



Metropolitan Water Reclamation District of Greater Chicago

**Welcome to the June
Edition of the 2023 M&R
Seminar Series**

NOTES FOR SEMINAR ATTENDEES

- Remote attendees' audio lines have been muted to minimize background noise. **For attendees in the auditorium, please silence your phones.**
- A question and answer session will follow the presentation.
- For remote attendees, Please use the “**Chat**” feature to ask a question via text to “Host”. **For attendees in the auditorium, please raise your hand and wait for the microphone to ask a verbal question.**
- The presentation slides will be posted on the MWRD website after the seminar.
- This seminar has been approval by the ISPE for one PDH and has been approved by the IEPA for one TCH. Certificates will only be issued to participants who attend the entire presentation.

Dr. Kathryn (Kate) B. Newhart
Assistant Professor of Environmental Engineering
United States Military Academy
West Point, New York



Dr. Kate Newhart is an Assistant Professor of Environmental Engineering with a specialization in data-driven monitoring and control for the water sector. Dr. Newhart holds a Ph.D. in Civil and Environmental Engineering from the Colorado School of Mines and previously worked at Metro Water Recovery, the largest wastewater utility in the Rocky Mountains, as an engineer in their Technology and Innovation Division, where she led data handling and data-driven modeling efforts. Dr. Newhart's current work includes soft-sensor development for disinfection, resource recovery, and disinfection by-product mitigation in collaboration with utilities and universities across North America.



UNITED STATES MILITARY ACADEMY
WEST POINT

Understanding and Embracing Machine Learning and Predictive Analytics

Dr. Kate Newhart

Assistant Professor of Environmental Engineering

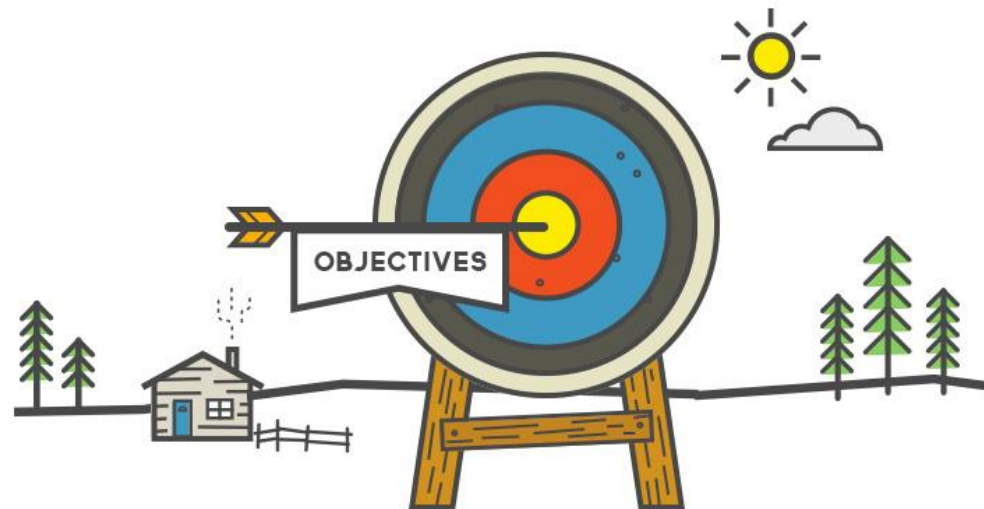
United States Military Academy

The views expressed herein do not necessarily represent the views of the U.S. Department of Defense, the Army, or the United States Government, and does not imply or constitute endorsement.



Objectives

- Define artificial intelligence (AI), machine learning (ML), digital twin, and soft sensor
- Explain the process of developing an ML model
- Understand how a machine learning model can be deployed, used, and maintained at a water utility





UNITED STATES MILITARY ACADEMY
WEST POINT

WHAT IS ARTIFICIAL INTELLIGENCE?



A Proposal for the

DARTMOUTH SUMMER RESEARCH PROJECT ON ARTIFICIAL INTELLIGENCE

June 17 - Aug. 16

We propose that a 2 month, 10 man study of artificial intelligence be carried out during the summer of 1956 at Dartmouth College in Hanover, New Hampshire. The study is to proceed on the basis of the conjecture that every aspect of learning or any other feature of intelligence can in principle be so precisely described that a machine can be made to simulate it. An attempt will be made to find how to make machines use language, form abstractions and concepts, solve kinds of problems now reserved for humans, and improve themselves. We



What is Artificial Intelligence?

“AI” is a computer system that can perform tasks that typically require human intelligence (e.g., learning, reasoning, perception, decision-making)

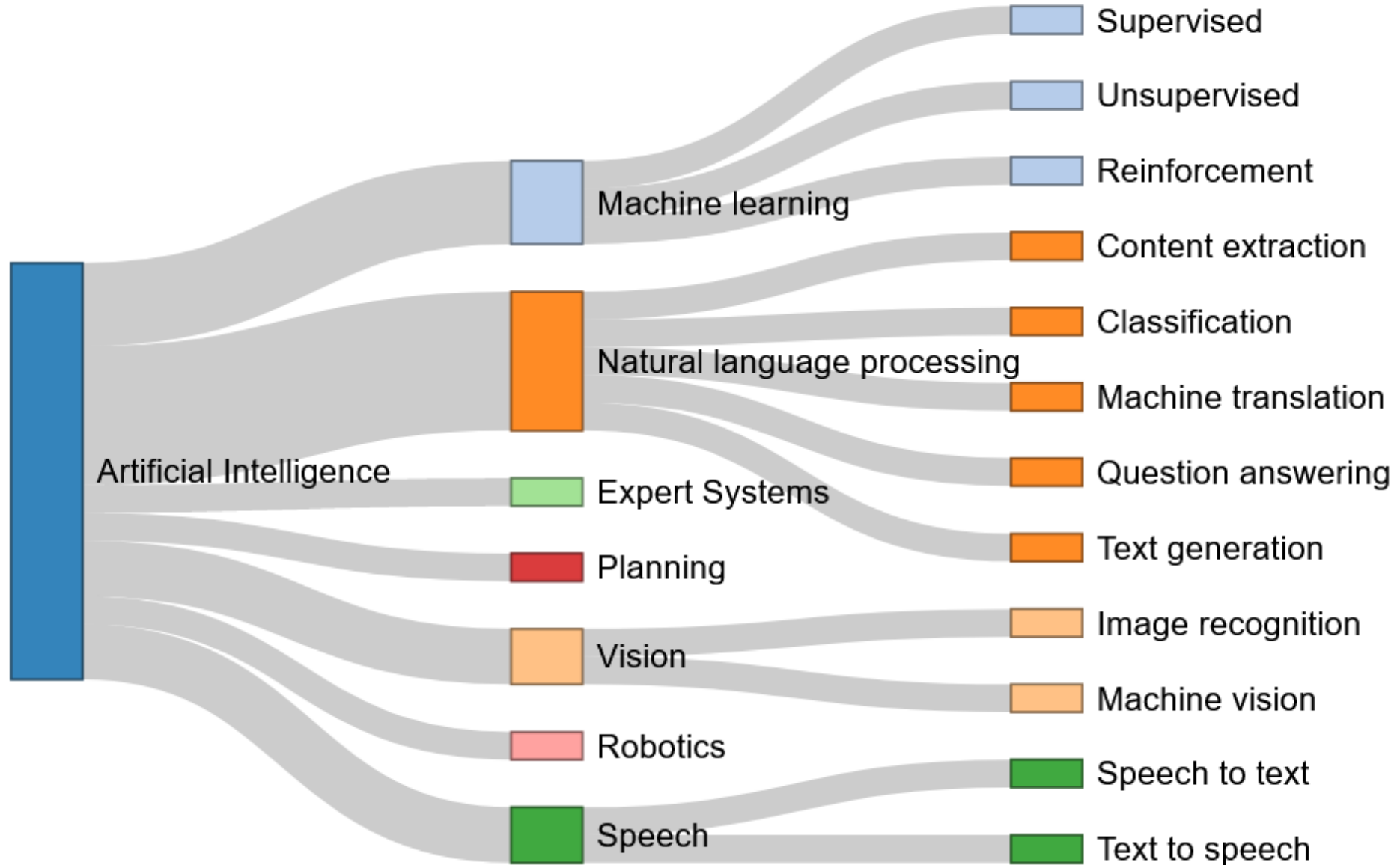


NETFLIX





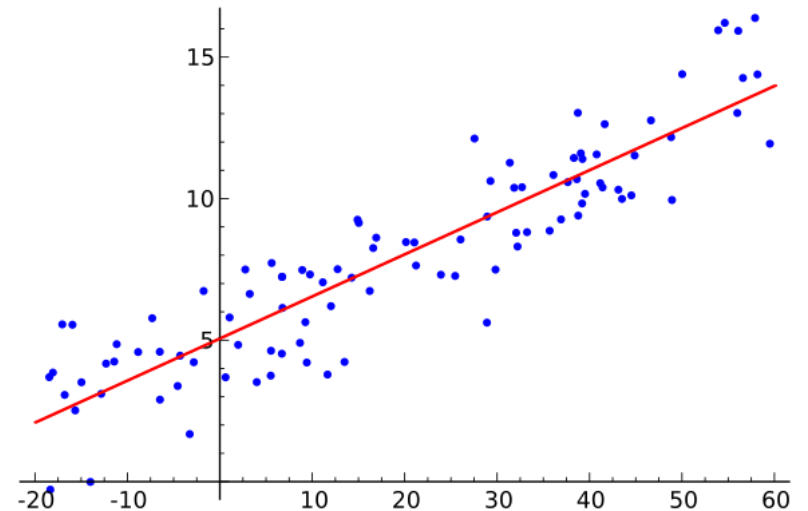
What is Artificial Intelligence?





What is Machine Learning (ML)?

- “ML” involves an algorithm that can infer how to do a specific task without explicit instruction from data
- Supervised → Explicit examples of results

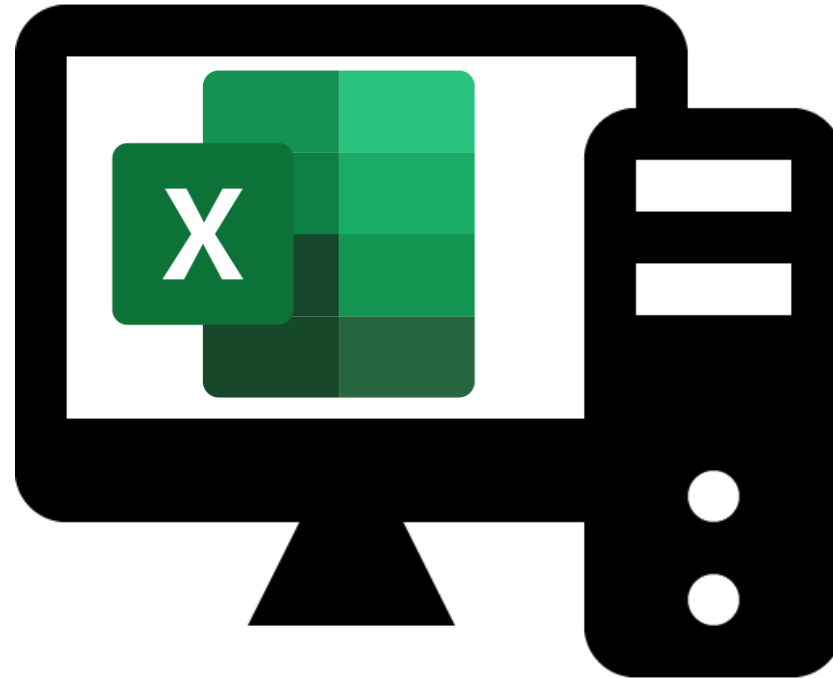
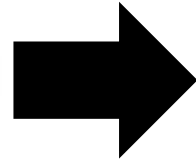




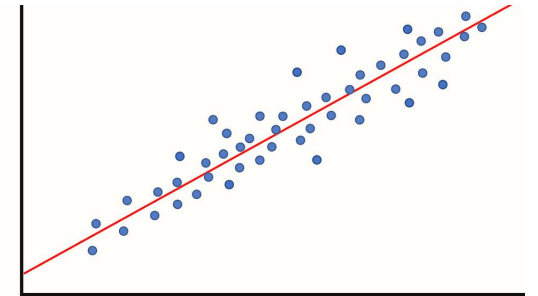
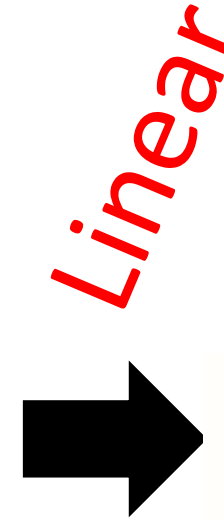
Model Training Process



Data



Machine



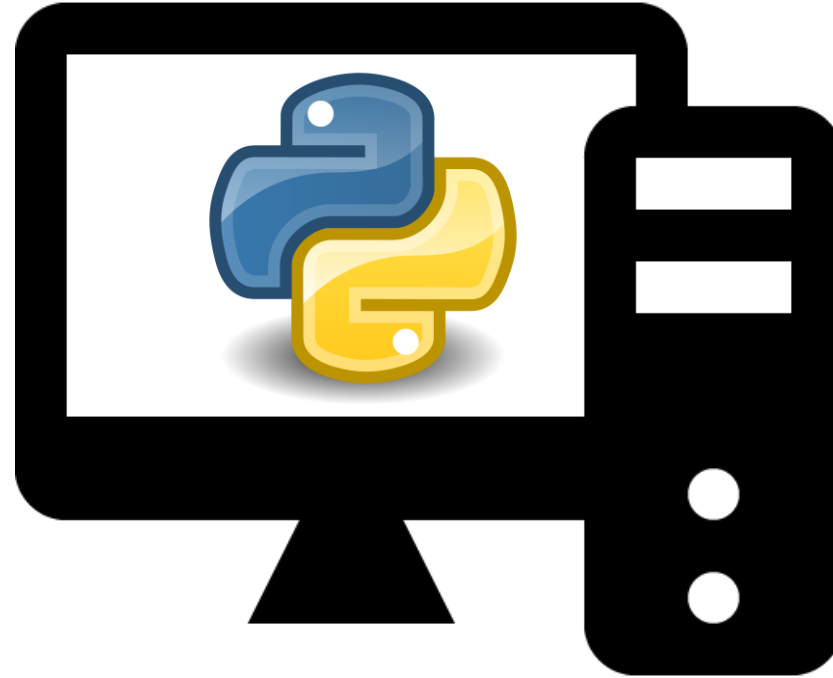
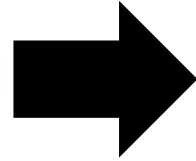
Model



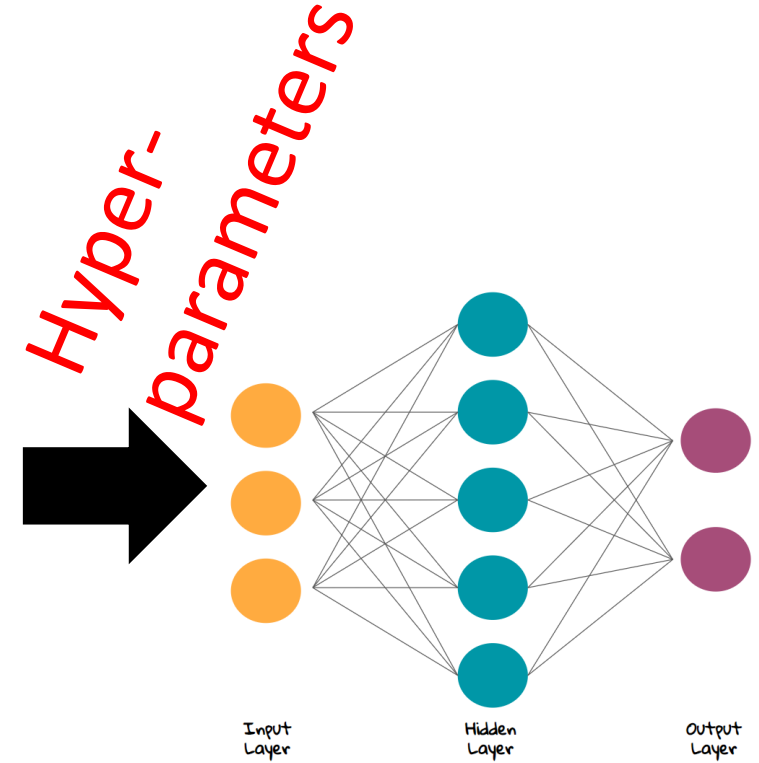
Model Training Process



Data



Machine



Model

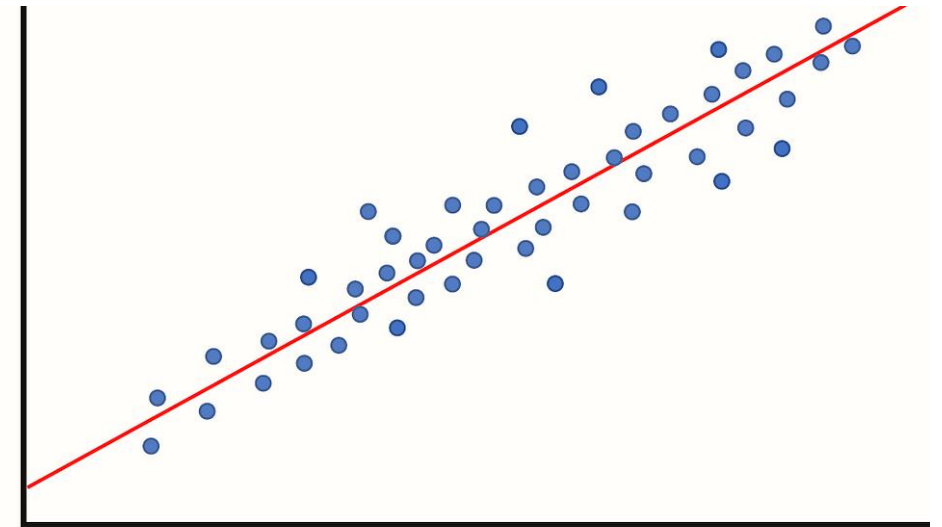


How is ML different?

Let's determine the treatment dose to remove chemical X, for real-time dose control.

1. Relationship between dose and removal **appears linear** → **fit linear regression** → **statistical**

Removal



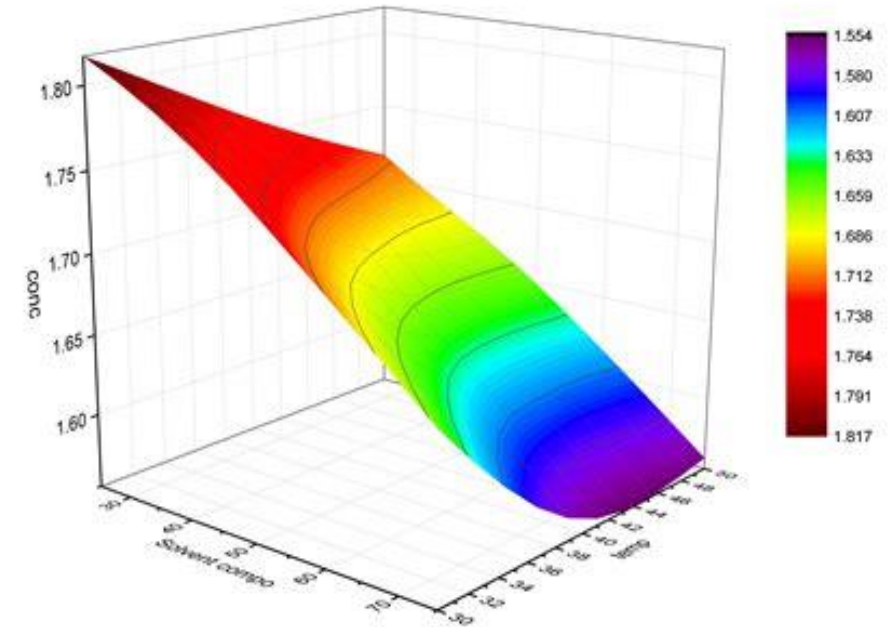
Dose



How is ML different?

Let's determine the treatment dose to remove chemical X, for real-time dose control.

1. Relationship between dose and removal appears linear \rightarrow fit linear regression \rightarrow statistical
2. **Known relationship** between dose, removal, and pH \rightarrow **fit equilibrium model** \rightarrow **mechanistic**

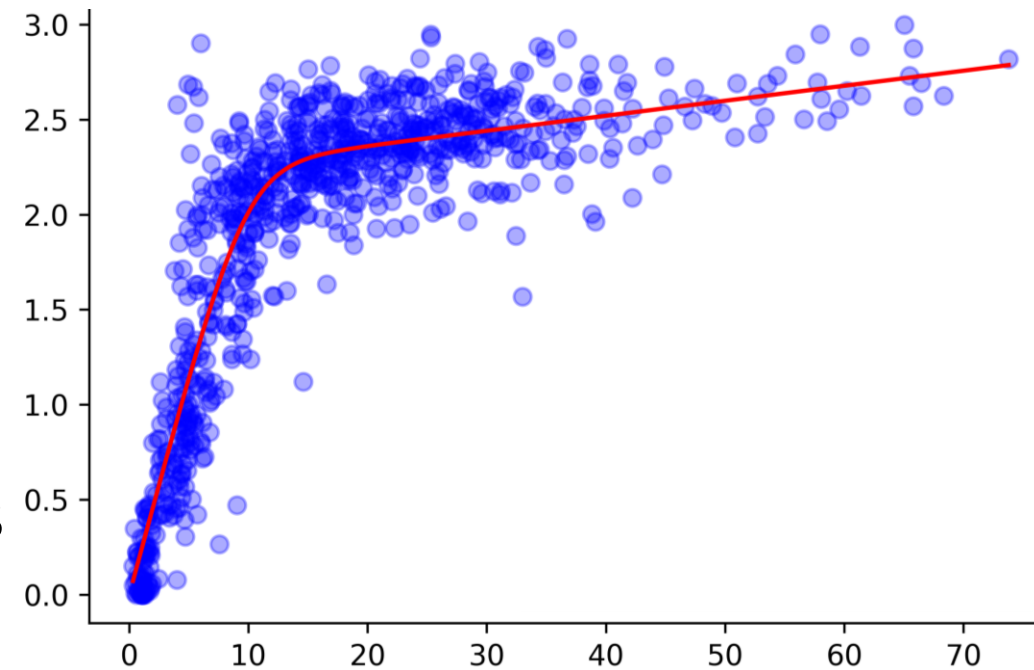




How is ML different?

Let's determine the treatment dose to remove chemical X, for real-time dose control.

1. Relationship between dose and removal appears linear → fit linear regression → statistical
2. Known relationship between dose, removal, and pH → fit equilibrium model → mechanistic
3. **Error** in statistical and mechanistic models is **too great** for the application...
Dose + removal + 10 other WQ variables
→ **fit incredibly flexible model** → **ML**

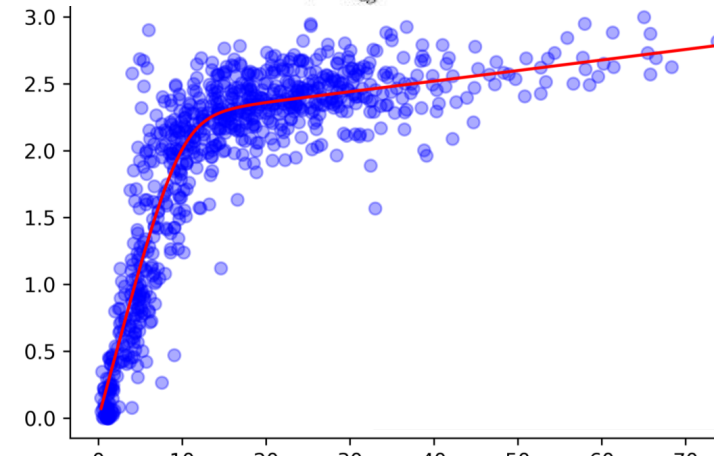
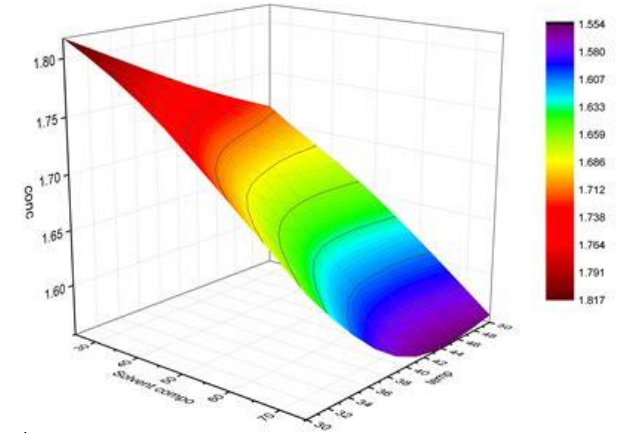
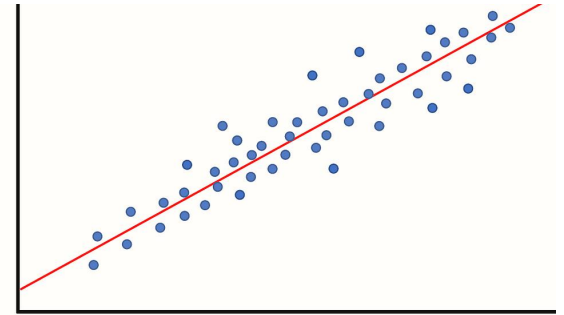




How is ML different?

Let's determine the treatment dose to remove chemical X, for real-time dose control.

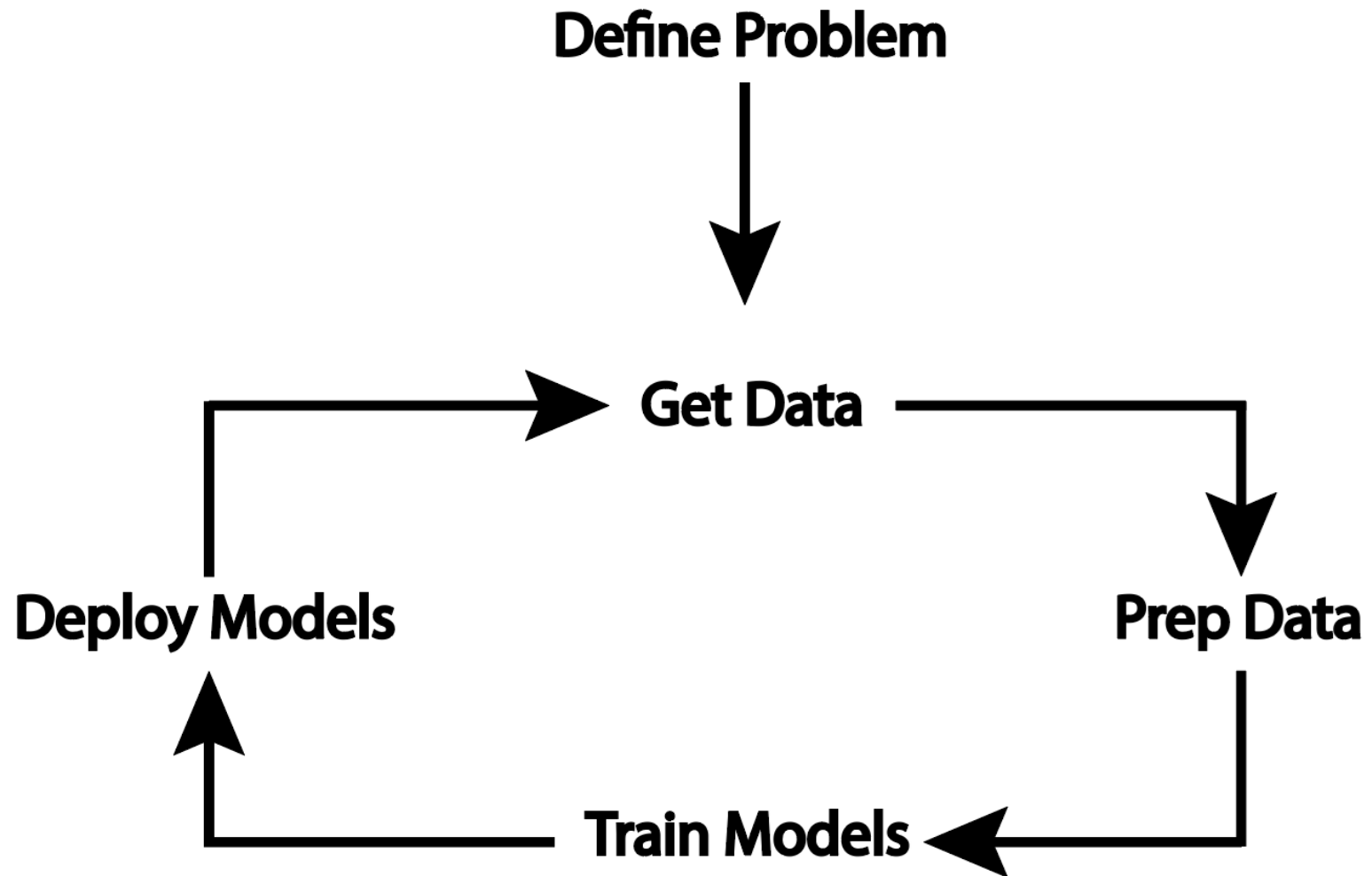
1. Statistical models = define relationship between variables, **analytical uncertainty estimates**
2. Mechanistic models = define relationship between variables, can **infer meaning** in coefficients
3. ML models = **no known relationship** between variables, experimental uncertainty estimates
 - More complex model building process
 - Answers different types of questions
 - Great flexibility requires great data!

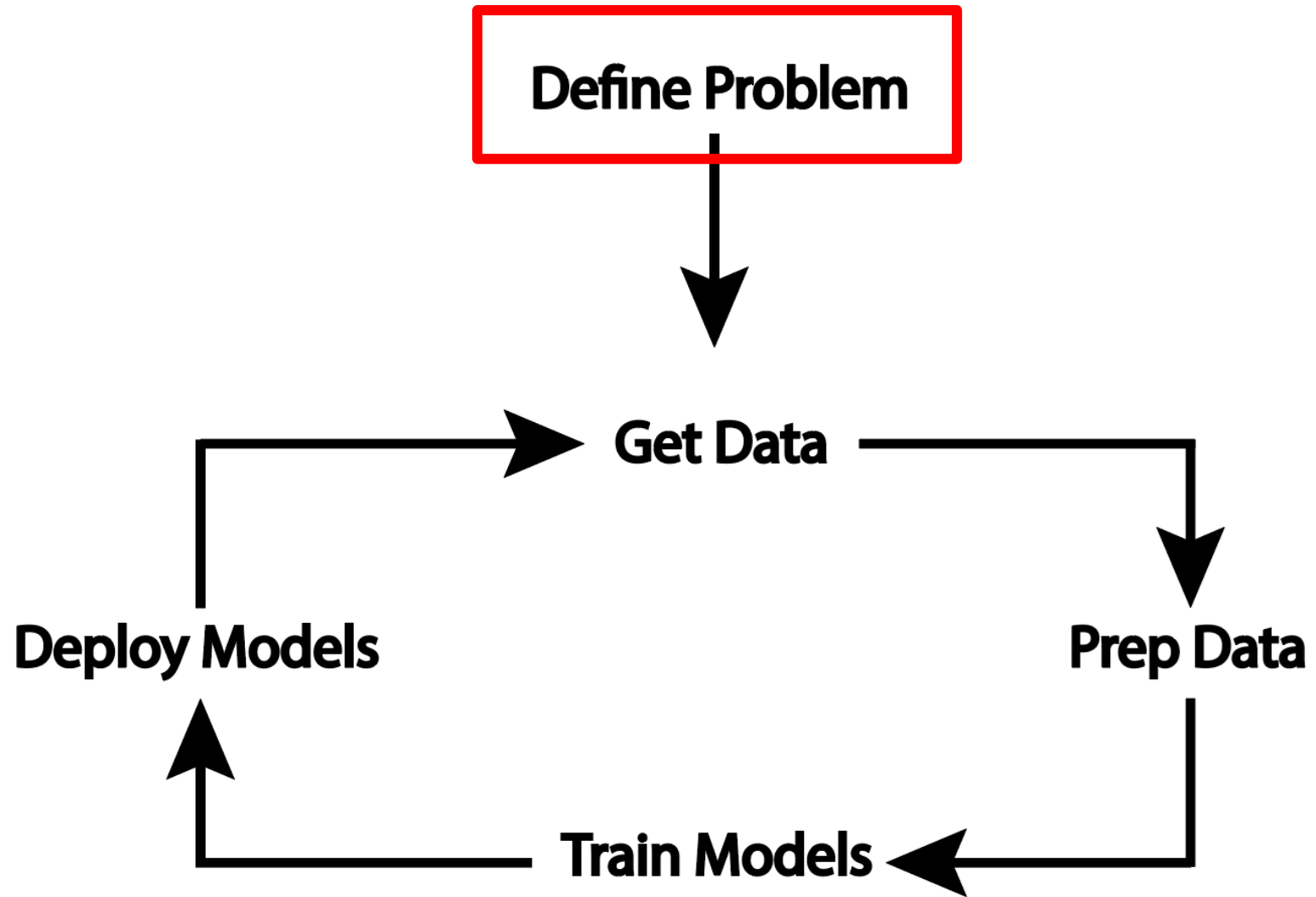




UNITED STATES MILITARY ACADEMY
WEST POINT

MACHINE LEARNING WORKFLOW







What types of problems?

- Water Quality and Flow Monitoring
 - “**Soft sensor**” uses an ML model to predict a constituent based on other online data. Often used for time-consuming laboratory measurements.
 - Early warning systems can detect faults in processes or changes in water quality, often before operator-set limits are exceeded.
 - Flow or load forecasting for early response
- Optimization
 - “**Digital Twins**”



What is a digital twin?

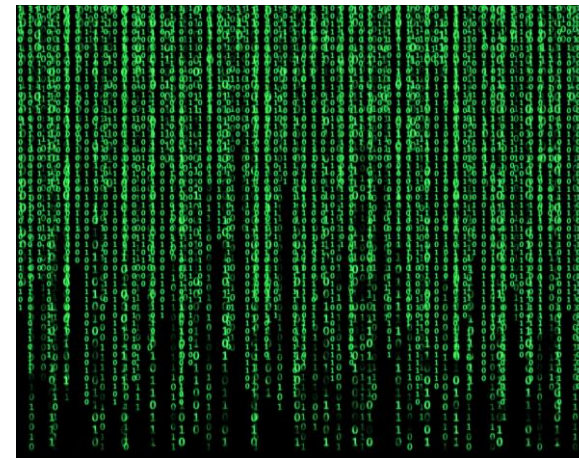
“Digital twin” is a model of a physical entity with an automated, bidirectional live data connection ... that allows for dynamic updating to maintain an accurate description ... as it evolves over time

–Peter Vanrolleghem + Torfs et al., 2022

Real Blower



Blower Model



Live data



Setpoints



Find Best Response





UNITED STATES MILITARY ACADEMY
WEST POINT

A “good” *ML* problem



Yishi Zuo, 2021



A “good” *ML* problem

- Problem statement is specific to an objective
 - Example: Objective may be to minimize chemical dose, not to understand chemical kinetics



Yishi Zuo, 2021

The biggest problem is keeping the original problem the problem!



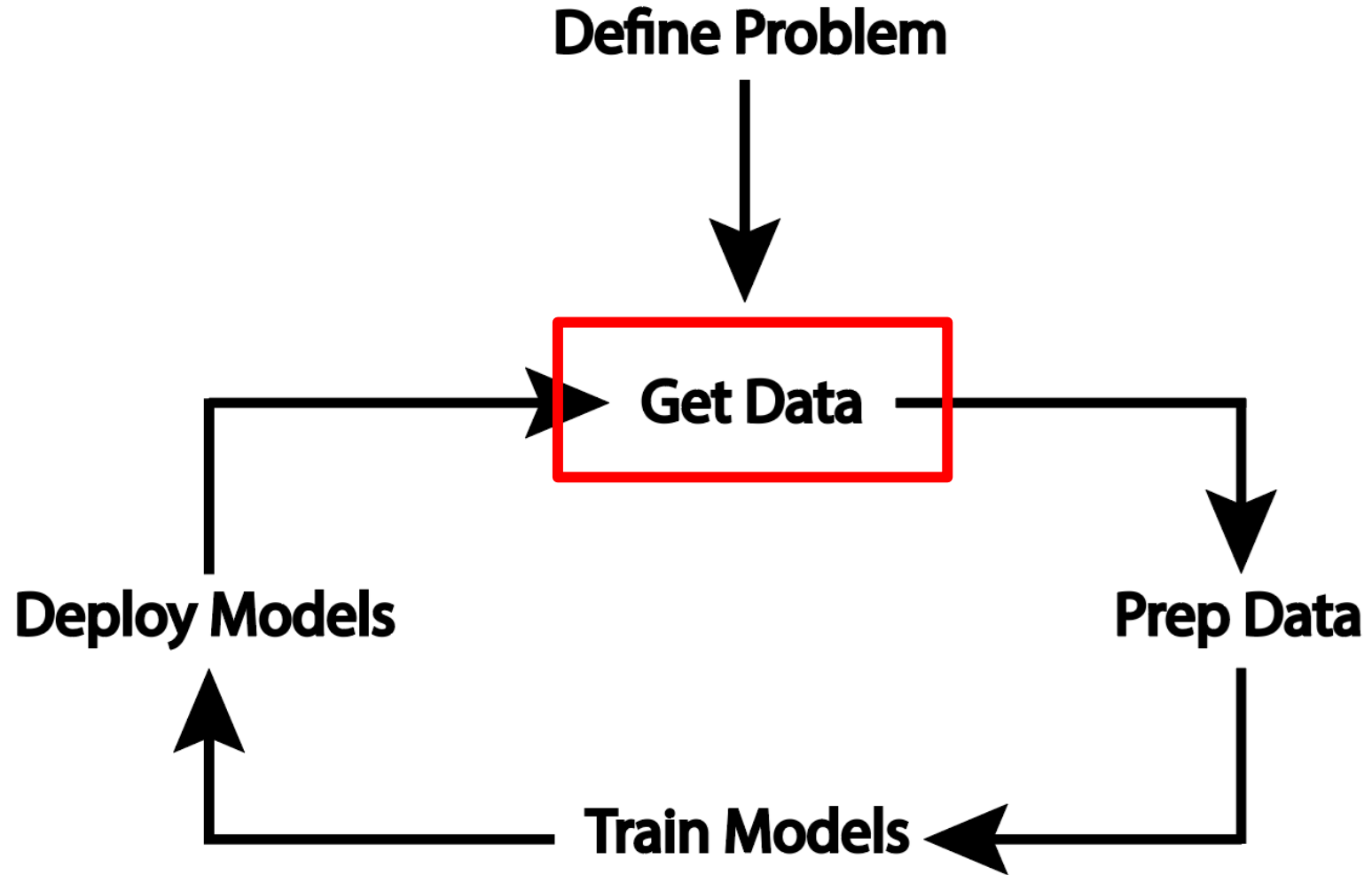
A “good” ML problem

- Problem statement is specific to an objective
 - Example: Objective may be to minimize chemical dose, not to understand chemical kinetics
- Simpler models cannot capture the observed trends (with sufficient accuracy for a given application)
- Sufficient, quality historical data is available for model development and will continue to be available



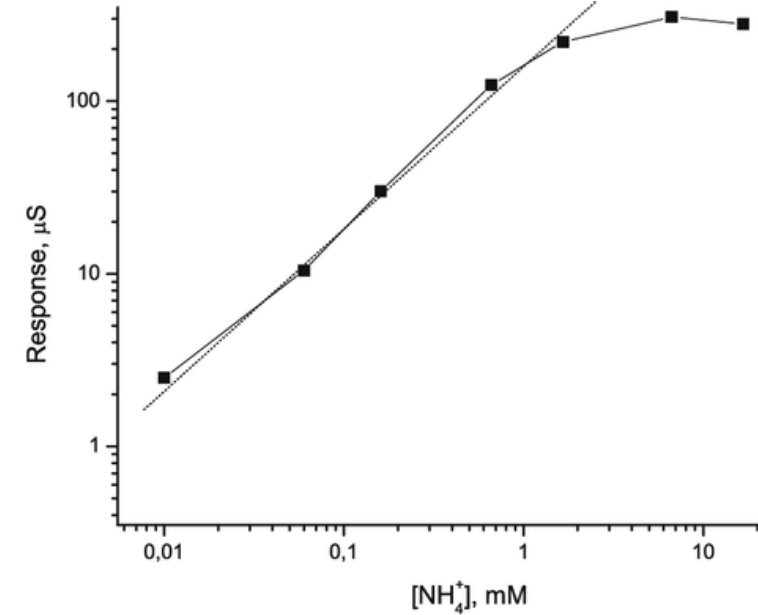
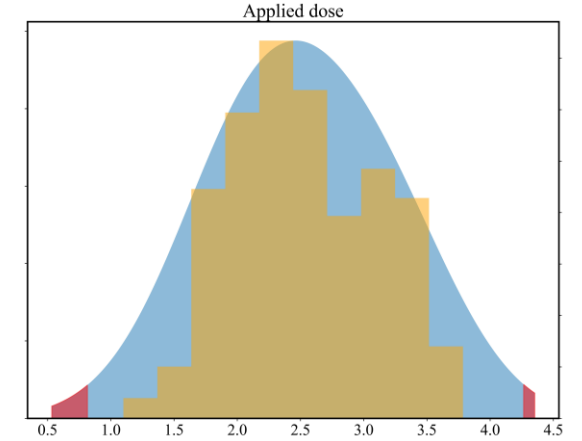
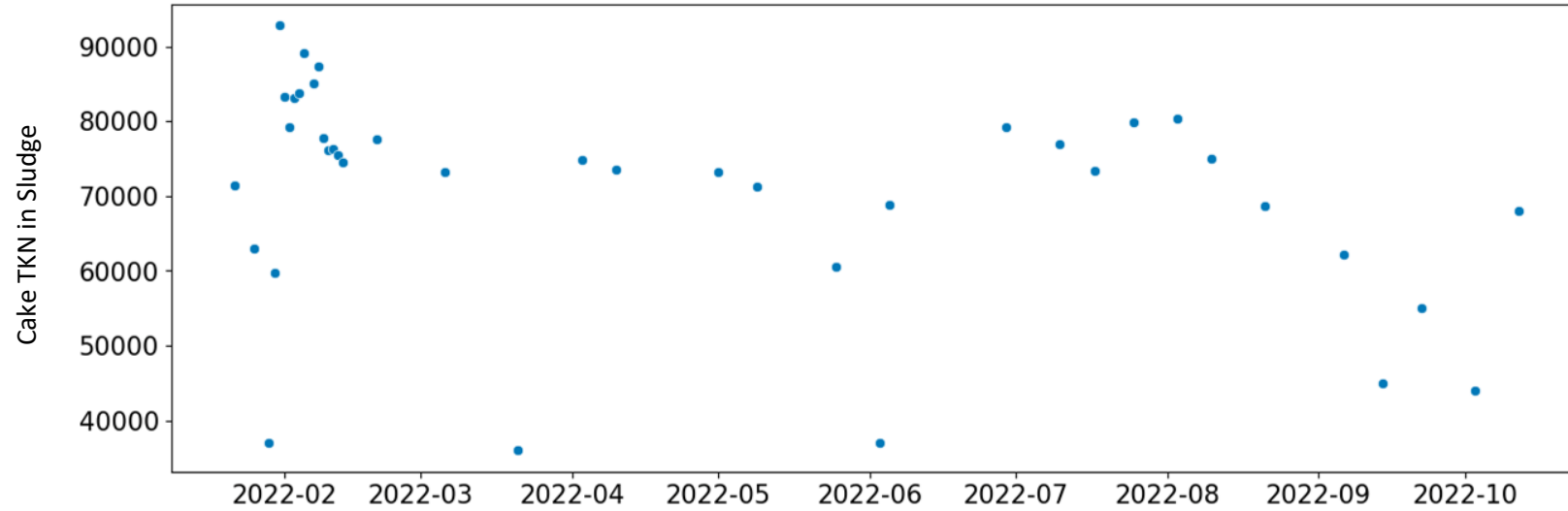
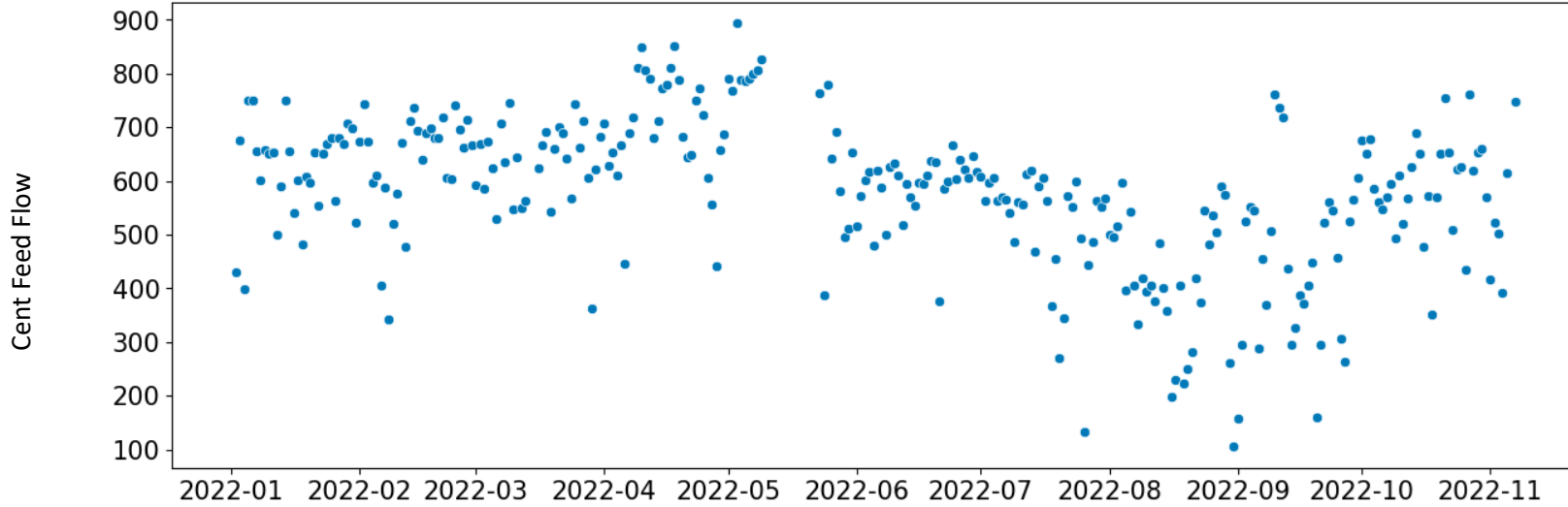
Yishi Zuo, 2021

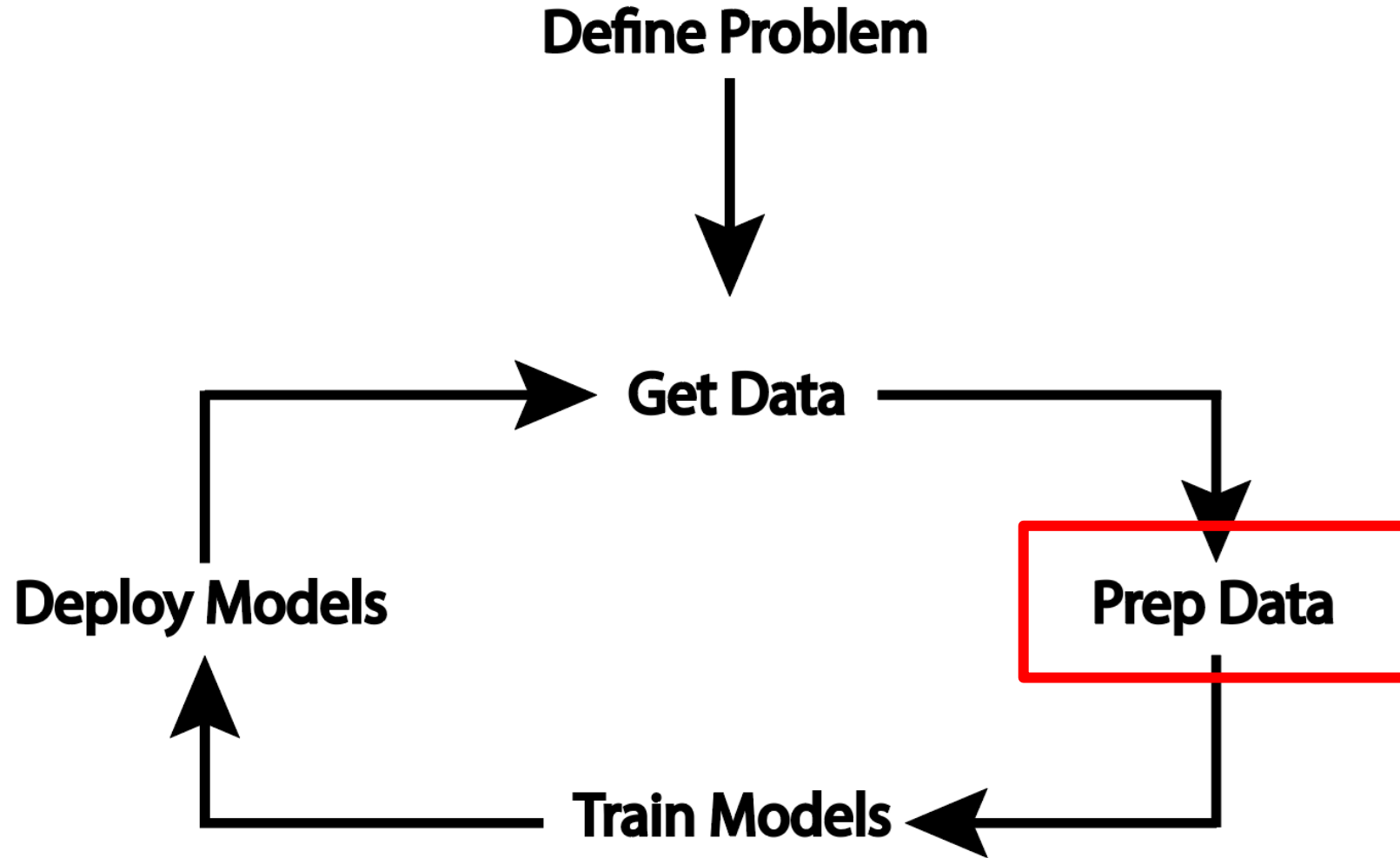
The biggest problem is keeping the original problem the problem!





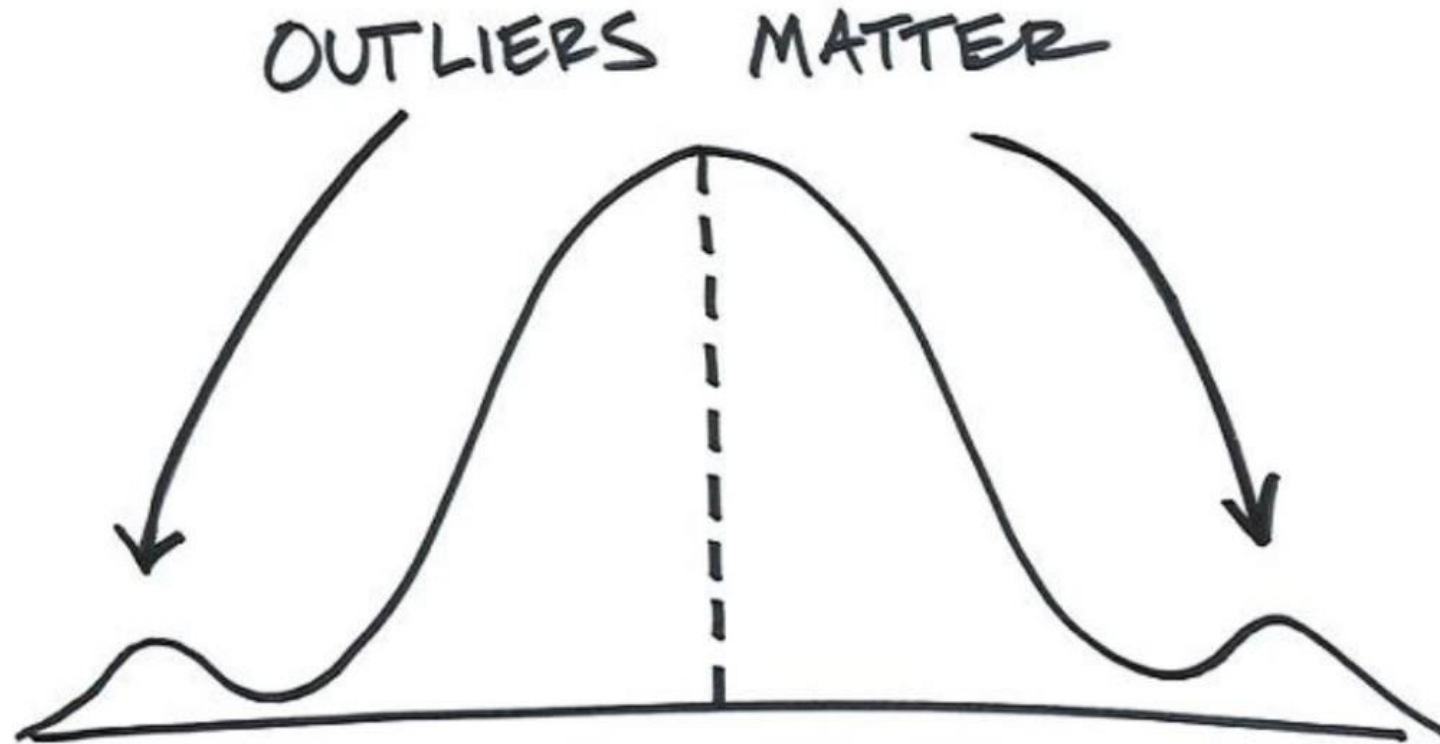
“I have SO much data”







Remove Outliers (?)





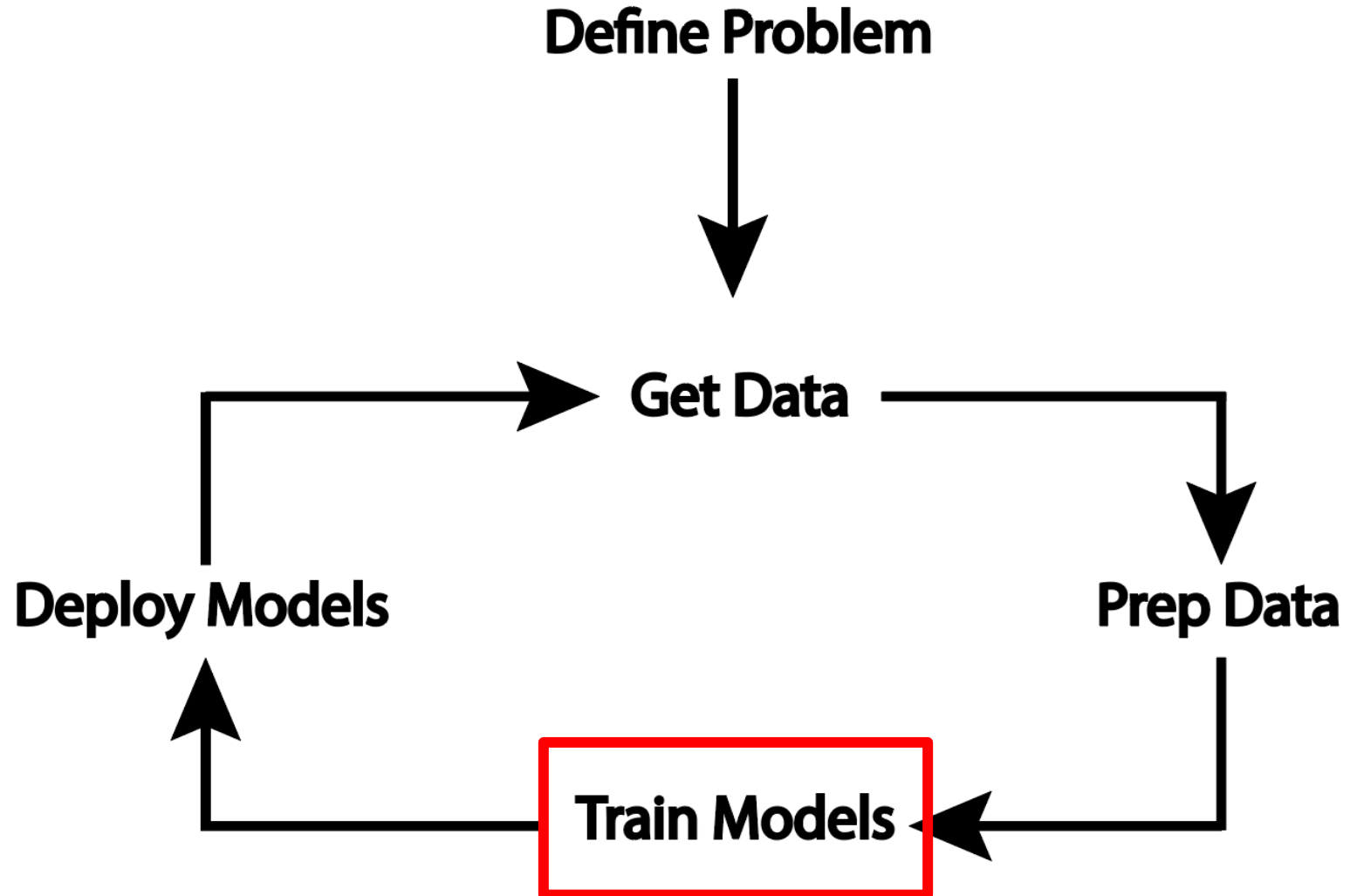
UNITED STATES MILITARY ACADEMY
WEST POINT®

Remove Irrelevant “Features”



<https://www.apartmenttherapy.com/before-and-after-messy-pantry-gets-realistic-transformation-37200882>

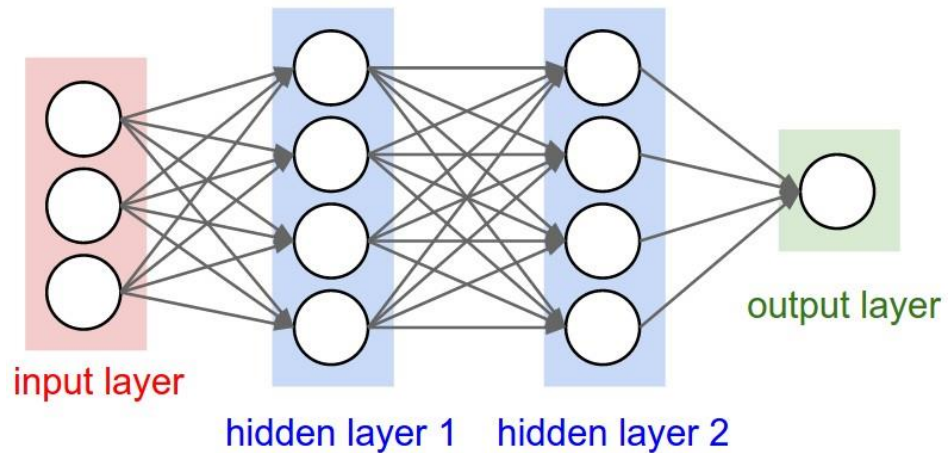
<https://sallysbakingaddiction.com/how-to-measure-baking-ingredients/>





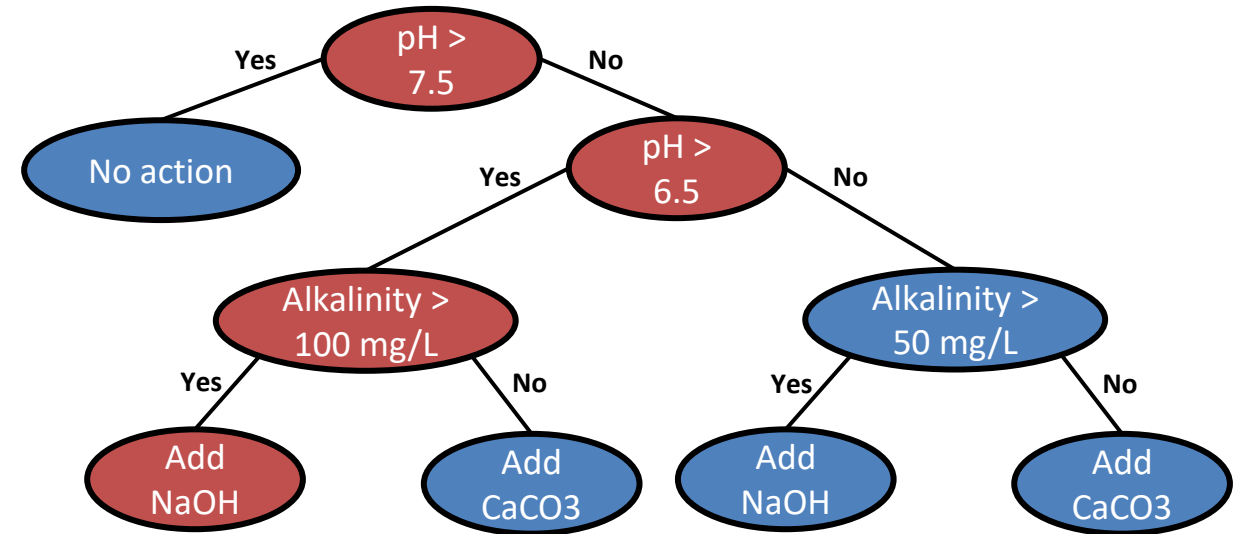
Neural Networks

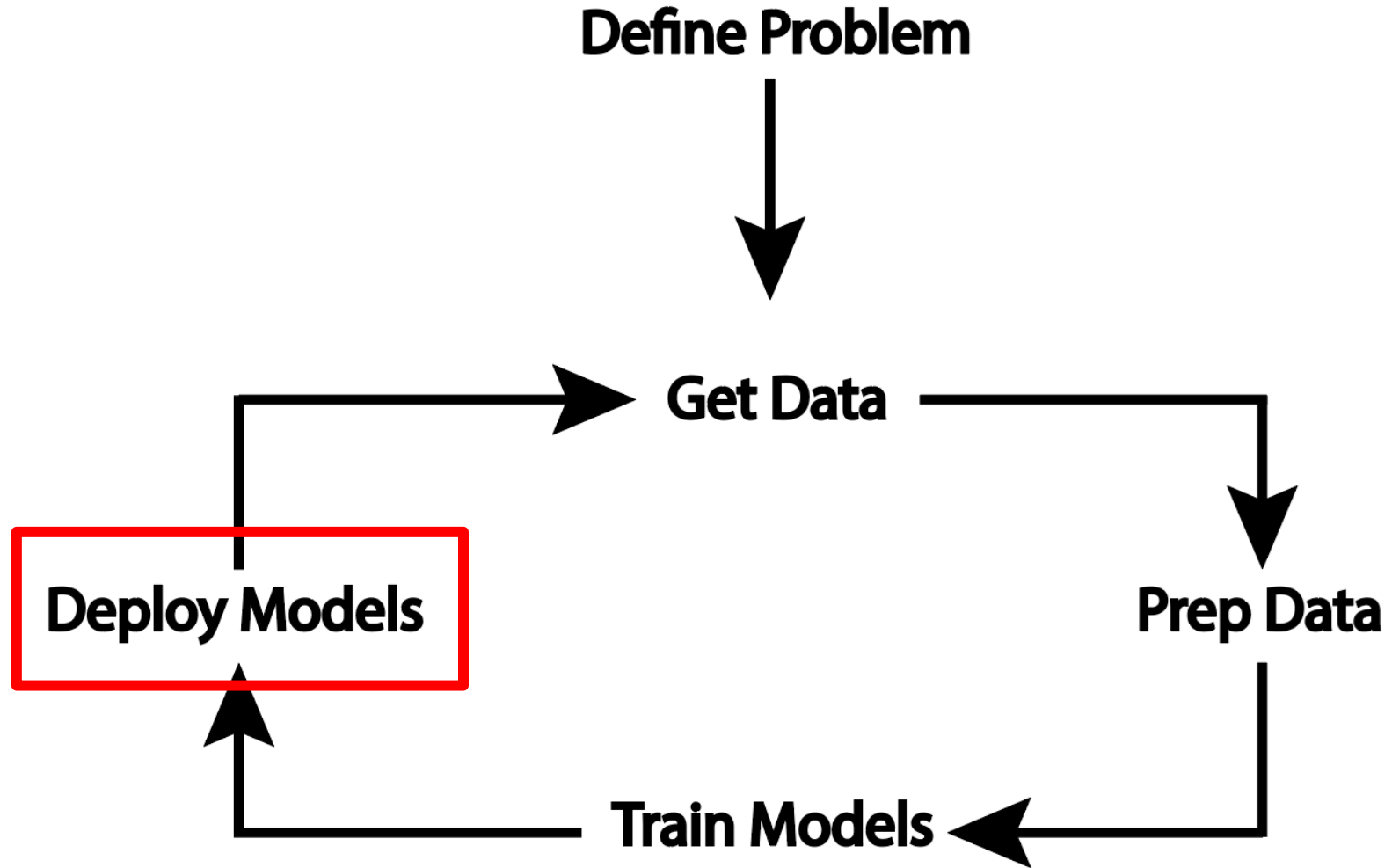
ANN, RNN, LSTM



Decision Trees

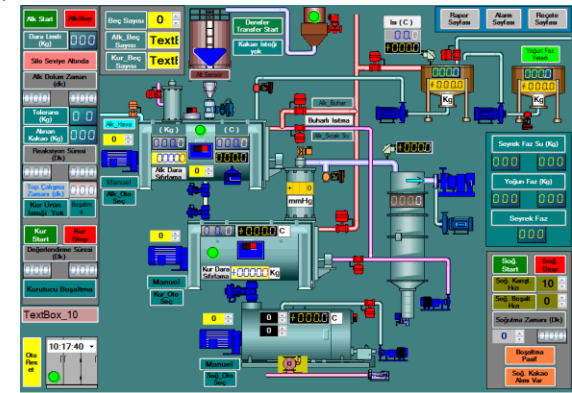
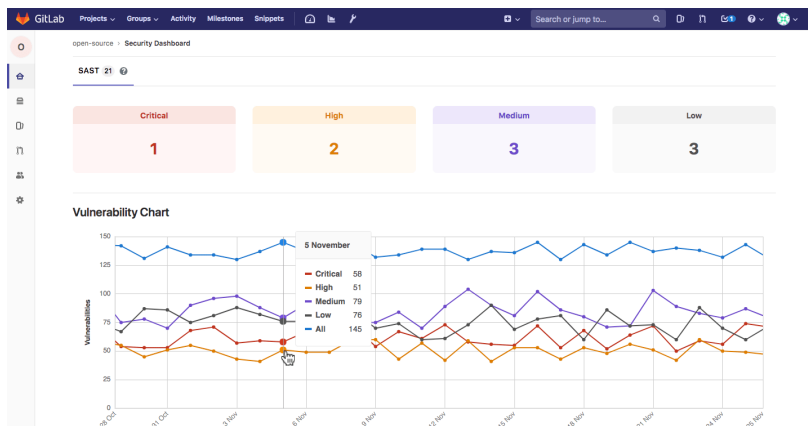
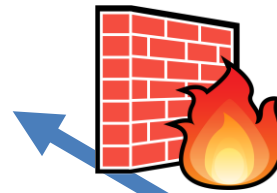
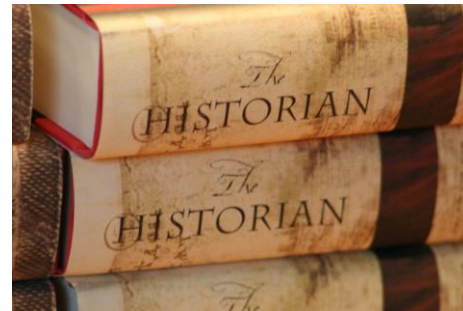
RF, AdaBoost, XGBoost





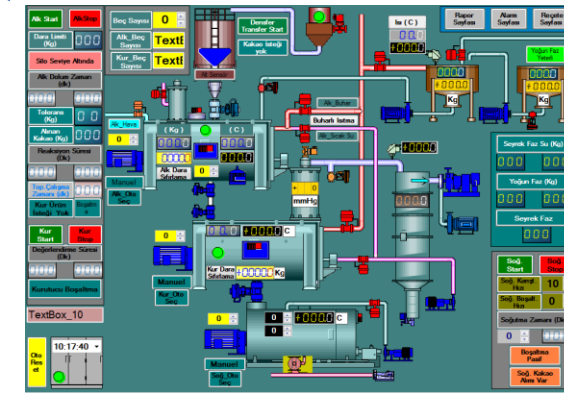
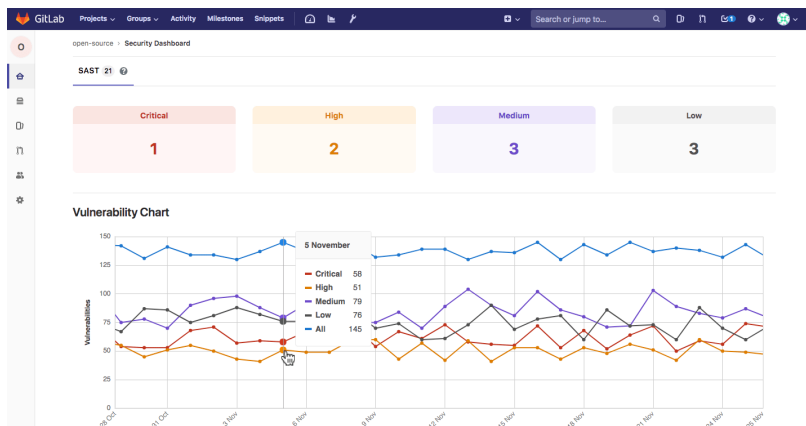
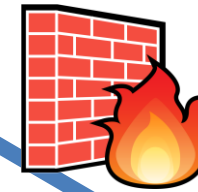
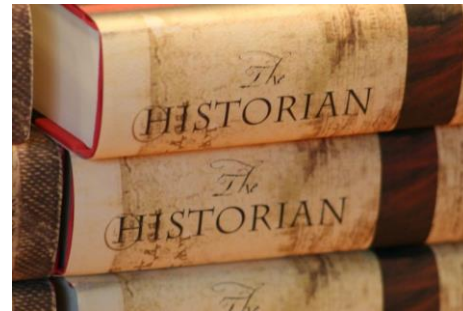


How do I use the model?



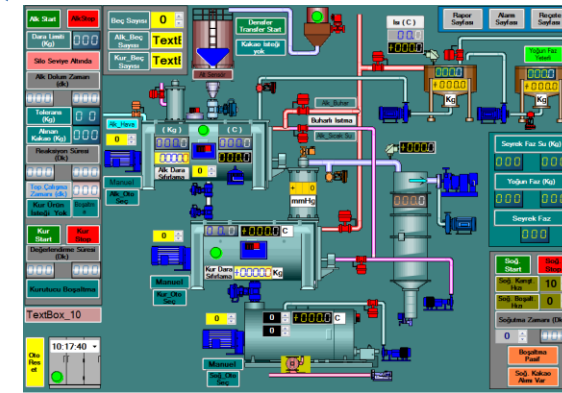
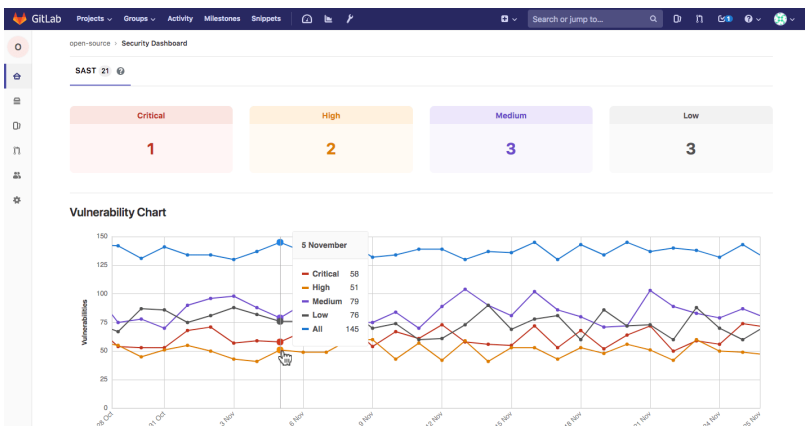
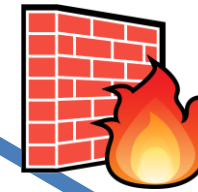
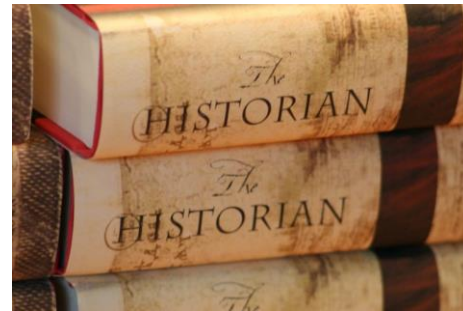


How do I use the model?





How do I use the model?





UNITED STATES MILITARY ACADEMY
WEST POINT

MACHINE LEARNING EXAMPLES



UNITED STATES MILITARY ACADEMY
WEST POINT



COLORADO SCHOOL OF
MINES



FORECASTING

City of Boulder, CO

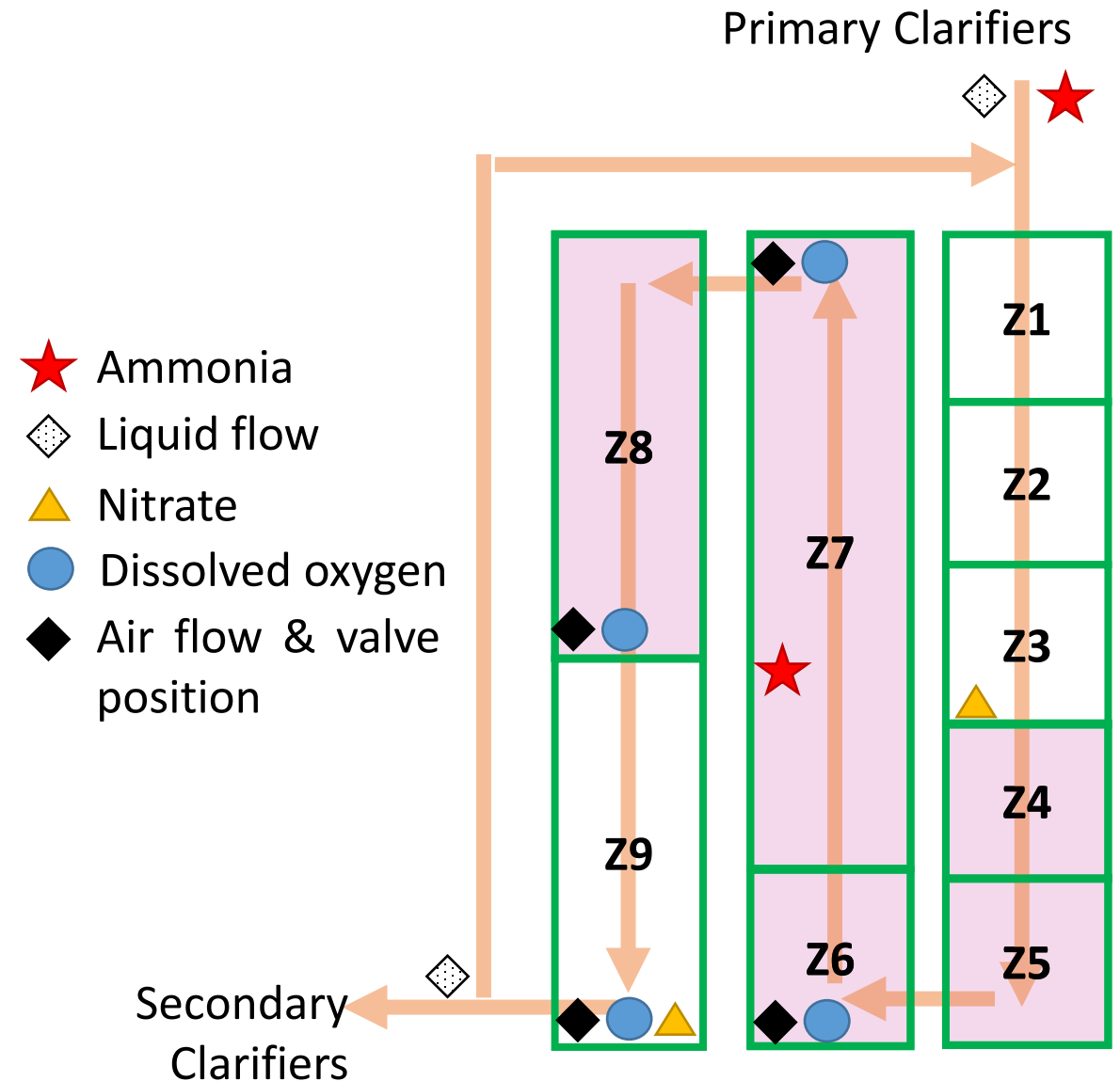
Water Resource Recovery Facility

- 25 MGD design, 12 MGD average
- Daily ammonia limits as low as 1.9 mg/L
 - High DO operation is trusted control strategy
- Multiple aeration control modes
 - Airflow
 - Dissolved oxygen (DO)
 - Ammonia-based aeration control (ABAC)





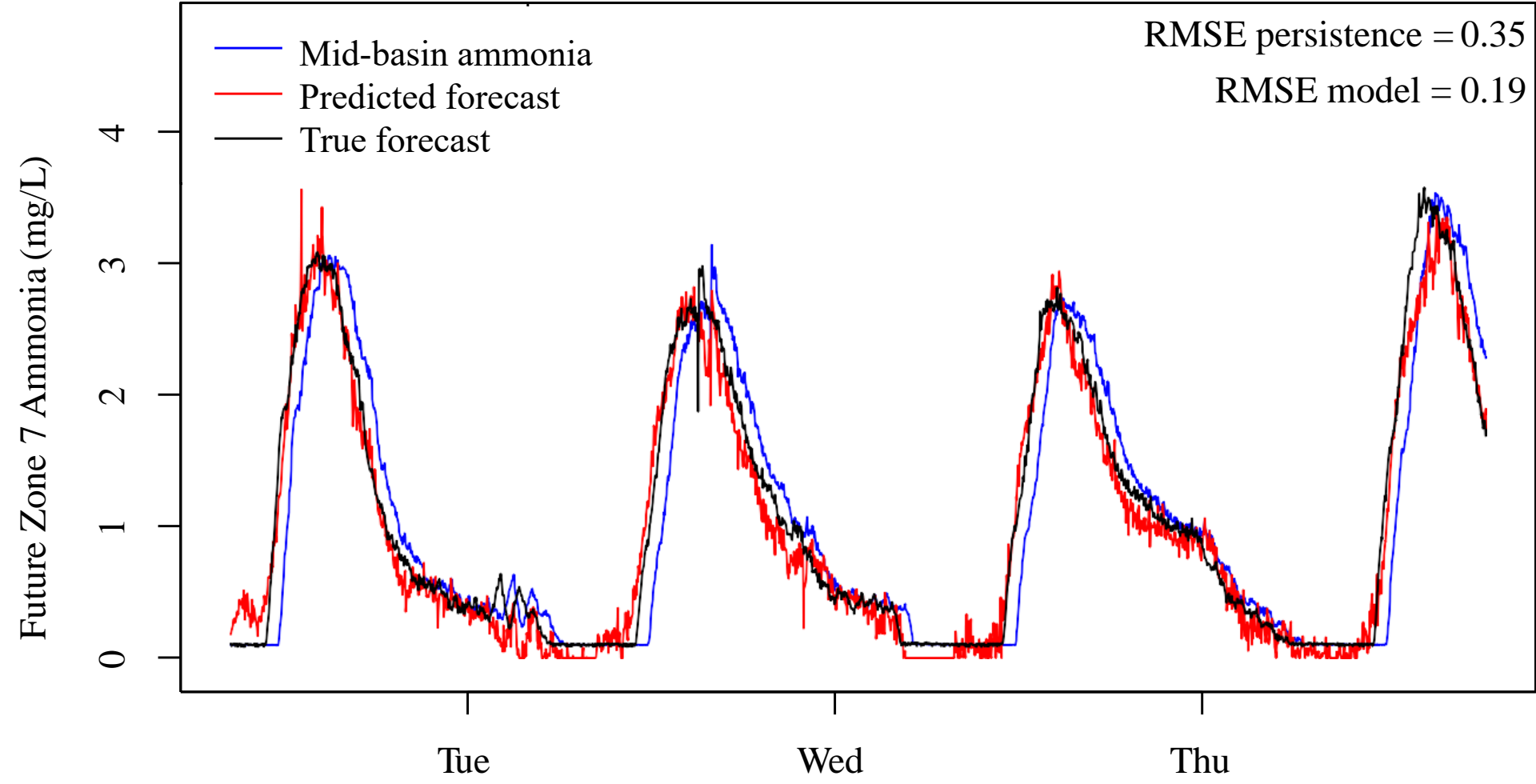
- Influent ammonia sensor was difficult to maintain
- Mid-basin ammonia was stable, but led to a delayed response
- **Replace ammonia in control strategy with a “forecasting” soft sensor**





Forecasting Aeration Demand

46%

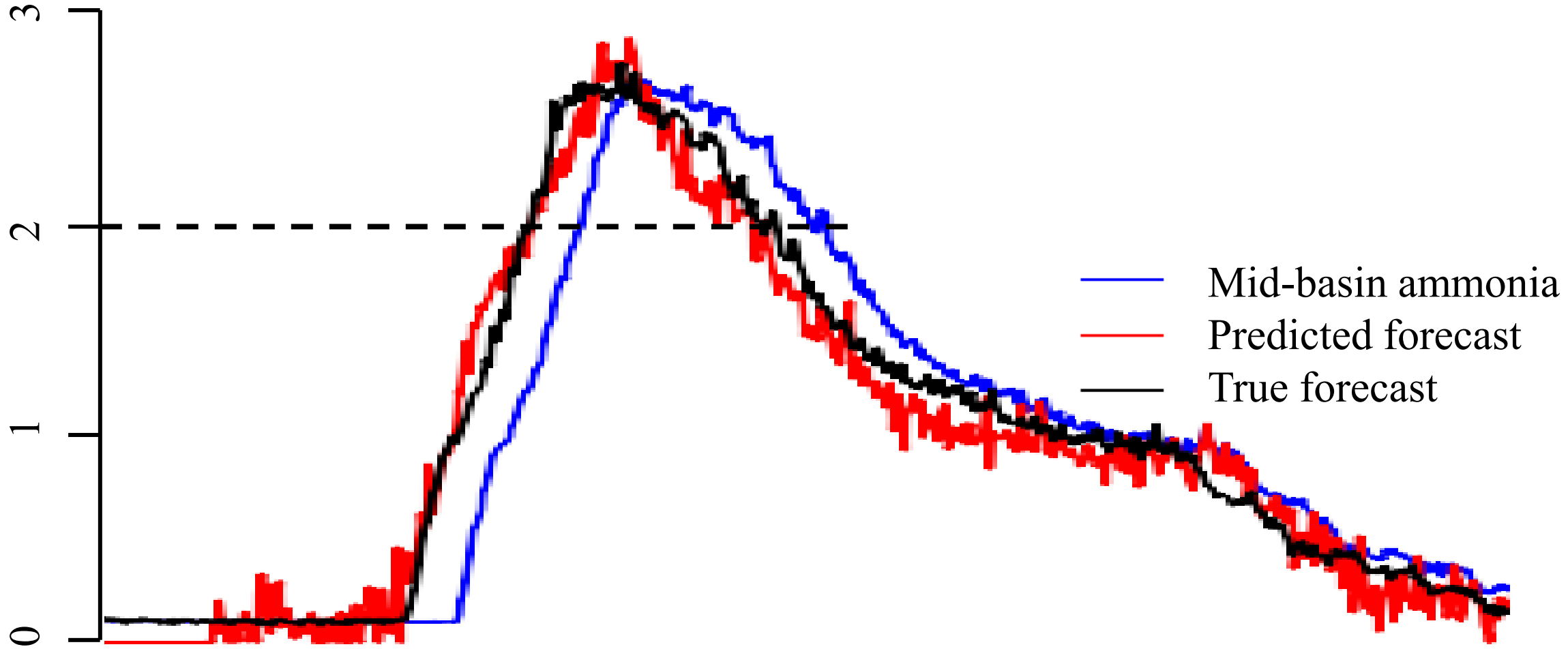


Diurnal, linear, and artificial neural network hybrid model.

2-min data frequency, 3-day training window.

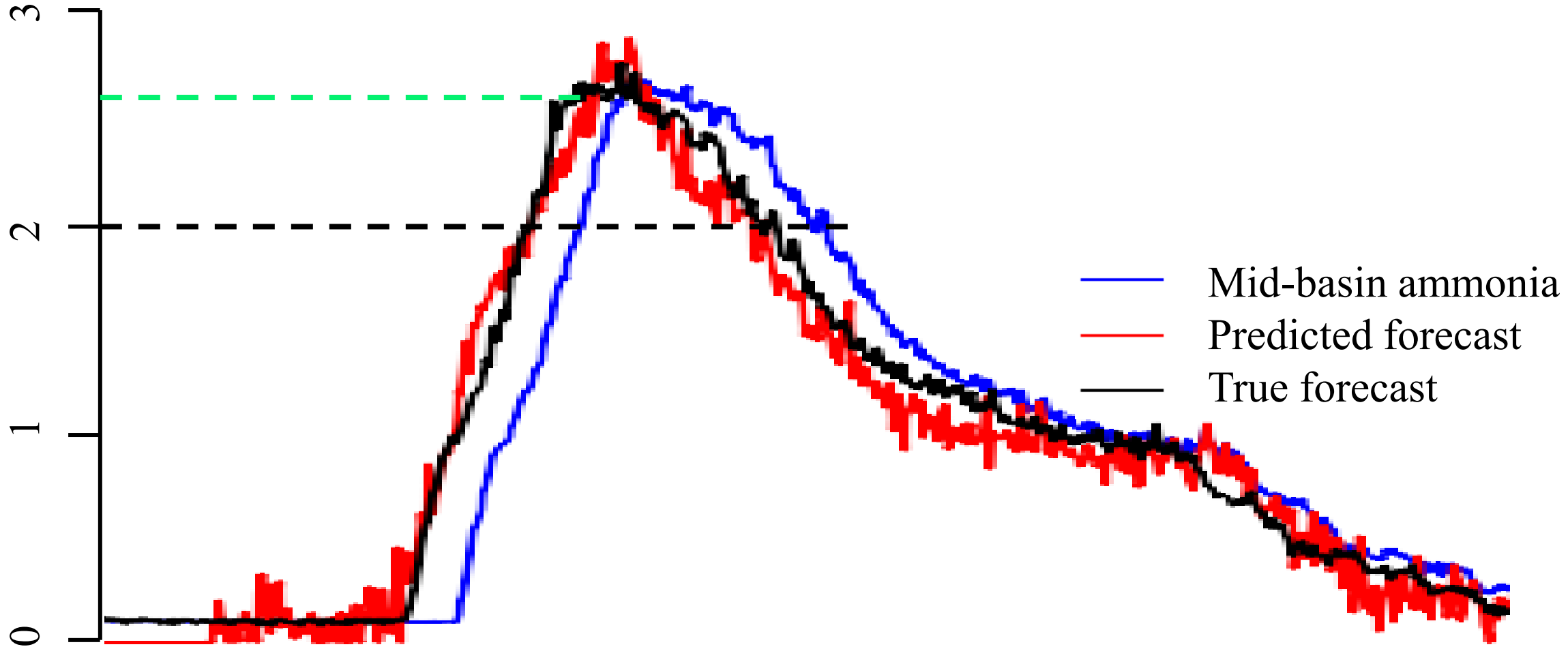


Future Zone 7 Ammonia





Future Zone 7 Ammonia

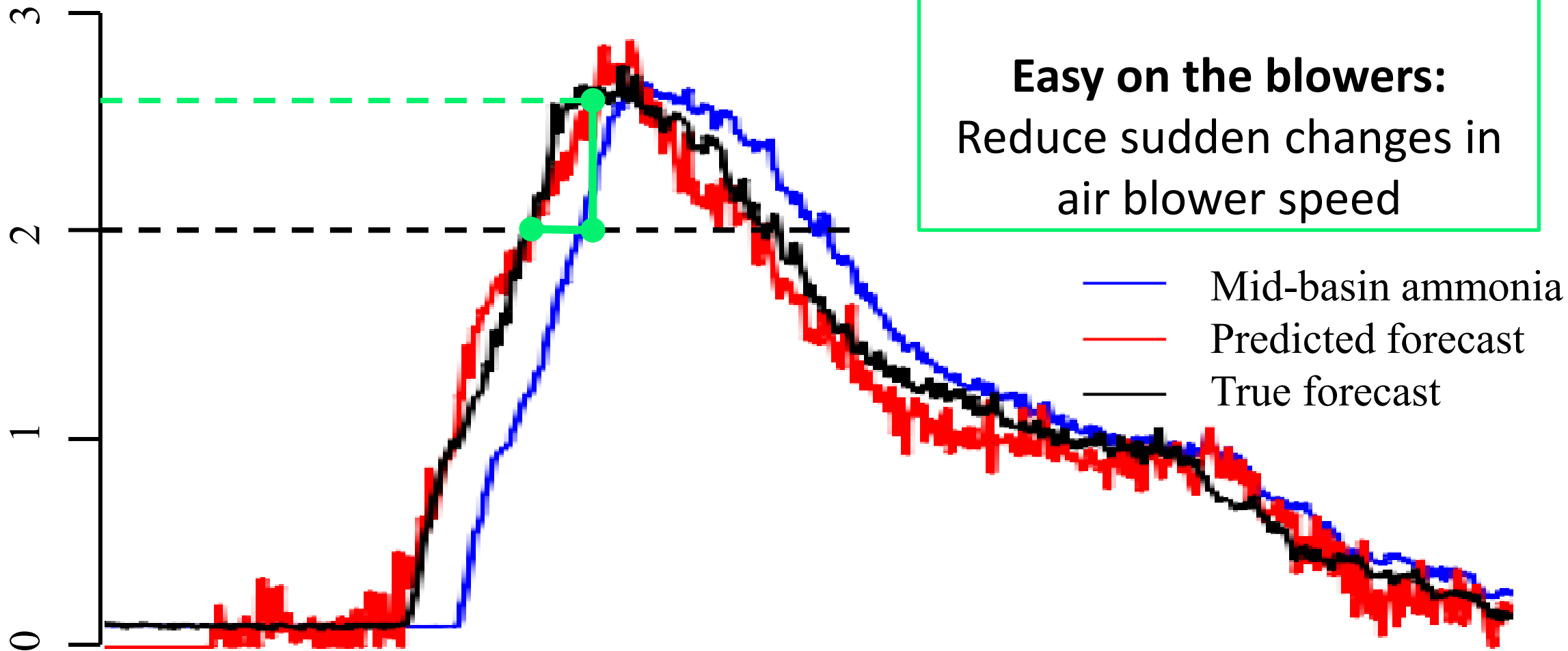




Prevent undertreatment:
Increase air output prior to
exceeding setpoint

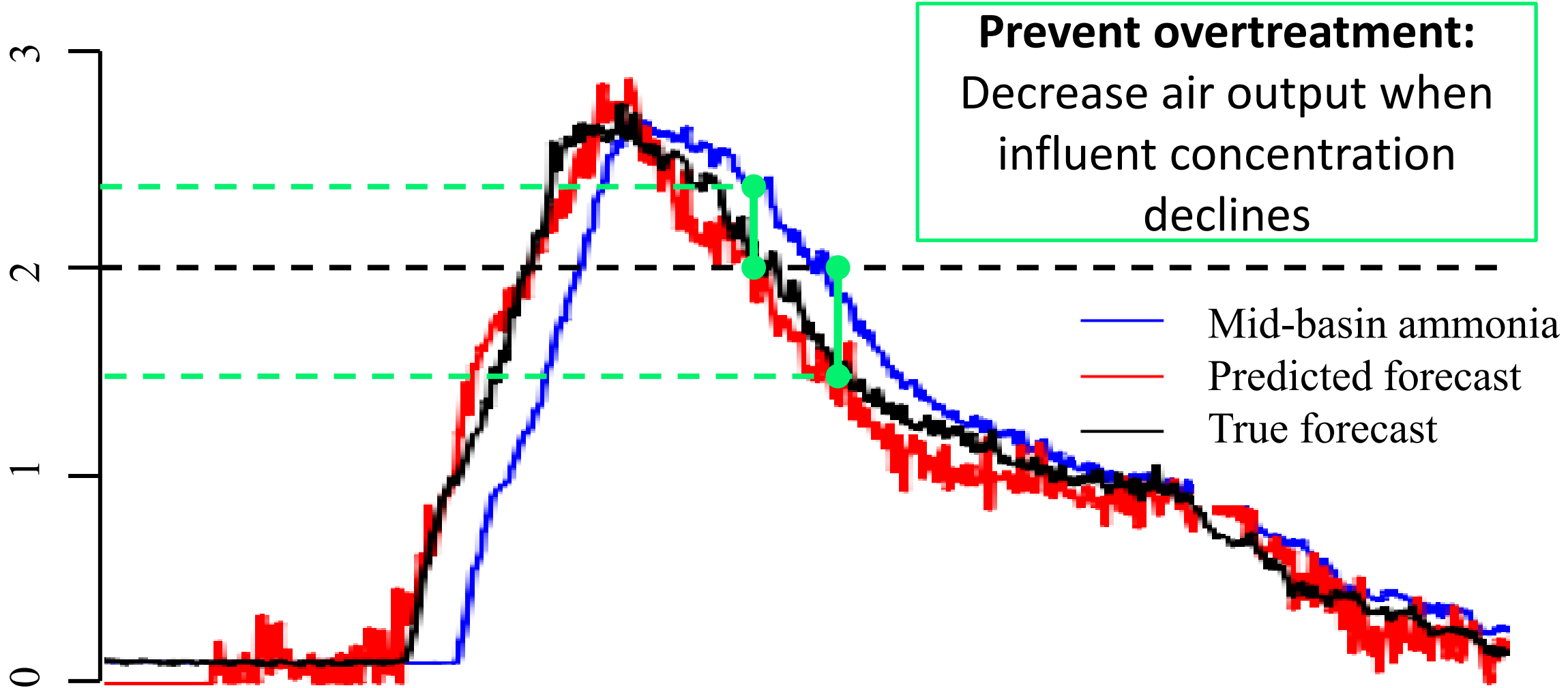
Easy on the blowers:
Reduce sudden changes in
air blower speed

Future Zone 7 Ammonia





Future Zone 7 Ammonia



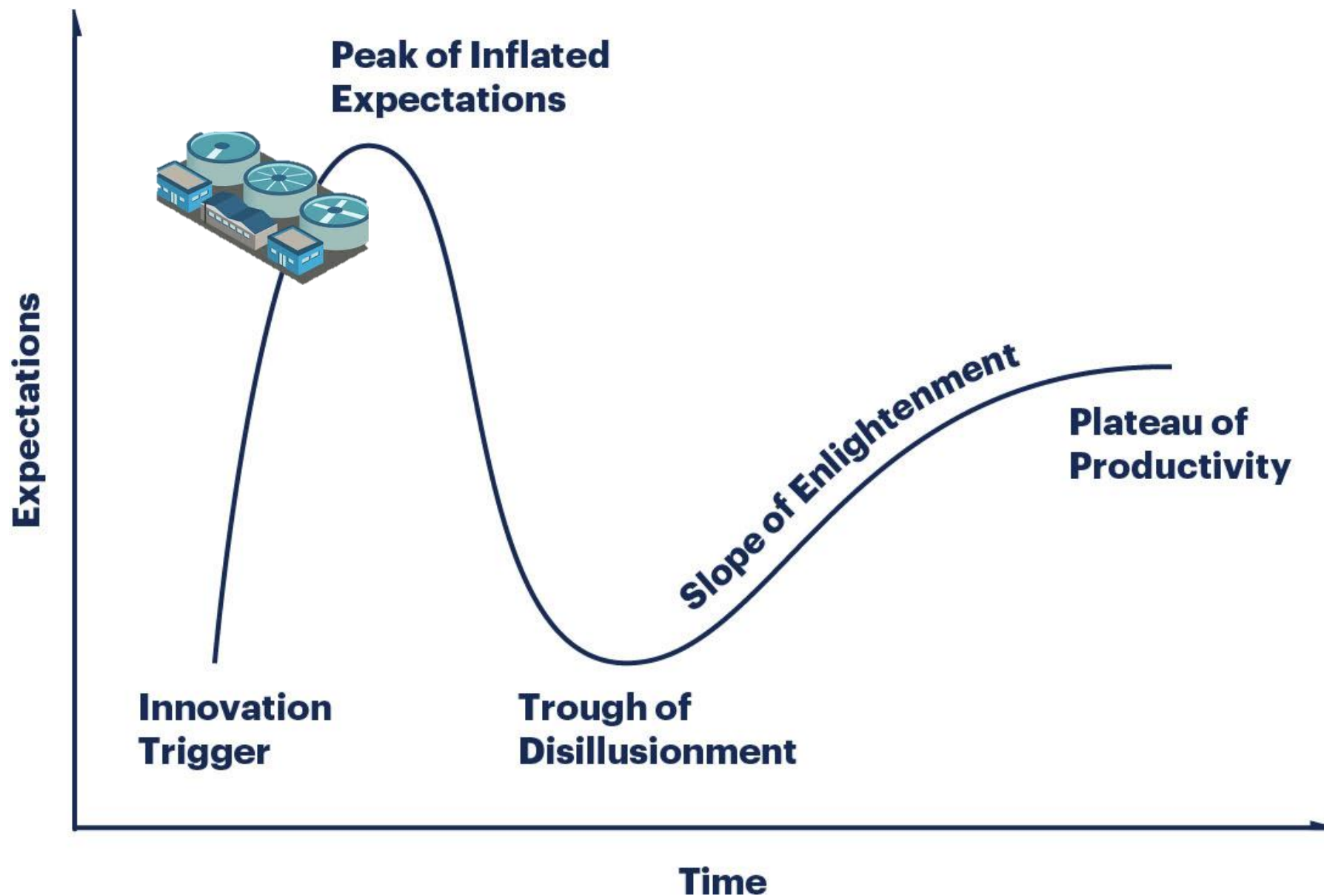


UNITED STATES MILITARY ACADEMY
WEST POINT

FINAL THOUGHTS

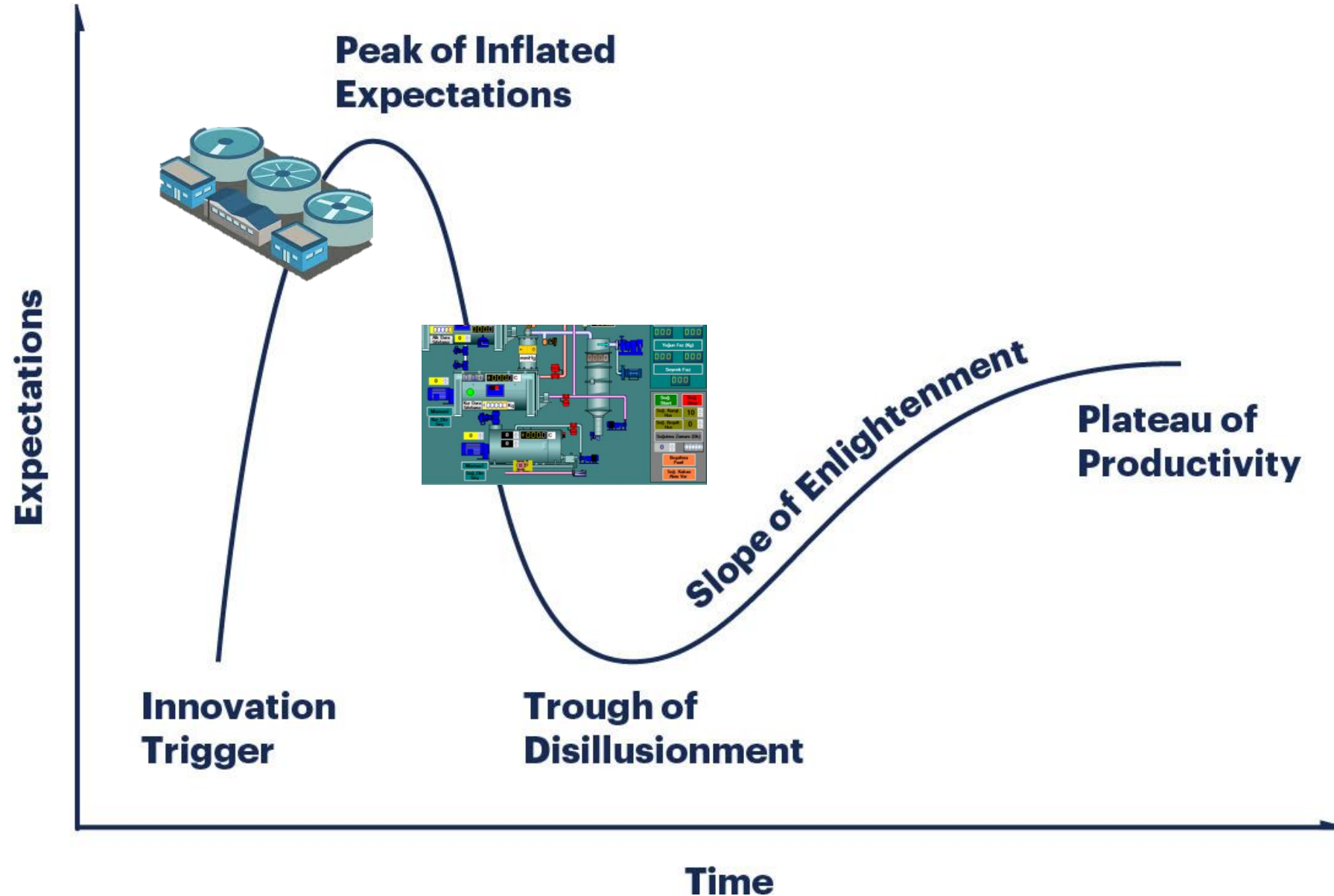


Hype Cycle: ML and Utilities



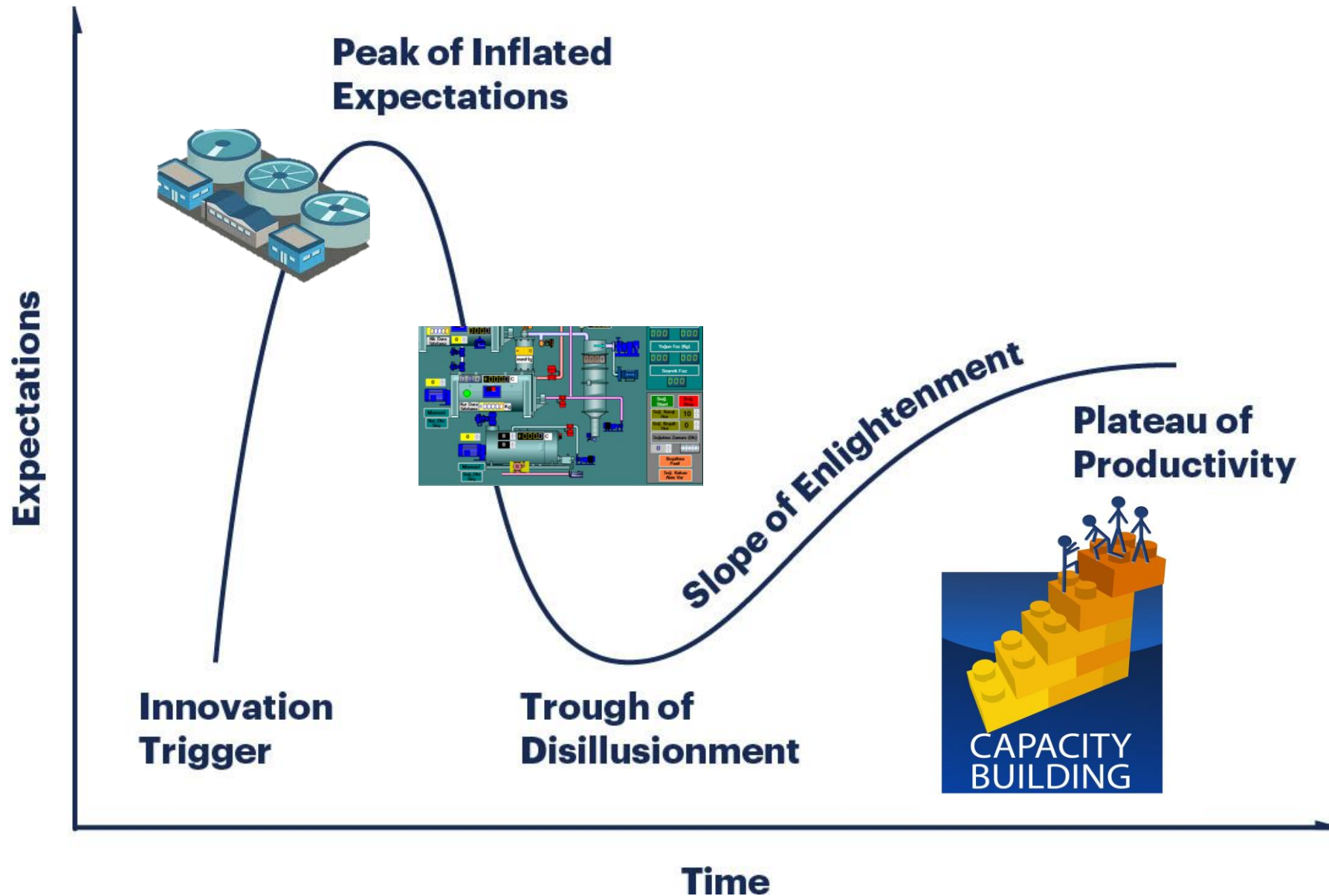


Hype Cycle: ML and Utilities



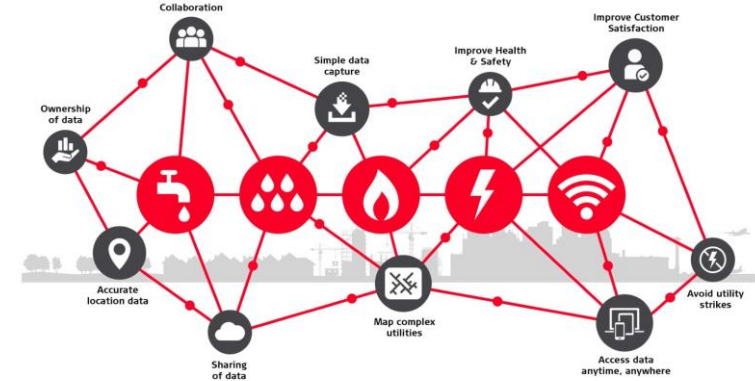
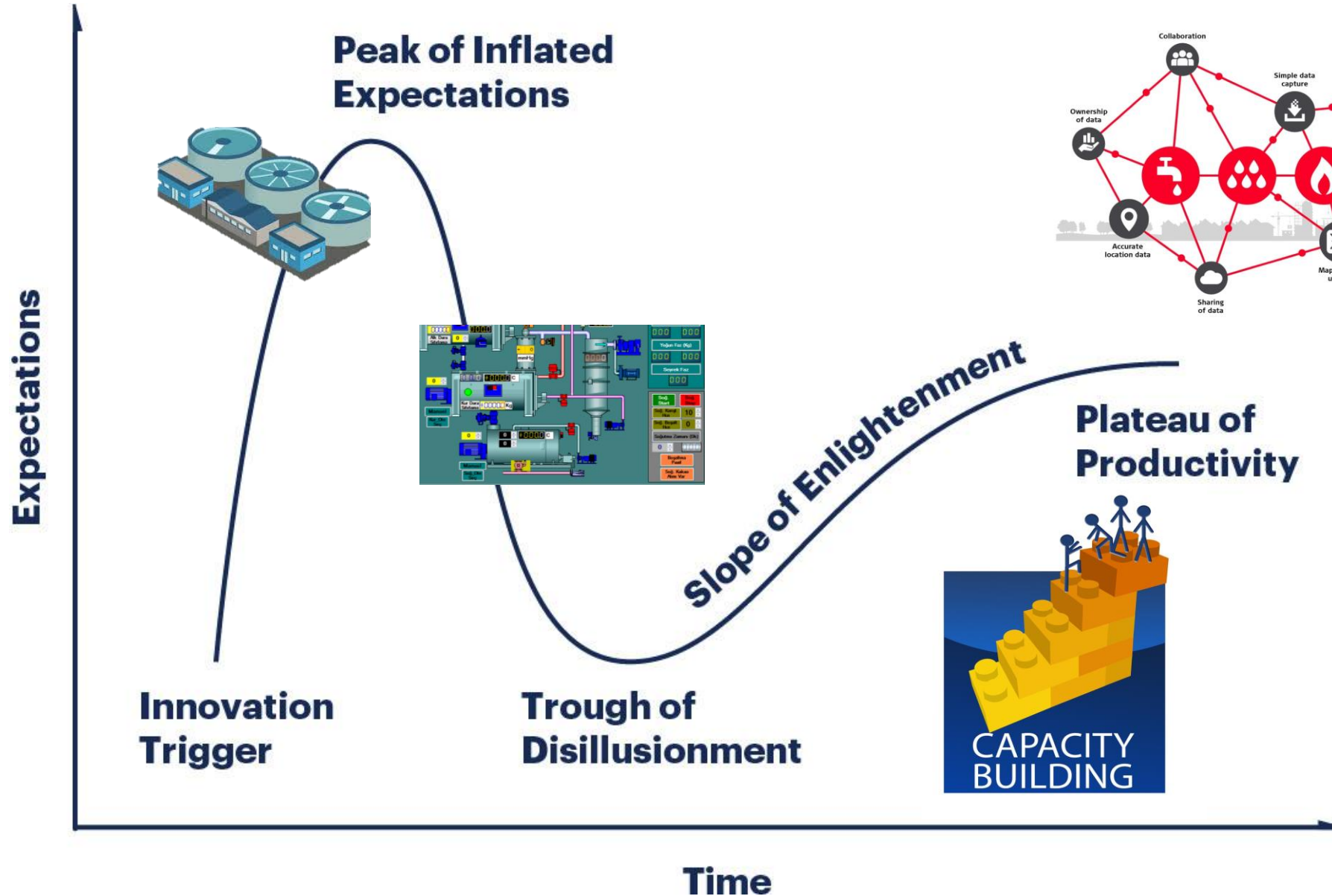


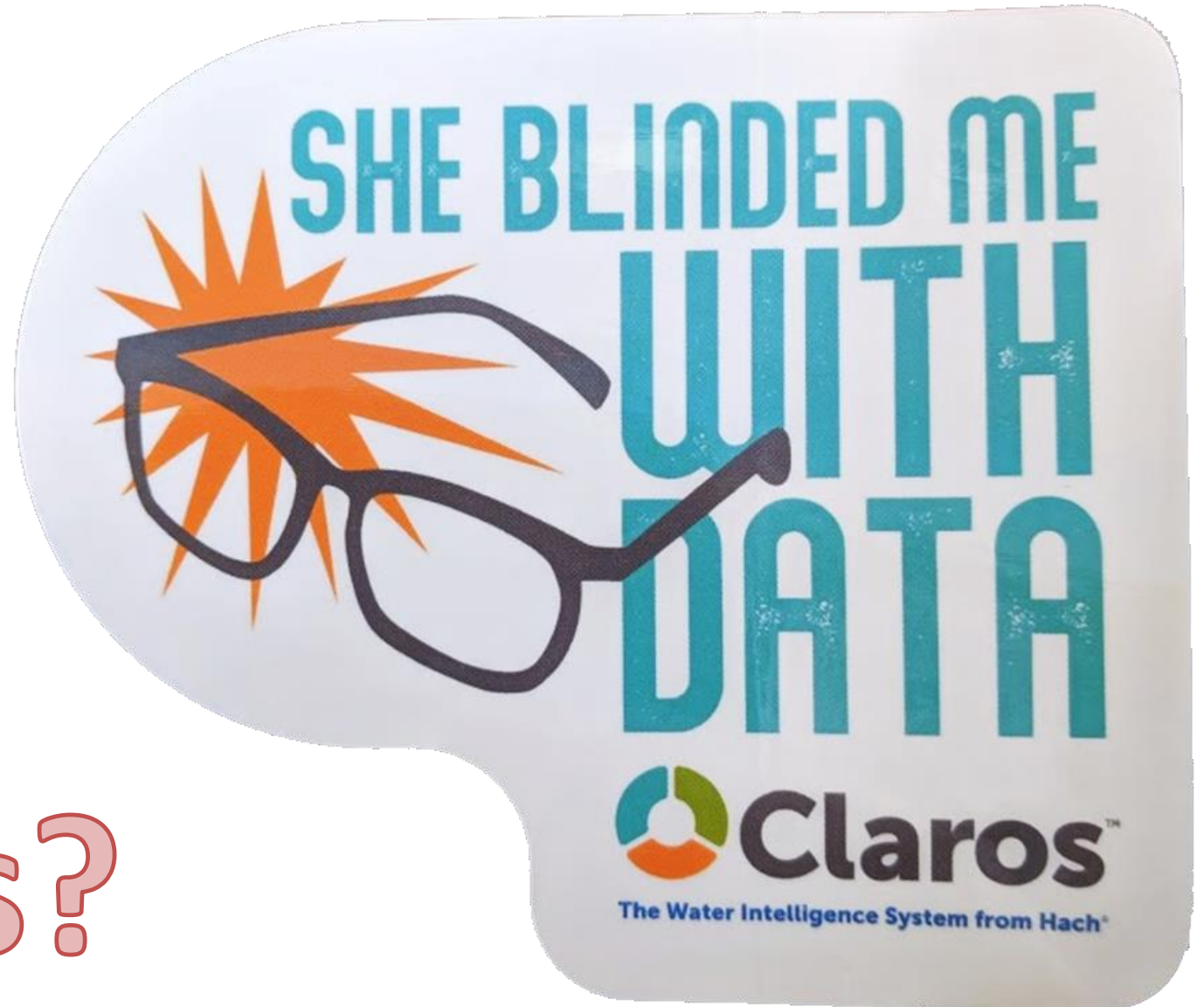
Hype Cycle: ML and Utilities





Hype Cycle: ML and Utilities





Questions?