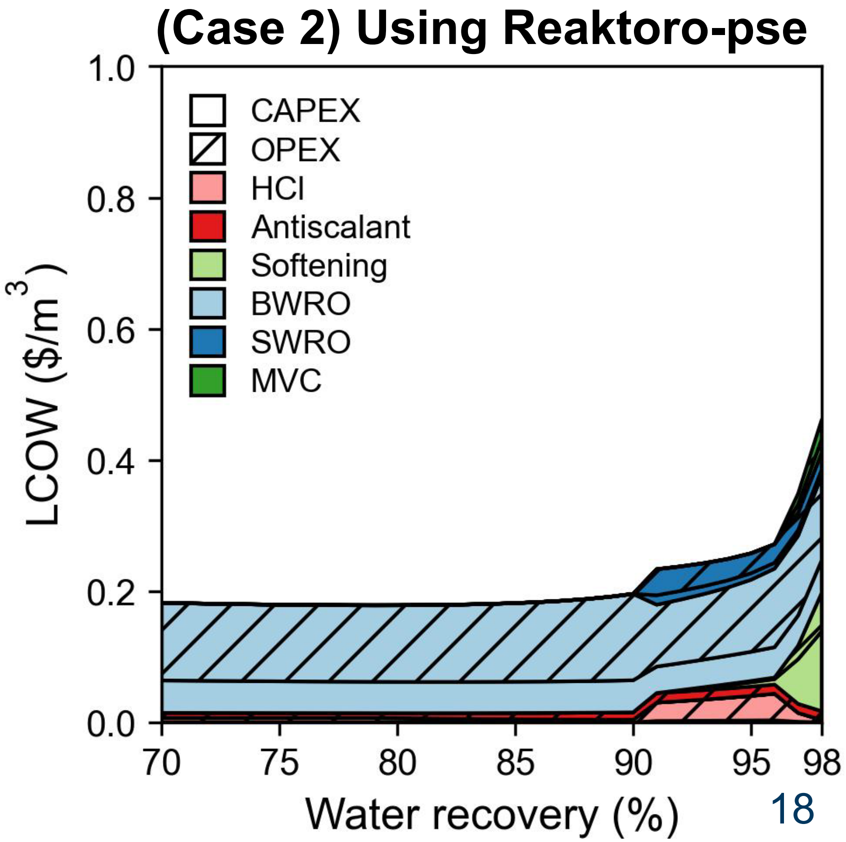


Reaktoro-PSE blocks are applied to estimate track changes in:

- pH
- Scaling tendencies
- Precipitation amount

Uses same database as ML models imitating them as closely as possible





NAWI

There is no “best” method, but Reaktoro-pse is a great starting point

	Narrow surrogates	Broad ML surrogates	Reaktoro-pse
Data quantity need	100-100,000 pts	500,000-1,000,000 pts	N/A
Data tailoring	None to high	High	N/A
Training time	10-600 seconds	>600 seconds	N/A
Computational intensity	Very low (1-2x increase)	Low to Mid (2-5x increase)	Mid to high (5-50x increase)
Stability in IPOPT	Medium (local minimum issues)	Medium (local minimum issues)	TBD (~preliminary stability is high, but sensitive to model and Jacobian scaling)
Error in estimates	~0-10% - depends on surrogates	~0-30% Depending on breadth of model and components Suffers from edge case errors	Exact solution

Thank you

- **National Energy Technology Laboratory:** David Miller, Tim Bartholomew, Markus Drouven, Andrew Lee, Andres Calderon-Vergara, Adam Atia, Chenyu Wang, Marcus Holly, Travis Arnold, Hunter Barber, Alejandro Garciadiego, Elmira Shamlou, Zhuoran Zhang, Savannah Sakhai

- **Lawrence Berkeley National Laboratory:** Deb Agarwal, Dan Gunter, Keith Beattie, Oluwamayowa Amusat, Jangho Park, Ludovico Bianchi, Jennifer Stokes-Draught, Xiangyu Bi, Michael Pesce

- **National Renewable Energy Laboratory:** Ben Knueven, Ethan Young, Jared Allen, Jordan Macknick, Kurby Sitterley, Kinshuk Panda, Zach Binger, Mukta Hardikar, Paul Vecchiarelli

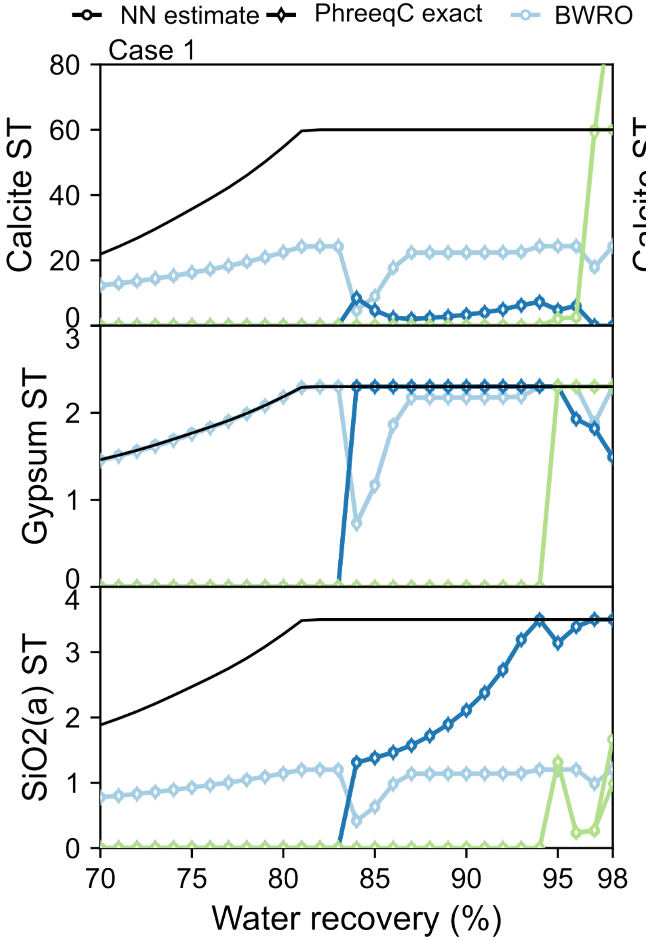
- **Oak Ridge National Laboratory:** Srikanth Allu, Austin Ladshaw, Johnson Dhanasekaran, Fahim Abdullah

- **SLAC National Accelerator Laboratory**: Alex Dudchenko

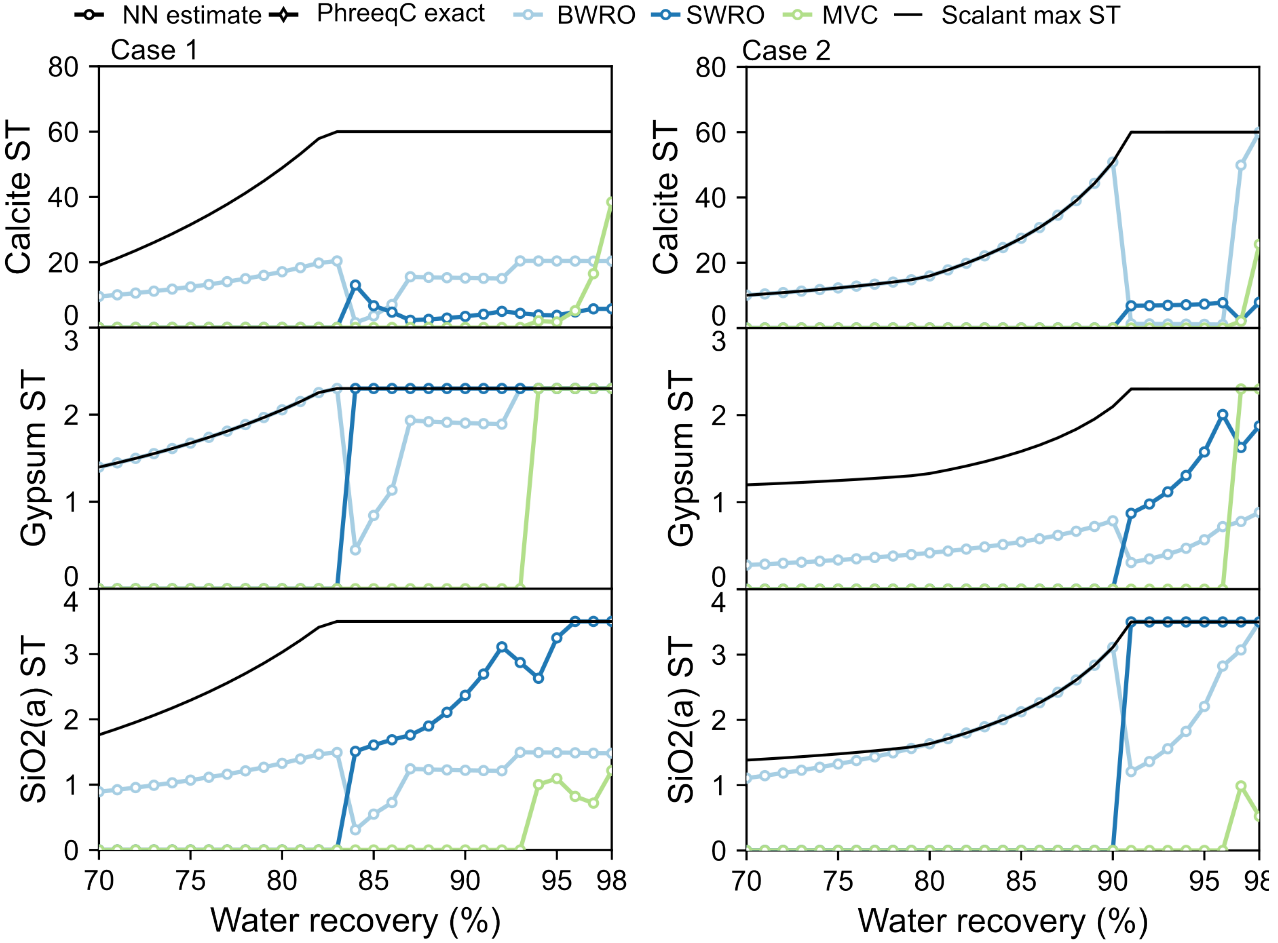
Thank you

NAWI

ML models



Reaktoro models



NAWI

Ion (mg/L)	Case 1	Case 2
Na	739	1120
Cl	870	1750
K	9	15
Ca	258	150
Mg	90	33
SO4	1011	260
HCO3	385	250
Sr	3	0.08
SiO2	25	30.5
TDS	3397	3609