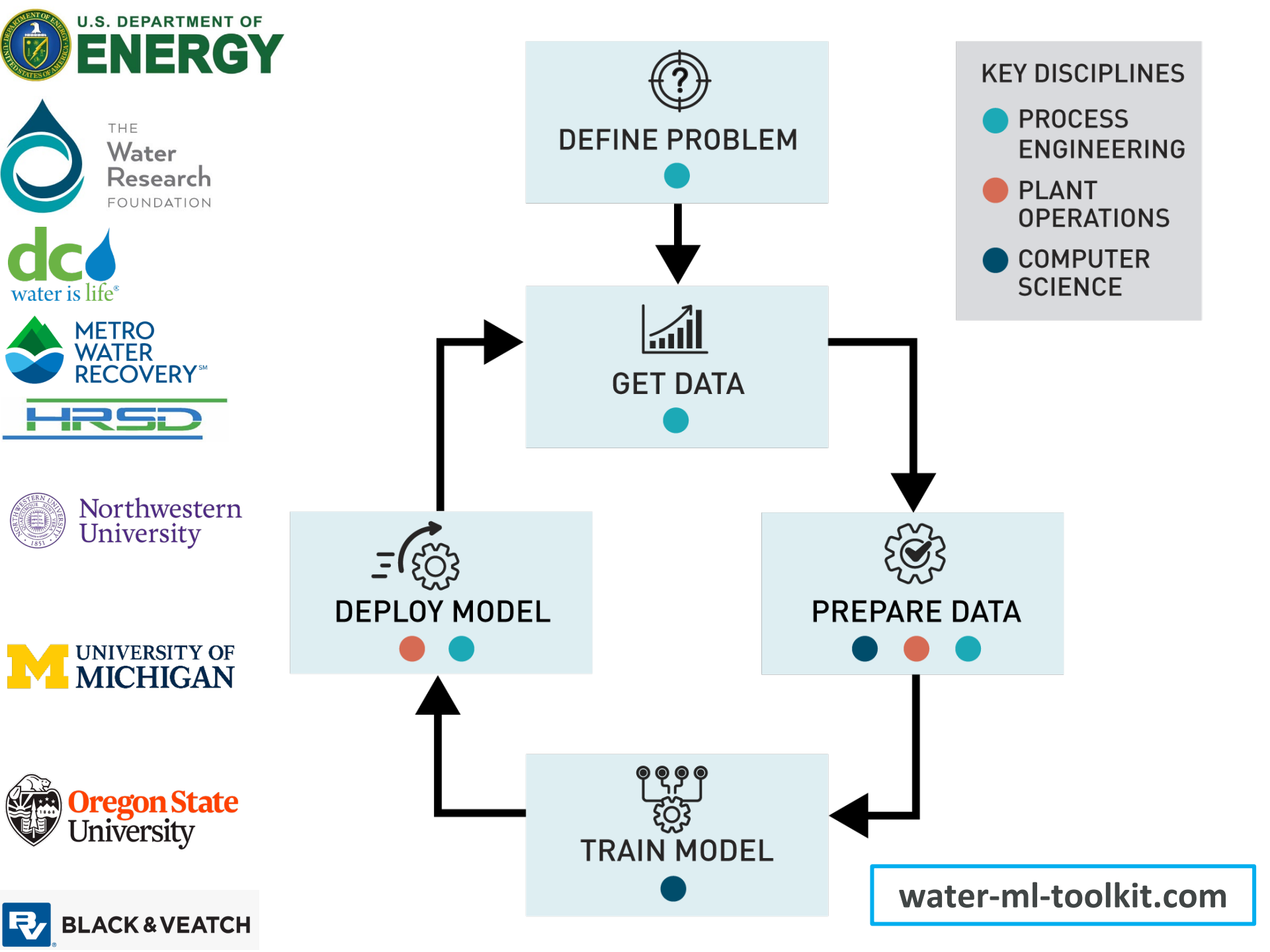
Modeling Phosphorous Removal Using Stakeholder-Centered Problem Solving

Dr. Kate Newhart Oregon State University

inspiring change

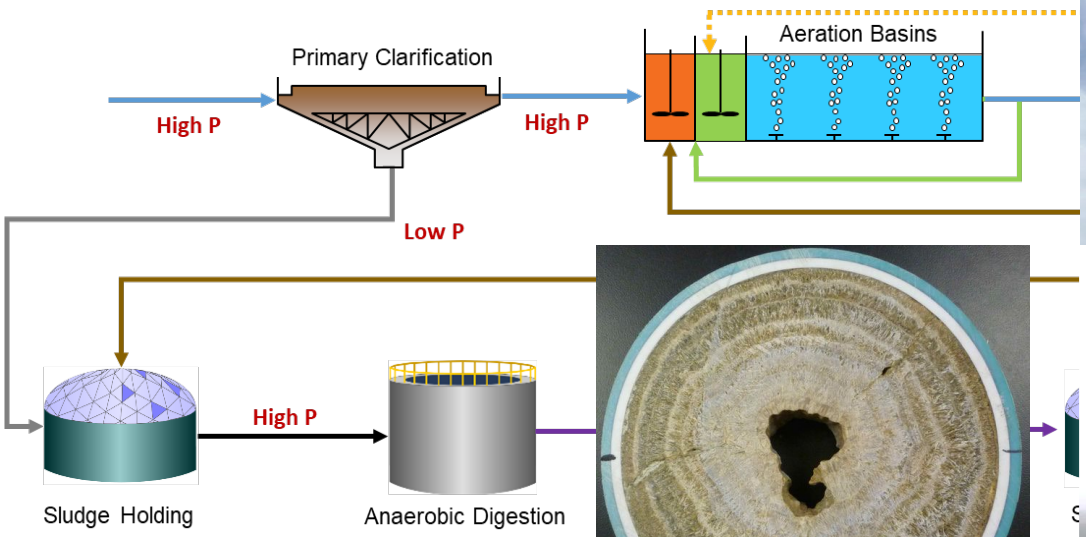

U.S. DEPARTMENT OF ENERGY
THE Water Research FOUNDATION
dc water is life
METRO WATER RECOVERY
HRSD
Northwestern University
UNIVERSITY OF MICHIGAN
Oregon State University
BLACK & VEATCH
DEFINE PROBLEM
GET DATA
PREPARE DATA
TRAIN MODEL
DEPLOY MODEL
KEY DISCIPLINES
PROCESS ENGINEERING
PLANT OPERATIONS
COMPUTER SCIENCE
water-ml-toolkit.com

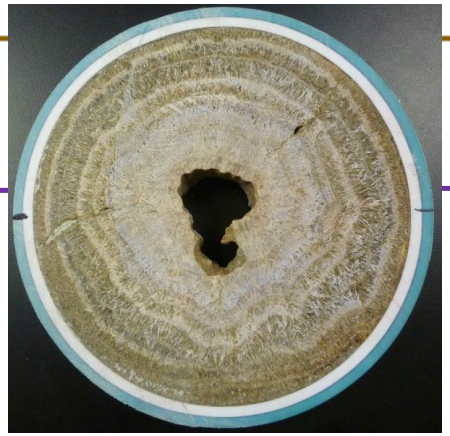
WHAT MAKES A "GOOD" ML PROBLEM?

- Specific to an **ML-appropriate objective**
  - Goal is to minimize energy, not to understand chemical kinetics
- **Measurable** performance criteria
  - X% improvement over status quo
- Implementation is feasible and **integrated into the model design**

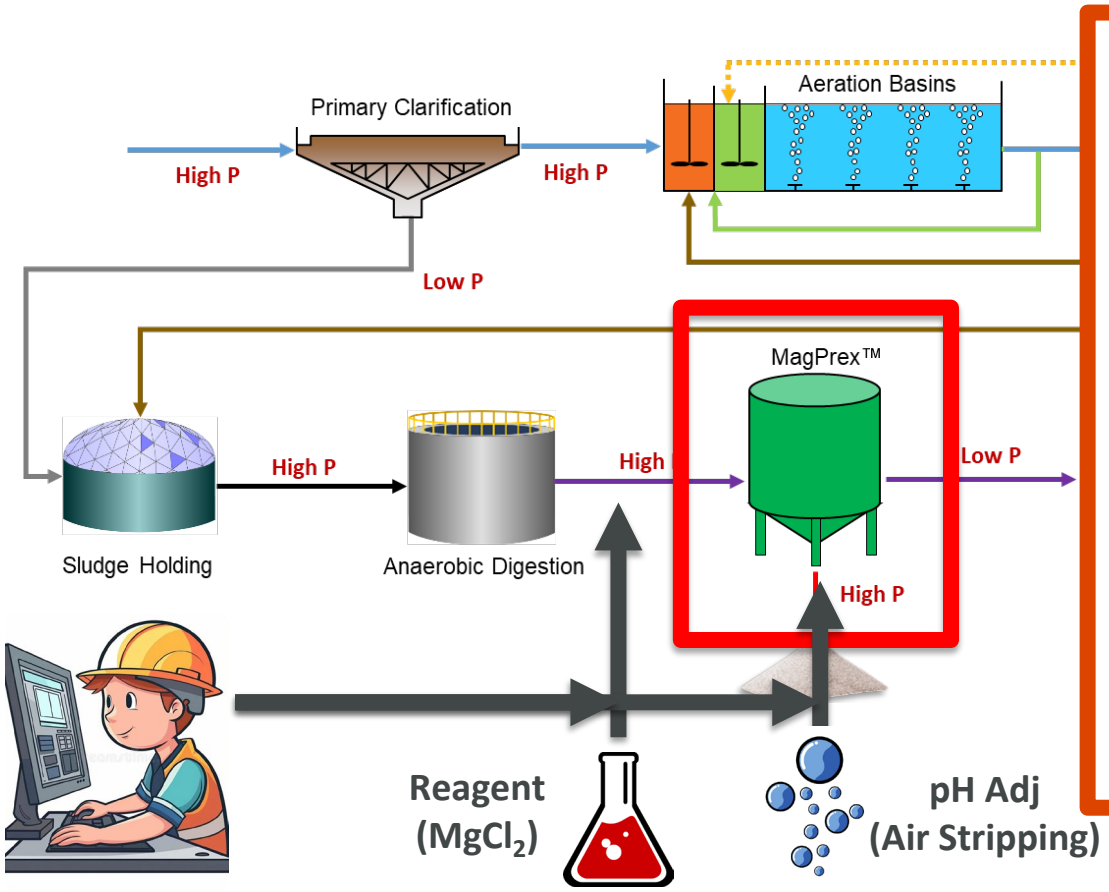
Yishi Zuo, 2021

PHOSPHORUS RECOVERY

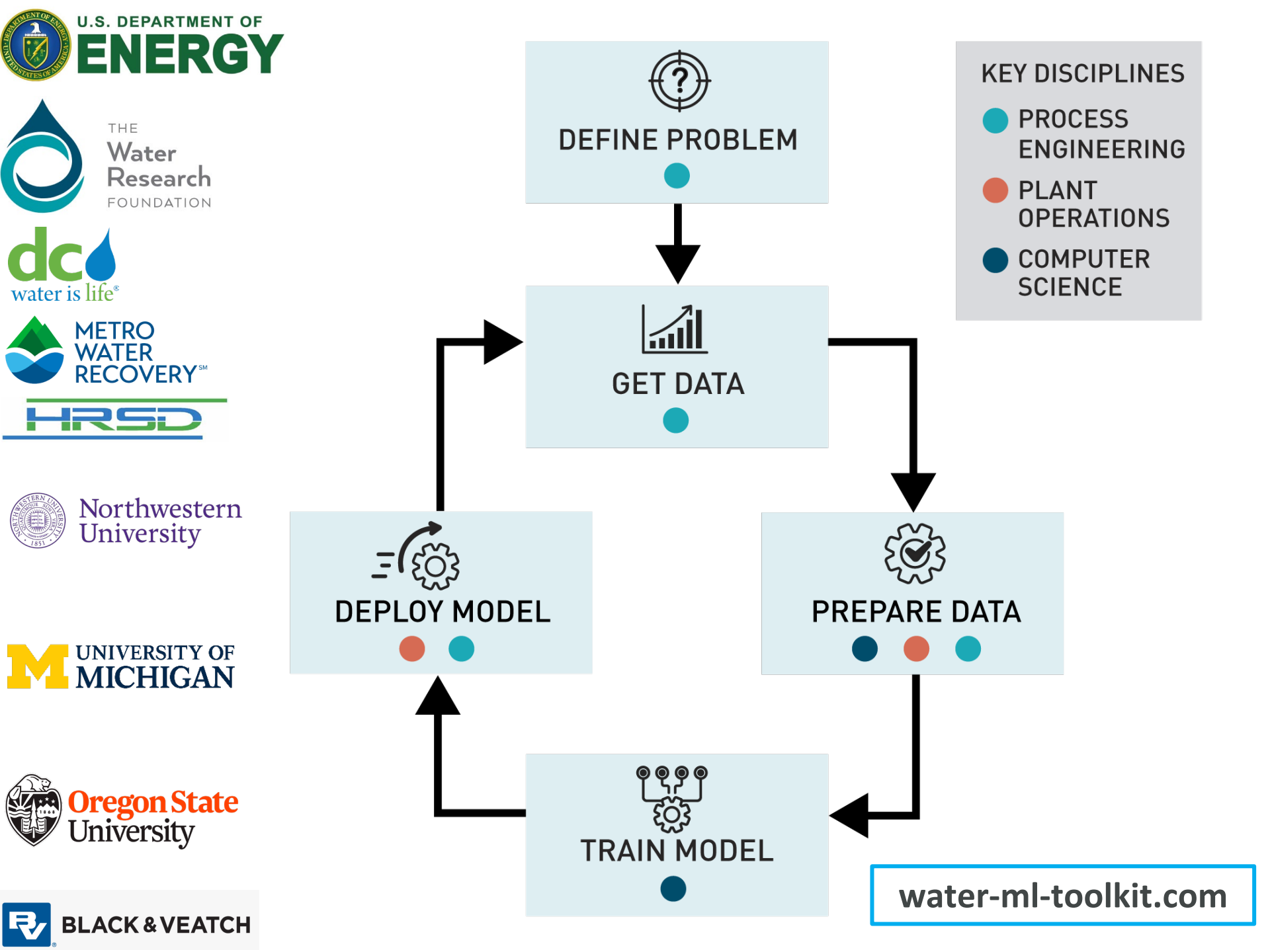




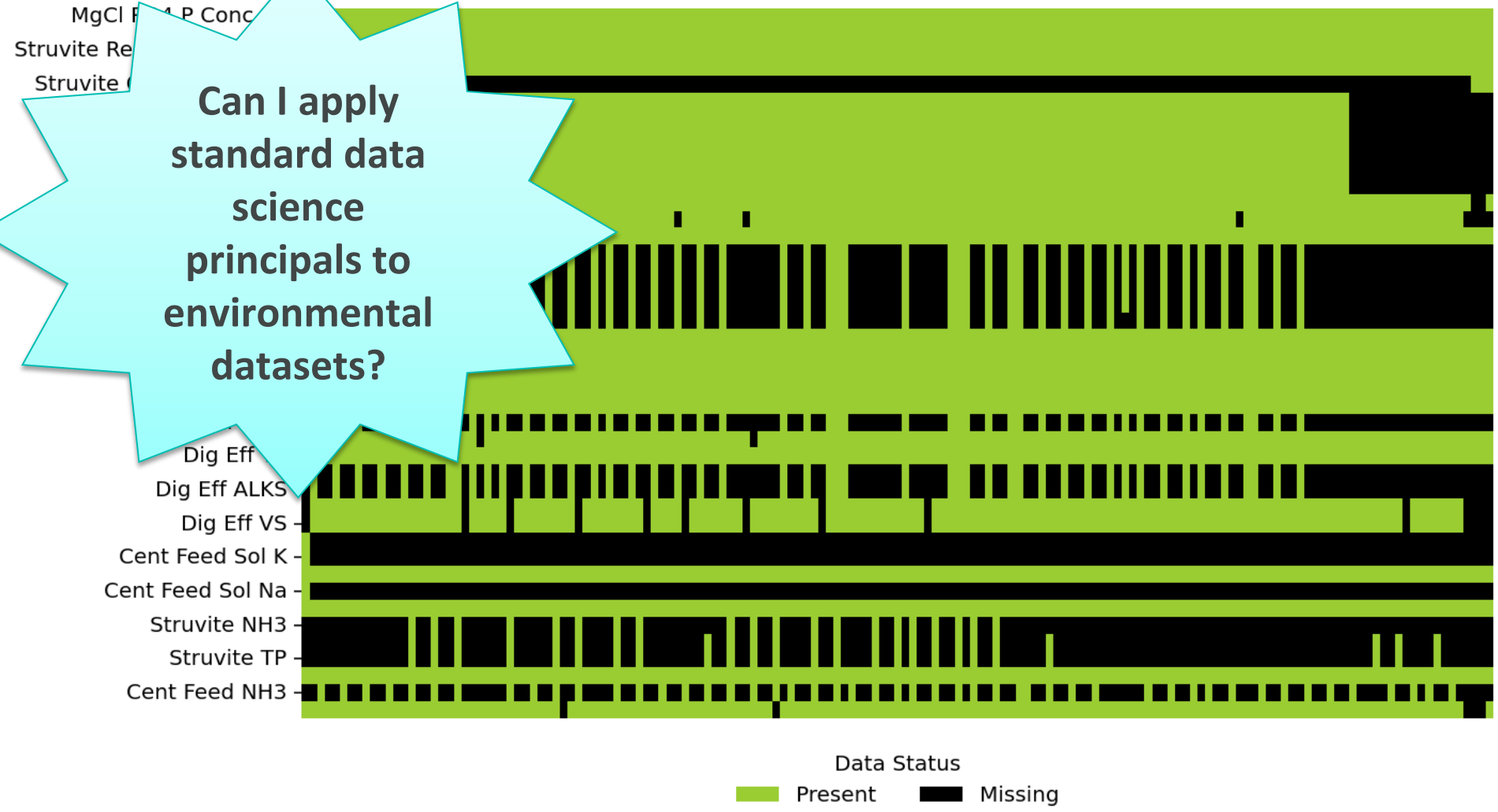
PHOSPHORUS RECOVERY



- Mass balance + reaction kinetics model is TERRIBLE
- Anaerobic digestate is NASTY. Sensors don't survive, samples are difficult to analyze
- Operators "guesstimate" reagent and pH setpoints

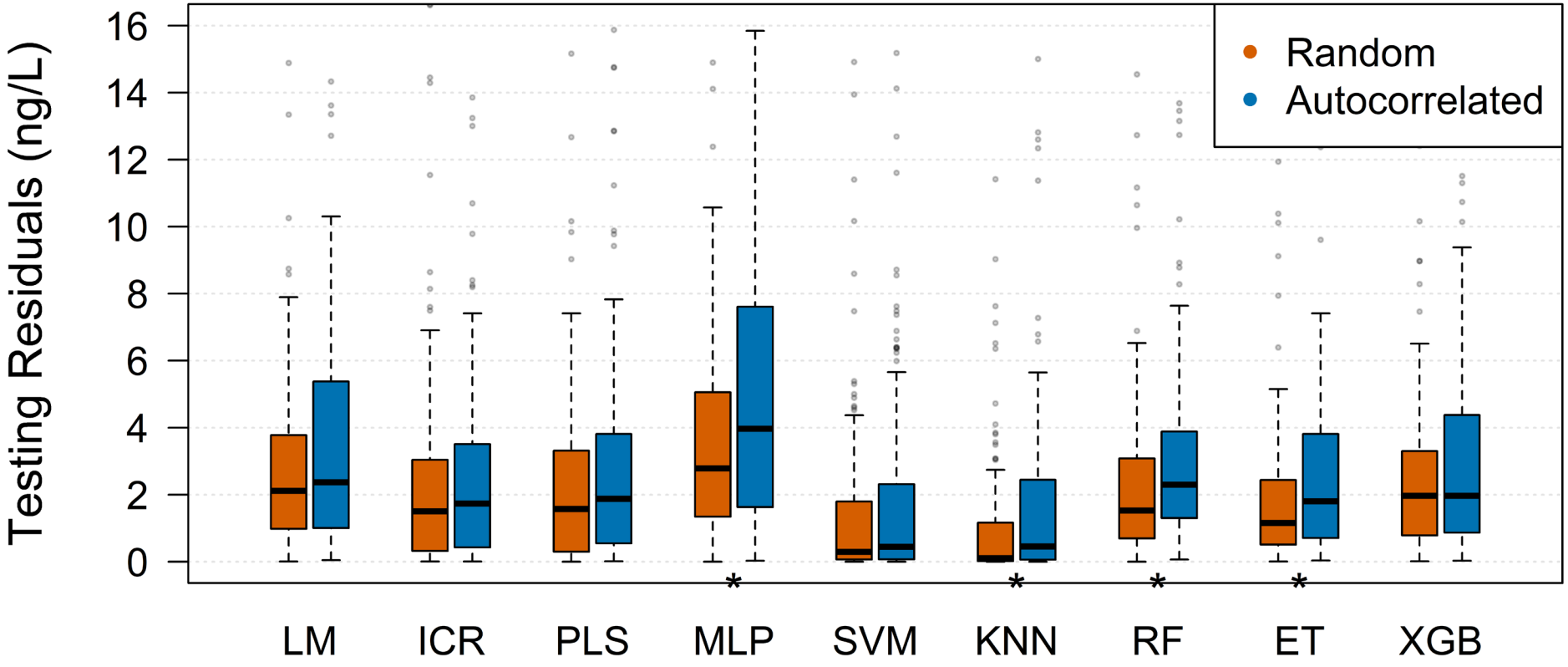

U.S. DEPARTMENT OF ENERGY
THE Water Research FOUNDATION
dc water is life
METRO WATER RECOVERY
HRSD
Northwestern University
UNIVERSITY OF MICHIGAN
Oregon State University
BLACK & VEATCH
DEFINE PROBLEM
GET DATA
PREPARE DATA
TRAIN MODEL
DEPLOY MODEL
KEY DISCIPLINES
PROCESS ENGINEERING
PLANT OPERATIONS
COMPUTER SCIENCE
water-ml-toolkit.com

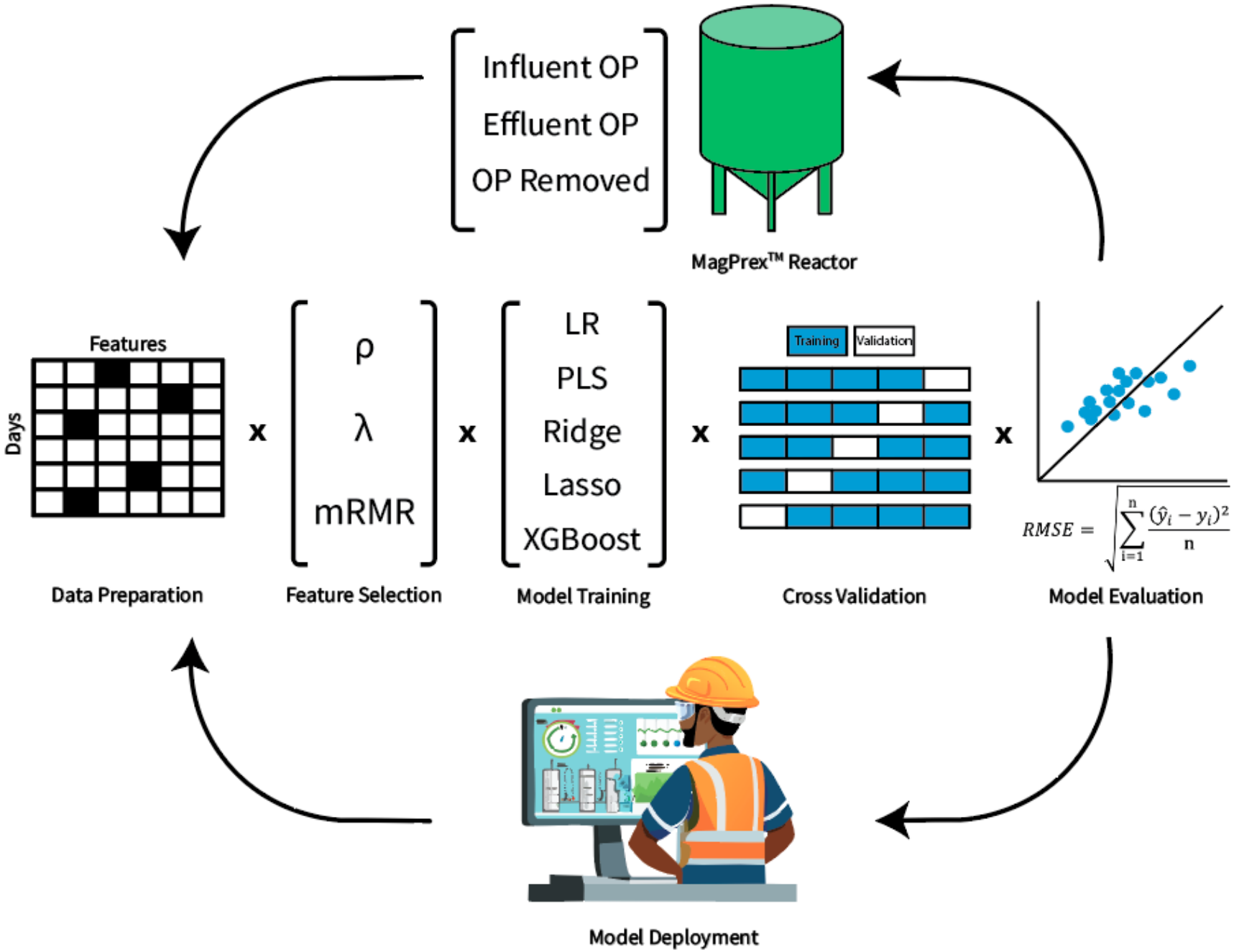
“I HAVE SOOOO MUCH DATA”

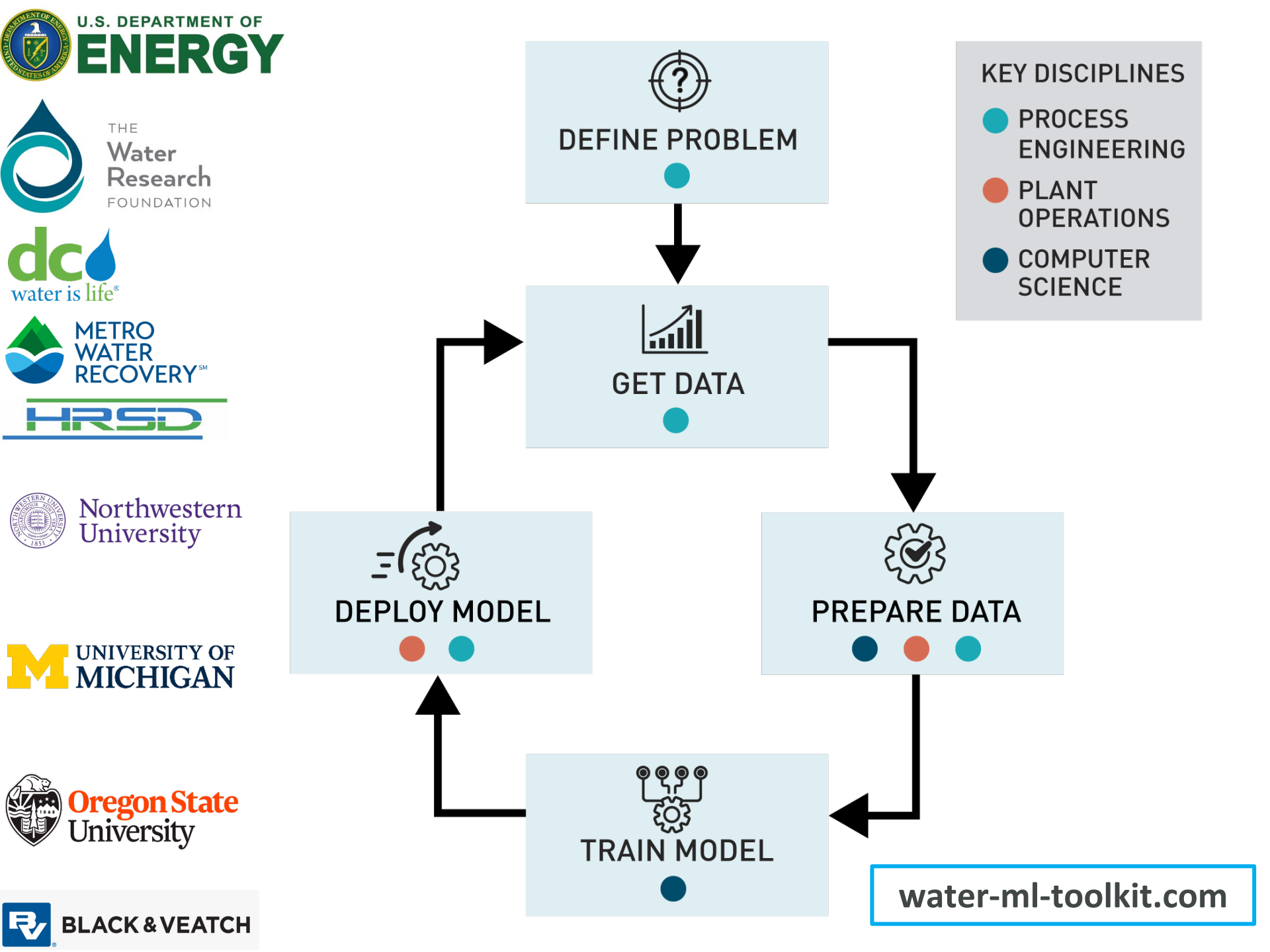


DIFFERENT PROJECT, SAME IDEA

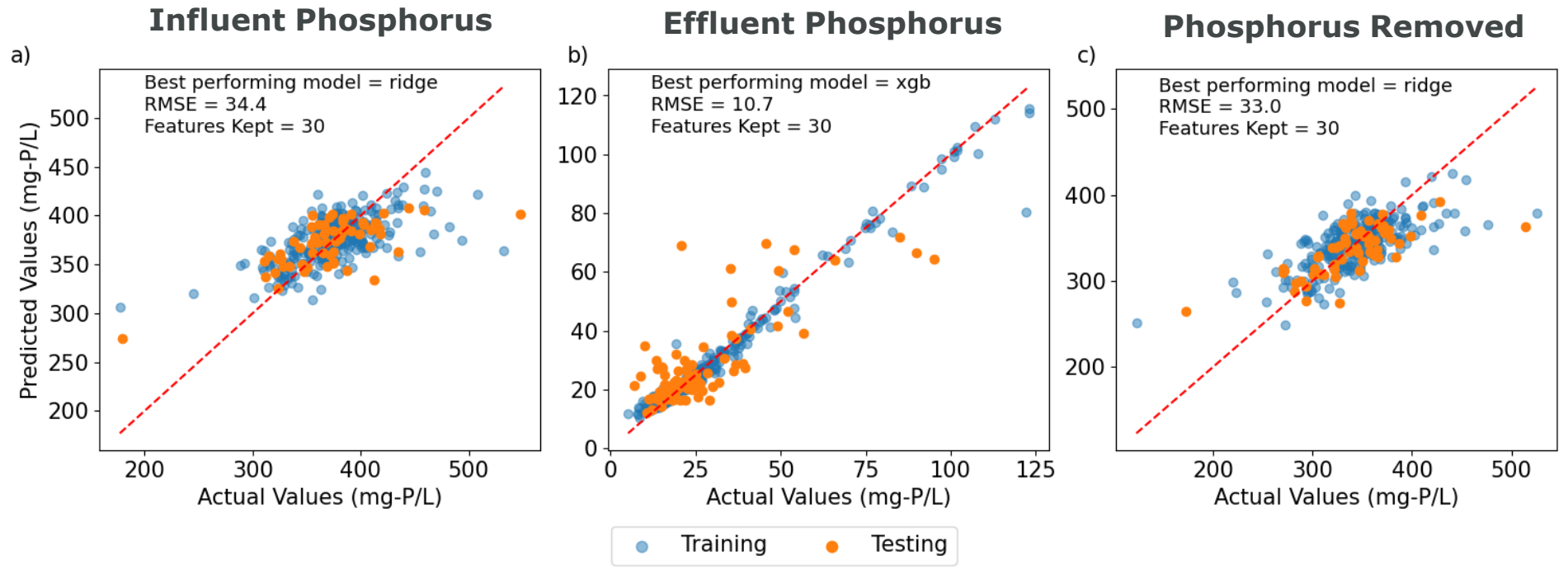
Environmental data is often time-series data. If taken frequently, data could be *autocorrelated* and standard data science practices like random splitting can lead to *data leakage* or artificially low errors.

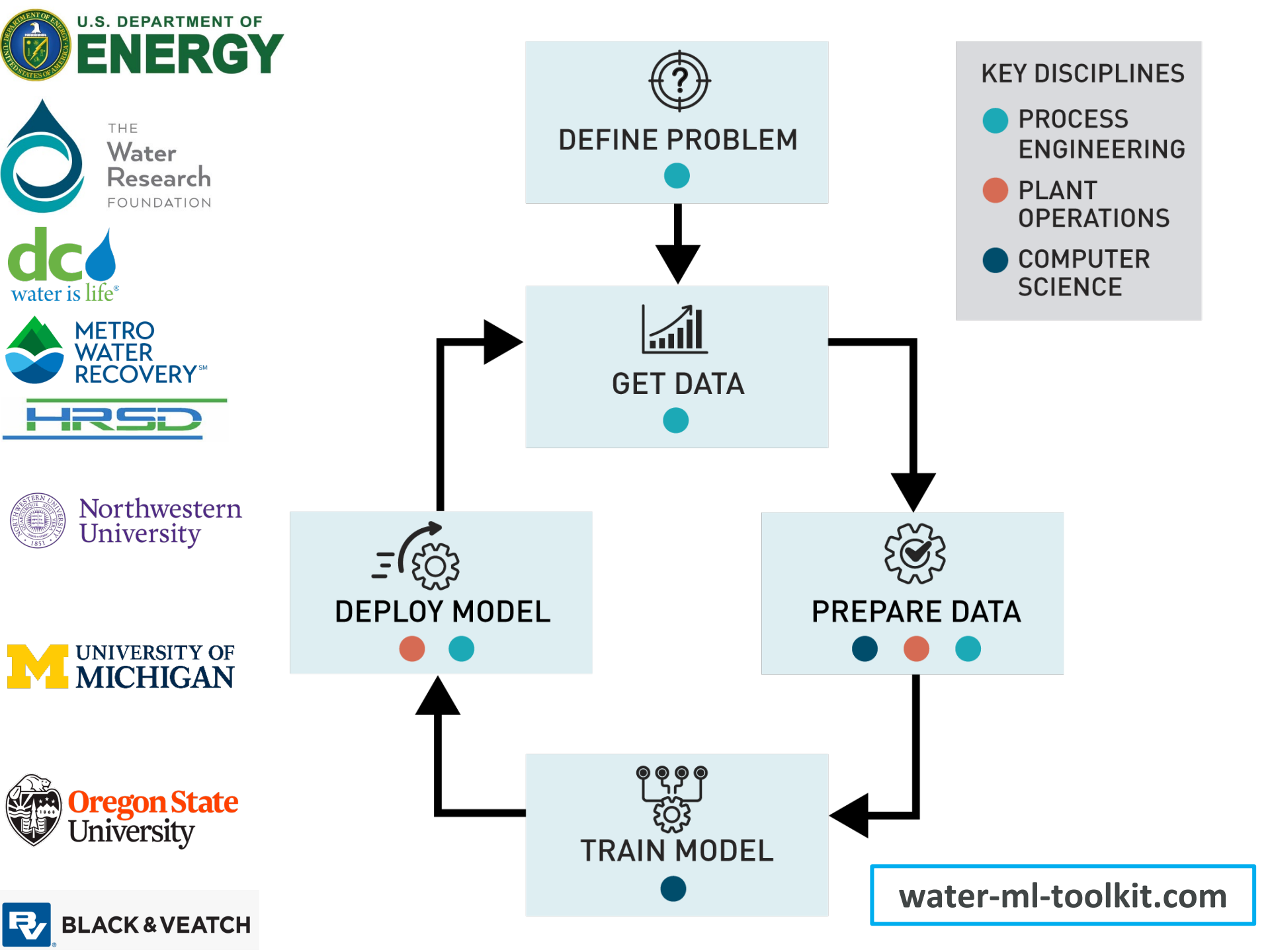


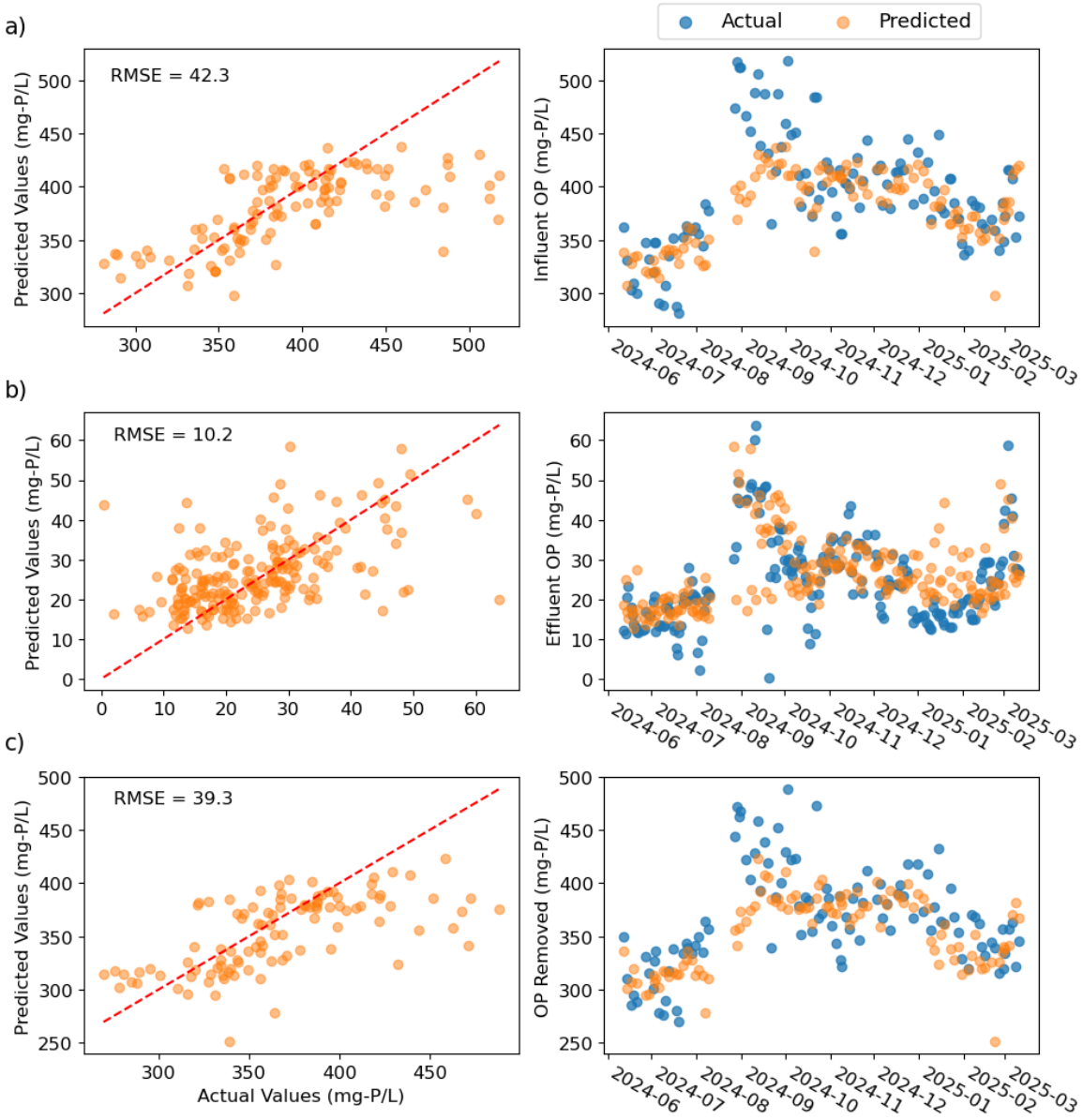

Influent OP
Effluent OP
OP Removed
MagPrex™ Reactor
Features
Days
x
ρ
λ
mRMR
x
LR
PLS
Ridge
Lasso
XGBoost
x
Training
Validation
x
$RMSE = \sqrt{\sum_{i=1}^{n} \frac{(\hat{y}_i - y_i)^2}{n}}$
Data Preparation
Feature Selection
Model Training
Cross Validation
Model Evaluation
Model Deployment


U.S. DEPARTMENT OF ENERGY
THE Water Research FOUNDATION
dc water is life
METRO WATER RECOVERY
HRSD
Northwestern University
UNIVERSITY OF MICHIGAN
Oregon State University
BLACK & VEATCH
DEFINE PROBLEM
GET DATA
PREPARE DATA
TRAIN MODEL
DEPLOY MODEL
KEY DISCIPLINES
PROCESS ENGINEERING
PLANT OPERATIONS
COMPUTER SCIENCE
water-ml-toolkit.com

BEST MODELS: 9% NRMSE




U.S. DEPARTMENT OF ENERGY
THE Water Research FOUNDATION
dc water is life
METRO WATER RECOVERY
HRSD
Northwestern University
UNIVERSITY OF MICHIGAN
Oregon State University
BLACK & VEATCH
DEFINE PROBLEM
GET DATA
PREPARE DATA
TRAIN MODEL
DEPLOY MODEL
KEY DISCIPLINES
PROCESS ENGINEERING
PLANT OPERATIONS
COMPUTER SCIENCE
water-ml-toolkit.com



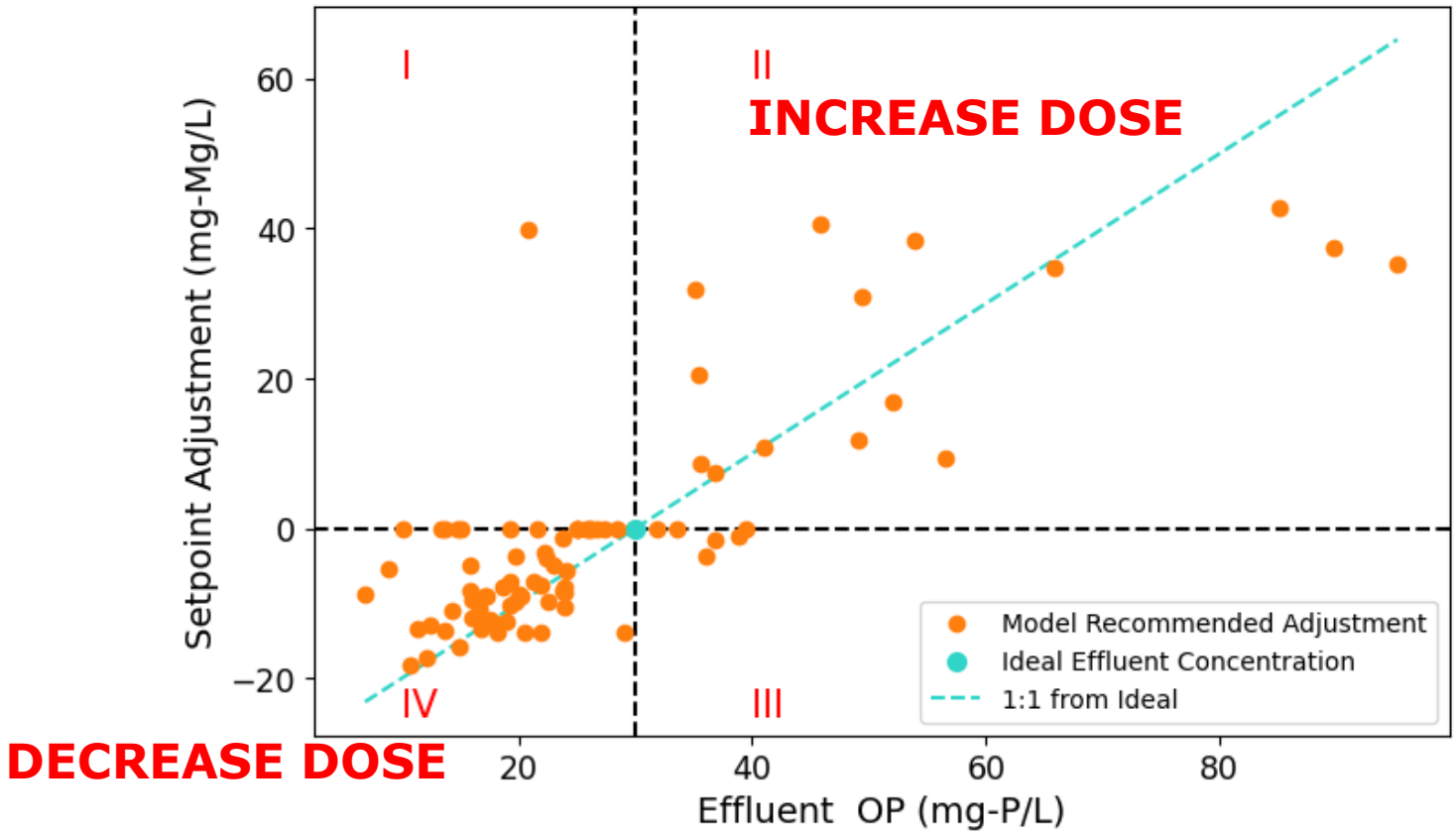
DEPLOY, IRL!

Influent Phosphorus

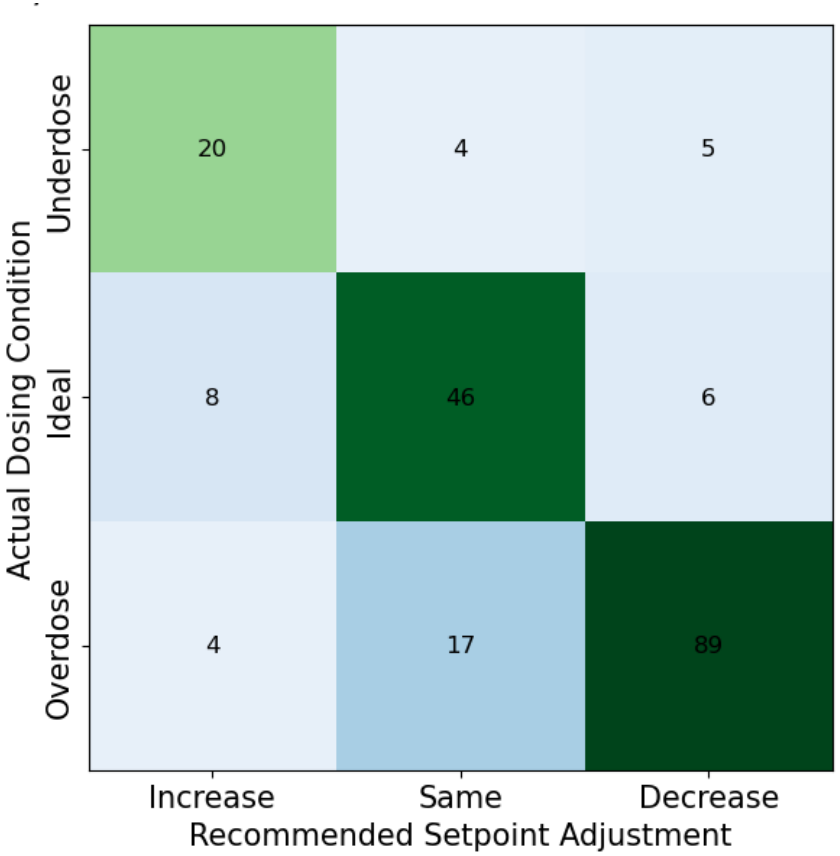
Effluent Phosphorus

Phosphorus Removed

BASED ON EFFLUENT PHOSPHORUS + INFLUENT CONDITIONS, WHAT SHOULD WE BE DOSING?



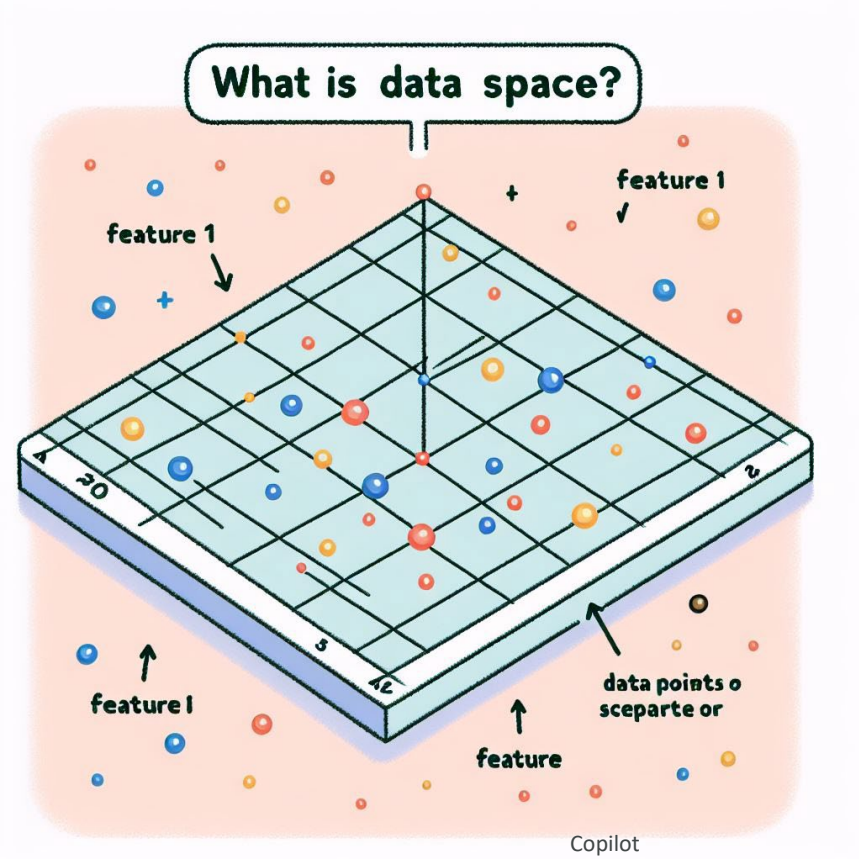
BASED ON EFFLUENT PHOSPHORUS + INFLUENT CONDITIONS, WHAT SHOULD WE BE DOSING?

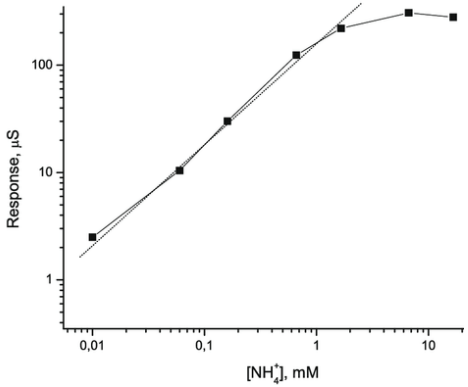


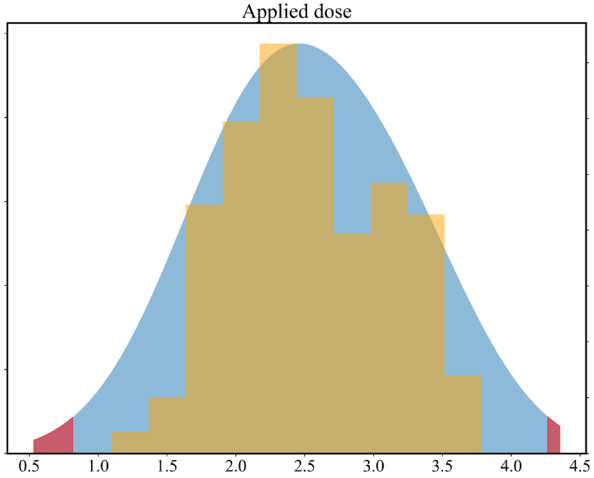
Bin setpoint adjustments: increase, decrease, same

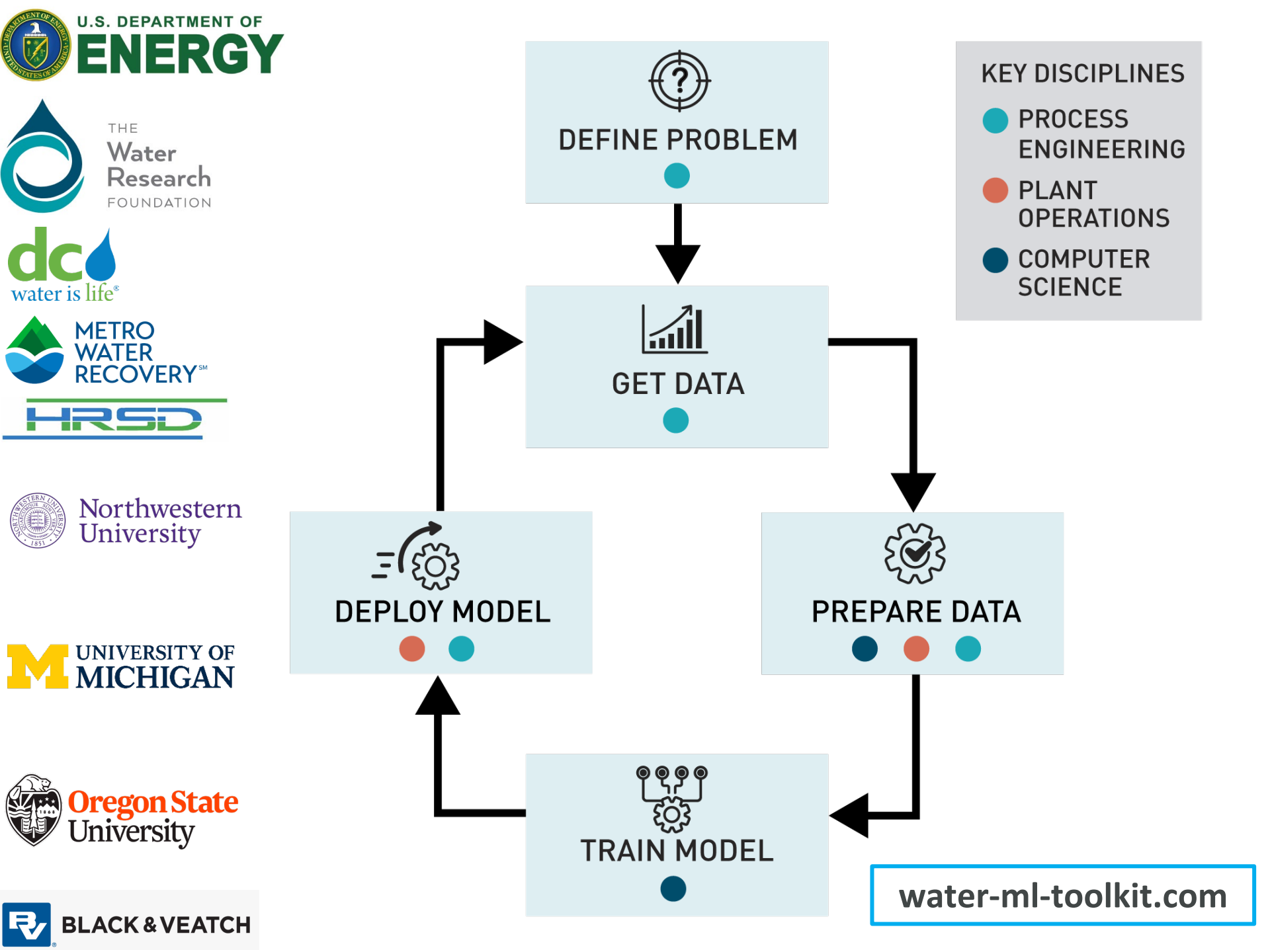
Correct 97% of the time

DATA SPACE


What is data space?
feature 1
feature 1
feature 1
feature
data points o
sceparte or
Copilot


Response, µS
100
10
1
0,01
0,1
1
10
[NH4+], mM


Applied dose
0.5
1.0
1.5
2.0
2.5
3.0
3.5
4.0
4.5


U.S. DEPARTMENT OF ENERGY
THE Water Research FOUNDATION
dc water is life
METRO WATER RECOVERY
HRSD
Northwestern University
UNIVERSITY OF MICHIGAN
Oregon State University
BLACK & VEATCH
DEFINE PROBLEM
GET DATA
PREPARE DATA
TRAIN MODEL
DEPLOY MODEL
KEY DISCIPLINES
PROCESS ENGINEERING
PLANT OPERATIONS
COMPUTER SCIENCE
water-ml-toolkit.com

WHAT HAVE WE LEARNED?

- Machine learning models can be used to capture non-ideal / highly non-linear process behavior
- Data are your solution… but also your problem
  - Environmental datasets are smaller than you think!
  - Beware artificially low errors / high accuracy!
- Put on your experimentalist hat!
  - No universal model, data preparation, error metric, etc.
  - Be creative and **user-centric!**

- This could be you!
- Now hiring PhD students!