Data Science for Gravity Main Cleaning Optimization

BACWA Collection System Committee Meeting

November 18, 2021

V&A

Agenda

1. Speaker Introduction
2. What is Data Science?
3. Data Science for Collection Systems
4. Data Science for optimization of system-wide gravity main cleaning
5. 3 phases of a Data Science approach
   - Cleaning Data Observations opportunity
6. Case Studies
   - High SSO rate – clean more efficiently to manage SSO's
   - Low SSO rate – potentially over cleaning system wide, optimize resources
7. Q&A

V&A

V&A Speaker Introduction

Lars Stenstedt

- BS Mechanical Engineering Stanford, MBA from UC Berkeley
- Previously with Fracta (www.fracta.ai) – pioneered use of Data Science and Machine Learning for potable water main LOF (EBMUD, SFPUC)
- High tech, computers/software background
- Joined V&A in 2019 to apply data science/machine learning to maintenance of collection system gravity mains
  - Since 1979, V&A has been providing field-based engineering solutions for water, wastewater and transit infrastructure
  - Flow Monitoring, Odor Control, Condition Assessment, Corrosion Services, Coating Systems

Chelsea Teall – Oakland Client Services Manager

Leo Ribeiro – Data Science Field Engineering

V&A

What is Data Science – and Machine Learning?

- Data Science = *process* of using data for Advanced Analytics
- Advanced Analytics ➔ Machine Learning/AI
- Machine Learning – enabled by Data Science process
  - Is the only way to mathematically optimize a large set of variables
  - "….use of data and algorithms to imitate the way that humans learn, gradually improving its accuracy"
  - Needs large sets (thousands) of labeled data points

Data Science for Collection Systems

- (You should…) *add* Data Science driven guidance into your decision-making
  - over the next few years…..
- One step at a time – with positive ROI for each step

➔For Collection Systems, V&A believes that optimizing the system-wide gravity main cleaning is the logical initial value proposition for Data Science and Machine Learning

Gravity main maintenance today

- CIP: repair/replace of gravity mains won't solve SSO problems for years
- Hot Spot/High Frequency Cleaning: maintain current approach
  - Don't need Data Science to tell you that locations of previous SSO's or known structural issues require high frequency cleaning
  - Best practice top-of-line CMMS enables rules based, reactive interval adjustment - "3 cleans in a row ➔ move to next lower frequency interval"
  - Data Science/Machine Learning can add value over time, but this is the smaller data set with higher risk, so not the logical place to start

Why System-wide Gravity Main Cleaning?

- Typically 90-98% of gravity main segments, all on single cleaning interval (once every 1-5 years)
  - Large data set: Machine Learning applicable for ~50+ mile system (1000+ segments)
  - Many variables to optimize

➔*With enough historical Condition data (how clean/dirty the segment was at a point in time), very good machine learning models can be developed*

System-wide Cleaning Challenges

- High SSO Collection Systems
  - Existing cohort segregation of segments not detailed enough
    - Some segments may be effectively self-cleaning
    - No residential FOG insights, Root progression not uniform
  - Pressing need to do a better job cleaning with available resources

- Low/no SSO Collection Systems
  - Potentially over cleaning – wastes water, degrades pipe, dependent on equipment availability and crew institutional knowledge
  - How to intelligently back off aggressive cleaning with minimal risk to hard-won SSO rate?

Adding Data Science Process to Decision Making

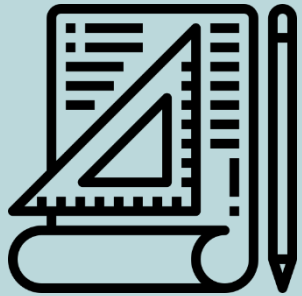
DATA

- GIS Data
- CMMS Data
- CCTV Data
- Cleaning Data
- Smartcover/sensor data
- SSO Data
- External Data

ANALYTICS

- Statistics
- Regression
- Machine Learning
- “Cohort of 1” analysis
- Leverage algorithms from larger data sets

IMPLEMENTATION

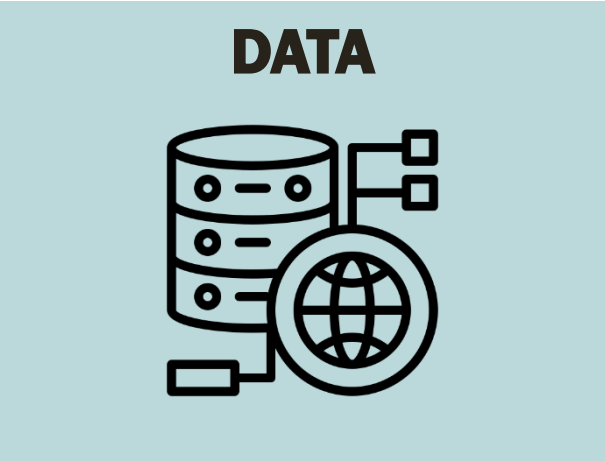


- Maintenance Scheduling
- CCTV Scheduling
- CIP Planning
- Data Collection Methods

$$ invested

V&A

V&A Data Science Services : DATA



Basics

- Digitizing and maintaining GIS data
- Finding missing / outlier values

Advanced

- Programmatic improvement of data – filling empty values with educated assumptions
- Joining Utility Data – CMMS, CCTV, lateral data
- Geospatially assigning and joining SSO data
- Joining External Data – trees, population, buildings, other plausible theories-of-causation

Use and build on all available data – V&A has no software or hardware/sensor agenda

Collection System DATA sets for Advanced ANALYTICS

1. GIS Asset Data –location, size, material, installation date, slope
2. CMMS Work Order Data – cleaning, rehab/repair
3. Historical Condition Data (lowest➔highest quantity)
   - SSO Data – lat/long location
   - Smartcover/sensor data – often used as alarms, limited locations
   - CCTV Data – mostly CIP focused (typ. clean before CCTV) and limited coverage
   - Cleaning Data – consistent data collected during cleaning process (clear/light/medium/heavy)
4. External Data to improve/fix GIS asset data (parcel data for install years etc) and to add variables for plausible-theories-of-causation

Why is Cleaning Data so important?

- CCTV coverage is limited, and much less than # of annually cleaned segments
- CCTV is CIP focused, typically segment cleaned prior to CCTV
- Observations collected during cleaning become thousands and thousands of Condition Data points over time if collected in consistent, scalable manner
- Engages the cleaning crews in the process

➔ Quantify the feedback/observations from Cleaning Crews

VANDA Gravity Main Cleaning Data Index

VANDA® Gravity Main Cleaning Data Index

Condition Rating	FOG (fat, oils, grease)	Representative FOG Photograph	ROOTS	Representative ROOTS Photograph	OTHER-SSO potential (debris, grit, rags)	Representative OTHER-SSO Photograph
Clear	-clear stream during cleaning, no FOG or root smells, no debris on hose					
Light						
Mod						
Heavy						
SSO	-SSO with cause(s) as reported on CIWQS website					

IN PROCESS

V&A Data Science Services : ANALYTICS

Basics

- Visualizing Data (Picture > Spreadsheet)
- Statistics

Advanced

- Cohort group building
- Regression analysis
- Segment level Likelihood calculations with Machine Learning
- Network effect benefits from multi-system algorithms
- Temporal per segment degradation curves

V&A Data Science Services : IMPLEMENTATION

Basics

- Recommendations for maintenance and inspection schedules adjustments
- Benchmarking and Best Practice Recommendations
- Data collection recommendations

Advanced

- Monthly, quarterly or annual schedule reviews to efficiency adjustments
- CIP Planning Recommendations

Adding Data Science Process to Decision Making

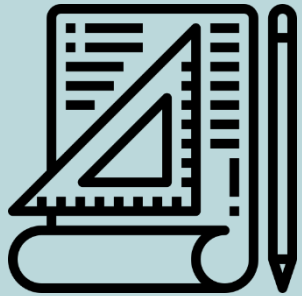
DATA

- GIS Data
- CMMS Data
- CCTV Data
- Cleaning Data
- Smartcover/sensor data
- SSO Data
- External Data

ANALYTICS

- Statistics
- Regression
- Machine Learning
- “Cohort of 1” analysis
- Leverage algorithms from larger data sets

IMPLEMENTATION



- Incremental gains
- Maintenance Scheduling
- CCTV Scheduling
- CIP Planning
- Improving Data Collection Methods

Continuous Improvement

V&A

Two Case Studies

- #1 High SSO rate
  - Just under 100 miles of gravity mains, mostly VCP
  - GIS missing all install years
  - Struggling to keep up with SSMP cleaning commitment
- #2 Low SSO rate
  - Just over 150 miles of gravity mains
  - Cleaning all gravity mains every 2 years, can safely reduce frequency if needs better understood

#1: High SSO rate, just under 100 miles

DATA Phase Summary

- Over 2000 gravity main segments
- GIS had no installation years – remedied by adding decade level assumption based on parcel data
- Numerous NULL values for pipe diameters – remedied by inheriting from ‘nearest neighbor’
- Digitally Joined all available data to create analytics-ready deep-rich data set

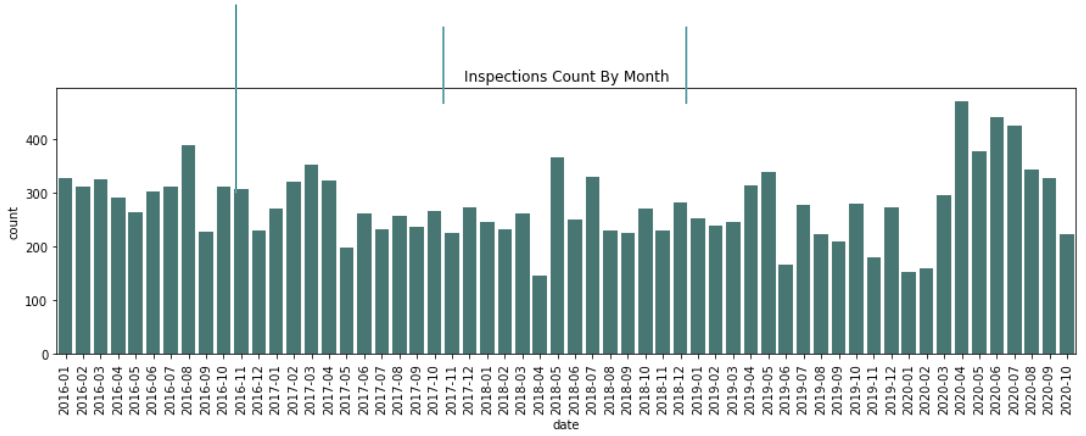
ANALYTICS & IMPLEMENTATION

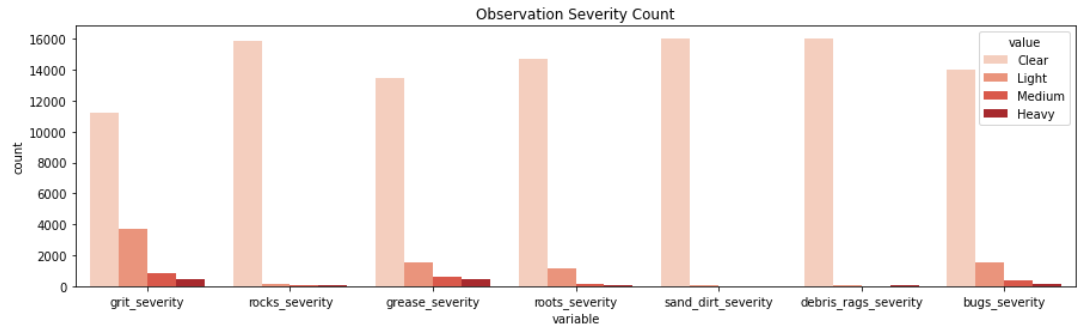
- Based on combined data sets, V&A generated segment predictions for “likelihood” of needing cleaning
- Customer has 3 year plan to increment to optimized frequency schedules with available resources
- Short-term reprioritizing order of basin cleaning
- Short-term considering likelihood when deciding on cleaning process
- Long term: use analytics to build optimized annual plans

#2: Low SSO rate, just over 150 miles

DATA Phase Summary

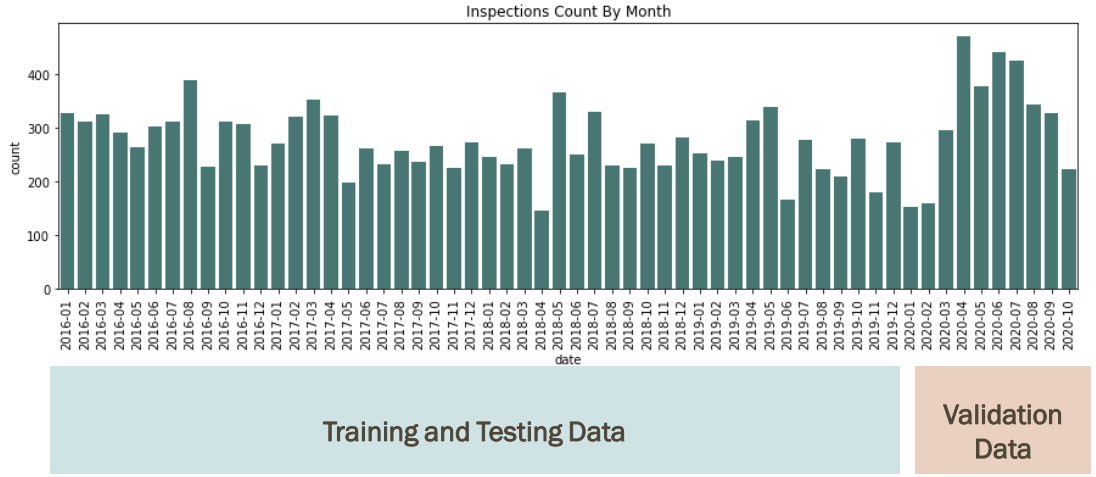
- V&A Analyzed the Cleaning Data results from 2016-2020
- 16127 observations
- 7 categories!
  - Grit, Rocks, Grease, Roots, Sand/Dirt, Rags and Bugs
- Three severities
  - Light, Medium, Heavy
- Coverage : 96.2%





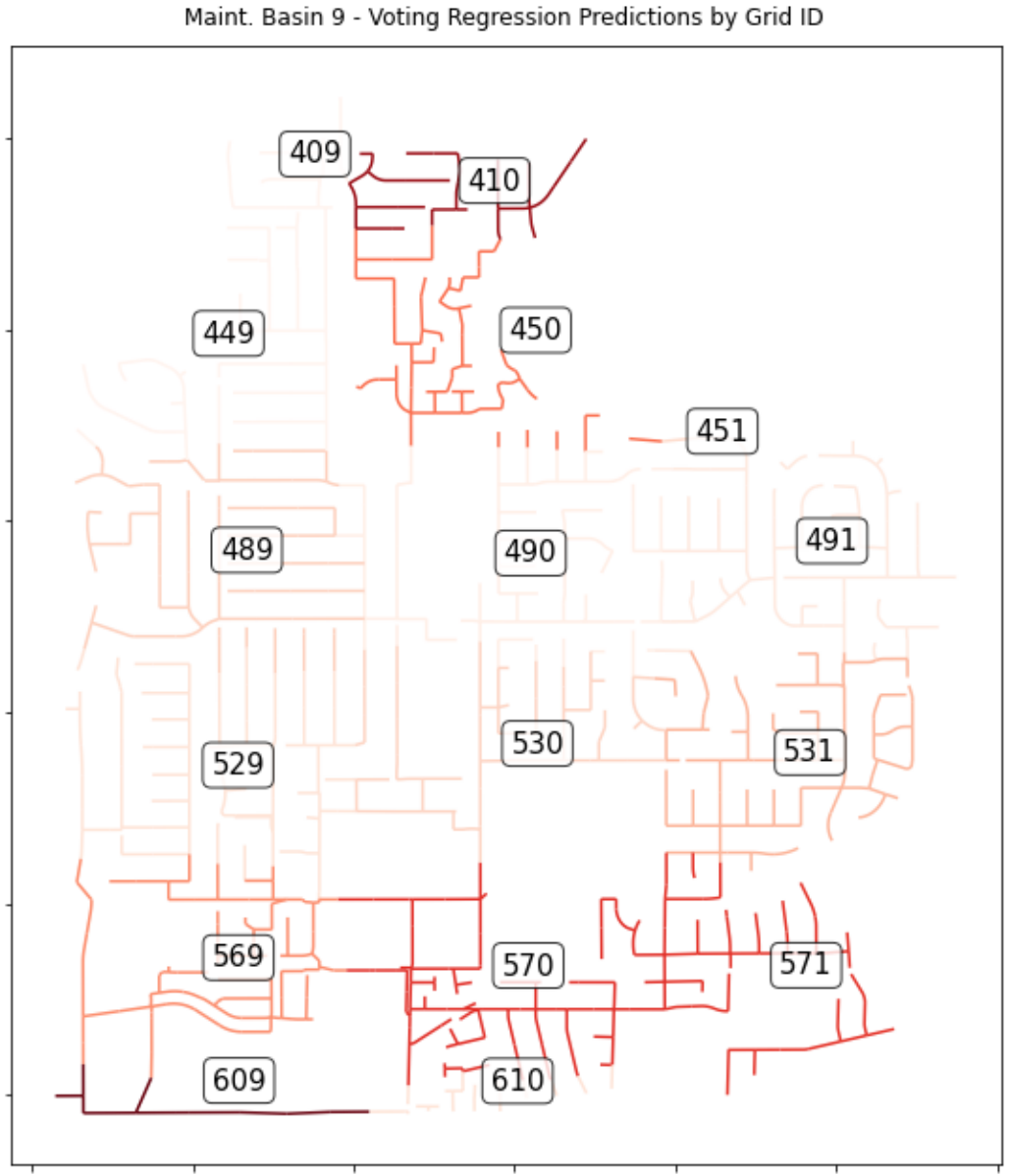
ANALYTICS - Prediction Modeling

- Typically, in Data Science various mathematical models are experimented with to find the best model for the data set
- V&A explored various models – regressions, random forests, gradient boosts etc.
- Cleaning data from 2016-2019 as the training / test data
- 2020 observations as validation data

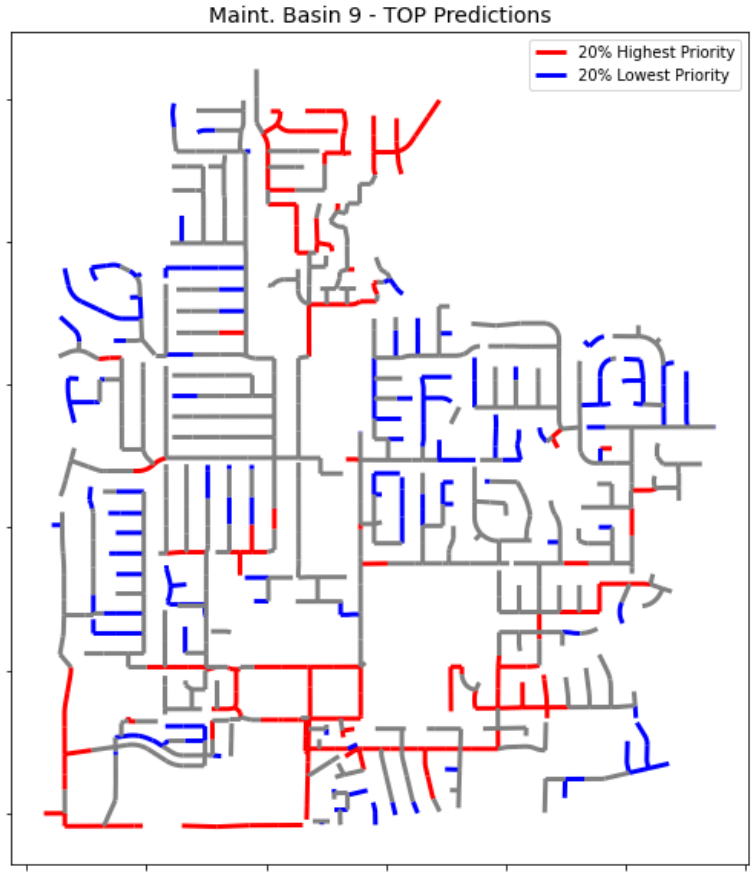


V&A analyzed various modeling algorithms to test predicting the 2020 cleaning observations

IMPLEMENTATION – provide crews a useful tool



grid_id	vr_prediction
609	1.64
410	1.48
570	1.12
571	1.11
450	0.84
569	0.69
531	0.50
489	0.36
530	0.33
451	0.29
610	0.24
529	0.24
491	0.18
490	0.13
409	0.08
449	0.08



Cleaning Crews leveraging in daily cleaning decisions

V&A Data Science Services Engagement Model

-Our goal is to be your long-term data science partner empowering your crews, engineers and planners to get the best return on % spent mitigating SSO's

Fixed Price for initial Data, Analytics, Implementation Service

(Negotiable - up to $20-30k typ.)

Monthly, Quarterly, or Annual: Analytics updates, cleaning schedule reviews and recommendations

(50k, 30k or 10k per year)

Q&A

Thank you!

- Agency specific presentations or lunch/learn on request

V&A Consulting Engineers

Lars Stenstedt: lstenstedt@vaengineering.com

Chelsea Teall: cteall@vaengineering.com

Leo Ribeiro: lribeiro@vaengineering.com

V&A
consulting engineers

Oakland | San Diego | Houston | Florida

Corrosion Engineering
Condition Assessment
Coating Systems Management
Flow Monitoring
Odor Control
Data Science
VANDA® Corrosion Rating System

www.vaengineering.com
510 . 903 . 6600