CROZIER

ARTIFICIAL INTELLIGENCE AND MACHINE LEARNING

To Troubleshoot Well and Pump Issues in a Municipal Setting

PRESENTED BY:
Chris Gerrits, M.Sc., P.Eng.
Manager, Hydrogeology

November 22, 2024 | Municipal Engineers Association Conference 2024

ABOUT US

Crozier is an Ontario, Canada-based consulting engineering firm in the land development and building industry

We are committed to growing careers and building communities by delivering multidisciplinary engineering services to the private sector.

19+

Years of organic and consistent growth since being founded in 2004.

300+

Workforce of entrepreneurial, energetic, and caring employees.

5

Offices in key Ontario markets: Collingwood, Milton, Toronto, Bradford and Guelph.

ABOUT US

Chris Gerrits, M.Sc., P.Eng.

- B.Sc. in Water Resource Engineering (2000).
- M.Sc. In Water Resource Engineering (2001).
- MECP Licensed Well Technician.
- Over 30 years of experience in water well exploration, construction and testing.

WHAT ARE ARTIFICIAL INTELLIGENCE & MACHINE LEARNING?

Artificial Intelligence (AI) is a field, which combines computer science and robust datasets, to enable problem-solving.

- **Strong AI vs Weak AI**
  - Weak AI – trained for specific tasks (Siri, Alexa, autonomous vehicles).
  - Strong AI – theoretical form of AI where a machine has intelligence equal to humans (AGI) or greater than humans (ASI).

- **Subfields of AI include Machine Learning (ML) and Deep Learning (DL)**
  - ML uses statistical methods to train algorithms to make predictions via data mining.
  - DL doesn't require a labeled dataset (ie. Supervised model), it can determine the different categories of data without human intervention.

ML MODEL

MACHINE LEARNING MODEL

We train different models such as Random Forest, Xgboost, and Artificial Neural Network

MODEL EVALUATION

Is a process of assessing the performance and effectiveness of a model by measuring various metrices

OPTIMIZATION

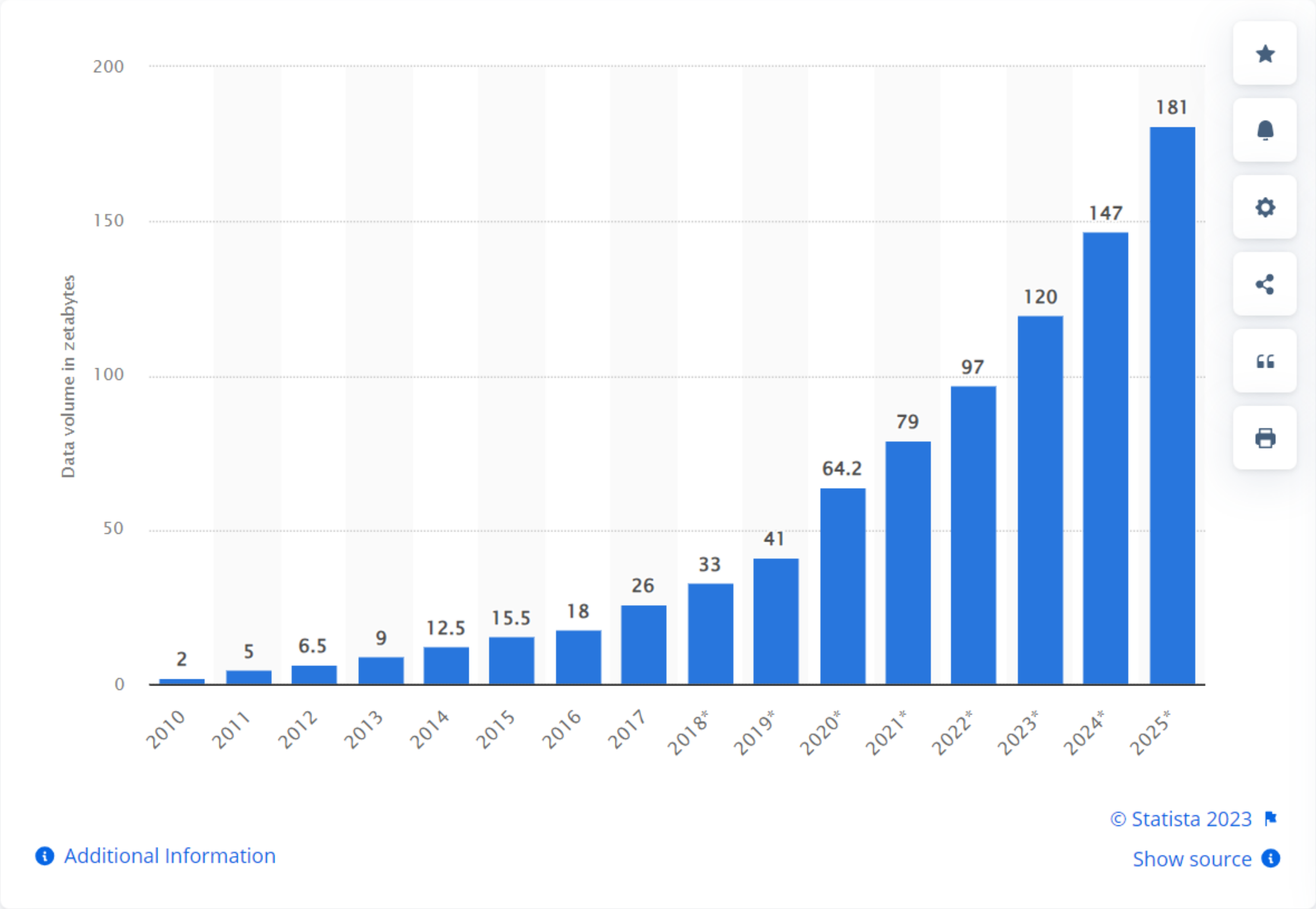
of model's predictive accuracy and generalization capabilities using various methods

WORLD DATA

- By 2025 there will be 181 zetabytes of data floating around the world.

- It would take a single user 181 years to download all that data.

- Canada as 269 data centres across the country (2021 stats).



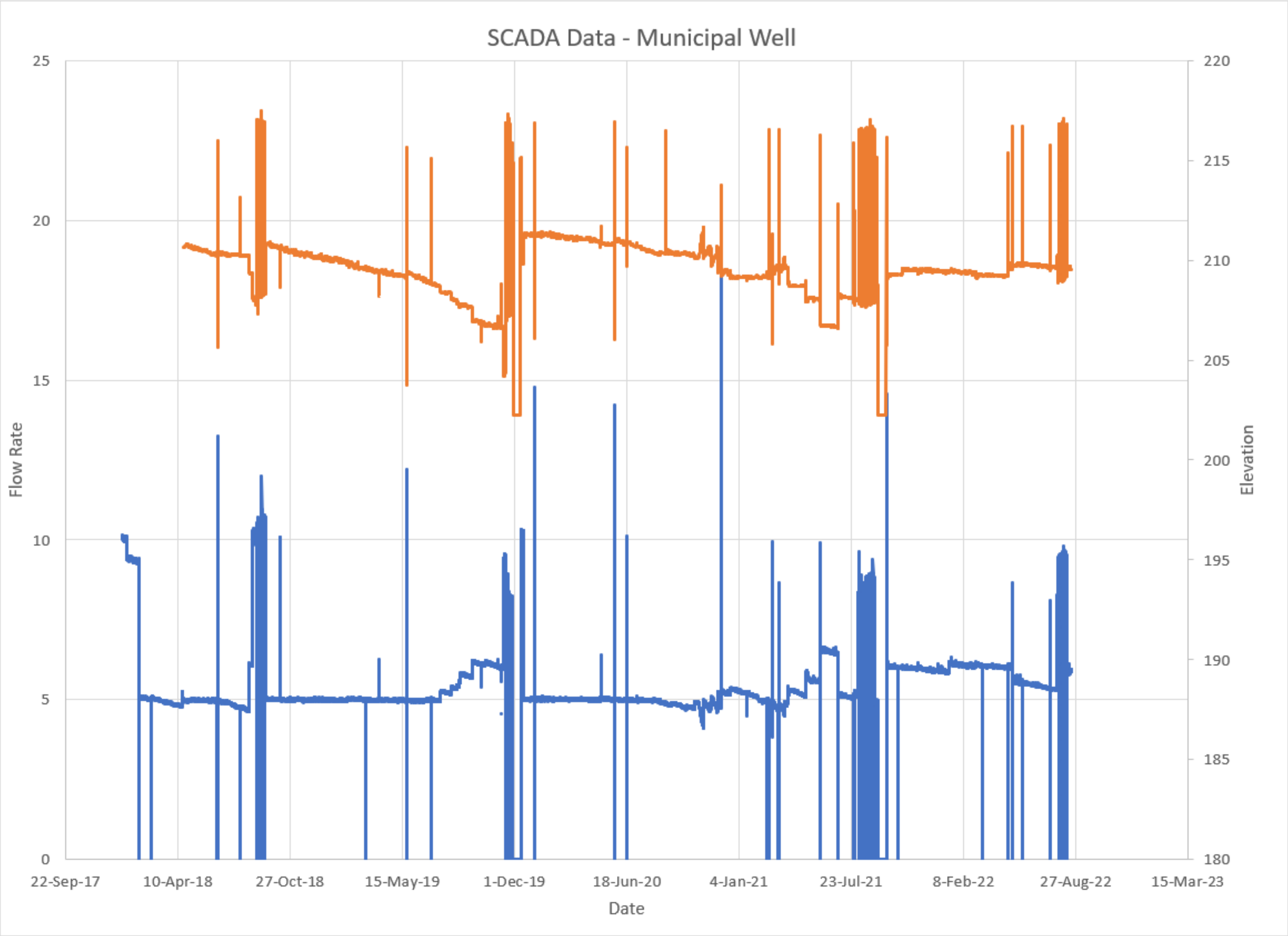
SUPERVISORY CONTROL AND DATA ACQUISITION TECHNOLOGY

- SCADA Systems have been in use in water & wastewater for decades.

- In Canada - **4,126** wastewater treatment plants and lagoon systems, **3,342** water treatment facilities, **472,488** km of underground pipes, **284,827** km of culverts and open ditches, many pump stations, storage facilities, and other assets.

- Precedence Research, Global SCADA market size;
  - $9.5 billion – 2022
  - $28.6 billion - 2032

DATA SCADA EXAMPLE



- Data from 01/01/2018 through 08/18/2022.
- 40583 data points per parameter.
- Water level, flow rate, pump speed other parameters collected.
- The data is used to some extent by the client.
- Could be used more to help with decision making.
- **TWO EXAMPLES OF HOW TO USE THIS DATA**

WHAT IS **PREDICTIVE MAINTENANCE?**

REACTIVE

Maintenance after breakdown

PREVENTIVE

Maintenance at regular intervals

PREDICTIVE

Predict breakdown before machine failure

PRESCRIPTIVE

Prescribes solutions

*"Predictive Maintenance is an **advanced technique** that helps us identify potential problems in machines before they completely stop working."*

CASE STUDY:

LEVERAGING OPERATIONAL DATA TO SCHEDULE REPAIR

To reduce unplanned operational downtime

PROBLEM STATEMENT

UNPLANNED DOWNTIME

- Unplanned downtime is a significant problem across multiple industries.
- Annual cost of unplanned downtime by industry;
    - FMCG/CPG - $35 billion
    - Oil/Gas - $47 billion
    - Heavy Industry - $225 billion
    - Automotive - $557 billion
- Fortune Global 500 firms lose $1.5 TRILLION globally to unplanned downtime.

PROBLEM STATEMENT

- Quantifying the hard costs of unplanned downtime in a communal or municipal system is difficult;
  - Billing models are all different.
  - Size & scale impact costs.
  - Unplanned downtime may not result in operating financial losses.
- 'Shadow costs' may be more significant to water system operators;
  - Lost staff time that has to be reallocated.
  - Poor public opinion of the operation and management.
  - Reduction in overall operational efficiency.

AI/ML FOR PREDICTIVE MAINTENANCE

MUTLIPLE SCENARIOS

MACHINE TYPE

- Case I – All machines are same
  - Single model can serve all machines.
- Case II – Machines are different
  - Model specific to each machine.

AI/ML FOR PREDICTIVE MAINTENANCE

MUTLIPLE SCENARIOS

FAILURE TYPE

- Case I – Single failure class
  - Predictive – problem.
- Case II – Multiple failures classes
  - Predictive + detecting multiple failure problems.

	THE CAUSES	THE PROBLEM: 1 Insufficient Discharge Pressure	2 Intermittent Operation	3 Insufficient Capacity	4 No Liquid Delivery	5 High Bearing Temperatures	6 Short Bearing Life	7 Short Mechanical Seal Life	8 High Vibration	9 High Noise Levels	10 Power Demand Excessive	11 Motor Trips	12 Elevated Motor Temperature	13 Elevated Liquid Temperature
1	Bent shaft					■	■	■	■		■			
2	Casing Distorted from excessive Pipe Strain					■	■	■	■		■		■	
3	Cavitations	■	■	■	■	■		■	■	■				■
4	Clogged impeller	■		■	■				■		■			
5	Driver imbalance						■	■	■					
6	Electrical Problems (Driver)						■	■	■		■	■	■	
7	Entrained Air (Suction or seal Leaks)	■	■	■					■	■			■	
8	Hydraulic Instability					■	■	■	■	■				
9	Impeller installed Backward (Double-suction Only)	■		■							■			
10	Impeller Mechanical Seal							■						
11	Inlet Strainer Partially Clogged	■		■					■	■				■
12	Insufficient Flow Through Pump													■
13	Insufficient suction Pressure (NPSH)	■	■	■	■				■	■				
14	Insufficient suction volume	■	■	■	■	■			■	■				■
15	Internal Wear	■		■					■		■			
16	Leakage in Piping, Valves, Vessels	■		■	■									
17	Mechanical Defects, Worn, Rusted, defective Bearings					■		■			■			
18	Misalignment					■	■	■	■		■		■	
19	Misalignment (Pump and Driver)								■		■	■		■
20	Mismatched Pumps in Series	■		■			■		■		■			
21	Non condensables Liquid	■	■	■					■	■			■	
22	Obstructions in lines or Pump Housing	■		■	■				■				■	■
23	Rotor imbalance						■	■	■					
24	Specific Gravity Too High	■									■		■	
25	Speed too High										■	■		
26	Speed too Low	■		■	■								■	
27	Total system Head Higher than Design	■	■	■	■	■		■					■	■
28	Total system Head Lower than Design					■		■	■	■	■	■		■
29	Unsuitable Pumps in Parallel Operation	■		■	■	■			■	■		■		■
30	Viscosity Too High	■		■							■		■	
31	Wrong Rotation	■			■						■		■	

Common failure modes of centrifugal pumping system

- International Journal of Modelling Identification and Control lists 31 different causes of failure in centrifugal pump.
- Thirteen different “problems”.

Sivaprakasam et all, 2008

AI/ML FOR PREDICTIVE MAINTENANCE

MUTLIPLE SCENARIOS

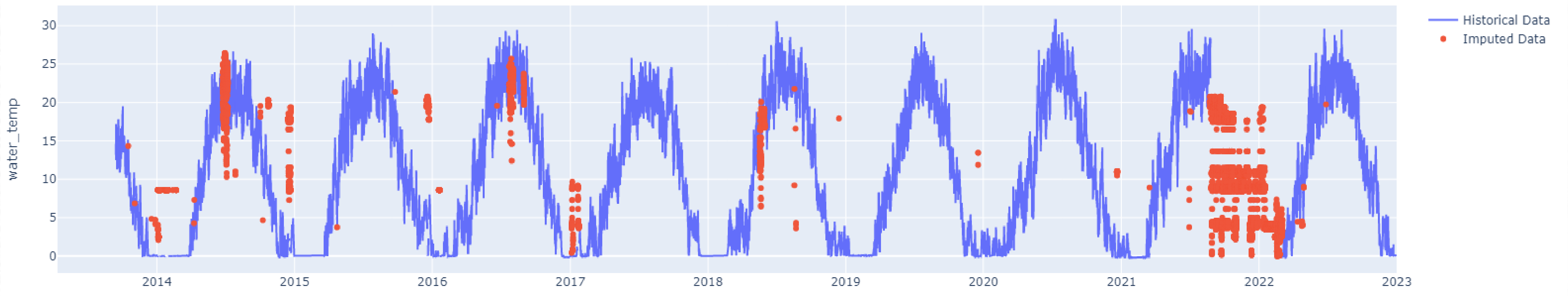
SENSOR

- Case I (Single Failure) – Sensors are calibrated
  - No data adjustment.
- Case II (Multiple Failures) – Sensors not calibrated
  - Data needs to be adjusted.

DATA QUALITY

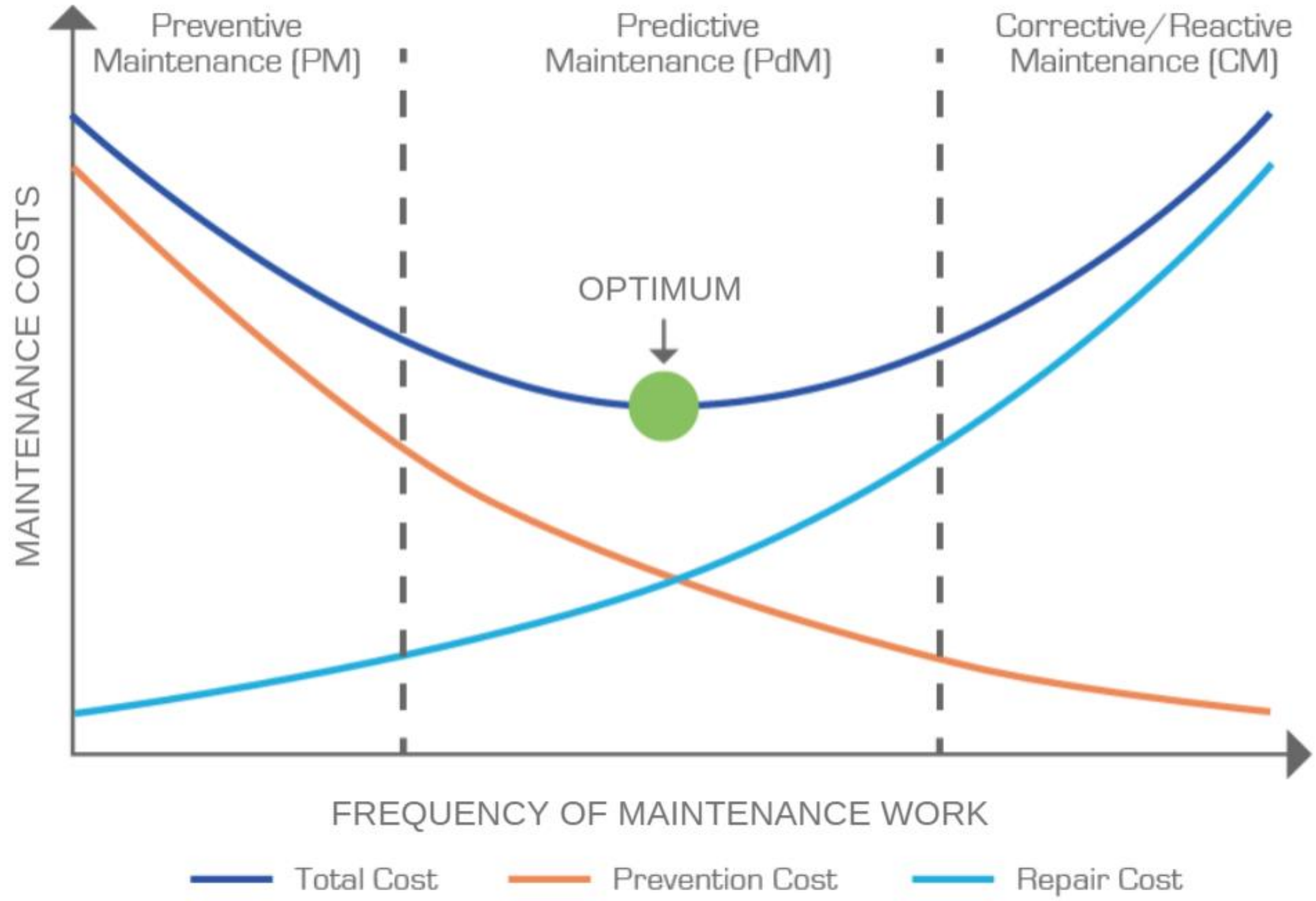
- Case I (Single Failure) – Data is ‘clean’ with no gaps
  - No data cleaning required.
- Case II (Multiple Failures) – Data has missing values
  - Need to strategize data cleaning and missing value **imputation** methods.

EXAMPLE: DATA IMPUTATION

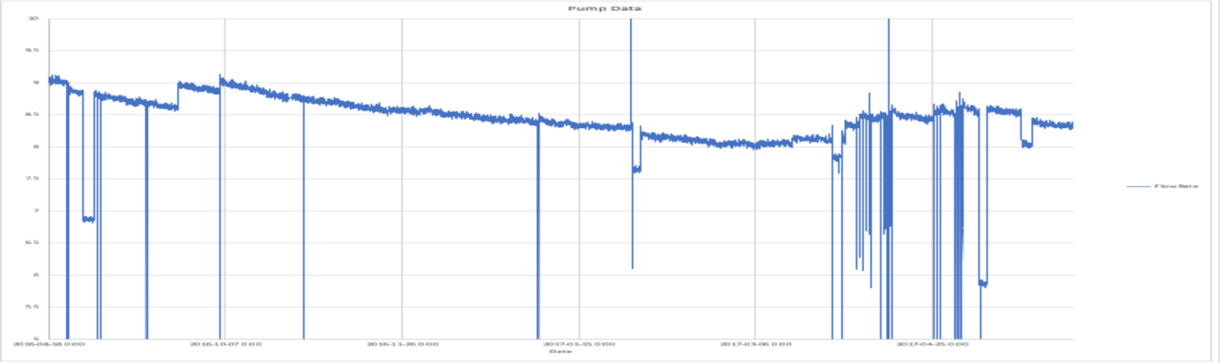


- Discarding high quality data because of data gaps is not optimal.
- We need to implement a sophisticated imputation method based on machine learning.
- Here, we have applied the Light Gradient Boosted Machine algorithm to impute missing values in the data.
- The imputation process almost recreates similar data by considering the trend and seasonality. However, some smoothing of the predicted data will be required before fitting any machine learning model for data forecasting.

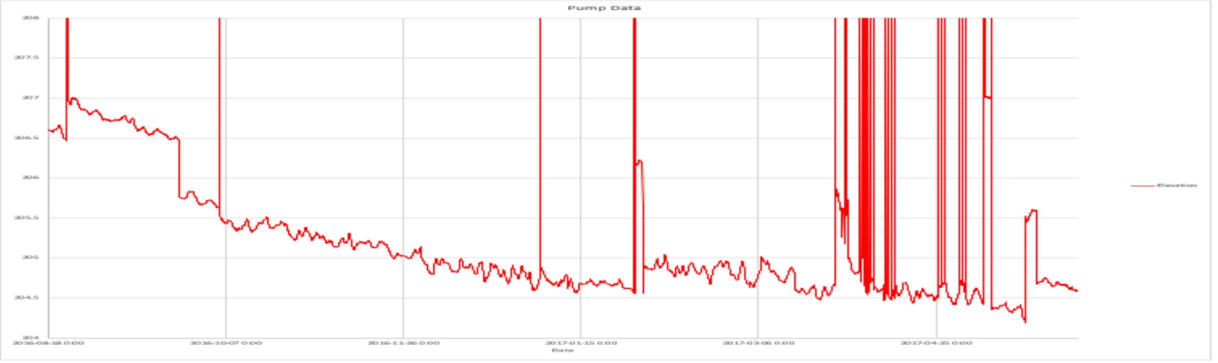
PROBLEM STATEMENT



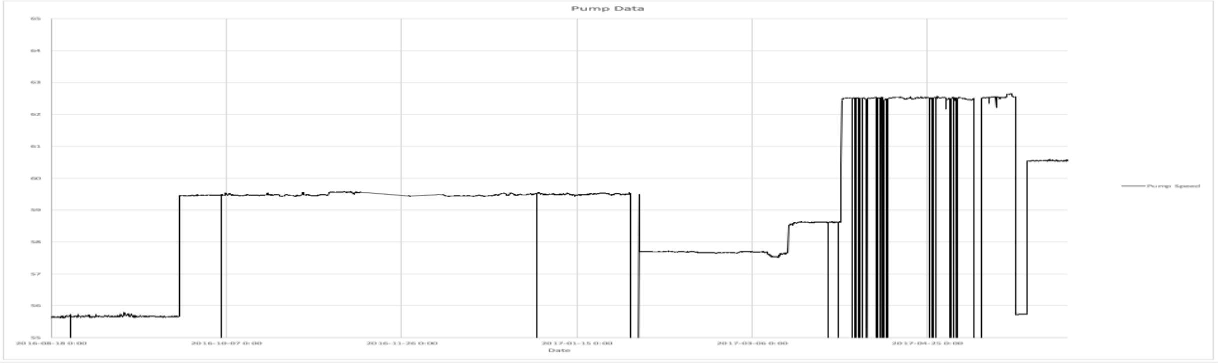
PREDICTION OF RUL – PUMP & MOTOR



Flow Rate



Groundwater Elevation



Pump Speed

PREDICTION OF RUL – PUMP & MOTOR

Flow Rate – decreasing slightly during the period

Groundwater Elevation – decreasing slightly during the period

Pump Speed – increasing during the period

Discharge Pressure – assumed constant

AI/ML predictive model can help predict when the pump will require maintenance or replacement so that it can be scheduled.

CASE STUDY

ARTIFICIAL INTELLIGENCE AND MACHINE LEARNING

to better predict scheduling of well rehabilitation and maintenance

1 YEAR LATER

CROZIER
CONSULTING ENGINEERS

THE SITE

LOCATION

- York Region (Richmond Hill).
- Approximately 12 miles (19 kilometers) northeast of the City of Toronto.
- York Region is a regional municipality that covers an area of 679 square miles (1,758 km²) between Lake Simcoe and Toronto .
- 41 Groundwater Wells (plus 3 in development).
- 2 Surface Water Treatment Plants and 2 Connections to Lake Ontario (via Peel and Toronto).

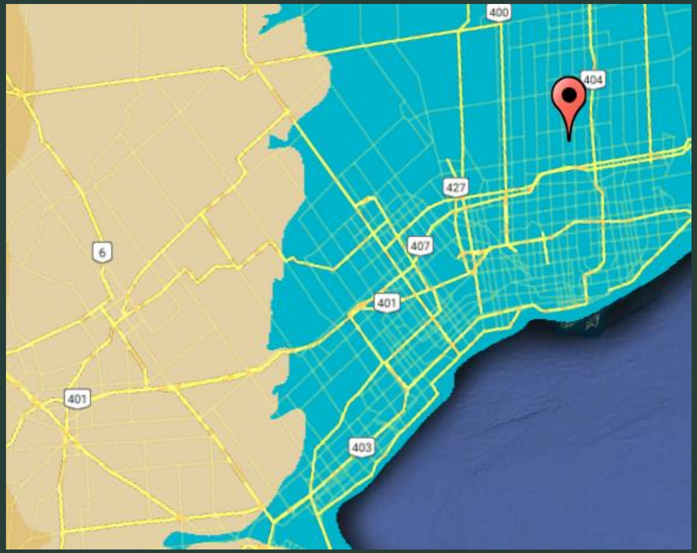
GEOLOGIC SETTING

PHYSIOGRAPHIC REGION

- Peel Plain covers 300 square miles across the Regional Municipalities of York, Peel, and Halton
- Gradual and uniform sloping topography

SURFICIAL GEOLOGY

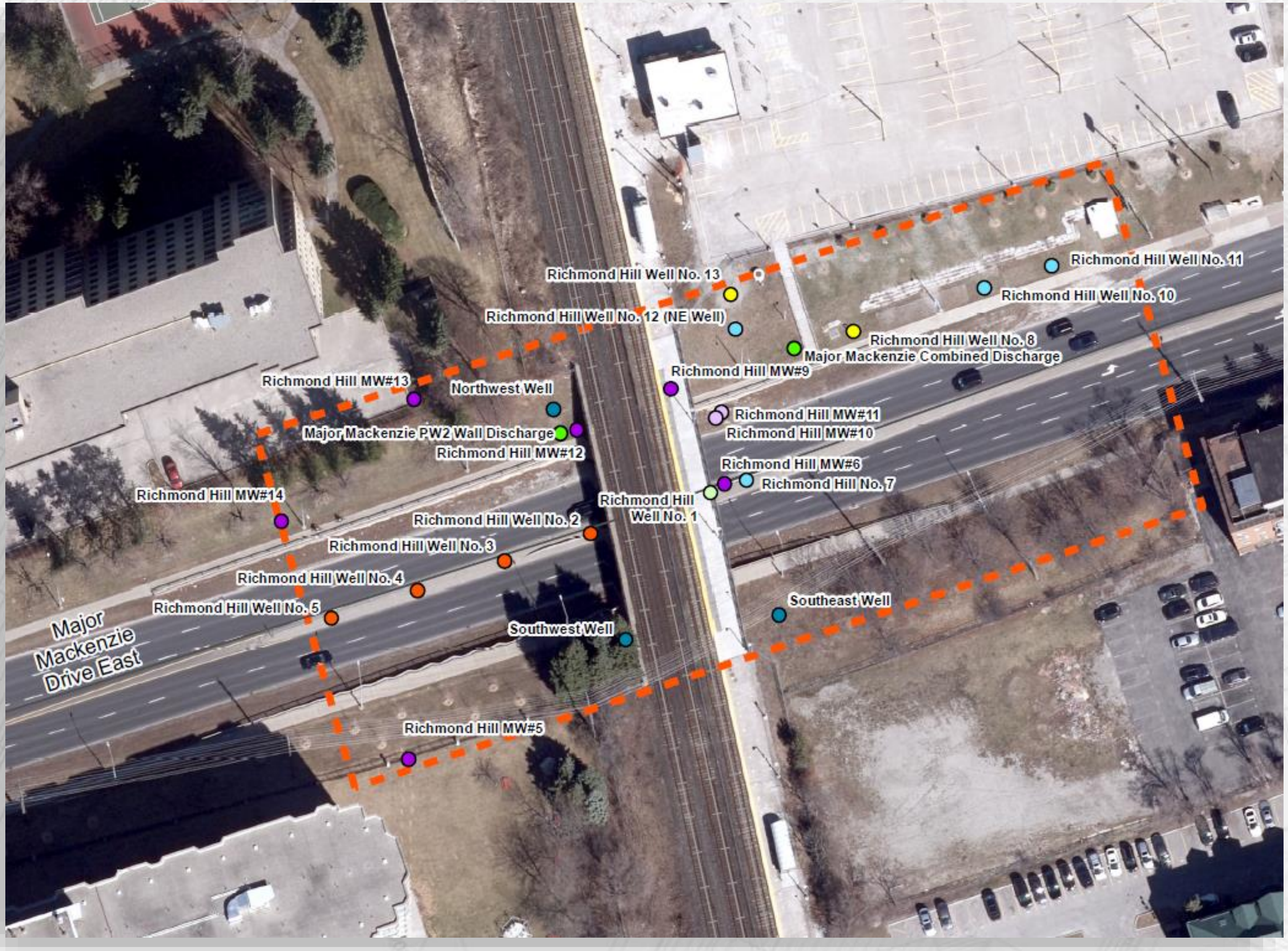
- Fine textured glaciolacustrine deposits
- Silt and clay, minor sand and gravel
- Some drumlin or drumlinoid features

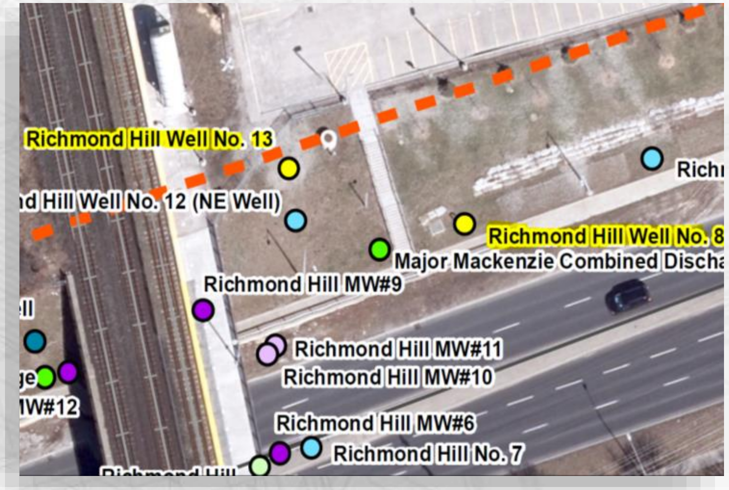


BEDROCK GEOLOGY

- Shale, limestone, dolostone, siltstone
- Georgian Bay Fm, Blue Mountain Fm, Billings Fm

THE SITE

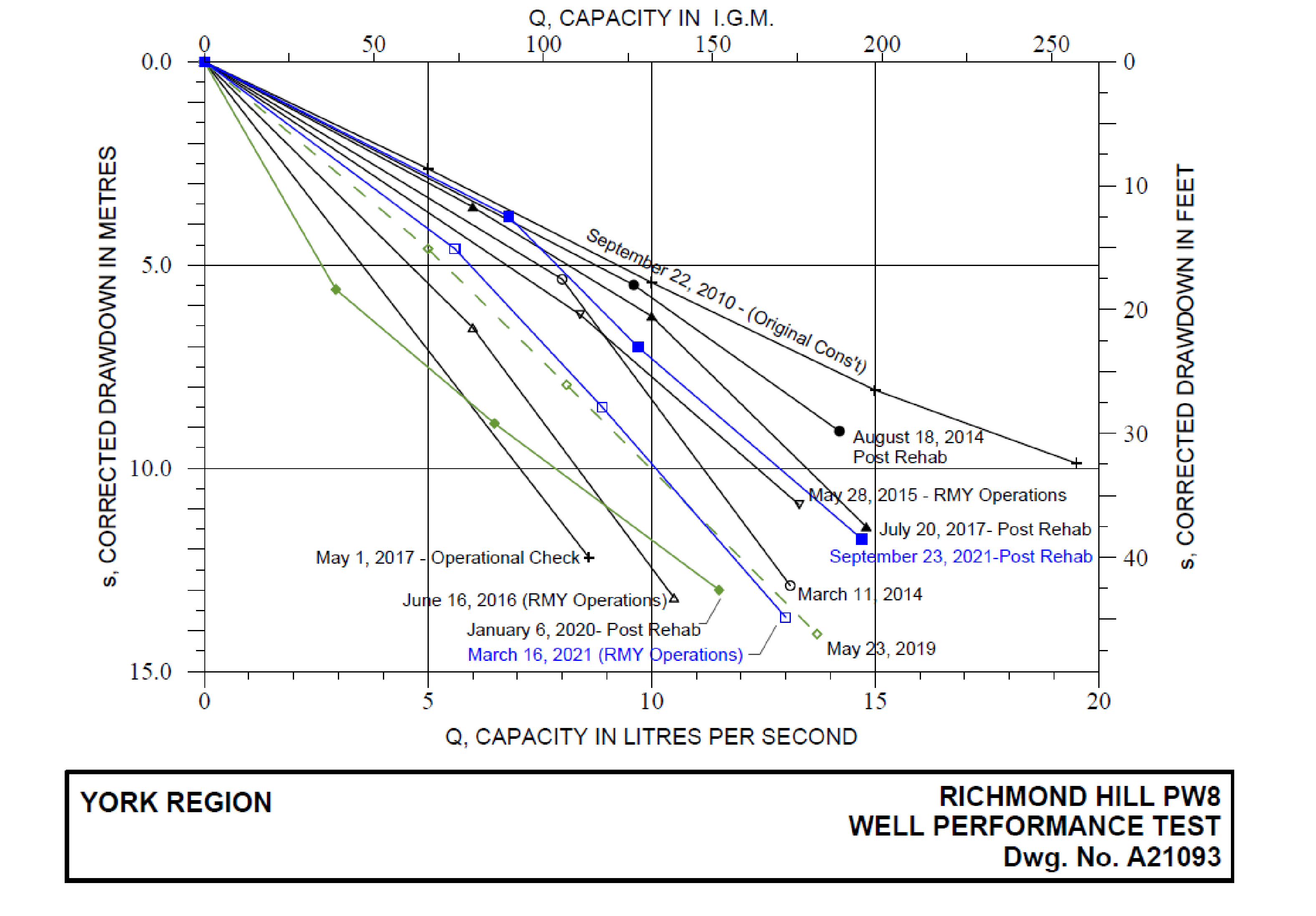




SITE LAYOUT

- 22 Wells total.
- 8 Monitoring Wells.
- 2 Active Dewatering Wells – PW8 and PW13.
- These wells were studied for the purpose of this project.

PROBLEM STATEMENT



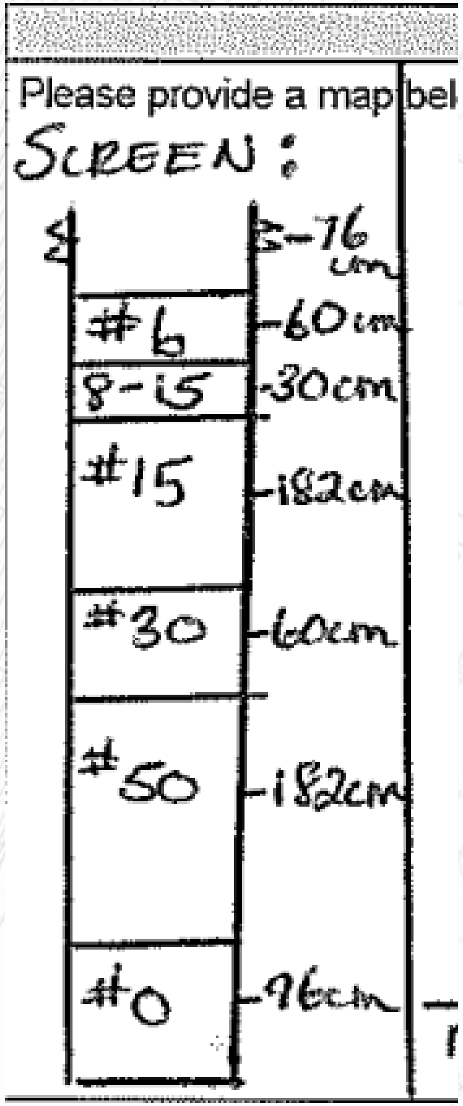
PROJECT

- Well Rehabilitation Diminishing Returns.
- New Water Supply Wells take 3-10 years to Develop.
- Rehabilitation is traditionally done on a REACTIVE basis.
- Once the need for rehabilitation is established there is a long delay to schedule work.

PROBLEM STATEMENT

- Delay between identification and work leads to additional decrease in well yield.
- Diminishing Returns means a % of the additional decrease cannot be recovered through rehabilitation.
- Rehabilitation is traditionally done on a **REACTIVE** basis.
- Goal is to schedule well rehabilitation on a **PROACTIVE** basis.
- Will lead to an increase in the useful life of the well **(ASSET MANAGEMENT)**.

PW8 WELL LOG



Screen Slot	Length
Lead Pipe	2.5 ft
#6 slot	2.0 ft
#8 - #15 slot	1.0 ft
#15 slot	6.0 ft
#30 slot	2.0 ft
#50 slot	6.0 ft
#0 slot	3.1 ft

- 12-inch diameter dewatering well with 12-inch telescoping Johnson screen.
- Screen extends from 60.9 – 83.0 ft below grade.
- Screen is adjacent Brown Medium/Coarse Sand overlying Brown Sand/Gravel.
- Bottom 3.1 ft of well screen is #0 slot tight wound to act as a sump for this application.

MACHINE LEARNING TECHNIQUES FOR PREDICTIVE MAINTENANCE

REGRESSION MODELS

Predict Remaining Useful Life

- Remaining Useful Life (RUL) prediction provides insights about when the machine will fail next time.
- This helps in scheduling maintenance in advance.
- Static and historic data is required for training purpose.

CLASSIFICATION MODELS

Classification of different types of Failures

- Classification models help in predicting types of failures in a machine or predicting if asset will fail within a certain time frame.
- Data needs to be accurately annotated.

ANOMALY DETECTION

Flagging Anomalous Patterns

- Anomaly detection models flag outliers or abnormal behavior of a machine.
- Target data is not required.
- False negative prediction can cause a huge loss.

PREDICT REMAINING USEFUL LIFE (RUL)

*We need to answer **when is the next rehabilitation** necessary for the well?*

DATA COLLECTION

- Water Flow.
- Pump Speed.
- Water Elevation.
- Specific Capacity (Calculated using Drawdown for each timestamp).

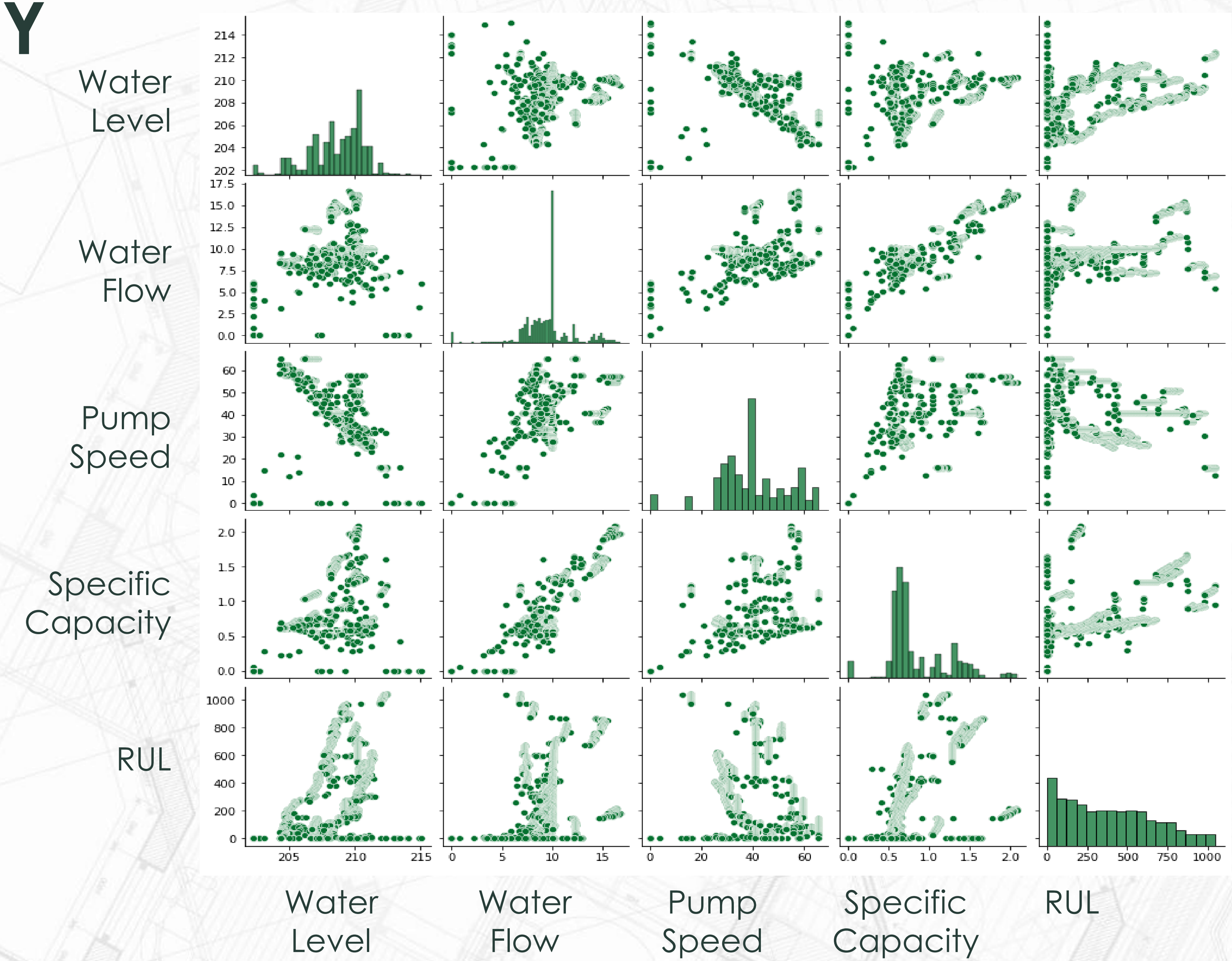
DATA CLEANING

- Converted data with 5 minutes granularity to hourly granularity.
- Replaced missing data points with rolling average of last 24 hours.

DATA ANNOTATION

- Assigned target values in terms of number of hours.

EXPLORATORY DATA ANALYSIS



MACHINE LEARNING **MODEL**

Train-Test Split with
Training Data: 75%
Testing Data: 25%.

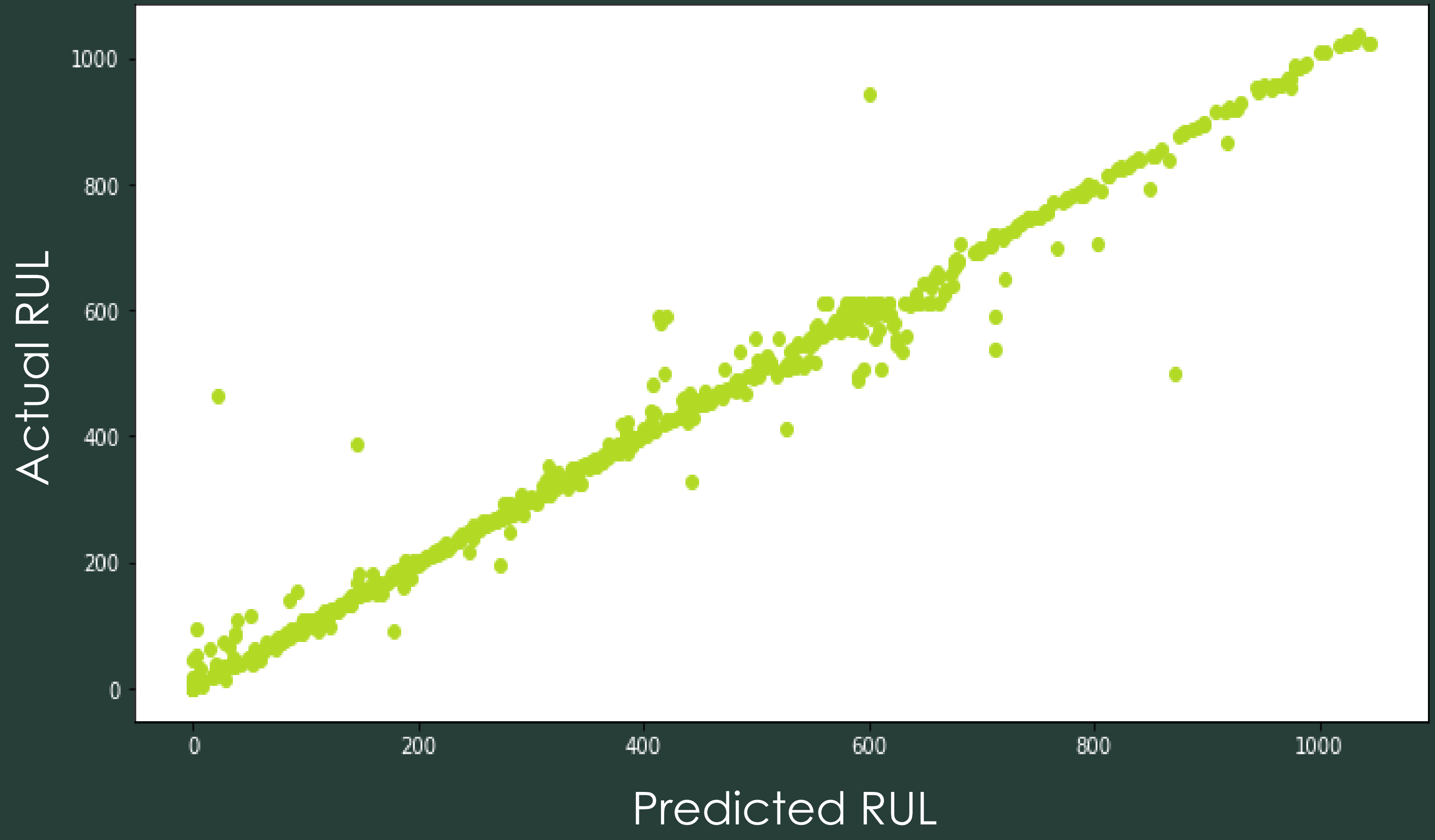
Supervised Models
XGBoost, RandomForest,
and Artificial Neural Network.

Standardization
of data before
fitting ML algorithms.

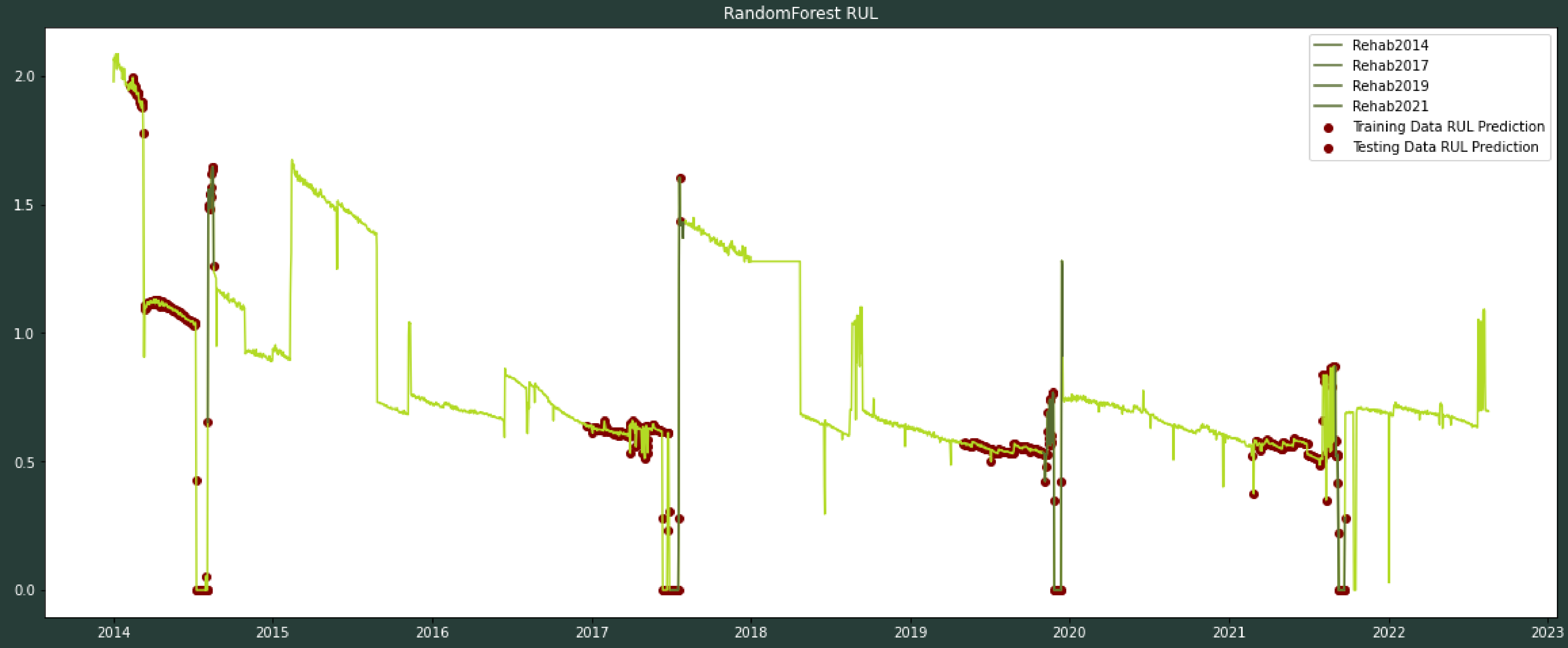
Optimization using
GridSearchCV with
10-fold cross validation.

SUPERVISED MACHINE LEARNING - RESULT

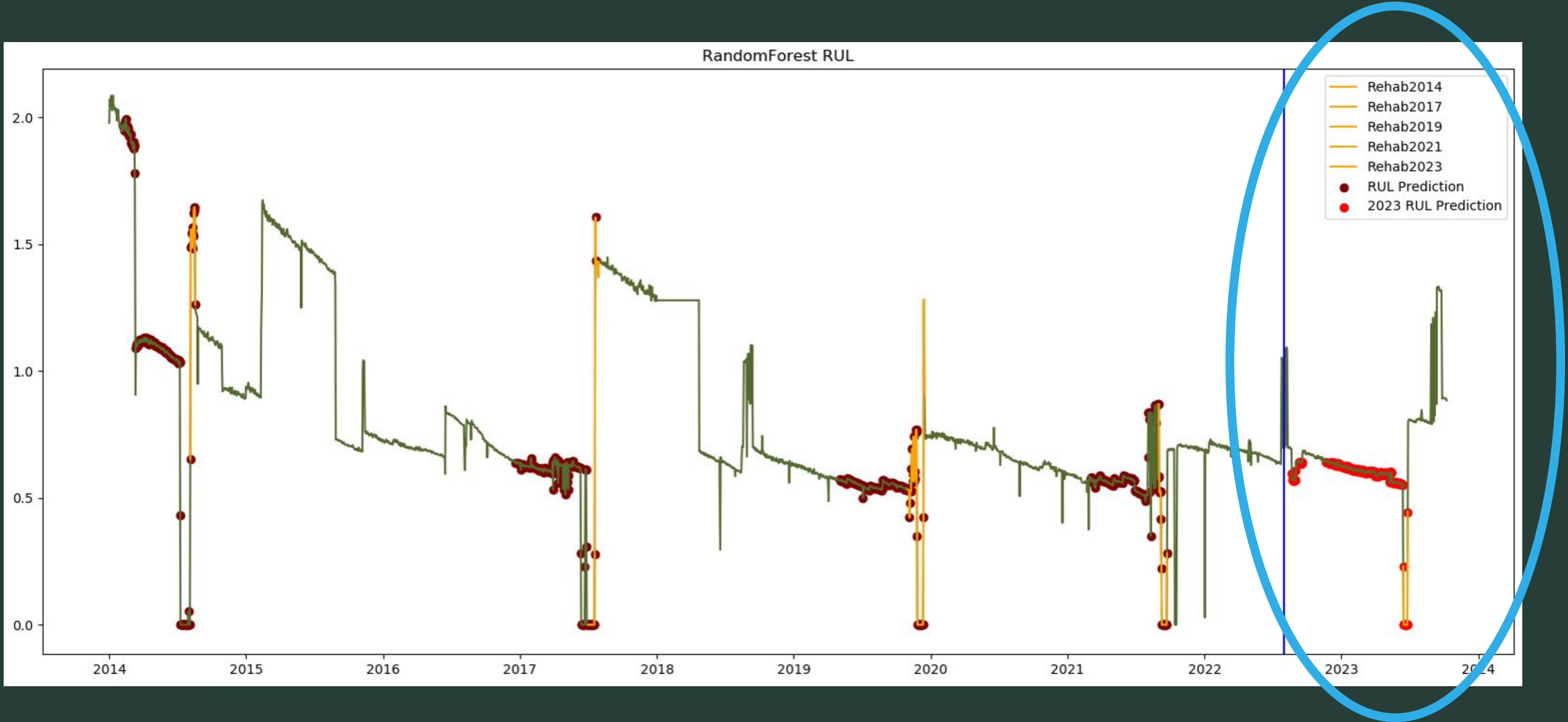
ML Regressor	RMSE	MAE	R2
XGBoost	1057.64	275.96	0.98
RandomForest	**980.6**	**200.81**	**0.98**
Neural Network	1329.96	577.64	0.96



NOV 2022 - PREDICTION OF REHAB BEFORE 180 DAYS

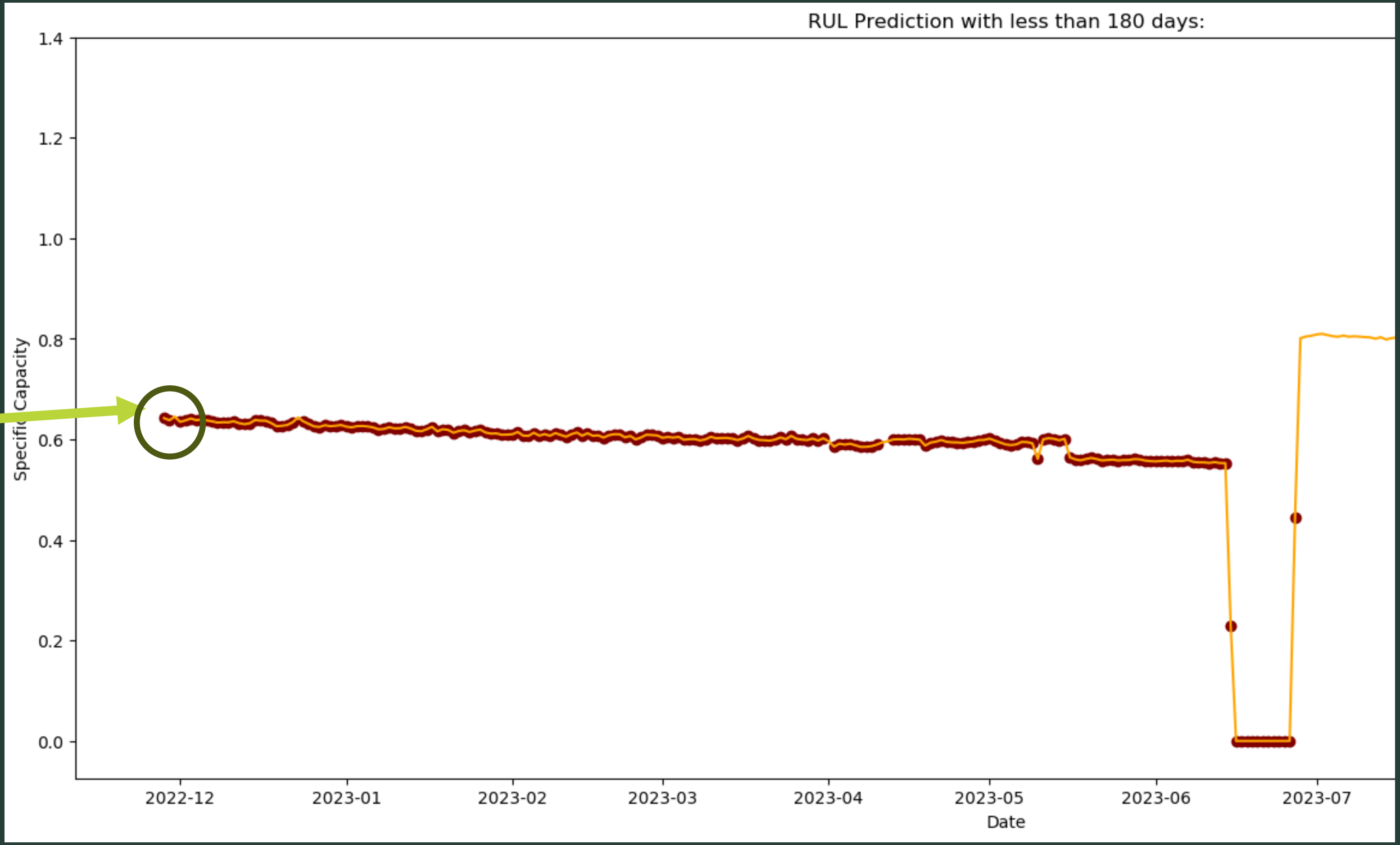


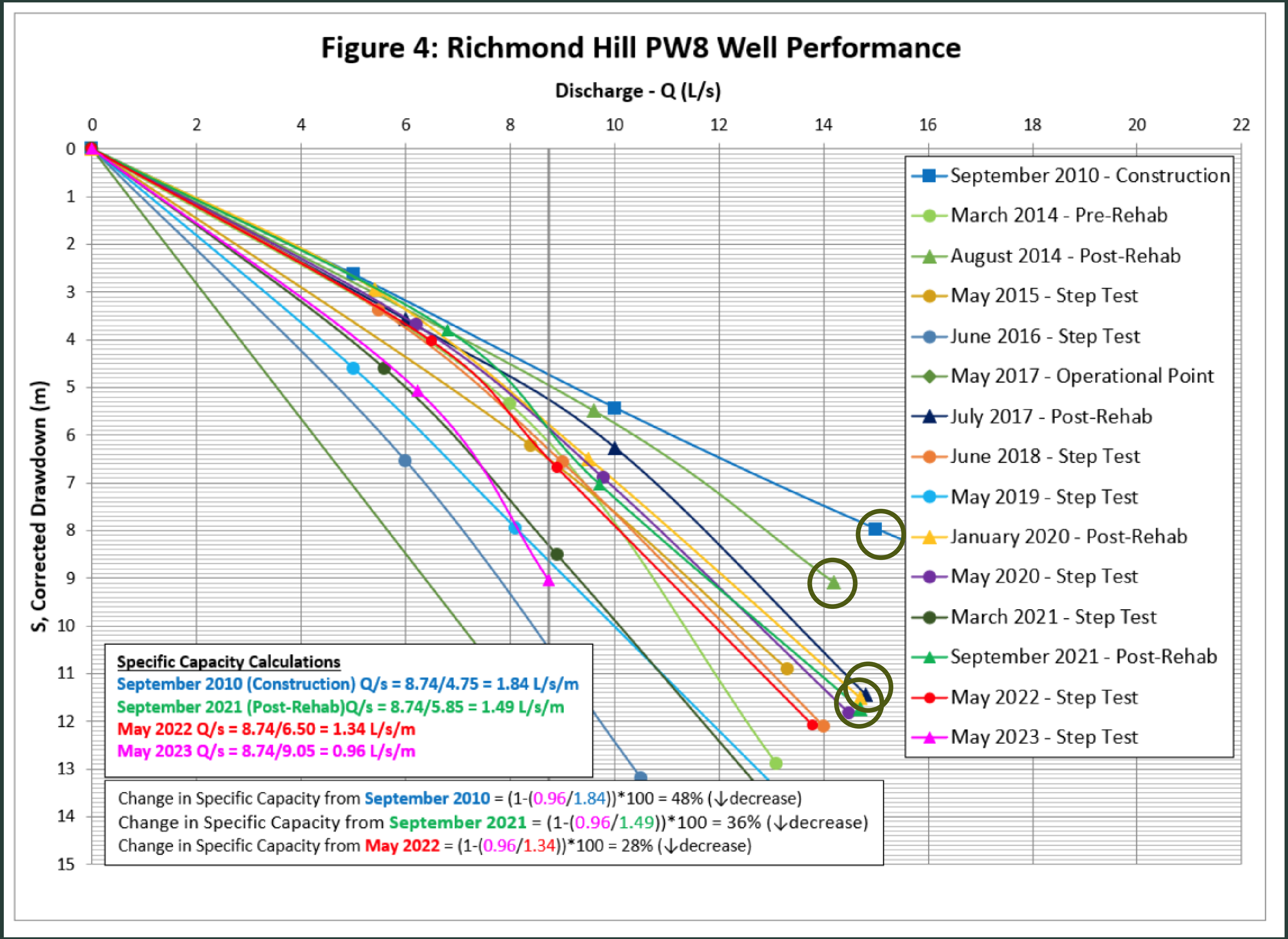
NOV 2023 - PREDICTION OF REHAB BEFORE 180 DAYS



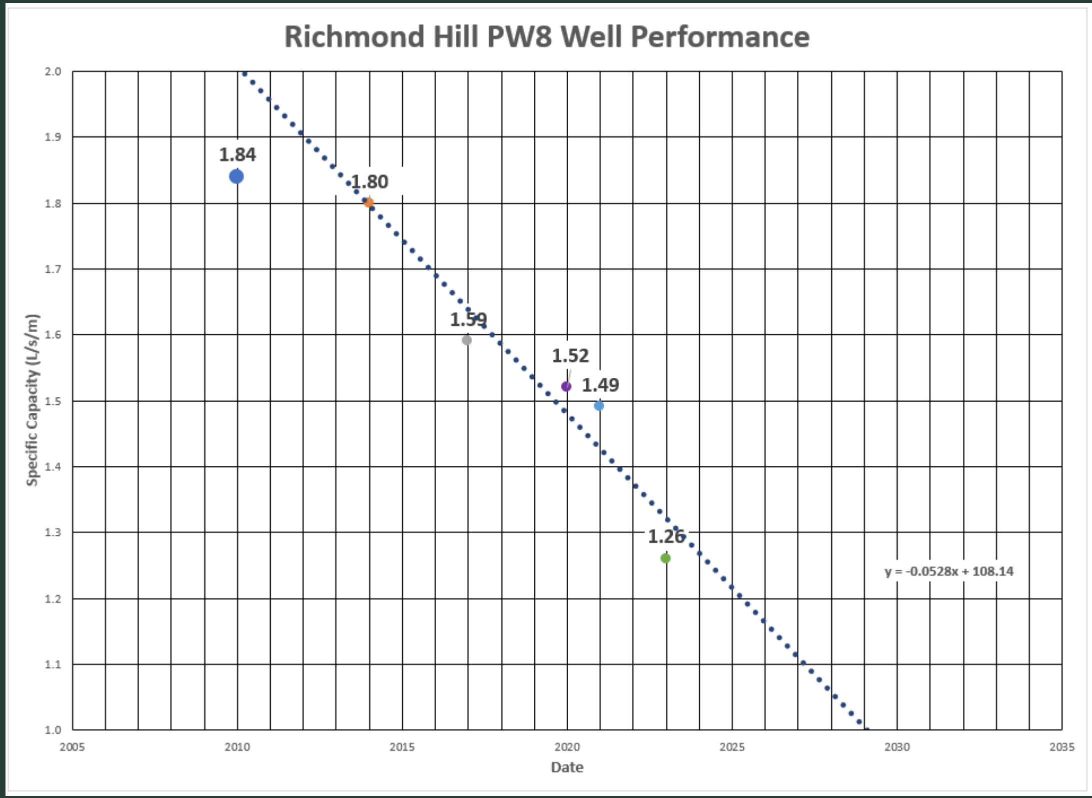
NOV 2022 - PREDICTION OF REHAB BEFORE 180 DAYS

- The model consistently indicated commending November 28, 2022 that well rehabilitation would be required within 180 days from that date.
- RUL < 180 days
- Therefore model predicts rehab on/ around May 26, 2023.





- Results of the May 2023 Step Test indicated that the well had reached the point where rehabilitation was required.
- Reduction >30% from last rehabilitation event.



- Linear regression indicates an average decrease of 0.0528 L/s/m per year within the system with rehabilitation regularly completed.
- That is 2.9% of the efficiency lost per year under ideal conditions.
- 0.0044 L/s/m per month lost (0.23%)
- In 2023 there were 2 months between when the rehabilitation was identified as being required and the commencement of the rehabilitation effort.
- Additional avoidable losses of 0.0088 L/s/m over than two month period.

- Post rehabilitation specific capacity continues to decrease overall. Each subsequent rehabilitation effort recovers less of the lost efficiency.
- At current rates the well will be operating at less than 50% of pre-construction conditions by 2030, less than 25% of pre-construction conditions by 2039.

WHAT DOES THIS ALL MEAN?

- Rehabilitation has been required 5 times over the 13 years of the life of the system with increasing frequency.
- Assuming rehabilitation is required every 2 years and there is a 2 month gap between identification of the need and completion of the work.
- 0.0044 L/s/m per month lost (0.23%).
- In 2023 there were two months between when the rehabilitation was identified as being required and the commencement of the rehabilitation effort.
- Additional avoidable losses of 0.008 L/s/m over a two-month period.
- **Effective prediction and proactive maintenance (rehabilitation) can add over 5% to the total life of the asset in this situation.**
- Each situation is different and in some cases the savings could be significant greater, some cases negligible.

LIMITATION OF USING SUPERVISED ML MODELS

NO GENERALIZATION

- We cannot implement same model to assets with different characteristics

NO EXTRAPOLATION

- It can only interpolate
- Accurate prediction on seen data only
- Cannot predict parameters that are unseen

CONTINUOUS UPDATES

- Model needs to retrain once we have new data to fit

LIMITATION OF USING SUPERVISED ML MODELS

NO GENERALIZATION

- We can not implement same model to assets with different characteristics.

NO EXTRAPOLATION

- It can only interpolate.
- Accurate prediction on seen data only.
- You cannot take the model and predict an unseen parameter

CONTINUOUS UPDATES

- Model needs to retrain once we have new data to fit.

THANK YOU

www.THECROZIERWAY.ca

2024 SME
CANADA'S TOP
SMALL & MEDIUM
EMPLOYERS

2024 WINNER
REPORT ON BUSINESS
CANADA'S TOP GROWING
COMPANIES

2024
SOUTHWESTERN ONTARIO'S
TOP EMPLOYERS

CELEBRATING 20 YEARS
SINCE 2004
CROZIER

CASE STUDY:

WATER QUALITY TREND ANALYSIS

CROZIER
CONSULTING ENGINEERS

Source: Minnesota Pollution Control Agency

CASE STUDY: Water Quality Trend Analysis

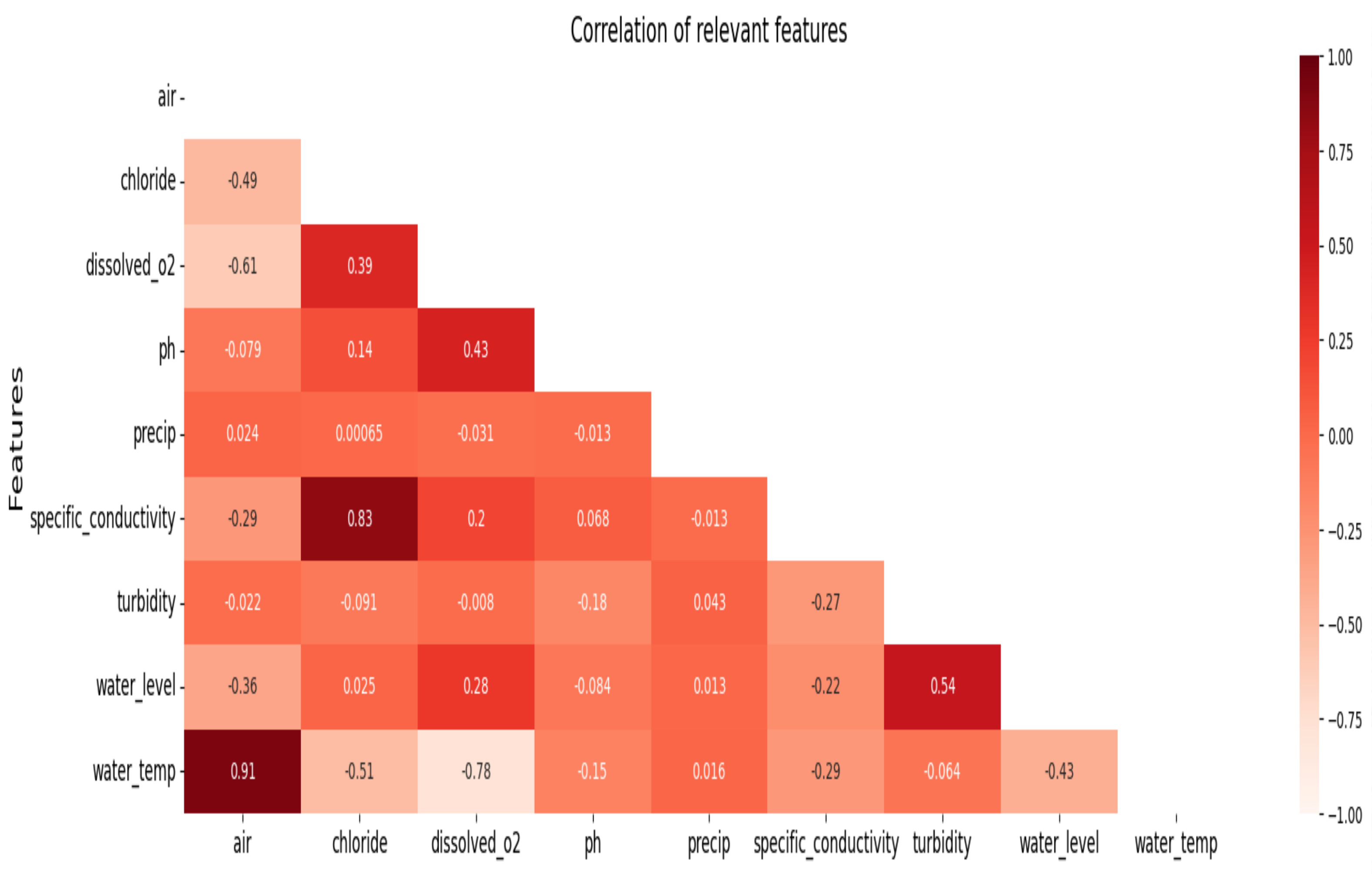
- Many Conservation Authorities and other regulatory bodies complete multi variable data collection
- Sample from one surface water location in a Southern Ontario stream
- Data collected in 15 minute intervals, millions of data points over the time period investigated

FUN FACT:

Chris failed Chem 1

Features	Start Datetime	End Datetime	Data Range
Air Temperature (deg c)	2010-02-20 07:00:00	2022-12-31 08:00:00	98943
Chloride (mg/L)	2010-02-20 07:00:00	2022-12-31 08:00:00	101647
Dissolved Oxygen (mg/L)	2010-02-20 07:00:00	2022-12-31 08:00:00	99171
pH	2010-02-20 07:00:00	2022-12-31 08:00:00	97149
Precipitation (mm/hr)	2010-02-20 07:00:00	2022-12-31 08:00:00	112730
Specific Conductivity (uS/cm)	2010-02-20 07:00:00	2022-12-31 08:00:00	98336
Turbidity (NTU)	2010-02-20 07:00:00	2022-12-31 08:00:00	94462
Water Level (masl)	2013-09-13 13:00:00	2022-12-31 08:00:00	81362
Water Temperature (deg c)	2010-02-20 07:00:00	2022-12-31 08:00:00	104818

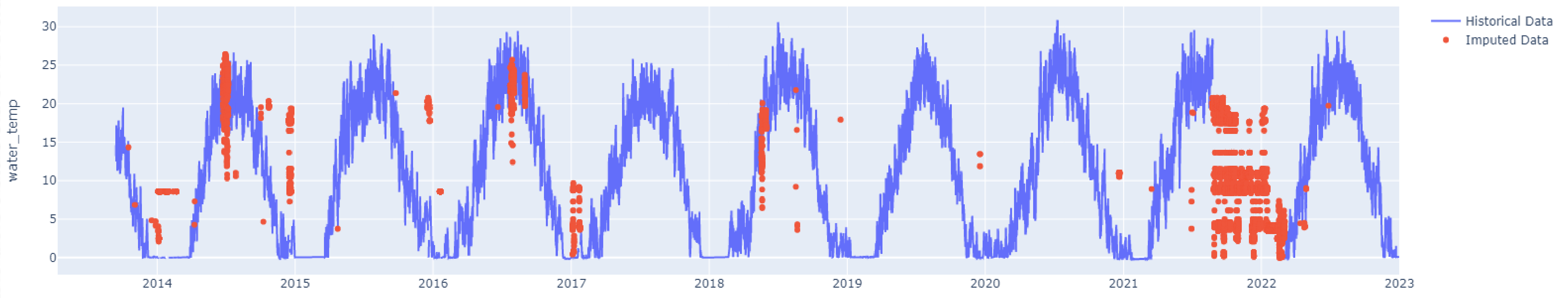
CASE STUDY: Water Quality Trend Analysis



The scatter plot below represents the correlation direction between features, while the correlation matrix represents the correlation factor between features.

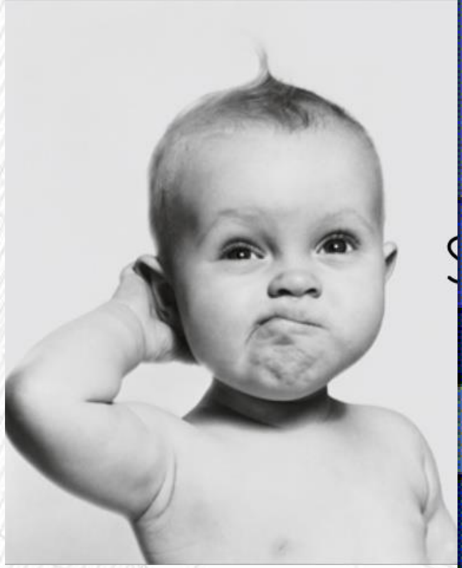
- High +ve correlations;
  - Water Temp - Air Temp (0.91)
  - Chloride - Conductivity (0.83)
- Moderate +ve correlations;
  - Water Level – Turbidity (0.54)
  - pH - Dissolved O2 (0.43)
  - Dissolved O2 - Chloride (0.39)
- High -ve correlation;
  - Dissolved O2 - Water Temperature (-0.78)
  - Dissolved O2 - Air Temperature (-0.61)
- Moderate -ve correlation;
  - Chloride - Water Temperature (-0.51)

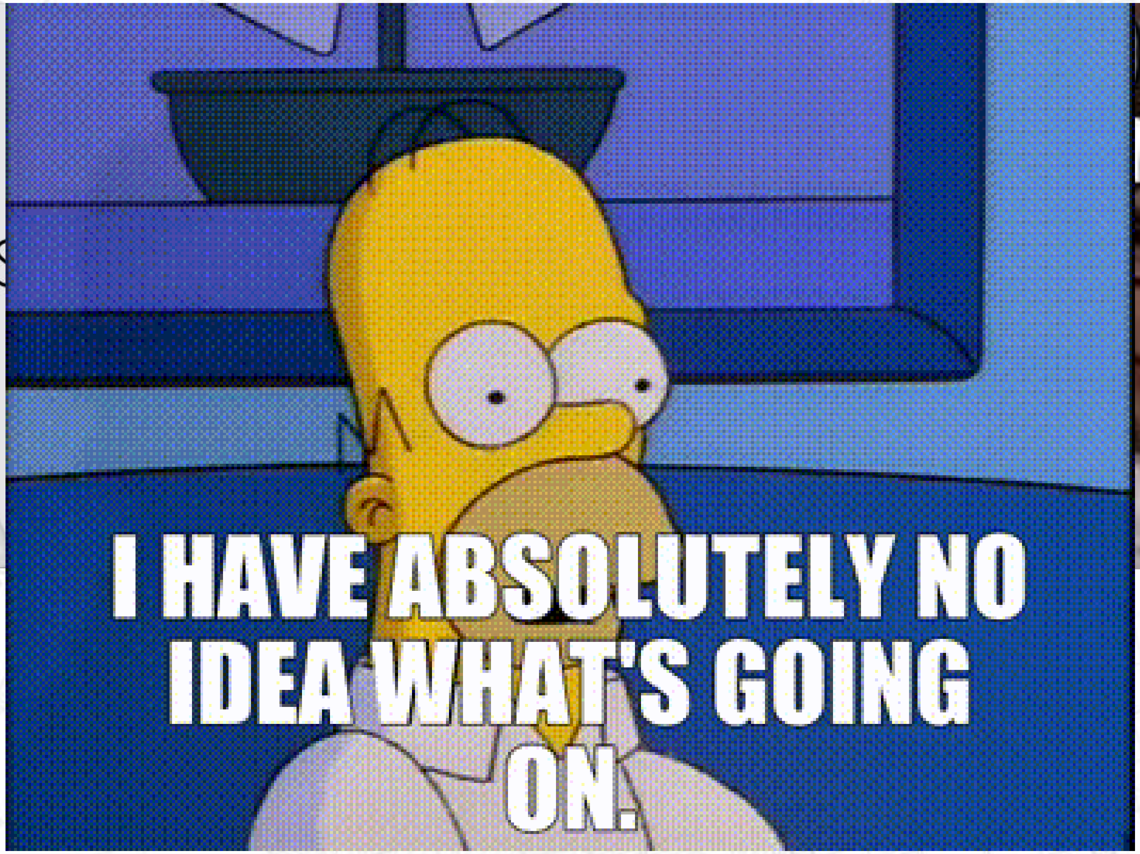
CASE STUDY: Water Quality Trend Analysis



- Discarding high quality data because of data gaps is not optimal.
- Considering the seasonal nature of the data, replacing missing values with just the mean or median value is not suitable.
- We need to implement a sophisticated imputation method based on machine learning
- Here, we have applied the Light Gradient Boosted Machine algorithm to impute missing values in the data.
- The imputation process almost recreates similar data by considering the trend and seasonality. However, some smoothing of the predicted data will be required before fitting any machine learning model for data forecasting.

CASE STUDY: Water Quality Trend Analysis

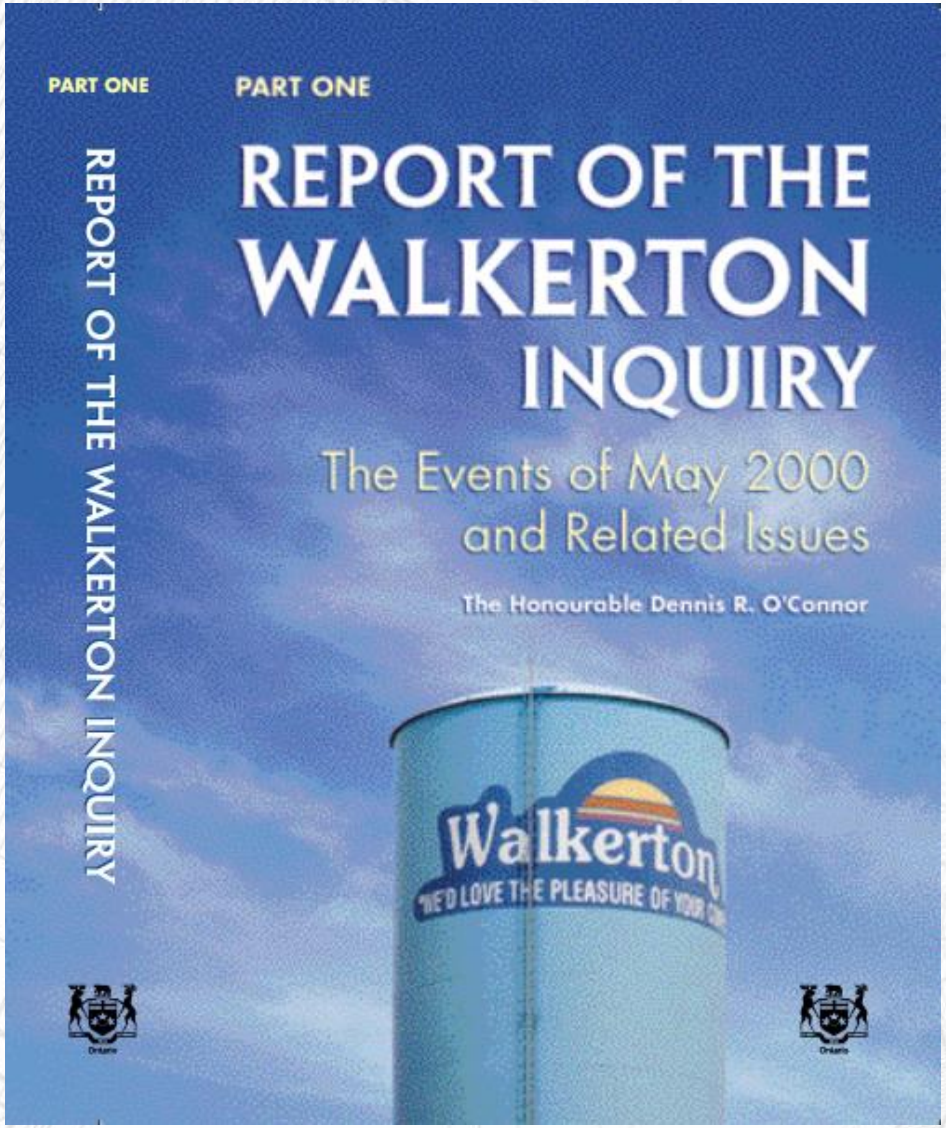




- Did not return a good predictive tool at this location.
- Every attempt to develop a model is not successful the first time.
- It takes many iterations

CASE STUDY

USING ARTIFICIAL INTELLIGENCE OBJECT DETECTION IN HYDROGEOLOGY



SOURCE WATER PROTECTION IN ONTARIO

- May 2000 an E.coli outbreak in Walkerton, a small rural town in Ontario, led to the death of seven people and over 2,000 cases of e.coli poisoning
- Following an inquiry and report on the findings (Walkerton Report), Ontario passed the Clean Water Act to ensure local areas protect their drinking water sources
- 19 multi-stakeholder source protection committees across the Province representing business, public, municipal and Indigenous interests.
  - 38 local source protection plans have been developed that identify actions to protect sources of municipal drinking water systems
  - Plans cover 450 municipal water systems in an area covering where 95% of the population live
- Source Water Protection Plan development required each Source Protection Authority (SPA) to complete a Drinking Water Threat Assessment to identify quality and quantity risks.

PROBLEM STATEMENT

- Identification of threats on a large scale requires a significant amount of effort
- Threats need to be constantly updated
- Training a model to identify potential threats through object detection could save considerable time and effort.

Gas Tank

SWM Facility

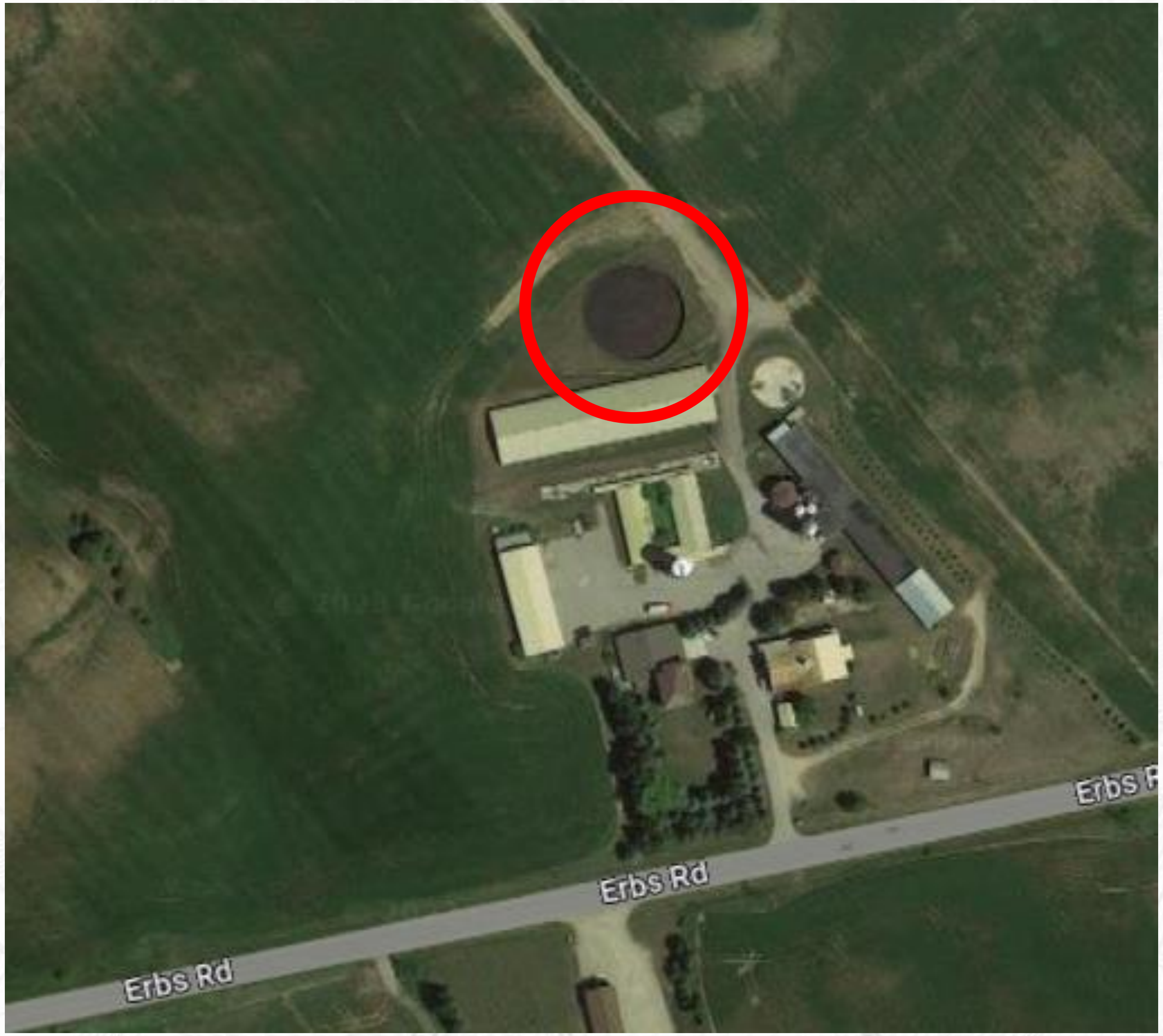
Salt Storage

DETECTING AGRICULTURAL SOURCE MATERIAL (ASM) STORAGE FACILITIES

- ASM – treated/untreated material capable of being applied to land as a nutrient
- Farm manure, runoff from feed lots and manure storage, milking wash water
- Not compost that meets Compost Guidelines, or commercial fertilizer
- Manure typically stored in large raised concrete facilities
- The Model will be trained to identify potential facilities from air photos and Lidar technology

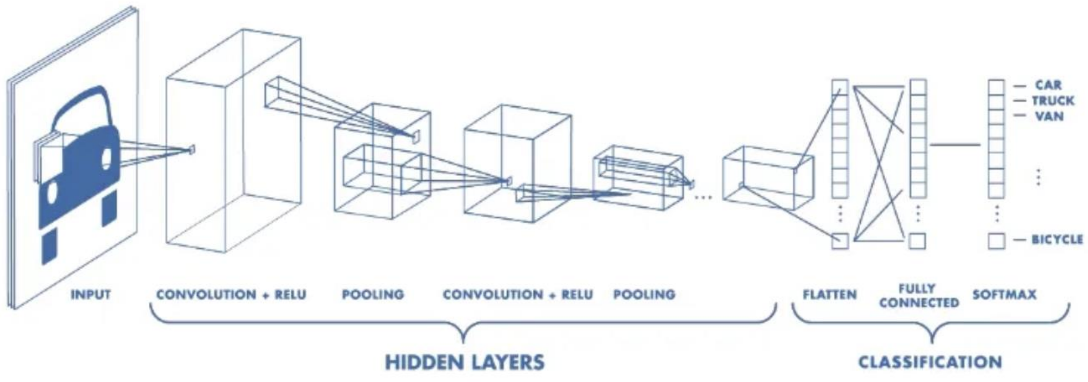
OBJECT DETECTION MODEL

- Image (or object) detection is a task in computer vision that involves identifying the presence, location and type of one or more objects in a given image.
- It is a challenging problem that involves building upon methods for object recognition (e.g. where are they), object localization (e.g. what are their extent), and object classification (e.g. what are they).
- A series of images of known storage facilities are used to train the model (training set) and the model is validated with a separate set of images (validation set).
- The selected images are on different scales and in different settings to ensure a variety of circumstances with which to train the model.



CONVOLUTIONAL NEURAL NETWORK (CNN)

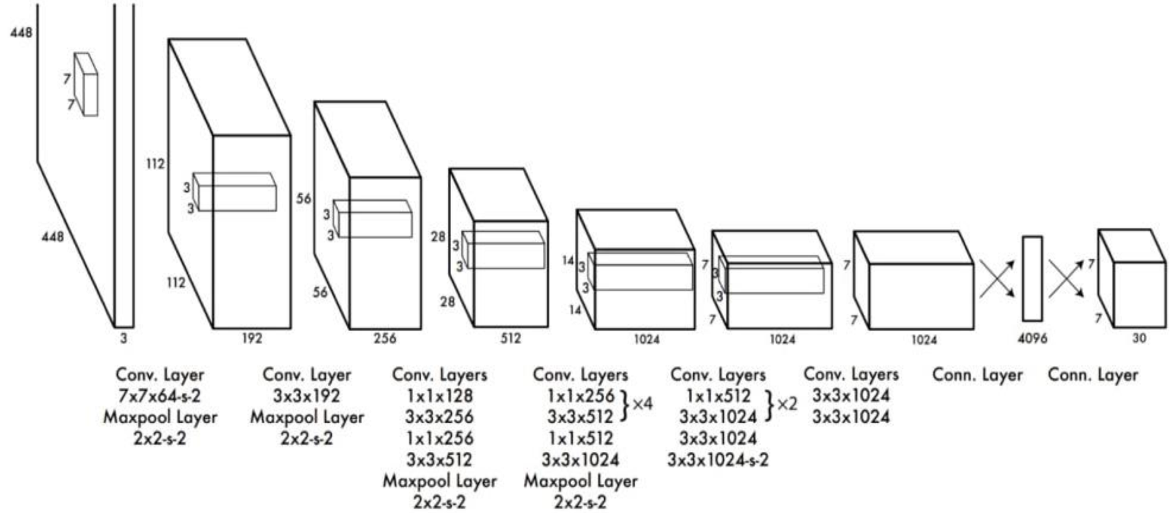
- CNN's are known for their superior performance with image, speech and audio signal detection.
- 3 layers; convolutional layer, pooling layer and fully connected (FC) layer.
- Convolutional layer is the main component and is comprised of input data, a filter and a feature map.



- Pooling layer reduces the number of parameters in the input via either mac pooling or average pooling which reduces the complexity and improves efficiency.
- The FC layer completes the task of classification based on features extracted through the previous layers

YOU ONLY LOOK ONCE (YOLOv8)

- YOLOv8 is an object detection model that performs object localization and classification in a single pass through the network – faster than traditional 2-step approach.
- Divides the input image into a grid and each grid cell is responsible for predicting objects within in – predicts bounding boxes, class probabilities and confidence scores for each cell.



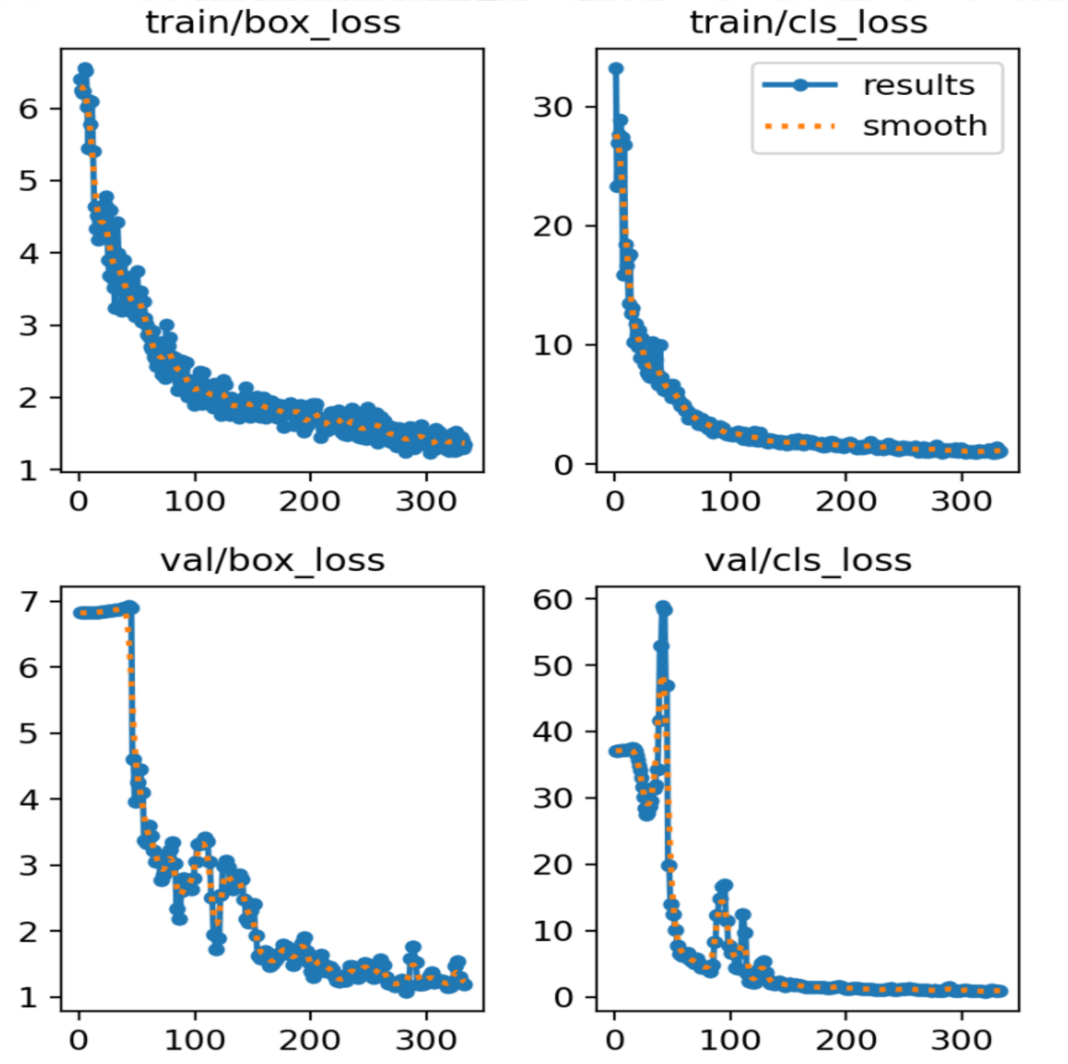
- Operates on multiple scales or resolutions of the image, allowing it to detect object of different sizes. Uses feature maps from different layers to achieve multi scale detection and assigns confidence scores to predictions.
- After predictions it uses Non Maximum Suppression to remove redundant bounding boxes and keep only the most confident ones – helps eliminate duplicate detections
- Optimized for real time applications due to its ability to process images quickly while maintaining good accuracy.

MODEL DEVELOPMENT

- First step in the development is to use Data Annotation (with CVAT Tools)
- Images are converted with annotation into YOLOv8 training data
- Manual process that relies on the model developer to input/annotate the known features

TRAINING

- Once the model has been developed and trained with the training data the model is validated
- The validation data is a separate data set used to validate the results. Data has not previously been "seen" by the model.
- The model runs through enough epochs until the losses are sufficiently low as shown in the graphs



PREDICTION

- The model is then used to locate the feature along with a confidence score (from 0 to 1)
- In this case, the model correctly identified twelve (12) manure storage tanks with a confidence from 0.3 – 0.8
- The model also failed to locate five (5) manure storage tank.
- These results are from a single model run.
    - It is expected that with further refinement the number of successful features located as well as the confidence score would both increase.


manure_tank 0.5
Google

CONCLUSIONS

Globally there is an unfathomable amount of data that can be used to inform decision makers

System operators should leverage existing data at their disposal to their benefit

There is a lot to be gained by better understanding data trends

NEXT STEPS

Actively working with York Region on their dewatering well PW8.

Apply Machine Learning to **operating** municipal drinking water supply well.

Apply Machine Learning techniques across multiple sectors

NEXT STEPS

Continue
to look for long
term SCADA data
from water/WW
system operators

Seek out additional
partnerships
to apply
Machine Learning
opportunities
in different areas.

THANK YOU

cgerrits@cfcrozier.ca