LIQUID FLOW RATE PREDICTIONS WITH MACHINE LEARNING FOR SCALE DETECTION

IMPERIAL COLLEGE LONDON
IN COLLABORATION WITH WINTERSHALL DEA

MSc APPLIED COMPUTATIONAL SCIENCE AND ENGINEERING
KEVIN FUNG

LIQUID FLOW RATE PREDICTIONS FOR SCALE DETECTION

AGENDA

1. Background and Project Definition
2. mlflowrate Software
3. Data Engineering
4. Exploratory Preparations
5. Results and Findings
6. Questions

LIQUID FLOW RATE PREDICTIONS FOR SCALE DETECTION

PROJECT BACKGROUND

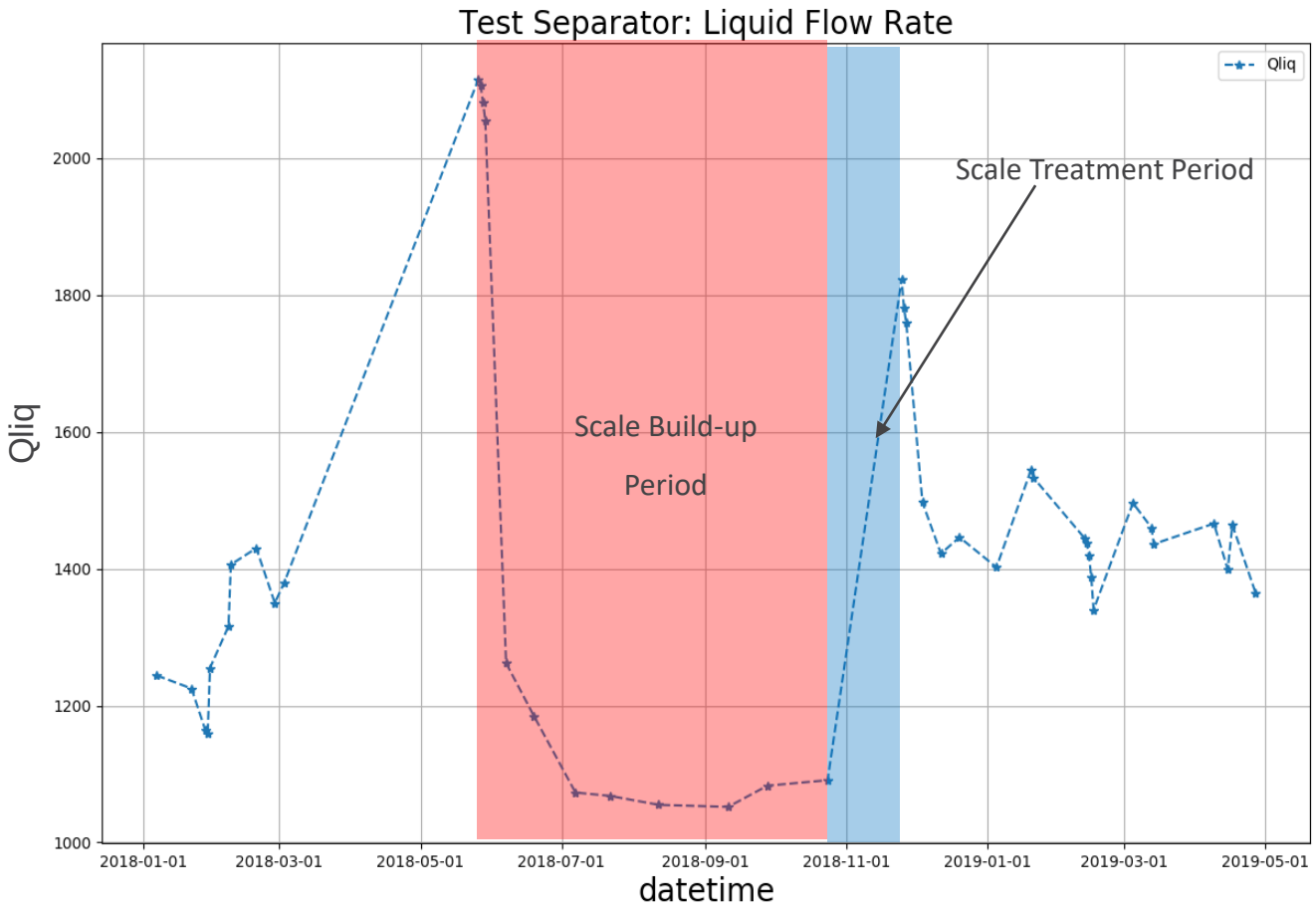
- Offshore oil wells are typically prone to scaling affecting the integrity and production efficiency of wells.
- Scaling is the build up of chemical precipitation due to a chemical reaction along the inner tubing.
- Scale treatments are expensive and operators must optimize this process.

Residual calcium carbonate scaling found within the tubing of a North Sea Oil Well after an acid wash.

LIQUID FLOW RATE PREDICTIONS FOR SCALE DETECTION

PROJECT BACKGROUND

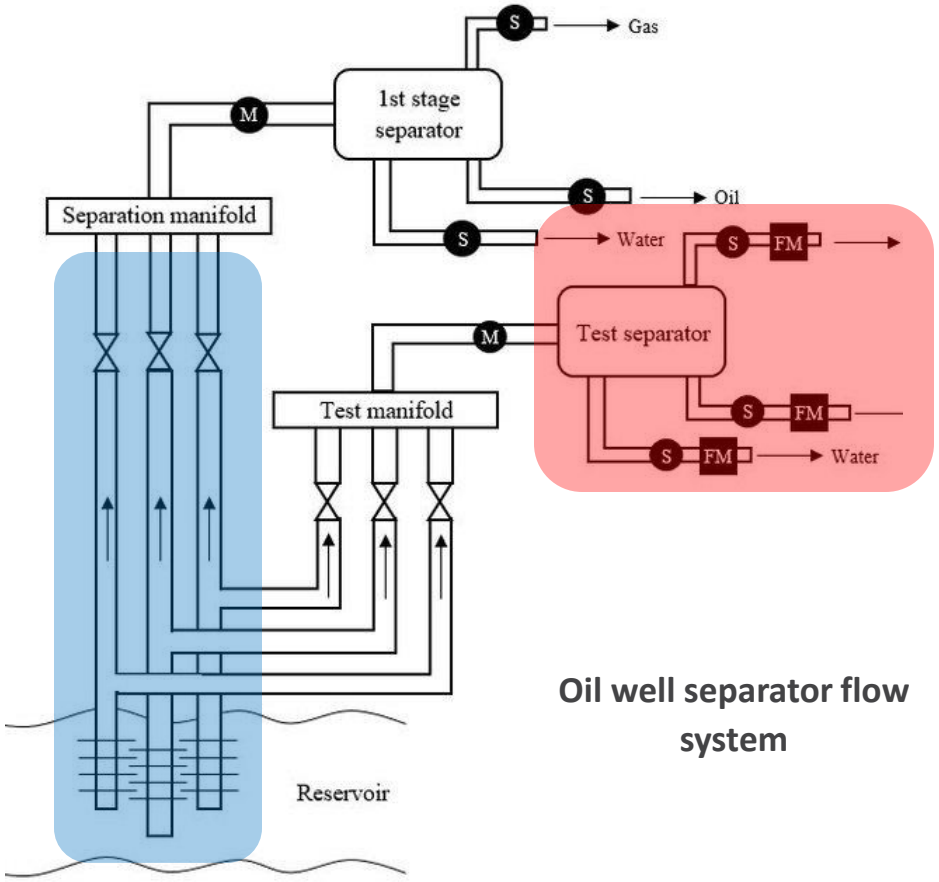
- Data provided for an oil well with significant calcium carbonate scaling.
- Liquid flow rate data best indicates the occurrence of scaling.
- Restricting tube diameter due to scaling decreases liquid flow rate.



LIQUID FLOW RATE PREDICTIONS FOR SCALE DETECTION

PROJECT BACKGROUND

Data Understanding and Overview



Oil well separator flow system

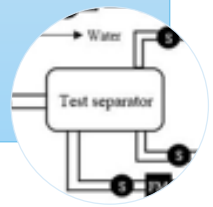
LIQUID FLOW RATE PREDICTIONS FOR SCALE DETECTION

PROJECT BACKGROUND

Data Understanding and Overview

Test Separator Data

- 117 Samples
- (Every few weeks)
- (2016-2019)



Field Data

- 30,000 to 86,000 Samples
- (Every hour)
- (2016-2019)

MIKON Liquid Rate Data (15% uncertainty range)

- 1204 Samples
- (Every day)
- (2016-2019)

$$Qoil = a_0 + a_1 whp$$

LIQUID FLOW RATE PREDICTIONS FOR SCALE DETECTION

PROJECT DEFINITION

Apply **machine learning models** to **predict daily liquid flow rates** and attempt to produce a model at a **similar prediction capacity** to the MIKON model.

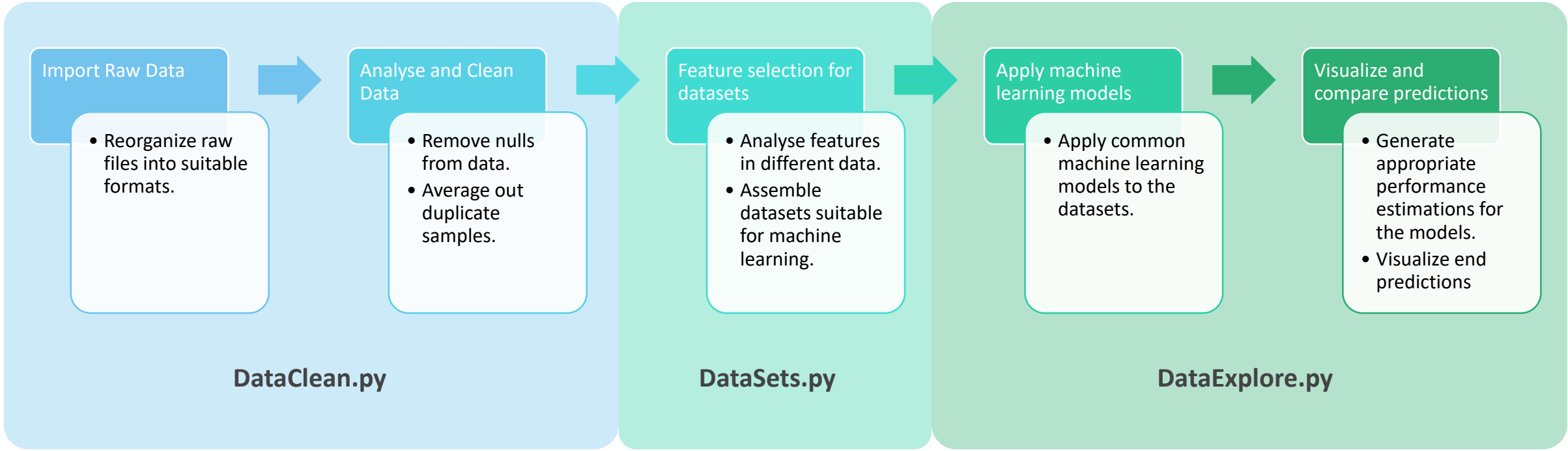
Project Objectives:

1. Produce a data integration and machine learning software for oil well data.
2. Perform feature selection and assemble suitable datasets for predictive exploration.
3. Apply and investigate the effects of different machine learning models to predict daily liquid flow rates.

LIQUID FLOW RATE PREDICTIONS FOR SCALE DETECTION

mlflowrate: SOLUTION SOFTWARE

Software Workflow and Key Features



LIQUID FLOW RATE PREDICTIONS FOR SCALE DETECTION

DATA ENGINEERING

Data Cleaning

- Daily average of all data
- Remove missing values, zero values, and time duplicates.

Test Separator Data	Field Data	MIKON Liquid Rate Data
❖ 112 Samples	❖ 1234 Samples	❖ 1180 Samples

Reduced samples after cleaning

LIQUID FLOW RATE PREDICTIONS FOR SCALE DETECTION

DATA ENGINEERING

Investigative Datasets

- Constructed 2 labelled datasets for modelling.
- First dataset includes all test separator features.
- Second dataset includes all field features excluding WHP.
- Pearson's Correlation Matrix show small correlation between WHP and liquid rate.

Dataset 1	Dataset 2
▪ Well head temperature (WHT)	▪ Well head temperature (WHT)
▪ Well head pressure (WHP)	▪ Down hole temperature (DHT)
▪ Down hole pressure (DHT)	▪ Down hole pressure (DHP)
▪ Gas lift rate (GLR)	▪ Gas lift rate (GLR)
▪ Choke	▪ Choke
	▪ Gas lift pressure (GLP)
	▪ Acoustic Sand Detection (ASD)

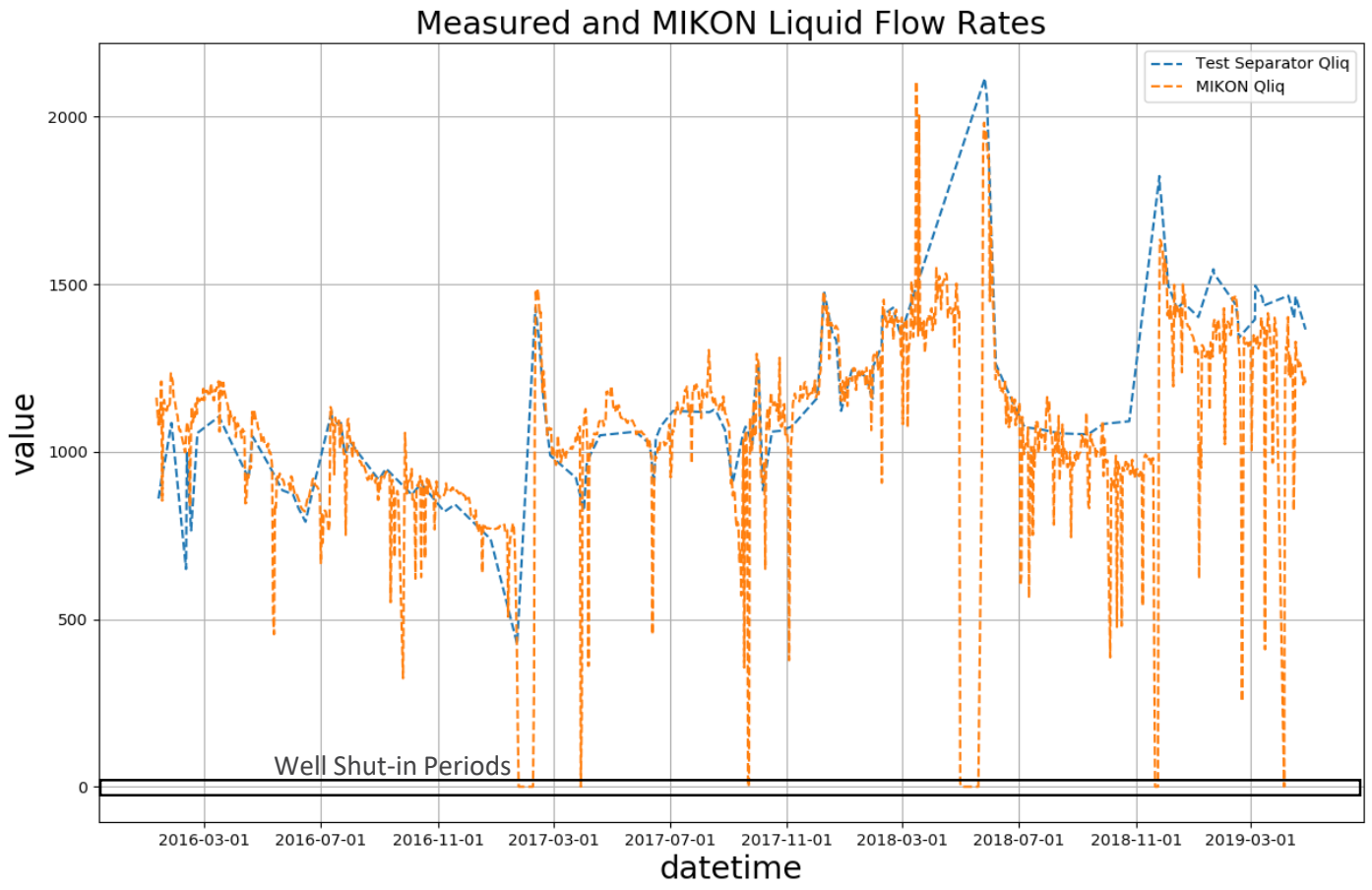
Features of constructed labelled datasets

LIQUID FLOW RATE PREDICTIONS FOR SCALE DETECTION

DATA ENGINEERING

Boosting Sample Sizes of Labelled Datasets

- Well shut-ins: where production flow is stopped.
- Identify well shut-ins in Field data and label those samples as 0 liquid rate.
- Increased dataset sizes from 112 to 144 samples.
- Models are able to learn well shut-ins.



LIQUID FLOW RATE PREDICTIONS FOR SCALE DETECTION

EXPLORATORY PREPARATIONS

Explorative Ideas

1. Naïve Approach: Hyper parameter tuned modelling with no data preprocessing.

2. Factor Analysis Approach: Hyper parameter tuned modelling on latent factors found with Factor Analysis.

Factor Analysis

- Form of dimensionality reduction.

- Assumes there are unknown variables which explain the multi-collinearity between features.

- **Preprocessing Approach:** generate all **polynomial and interaction features** to the second order degree, then **reduce the number of features** down using factor analysis.

LIQUID FLOW RATE PREDICTIONS FOR SCALE DETECTION

EXPLORATORY PREPARATIONS

Machine Learning Models Used

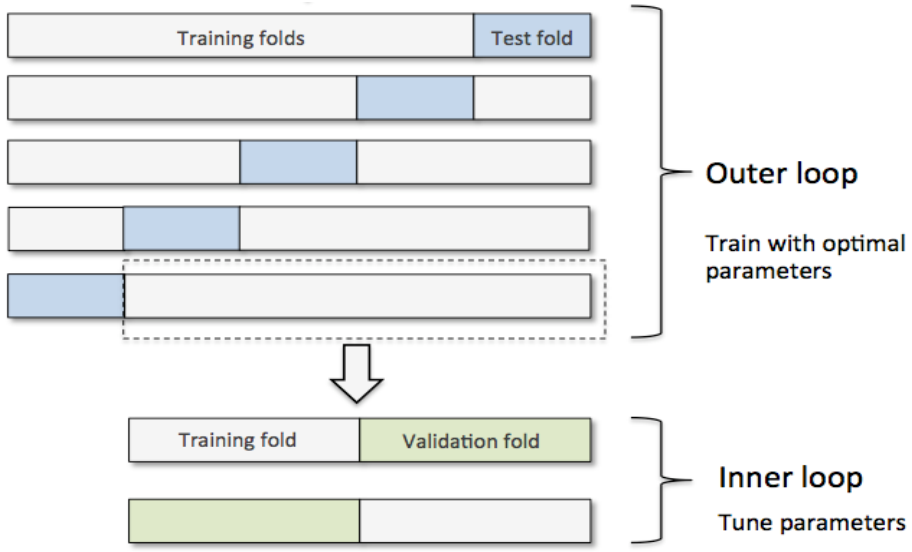
Regression Models	Linear Regression
	Lasso
	Ridge
	Elastic Net
	PLS Regression
Non-Regression Models	Random Forests

LIQUID FLOW RATE PREDICTIONS FOR SCALE DETECTION

EXPLORATORY PREPARATIONS

Model Performance Metrics

- Error metrics: R2 and RMSE.



Nested K-Fold Cross Validation
(Sebastian Raschka, 2019)

Validation: Nested K-Fold Cross Validation

- Uses entire labelled dataset for estimation.
- Includes hyperparameter tuning of models into estimation.
- Removes biasing from hyperparameter tuning.
- 10 outer folds, 5 inner folds used.

Evaluation: MIKON Comparison

- Calculate RMSE away from the MIKON uncertainty boundaries.
- Calculate percentage of predictions that lie within MIKON uncertainty.
- Visualize against the MIKON model.

LIQUID FLOW RATE PREDICTIONS FOR SCALE DETECTION

RESULTS AND FINDINGS

Naïve Approach:

General Model Error Estimations

Model	R2	RMSE
MIKON	**0.969**	**106.521**
Linear Regression	0.780	265.338
ElasticNet	0.776	268.138
Lasso	0.793	258.953
Ridge	0.783	264.884
PLS Regression	0.779	266.237
Average of Regression Models	**0.782**	**264.710**
Random Forests	**0.881**	**174.538**

Dataset 1: Test Separator Features

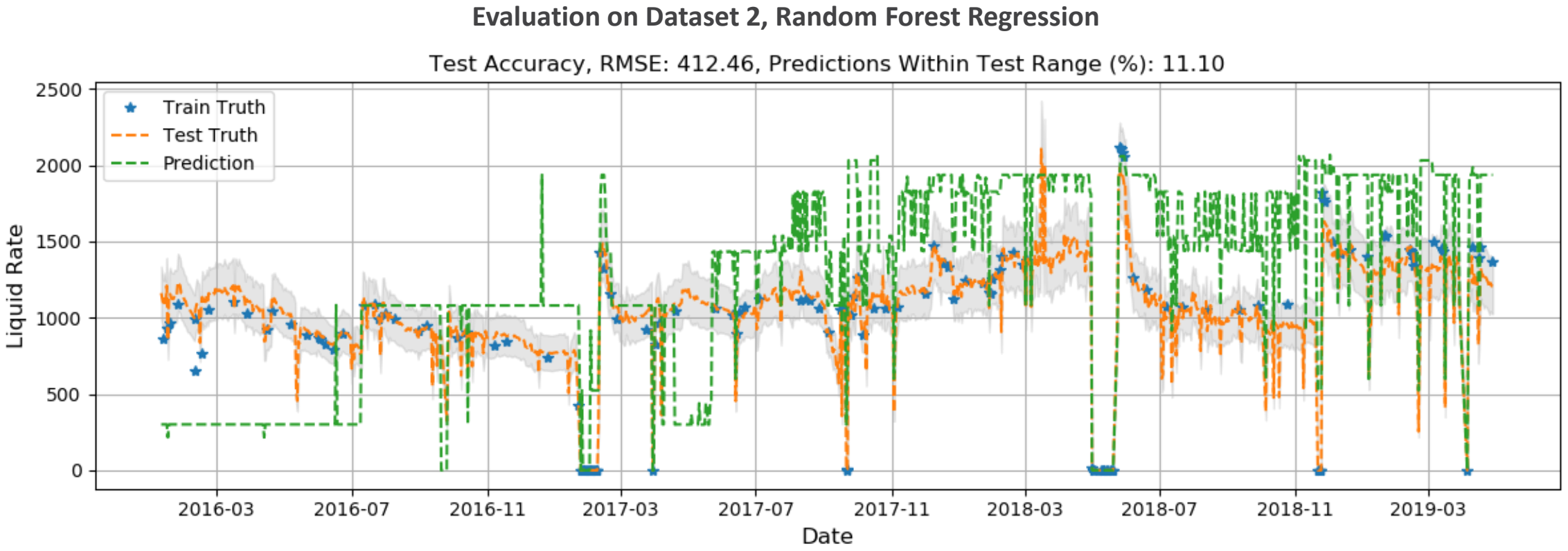
Model	R2	RMSE
MIKON	**0.969**	**106.521**
Linear Regression	0.752	273.191
ElasticNet	0.737	281.154
Lasso	0.744	277.575
Ridge	0.738	281.305
PLS Regression	0.712	286.511
Average of Regression Models	**0.736**	**279.947**
Random Forests	**0.936**	**134.760**

Dataset 2: Field Features without WHP

LIQUID FLOW RATE PREDICTIONS FOR SCALE DETECTION

RESULTS AND FINDINGS

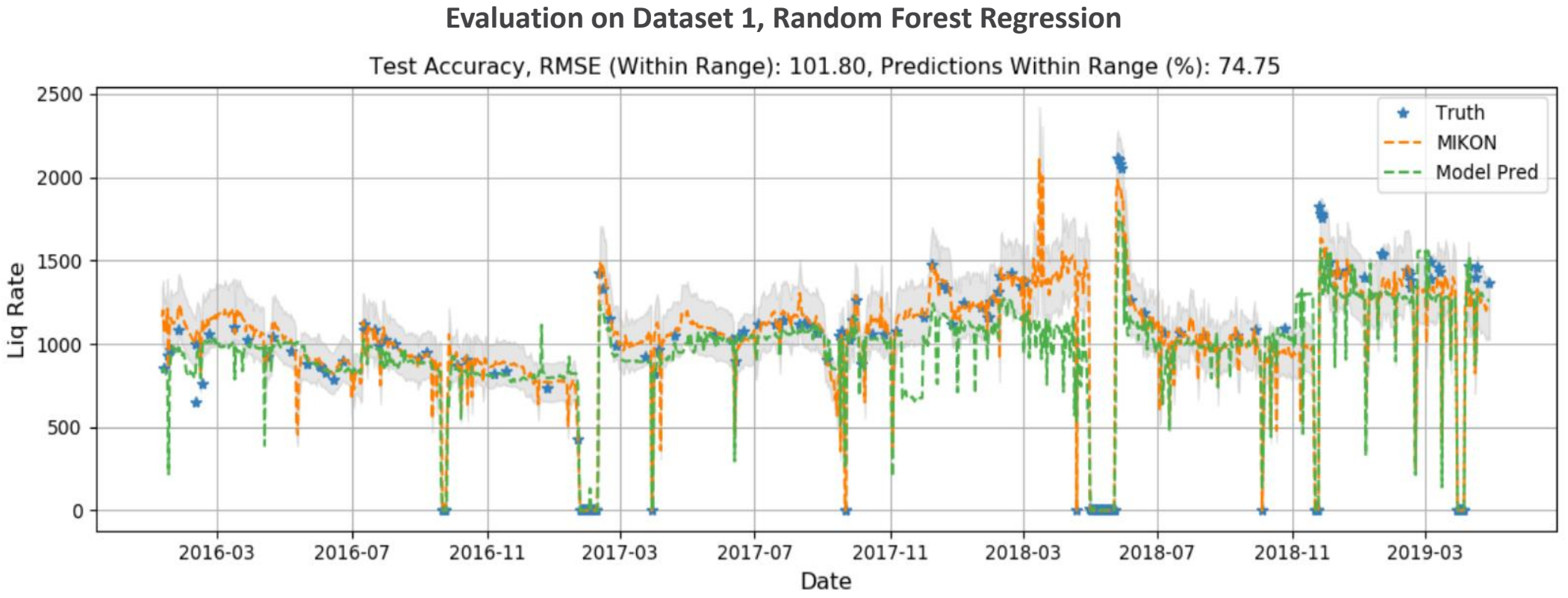
Naïve Approach:



LIQUID FLOW RATE PREDICTIONS FOR SCALE DETECTION

RESULTS AND FINDINGS

Naïve Approach:

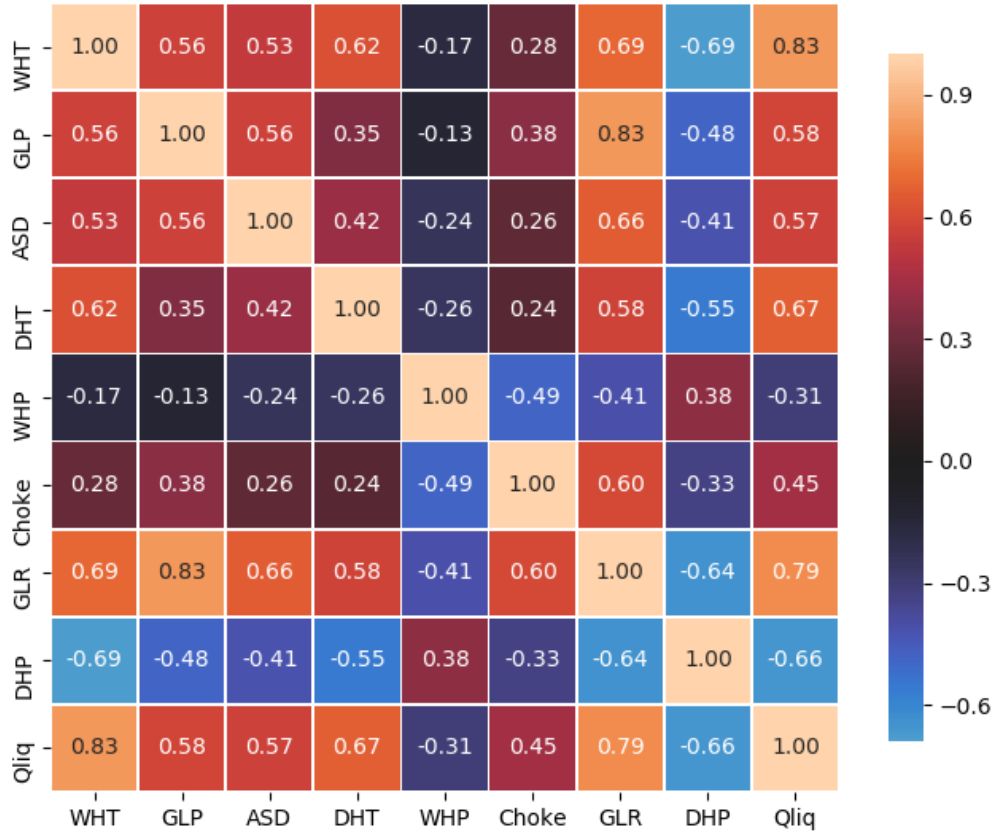


LIQUID FLOW RATE PREDICTIONS FOR SCALE DETECTION

RESULTS AND FINDINGS

Naïve Approach Findings

- Dataset 1 produces better model results: less features means lower model variance.
  - Small dataset with uneven distribution is another reason.
- The differences may be explained by multi-collinearity of features.
- Validation method produces optimistic performance estimations:
  - May be due to the limited data provided.



Pearson's Correlation Matrix of Field Data

LIQUID FLOW RATE PREDICTIONS FOR SCALE DETECTION

RESULTS AND FINDINGS

Factor Analysis Approach:

General Model Error Estimations

Model	R2	RMSE
MIKON	**0.969**	**106.521**
Linear Regression	0.752	273.191
ElasticNet	0.737	281.154
Lasso	0.744	277.575
Ridge	0.738	281.305
PLS Regression	0.712	286.511
Average of Regression Models	**0.736**	**279.947**
Random Forests	**0.938**	**134.696**

Dataset 1

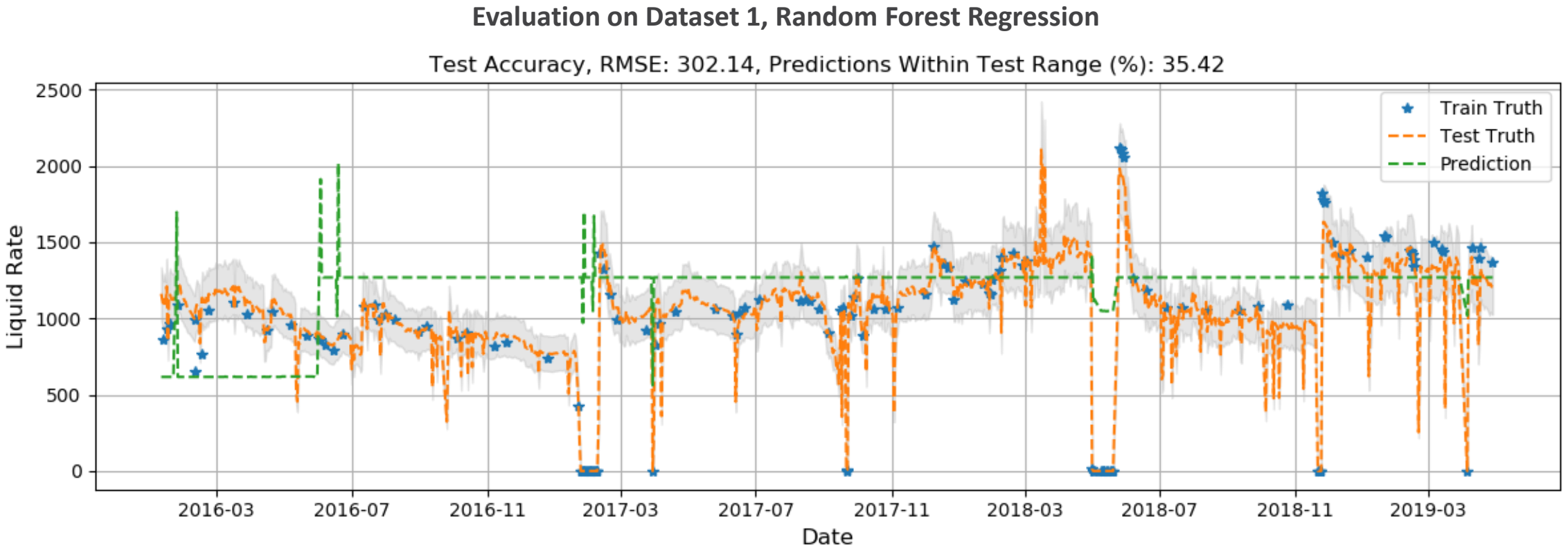
Model	R2	RMSE
MIKON	**0.969**	**106.521**
Linear Regression	0.752	273.191
ElasticNet	0.737	281.154
Lasso	0.744	277.575
Ridge	0.738	281.305
PLS Regression	0.712	286.511
Average of Regression Models	**0.736**	**279.947**
Random Forests	**0.937**	**134.314**

Dataset 2

LIQUID FLOW RATE PREDICTIONS FOR SCALE DETECTION

RESULTS AND FINDINGS

Factor Analysis Approach:



LIQUID FLOW RATE PREDICTIONS FOR SCALE DETECTION

SUMMARY

- Developed a solution software for integrating raw oil well data to predict liquid rates.
- Produced reasonable daily liquid rate predictions with Random Forest Regression with 75% of predictions inside the MIKON model: Scale detection is possible.
- Models are able to recognize well shut-ins.
- More test separator data must be acquired to improve predictions.
  - This may be why our larger featured dataset produced lower accuracies.

LIQUID FLOW RATE PREDICTIONS FOR SCALE DETECTION

EXPLORATORY MENTIONS AND FUTURE WORKS

- ZCA pre-whitening of data to remove multi-collinearity of features: insufficient results.
  - Features had non-gaussian distributions.
- Investigated 2 layer neural network to model liquid flow rate predictions.
  - Limited amount of data which may not represent the unseen field data
- Investigated several 2 layer neural networks to predict highly correlated features to liquid flow rate, then transfer learn gradients and pretrain another neural network to predict liquid rate.
- Potential to utilize autoencoders to train on the large unlabeled dataset, then pretrain a neural network to predict liquid flow rates.
  - May alleviate the small dataset problem.

wintershall dea

QUESTIONS

LIQUID FLOW RATE PREDICTIONS FOR SCALE DETECTION

DATA ENGINEERING

Data Cleaning

We prepare:

- Daily average of all data

We remove:

- Missing values across time samples.
- Zero values due to malfunctioning sensors.
- Daily time duplicates due to repeated measurements made in a day.

Test Separator Data	Field Data	MIKON Liquid Rate Data
❖ 112 Samples	❖ 1234 Samples	❖ 1180 Samples

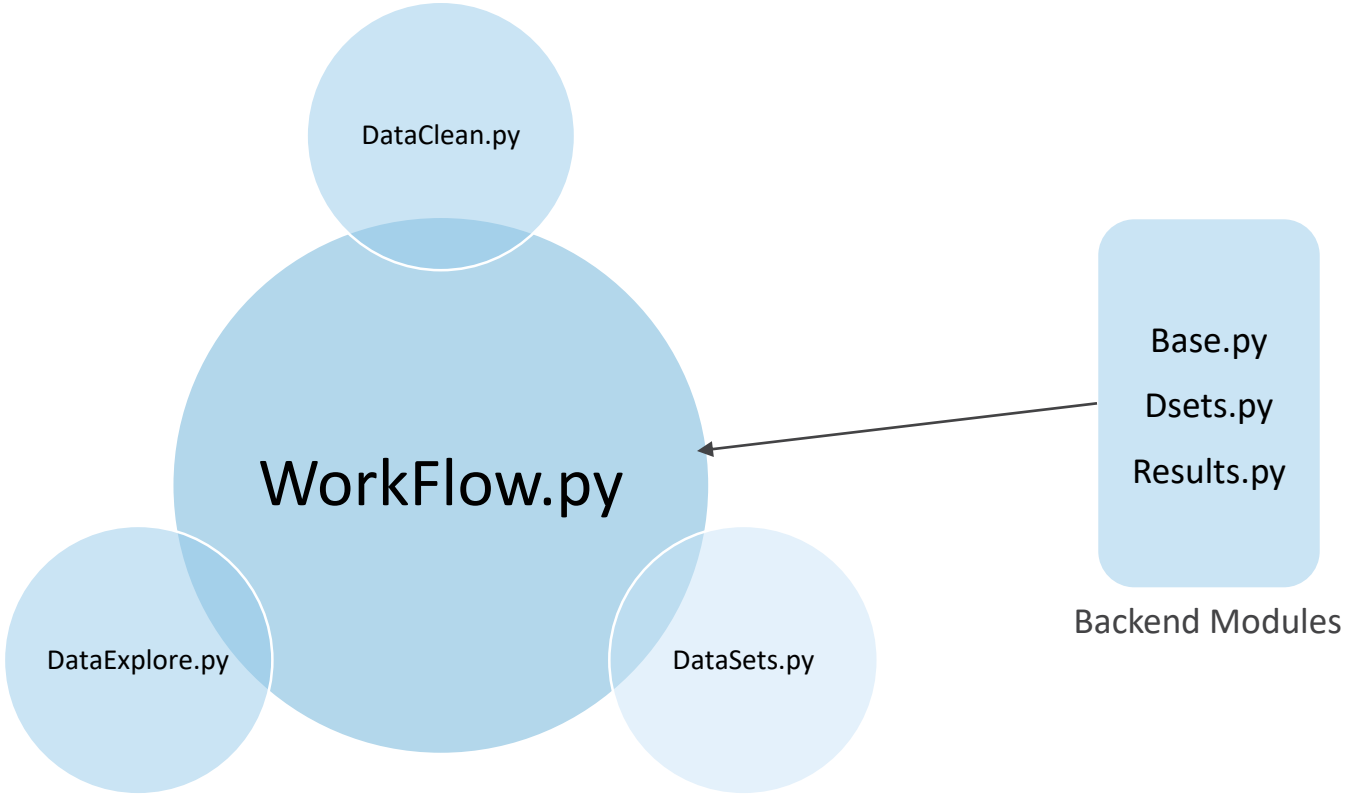
Reduced samples after cleaning

LIQUID FLOW RATE PREDICTIONS FOR SCALE DETECTION

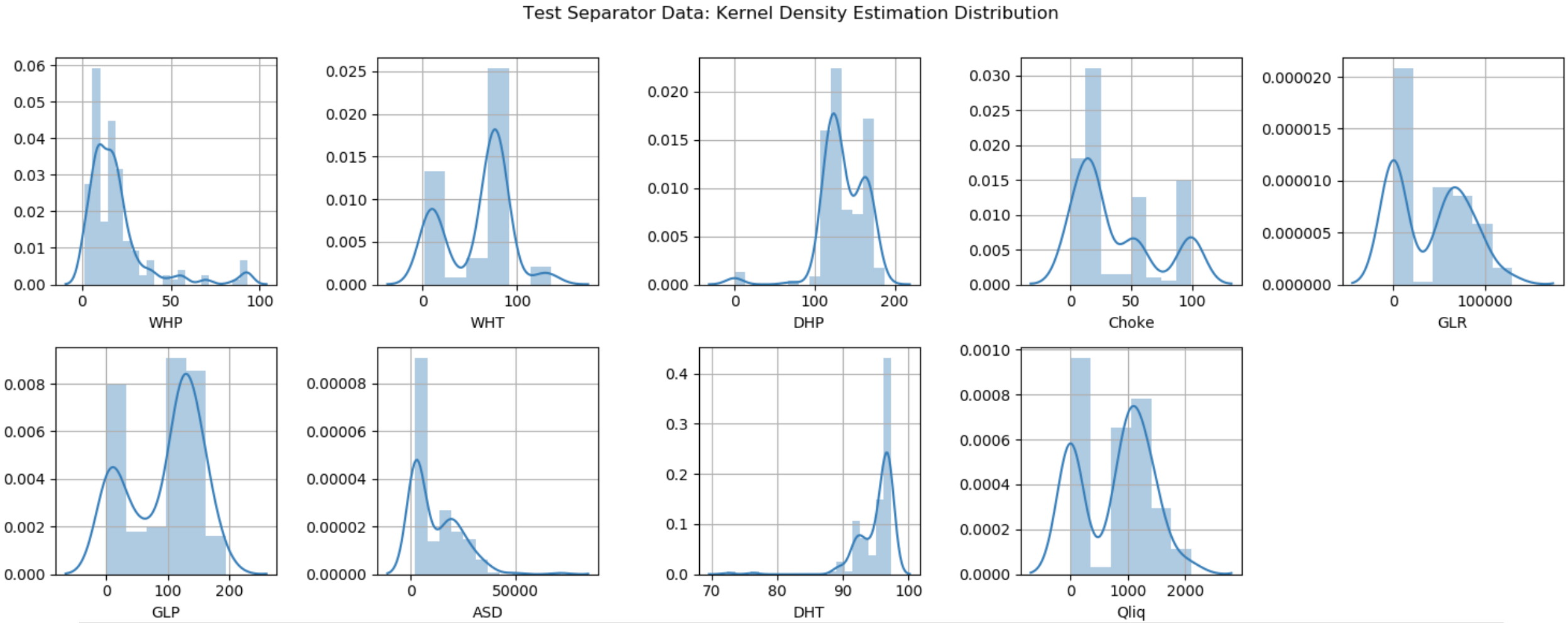
mlflowrate: SOLUTION SOFTWARE

Design Architecture

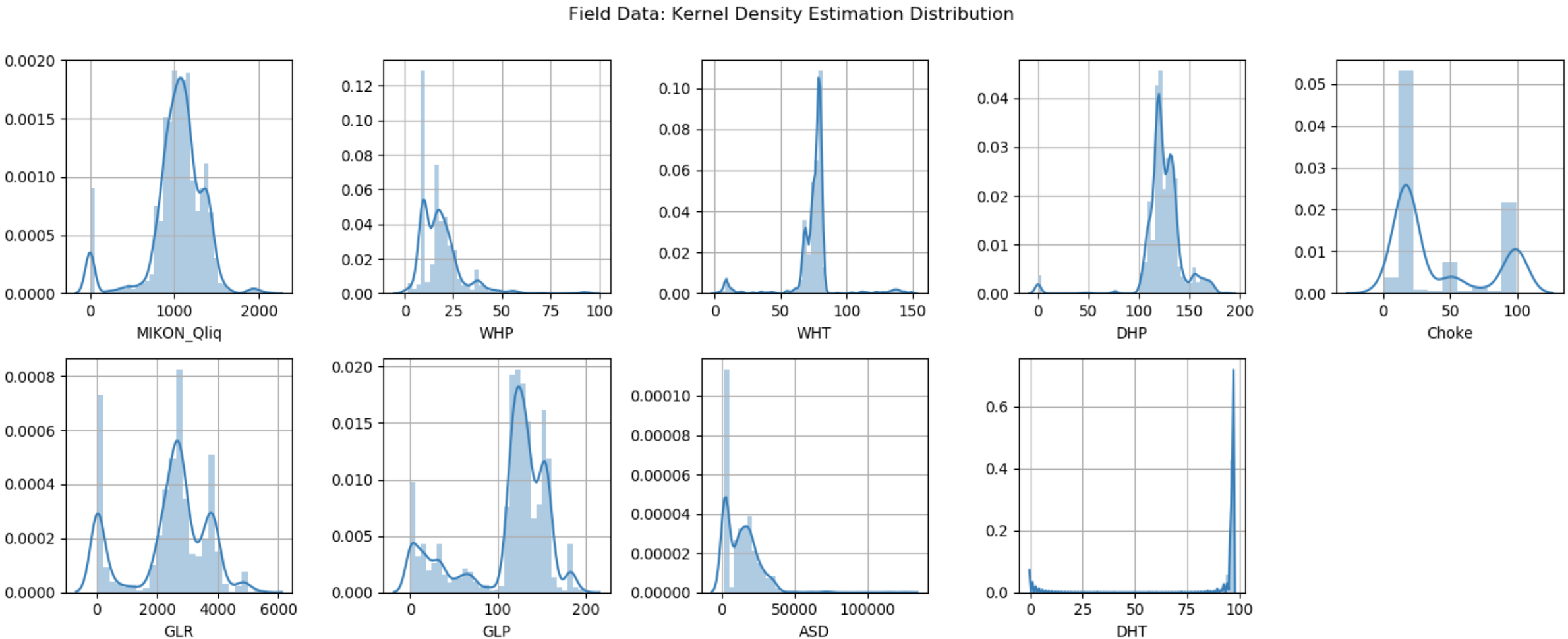
- High level software for new users.
- Based on Spark for big data processing.
- Integration of raw messy data into data science pipeline.
- Supportive machine learning tools for model performance estimations and visualisation.



LIQUID FLOW RATE PREDICTIONS FOR SCALE DETECTION

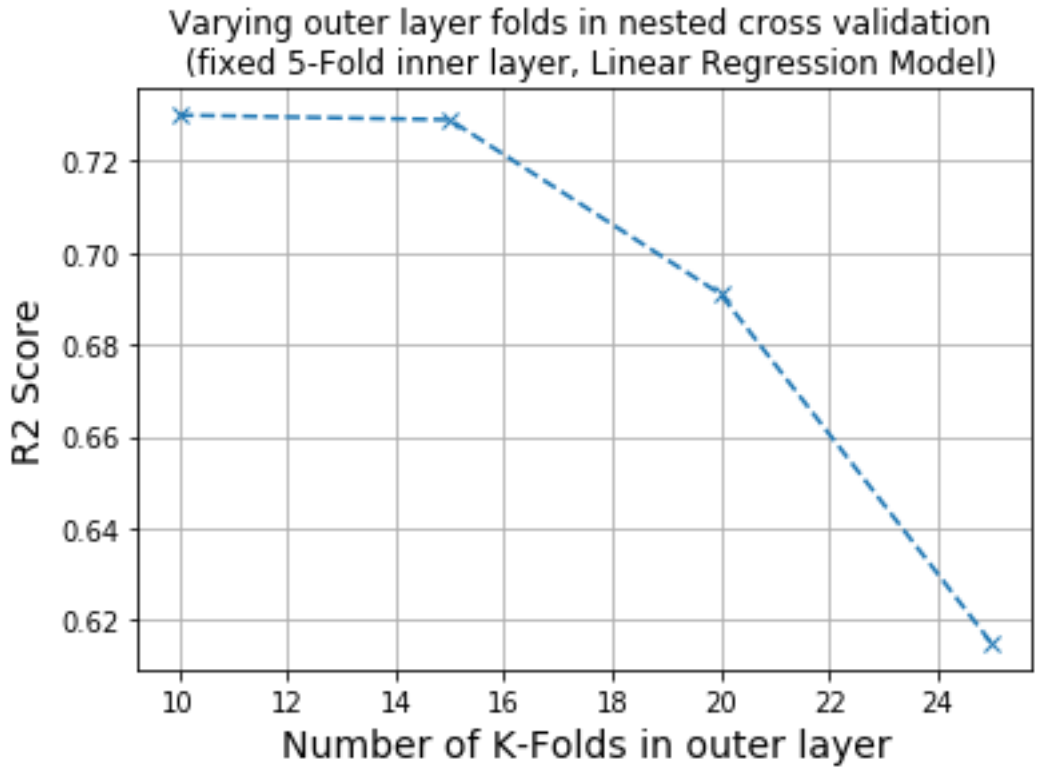


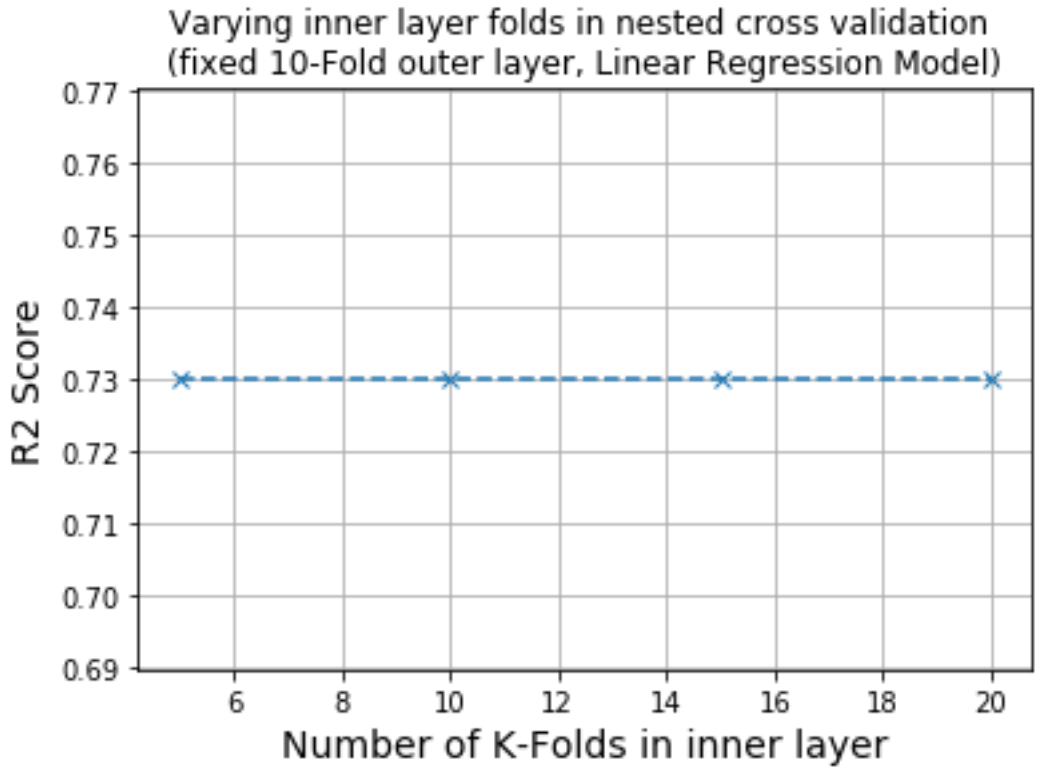
LIQUID FLOW RATE PREDICTIONS FOR SCALE DETECTION



LIQUID FLOW RATE PREDICTIONS FOR SCALE DETECTION

INVESTIGATING VALIDATION TECHNIQUE

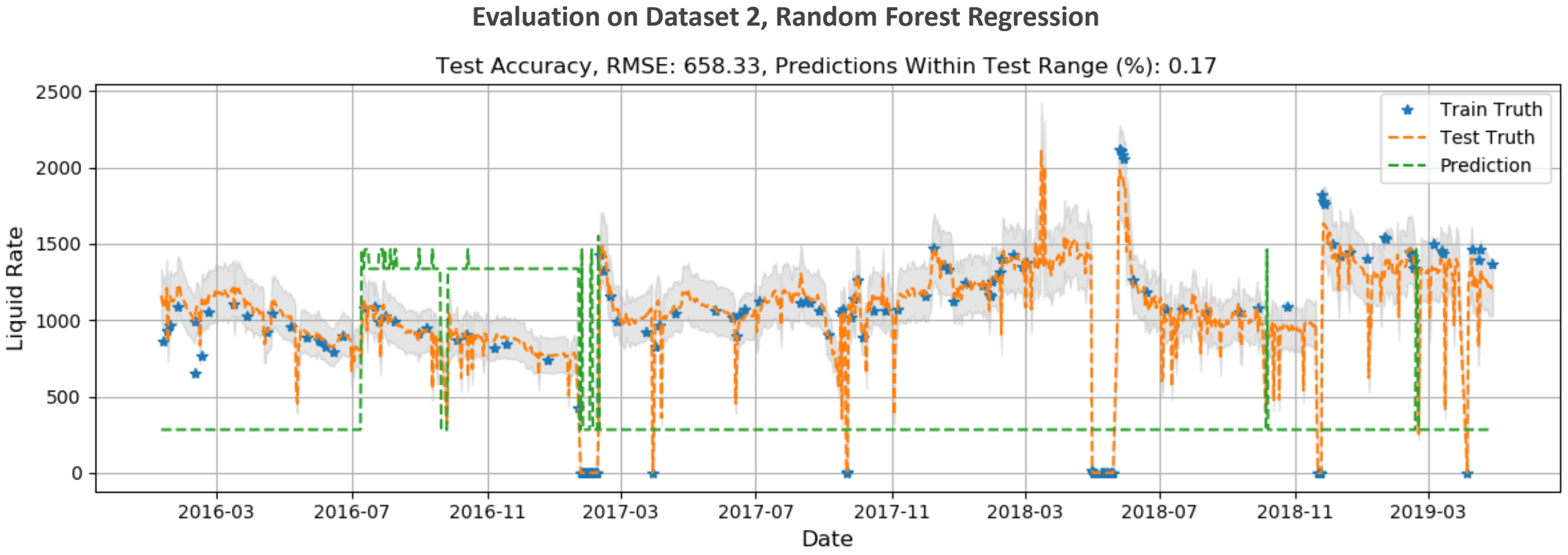




LIQUID FLOW RATE PREDICTIONS FOR SCALE DETECTION

RESULTS AND FINDINGS

Factor Analysis Approach:



RESULTS AND FINDINGS

Factor Analysis Approach Findings

- Results for both datasets do not predict the trends correctly: Scale build-up cannot be detected.
- Dataset 1 evaluations are better than Dataset 2.
- Latent variables are not enough to accurately predict liquid flow rates.
- Feature selection is important to producing better accurate models, may also be because of the limited number of samples.