

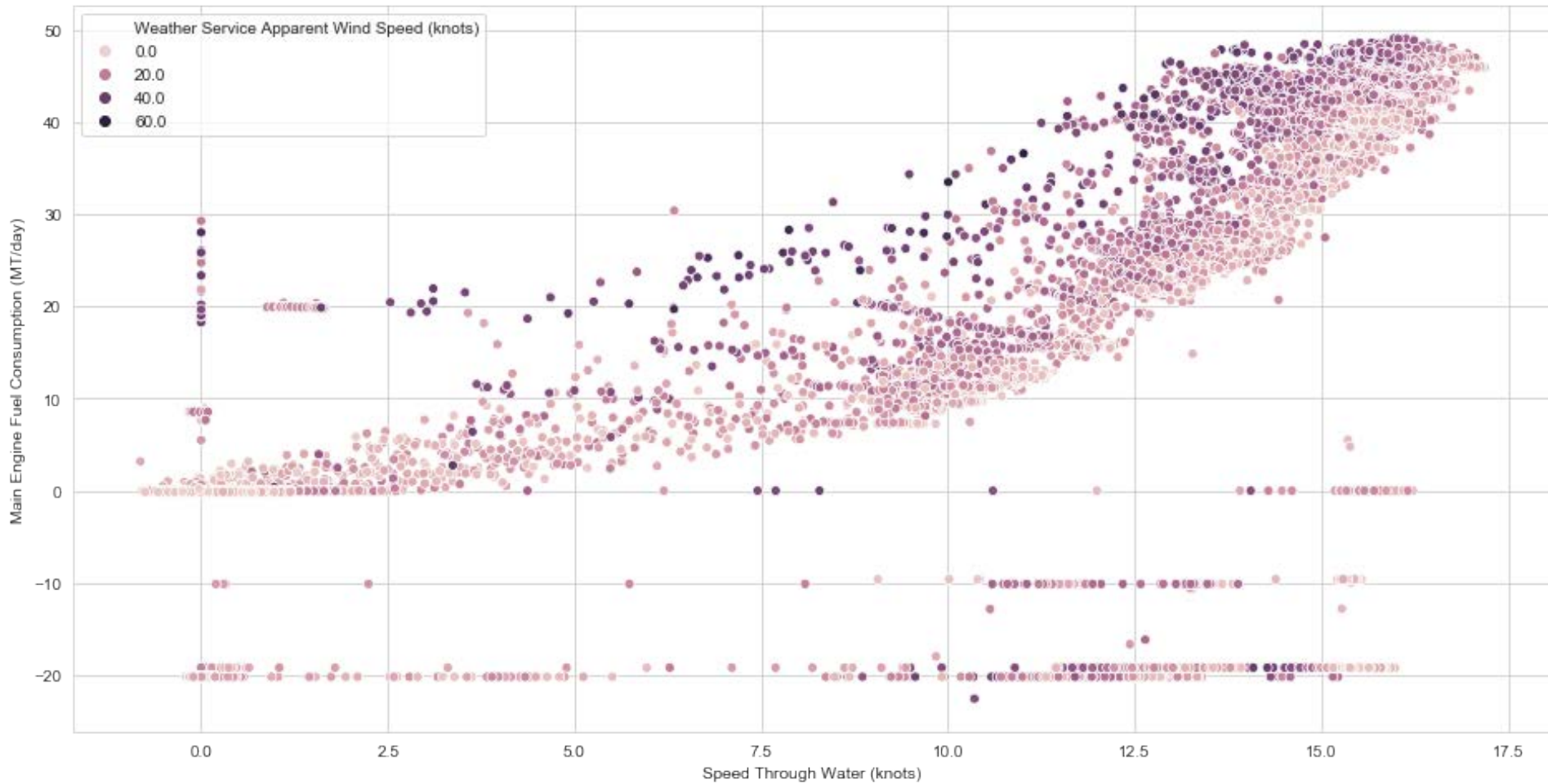
# Ship Performance

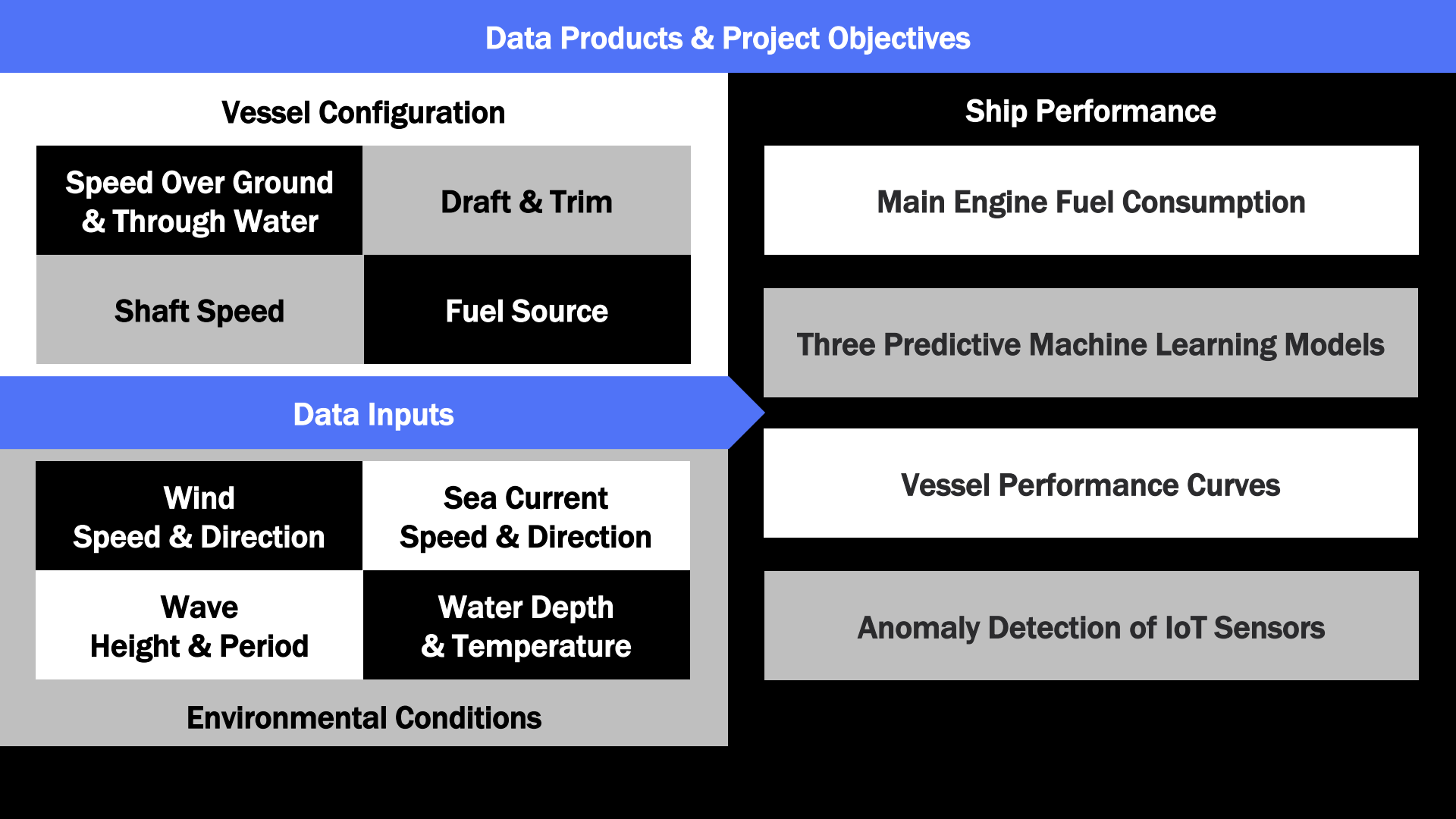


Adam C Dick

[www.linkedin.com/in/adamcdick](https://www.linkedin.com/in/adamcdick)

# Raw Data: How Do Speed and Environment Impact Fuel Consumption?





```
graph TD; A[Data Wrangling  
Record Cleaning] --> C[Iterative Cycle]; B[Outlier Detection  
Feature Selection] --> C; D[Vessel Performance Curves] --> C; C --> E[Feature Engineering  
Exploratory Data Analysis]; C --> F[Predictive Modeling  
Multi-Linear Regression];
```

**Data Science Process**

**Data Wrangling**  
Record Cleaning

**Outlier Detection**  
Feature Selection

**Vessel Performance Curves**

**Iterative Cycle**

**Feature Engineering**  
Exploratory Data Analysis

**Predictive Modeling**  
Multi-Linear Regression

## Data Wrangling Record Cleaning

**Outlier Detection**  
**Feature Selection**

## Vessel Performance Curves

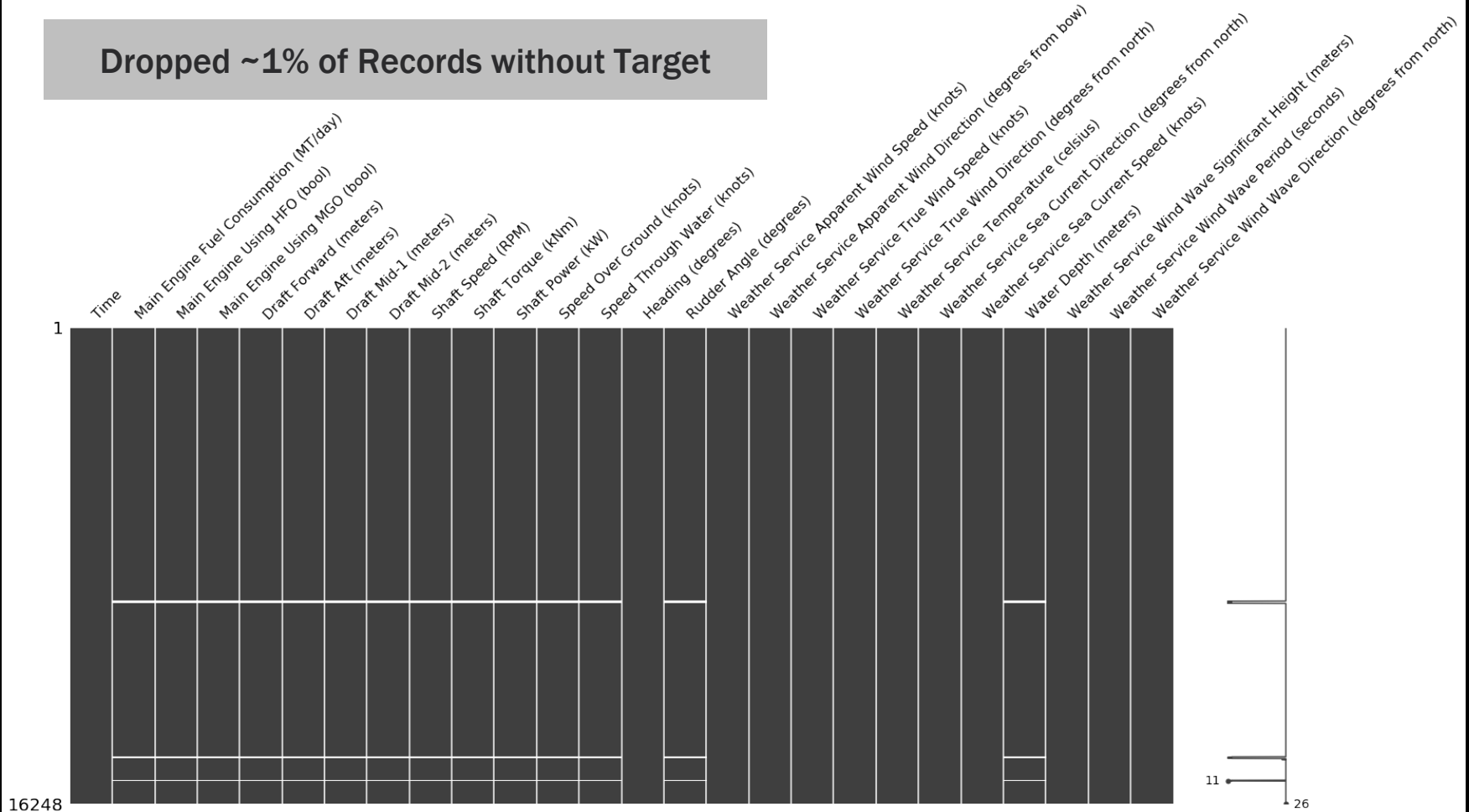
**Iterative Cycle**

## Feature Engineering Exploratory Data Analysis

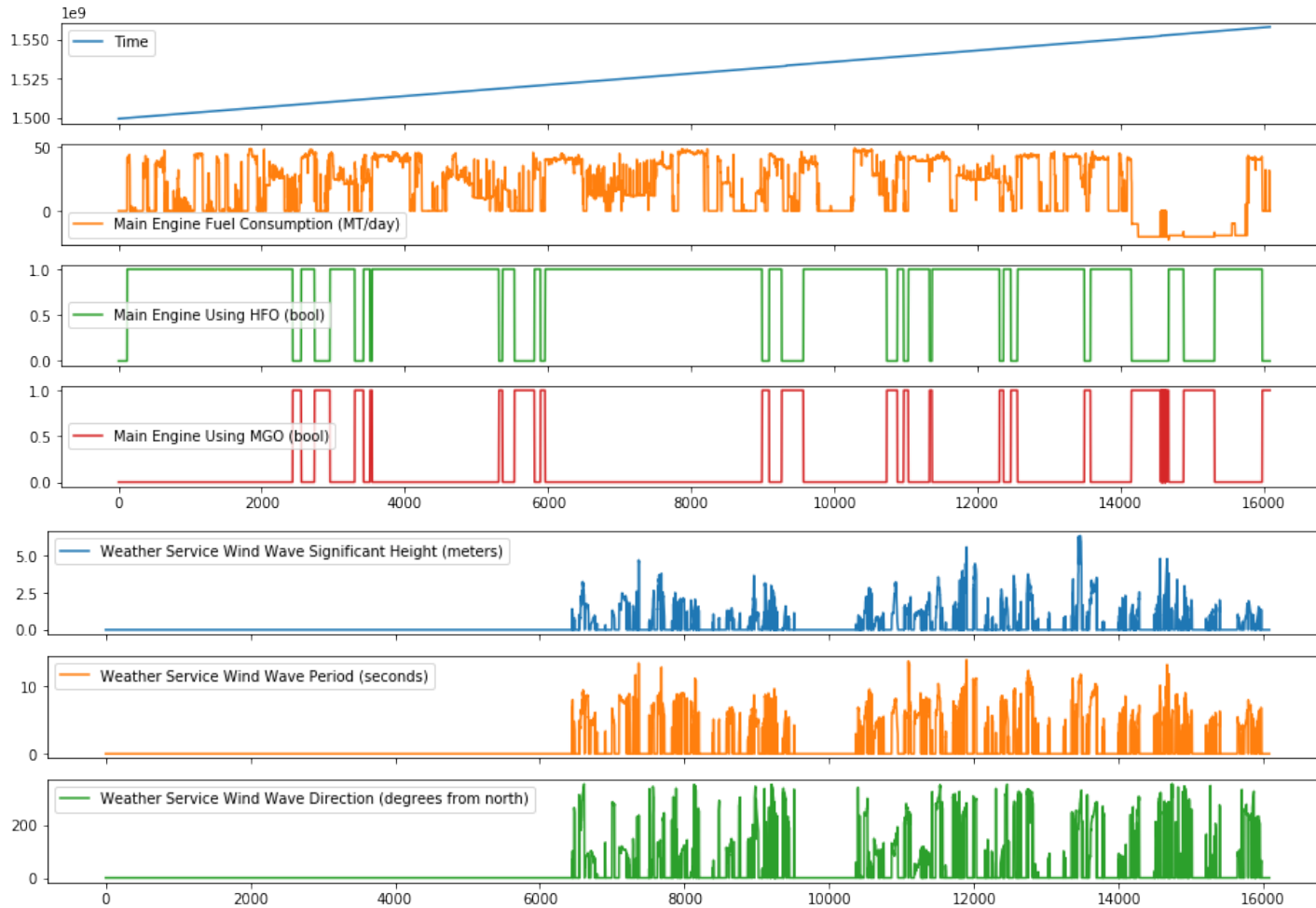
## Predictive Modeling Multi-Linear Regression

# Data Wrangling: Missing Records

Dropped ~1% of Records without Target



# Data Cleaning: Time, Main Engine & Wave Conditions



**DateTime Conversion**

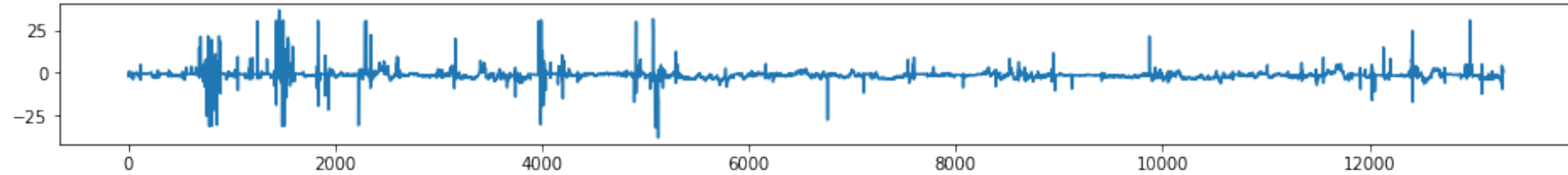
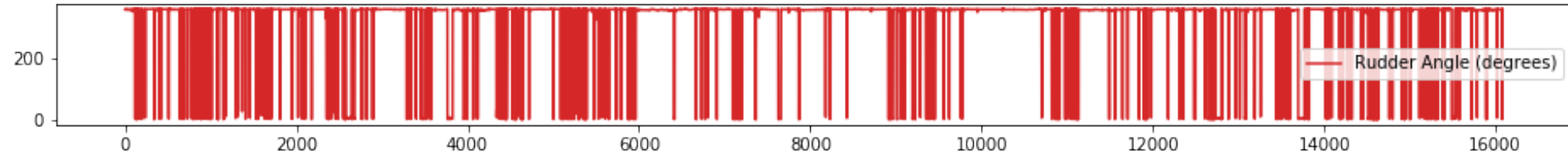
**Calibration / Negative  
Fuel Consumption**

**Reversion  
to Boolean Values**

**Dropped Columns  
with Excessive  
Missing Records**

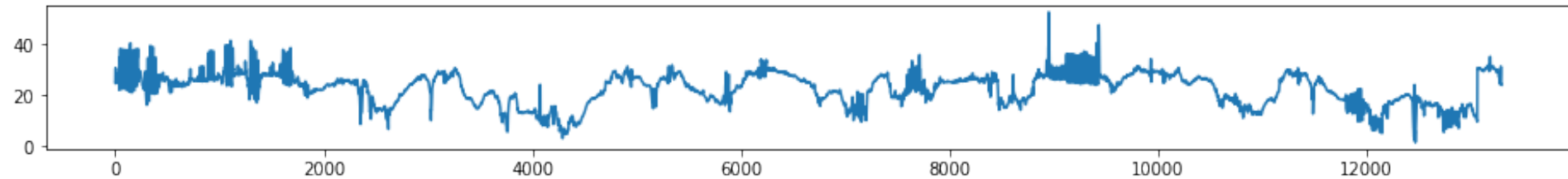
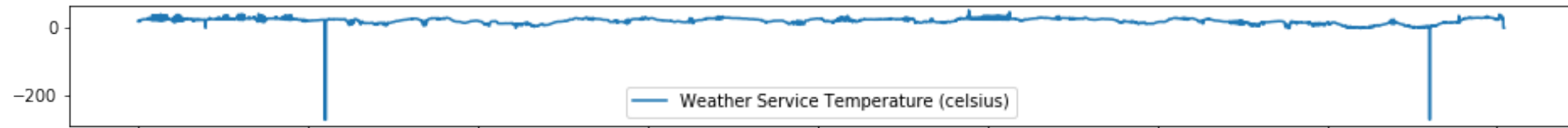


# Data Cleaning: Rudder Angle & Water Temperature

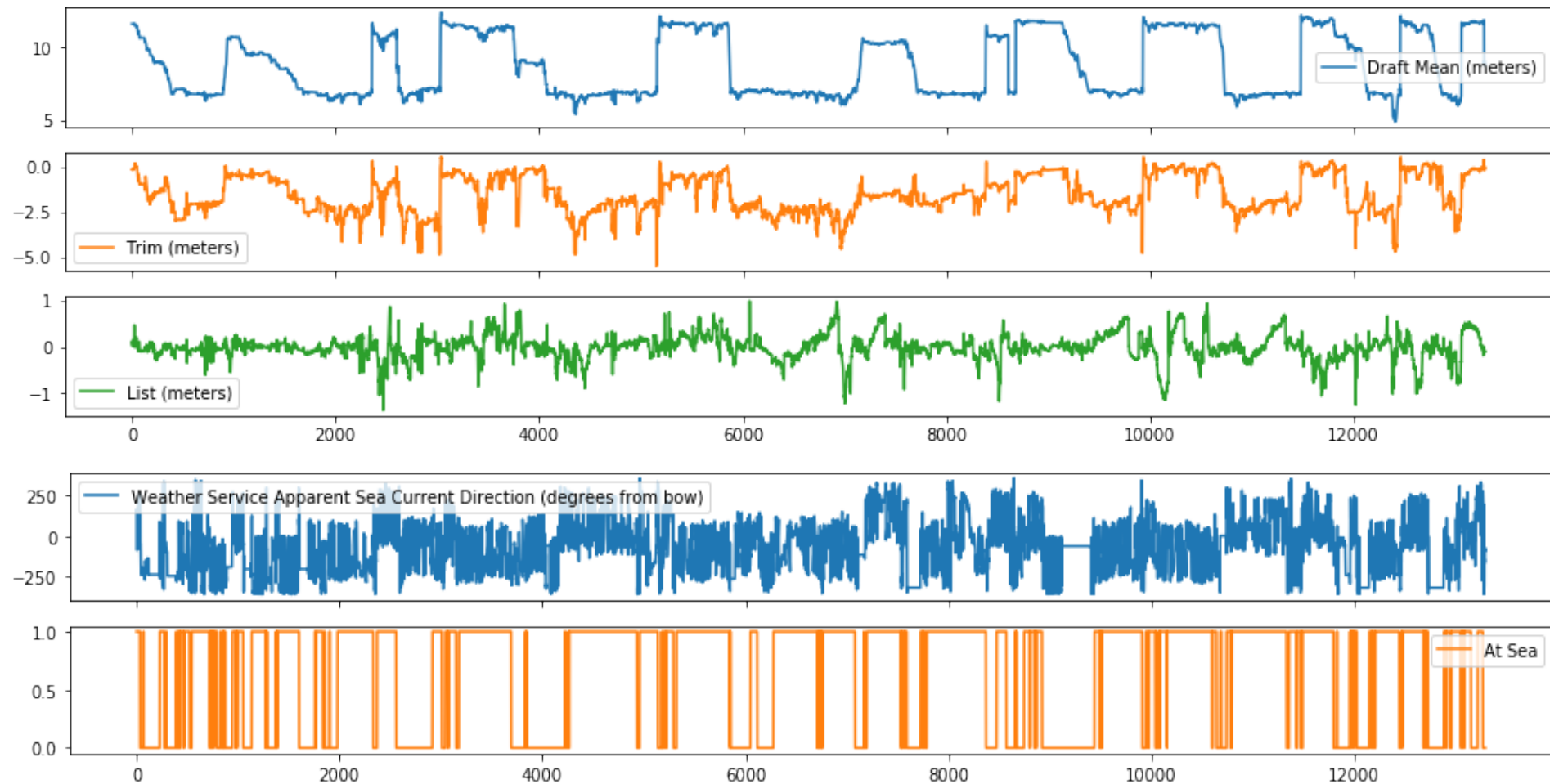


Rotated Rudder Angle for Continuity

Replaced Absolute Zero Temperatures

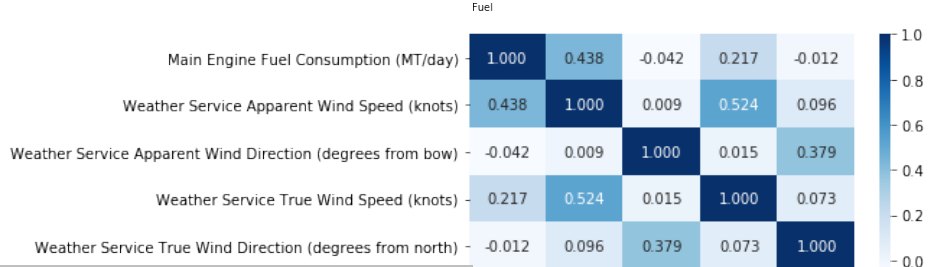
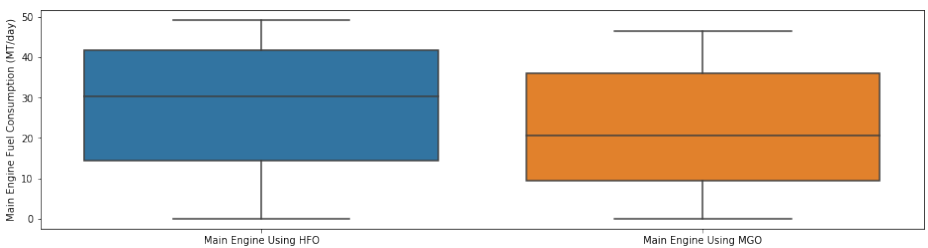


## Feature Engineering: Mean Draft, Trim, List, Apparent Sea Direction, At Sea & Speed Squared



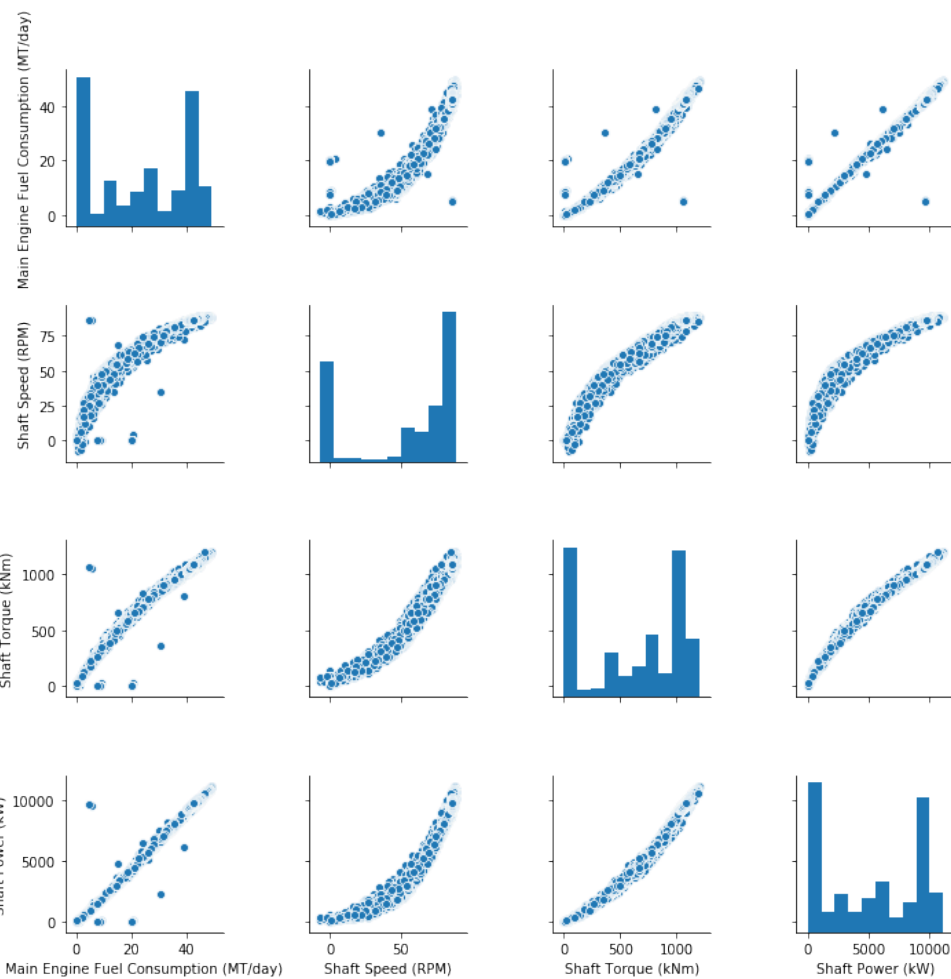


# Exploratory Data Analysis: Fuel Source, Wind Conditions, Shaft Performance

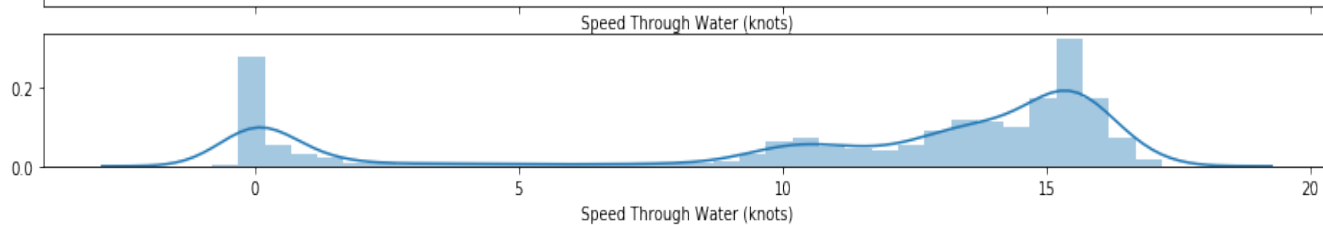
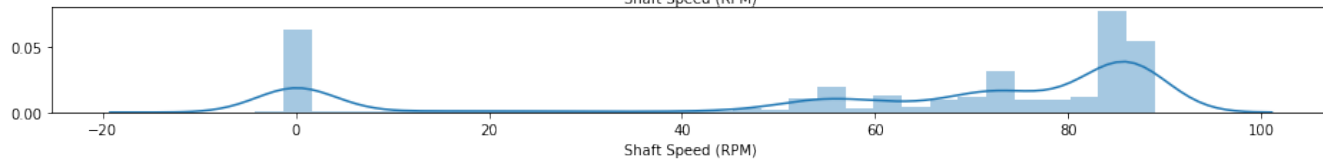
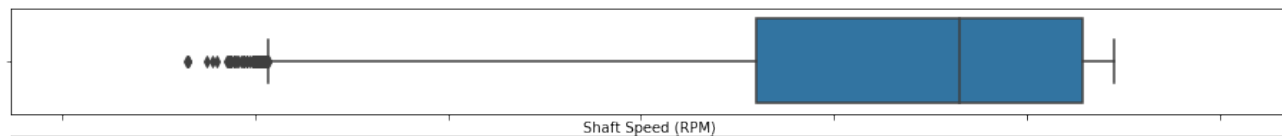
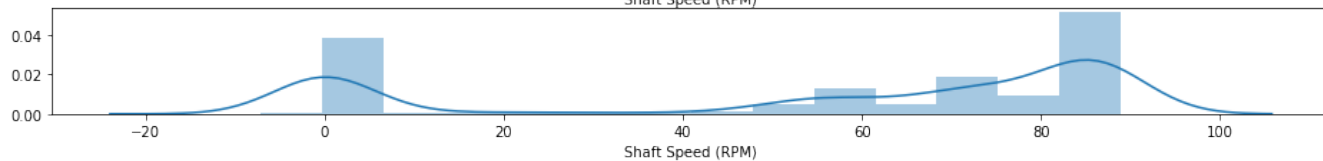
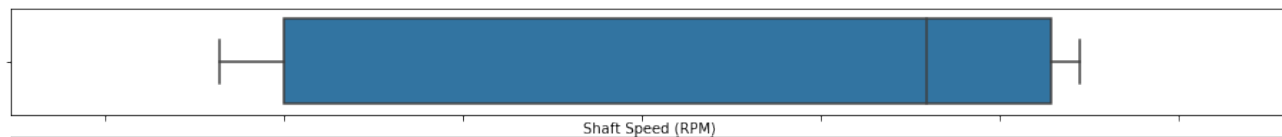


Apparent Wind Speed  
Most Correlated  
with Consumption

Free-Body Diagram  
Vector Component  
along Line of Action



# Outlier Detection: Shaft Speed & Speed through Water (Possibly Towed & Anchored)



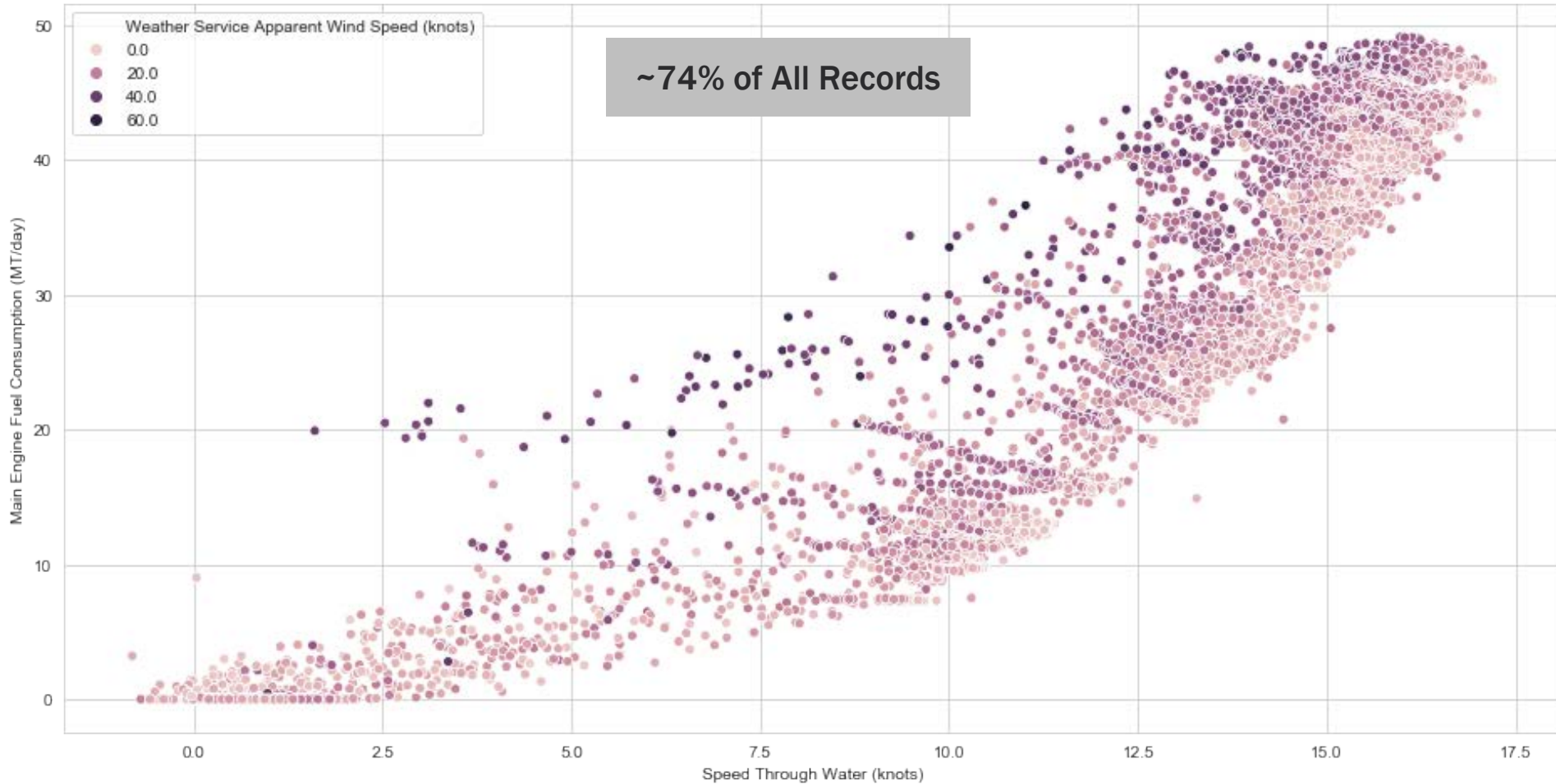
**Dropped 7.4% of Records:  
Zero Shaft Speed  
& Non-Zero Consumption**

**High Shaft Speed > 85 RPM  
Using Low Fuel < 6 MT/Day**

**Low Shaft Speed < 40 RPM  
Using High Fuel > 85 MT/Day**

**Dropped 0.2% of Records:  
Zero Speed through Water  
& Non-Zero Consumption**

# Data Set After Cleaning and Outlier Removal



# Multi-Linear Regression Modeling: Coefficients, Training Set & Predictions

	Coefficient	Min Limit	Max Limit	Min Effect	Max Effect
Feature					
Speed Through Water Squared (knots**2)	0.205837	0.000000	292.832568	0.000000	60.275682
Draft Mean (meters)	1.449527	4.999963	12.385497	7.247580	17.953109
Weather Service Apparent Wind Speed (knots)	0.316708	0.216047	51.955335	0.068424	16.454660
Trim (meters)	-1.593494	-5.379965	0.547380	8.572944	-0.872247
Weather Service Sea Current Speed (knots)	-0.382352	0.000000	3.184603	-0.000000	-1.217640
Speed Through Water (knots)	-0.845935	-0.812132	17.112351	0.687011	-14.475938
Intercept	-16.964856	NaN	NaN	NaN	NaN

80 / 20 Train Test Split

Training Set:

9,640 Records

R-Squared: 96.8%

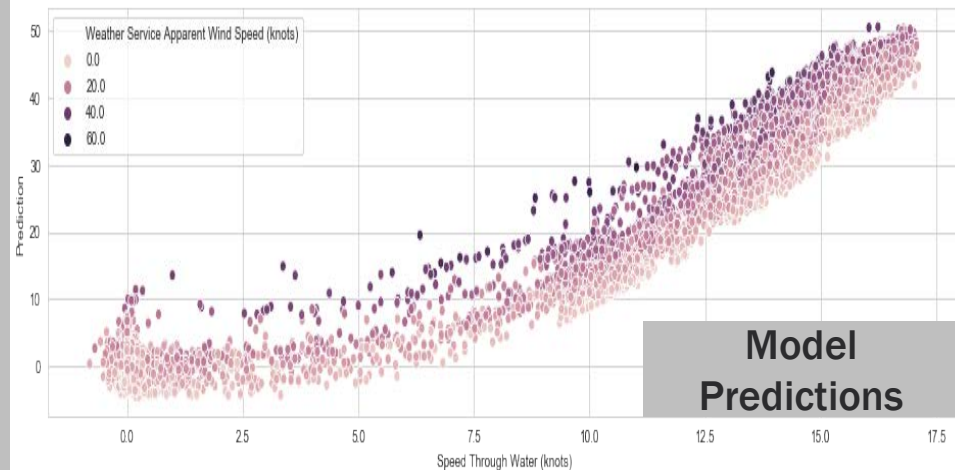
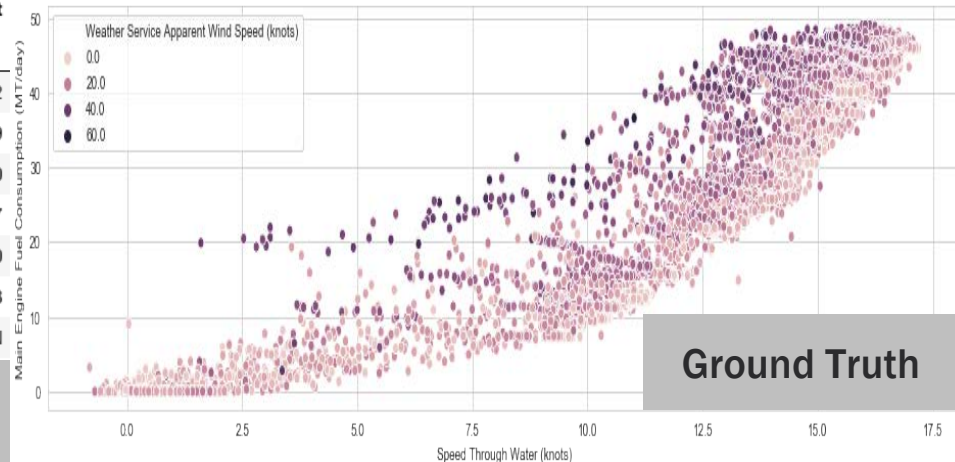
Root Mean Squared Error: 2.98 MT/Day

Test Set:

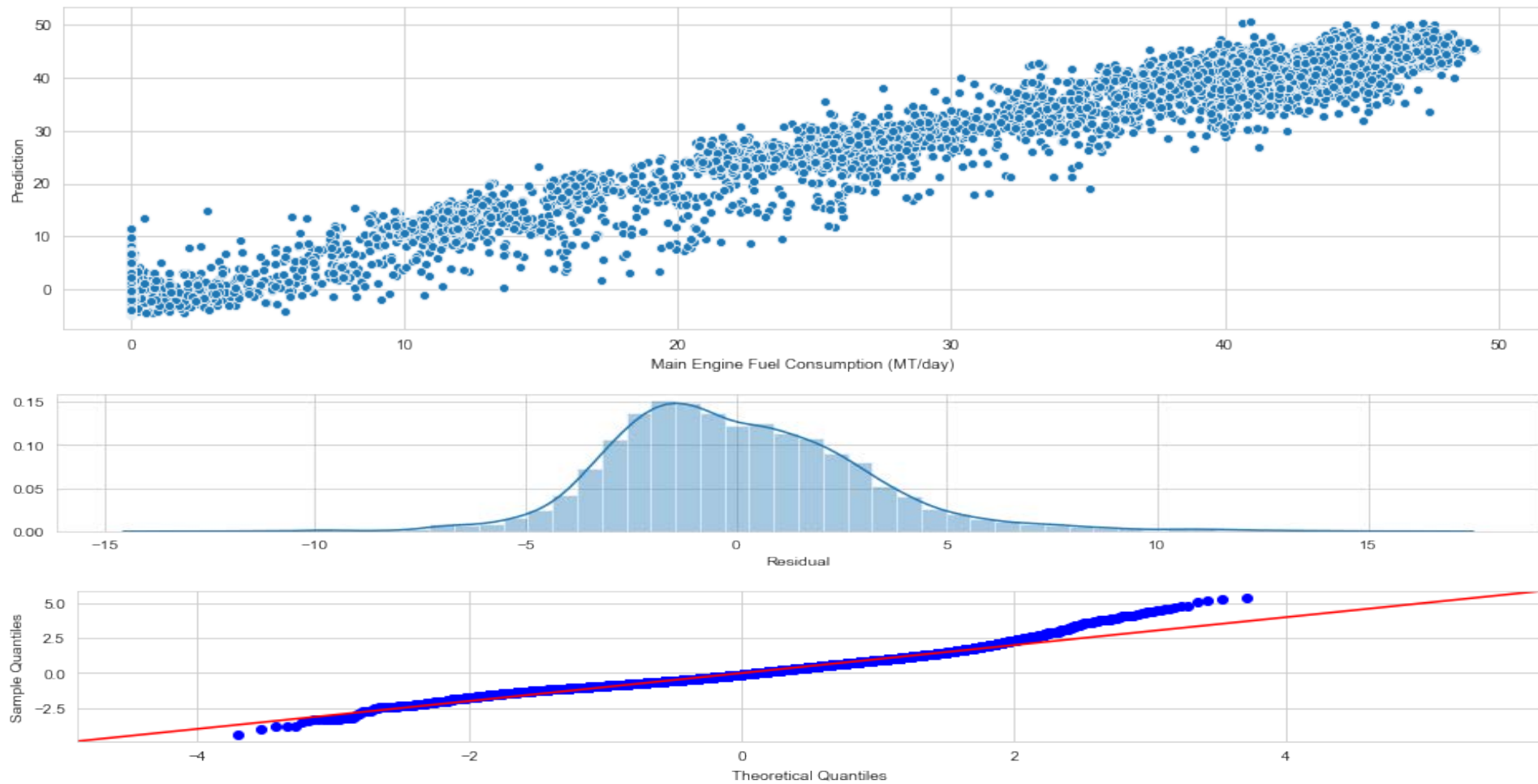
2,410 Records

R-Squared: 96.9%

Root Mean Squared Error: 2.95 MT/Day



# Multi-Linear Regression Modeling: Residual Linearity, Distribution & Normality



# Nearest Neighbor Regression Modeling: k-Neighbor Selection, Training Set & Predictions

Neighbors	Train R-Squared	Train RMSE	Test R-Squared	Test RMSE	Difference R-Squared	Difference RMSE
1	1.000000	0.000000	0.976815	2.522182	-0.023185	2.522182
2	0.994349	1.258252	0.981268	2.267098	-0.013082	1.008846
3	0.991852	1.510941	0.981896	2.228771	-0.009956	0.717830
4	0.990013	1.672826	0.982455	2.194072	-0.007557	0.521245
5	0.988737	1.776420	0.982685	2.179643	-0.006052	0.403223
6	0.987691	1.857088	0.982324	2.202263	-0.005367	0.345174

**80 / 20 Train Test Split**

**Training Set:**

**9,640 Records**

**R-Squared: 98.9%**

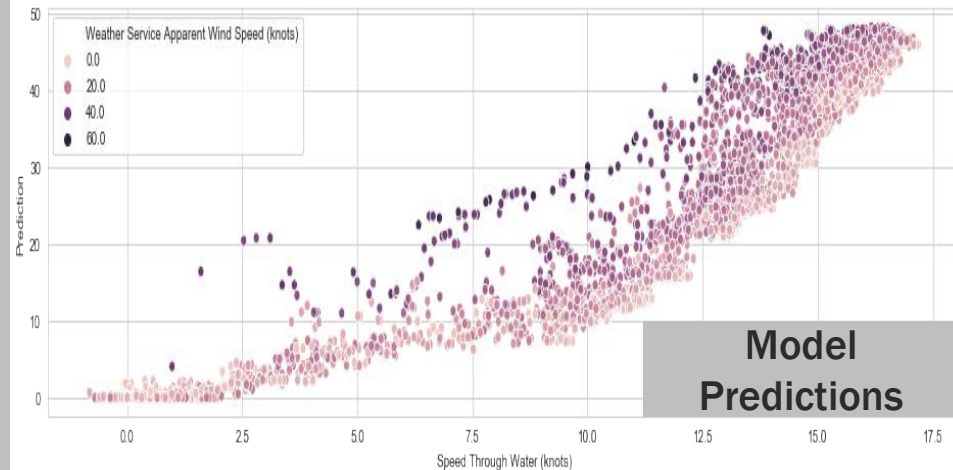
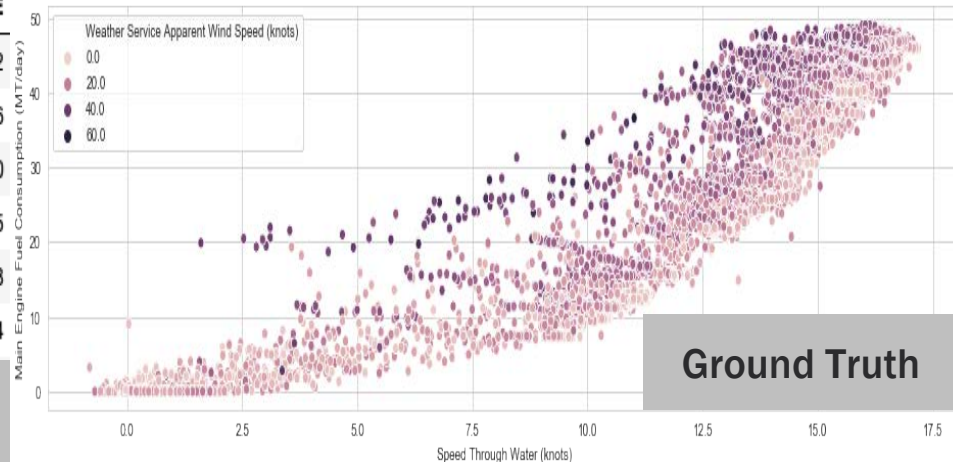
**Root Mean Squared Error: 1.78 MT/Day**

**Test Set:**

**2,410 Records**

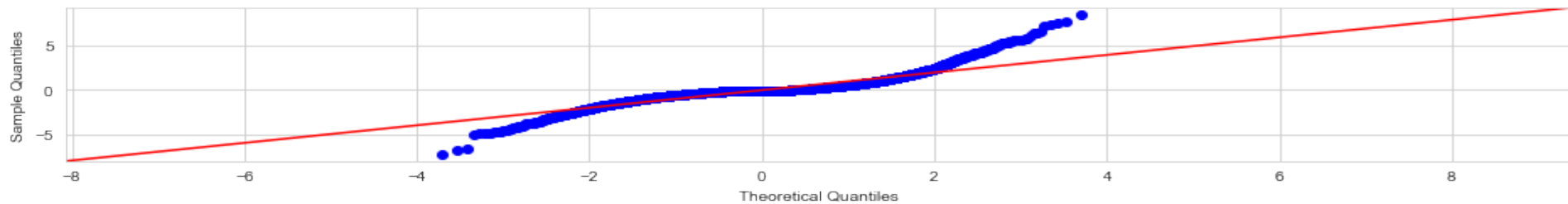
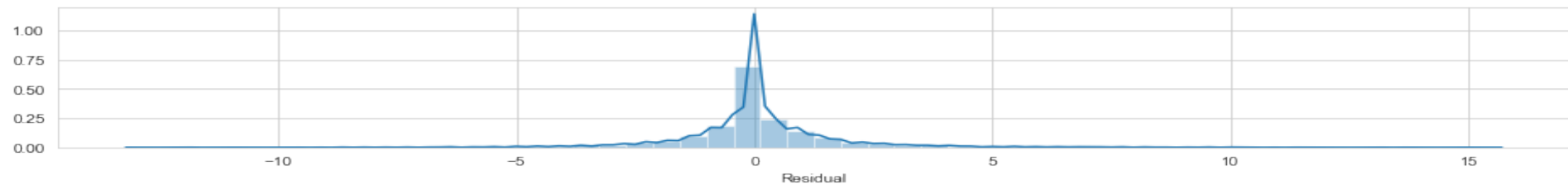
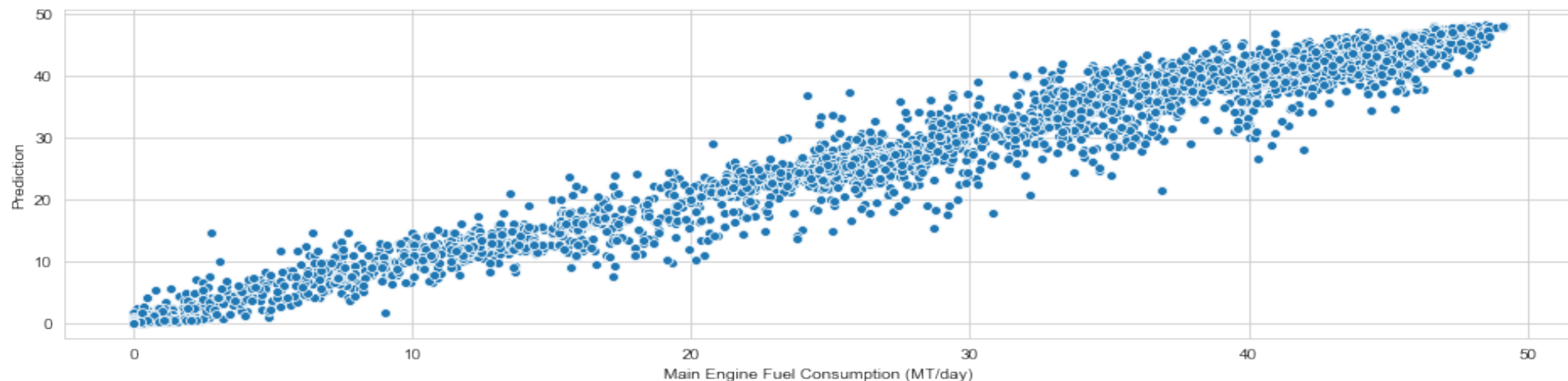
**R-Squared: 98.3%**

**Root Mean Squared Error: 2.18 MT/Day**





# Nearest Neighbor Regression Modeling: Residual Linearity, Distribution & Normality



# Random Forest Regression Modeling: Feature Importance, Training Set & Predictions

Feature	Importance
Speed Through Water Squared (knots**2)	0.490366
Speed Through Water (knots)	0.462709
Draft Mean (meters)	0.021708
Weather Service Apparent Wind Speed (knots)	0.020002
Trim (meters)	0.004443
Weather Service Sea Current Speed (knots)	0.000771

80 / 20 Train Test Split

Training Set:

9,640 Records

R-Squared: 99.3%

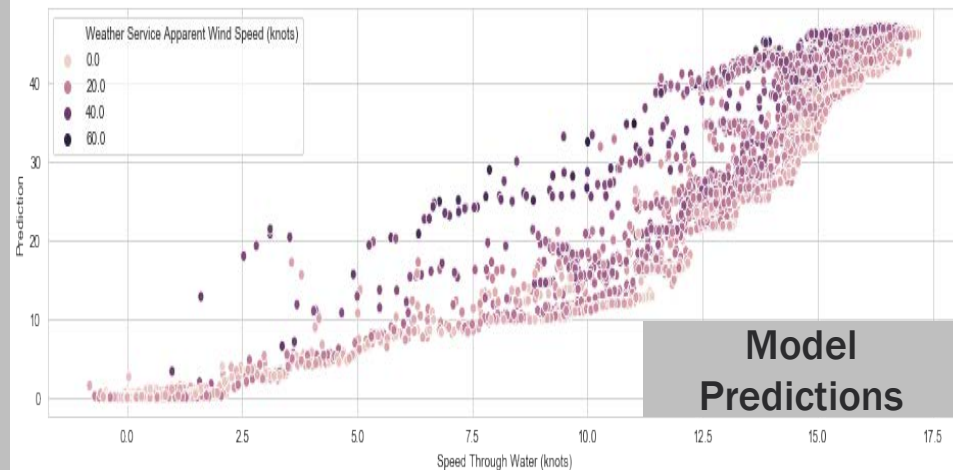
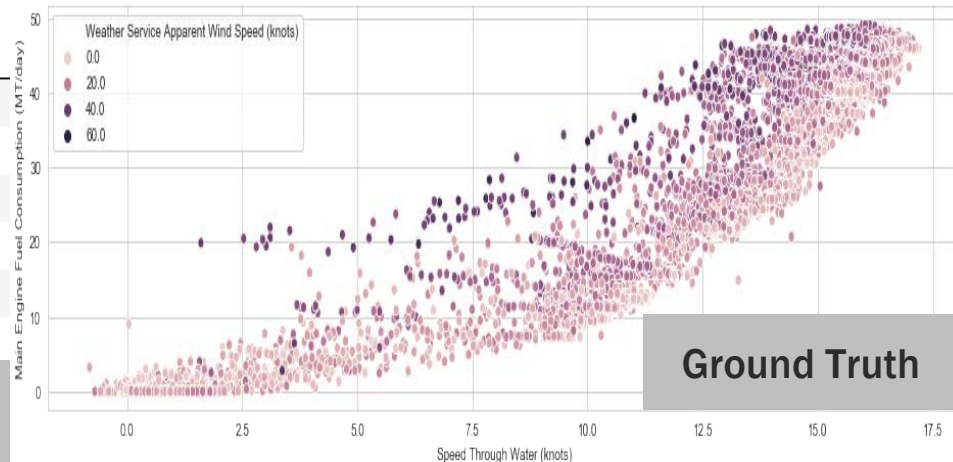
Root Mean Squared Error: 1.42 MT/Day

Test Set:

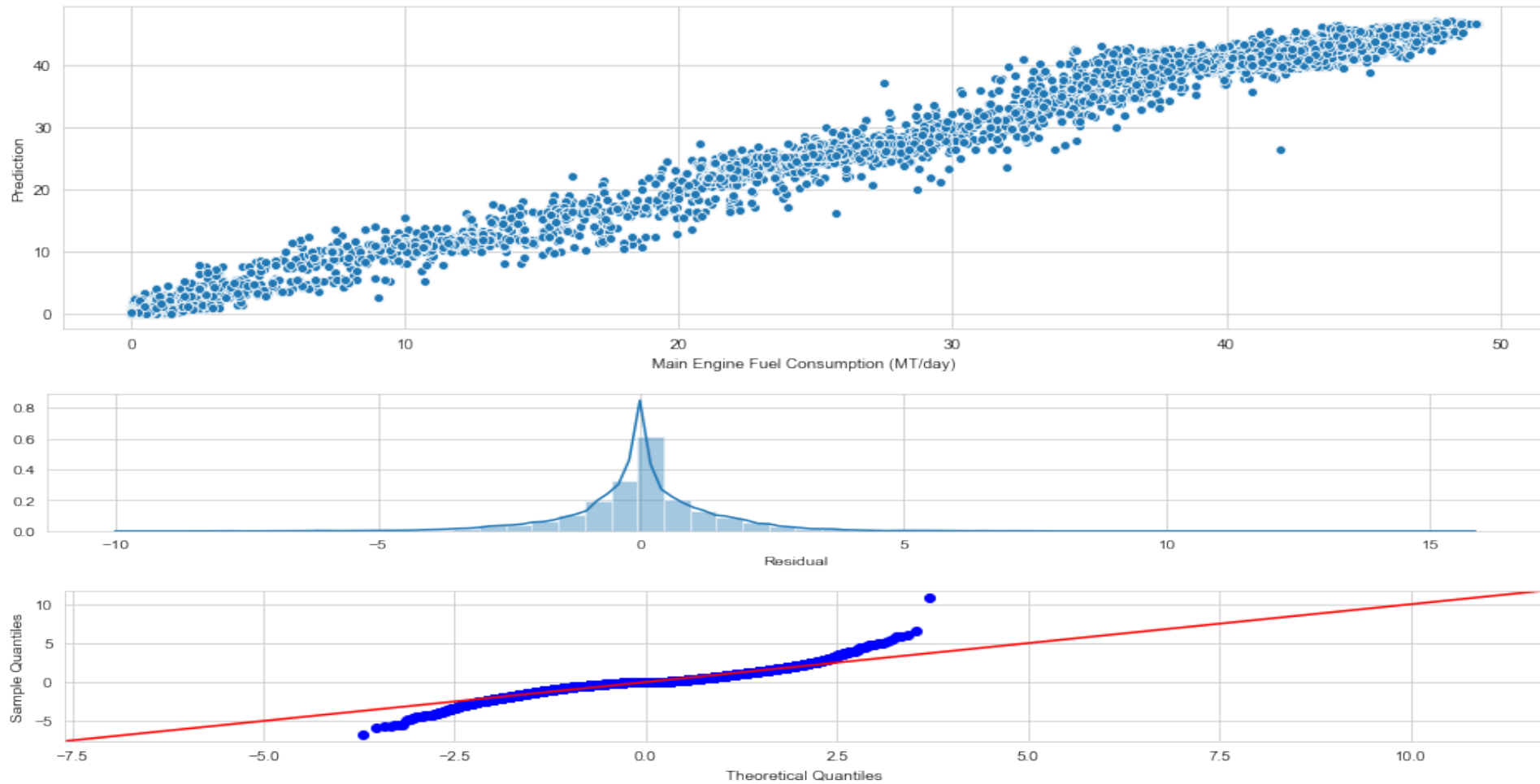
2,410 Records

R-Squared: 98.9%

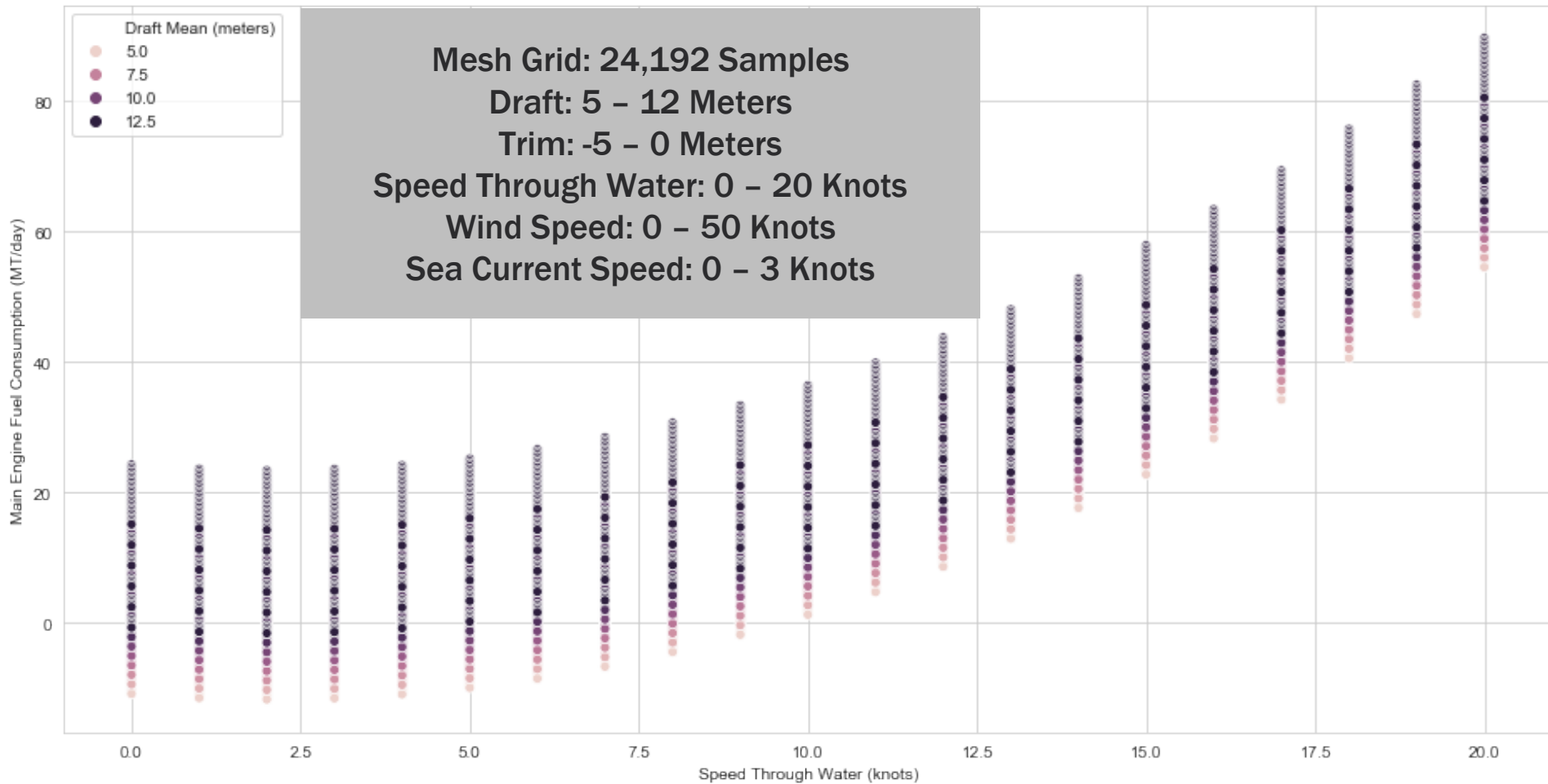
Root Mean Squared Error: 1.74 MT/Day



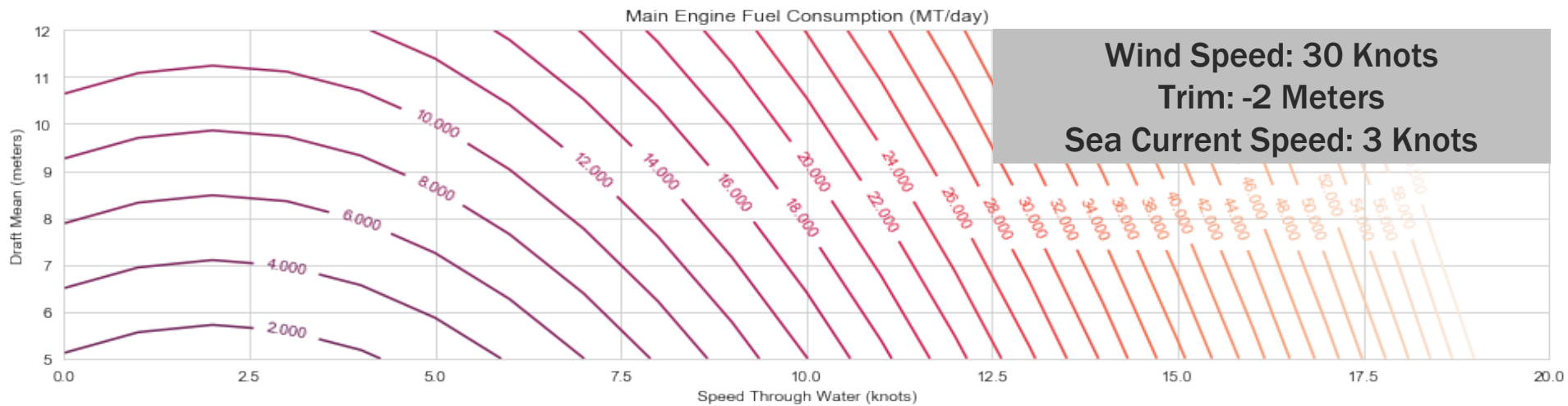
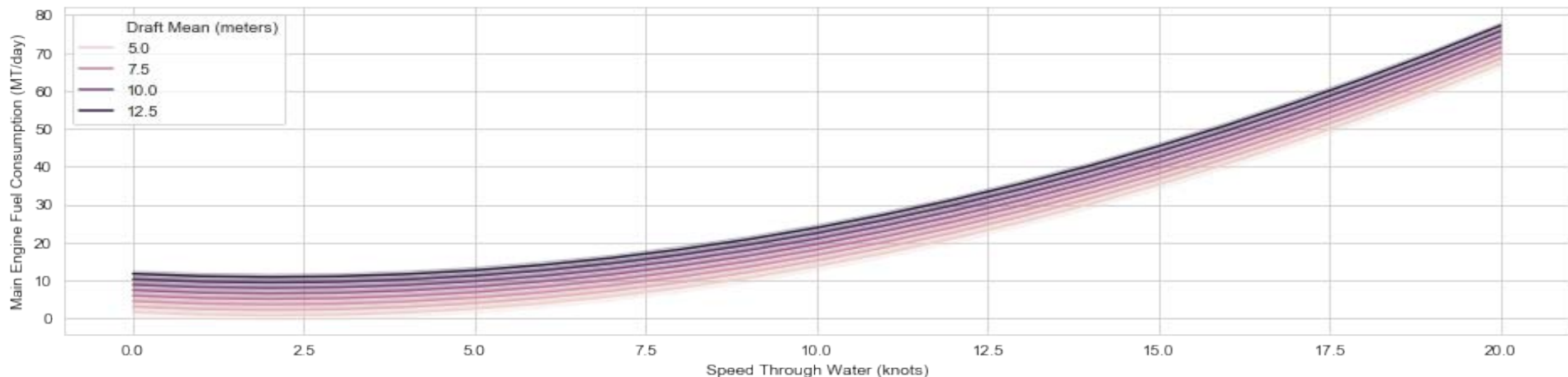
# Random Forest Regression Modeling: Residual Linearity, Distribution & Normality



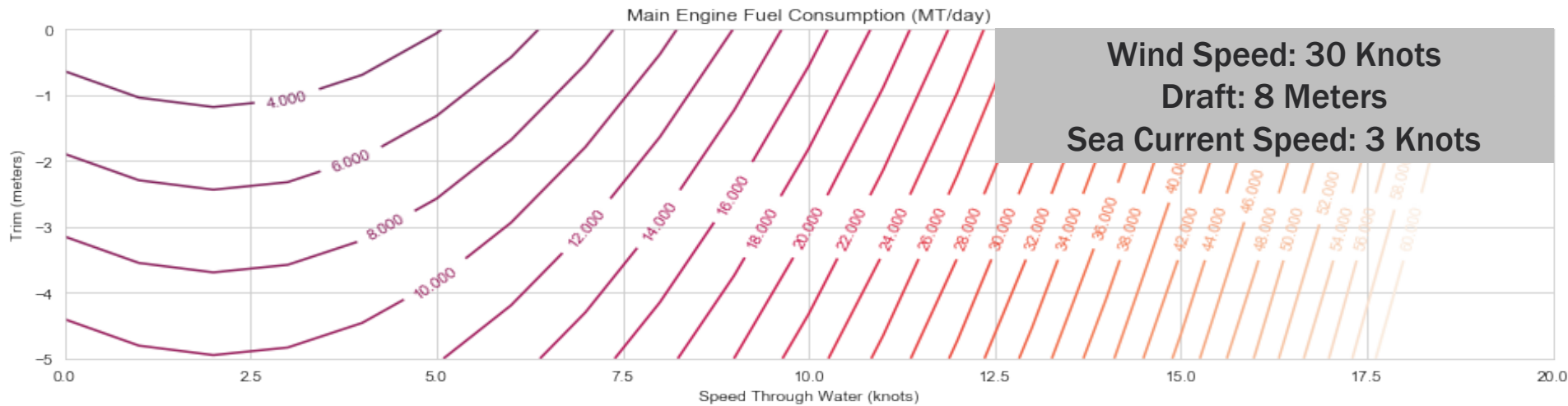
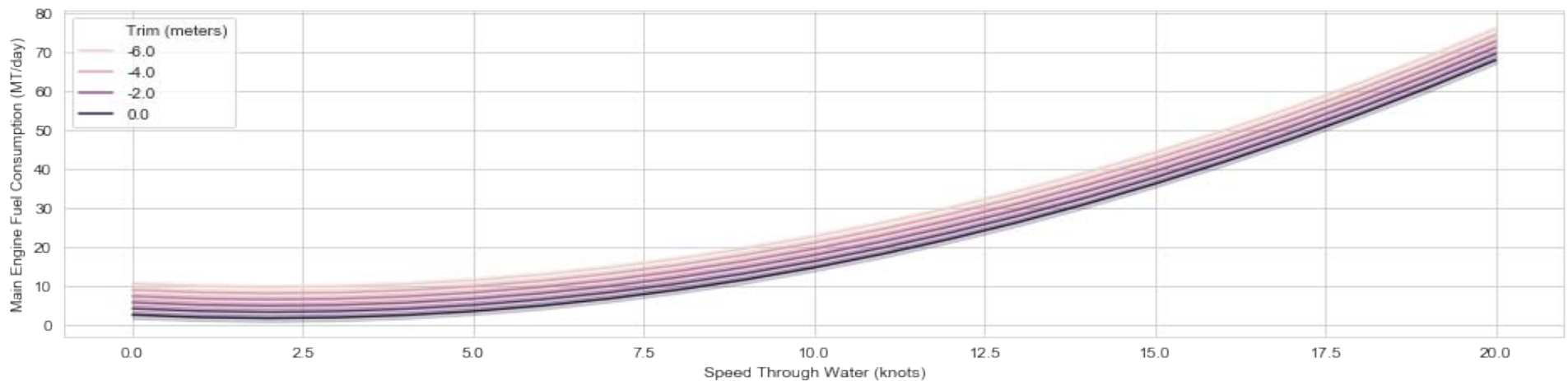
# Multi-Linear Regression Vessel Performance Curves: Prediction Sampling



# Multi-Linear Regression Vessel Performance Curves: Sensitivity to Draft

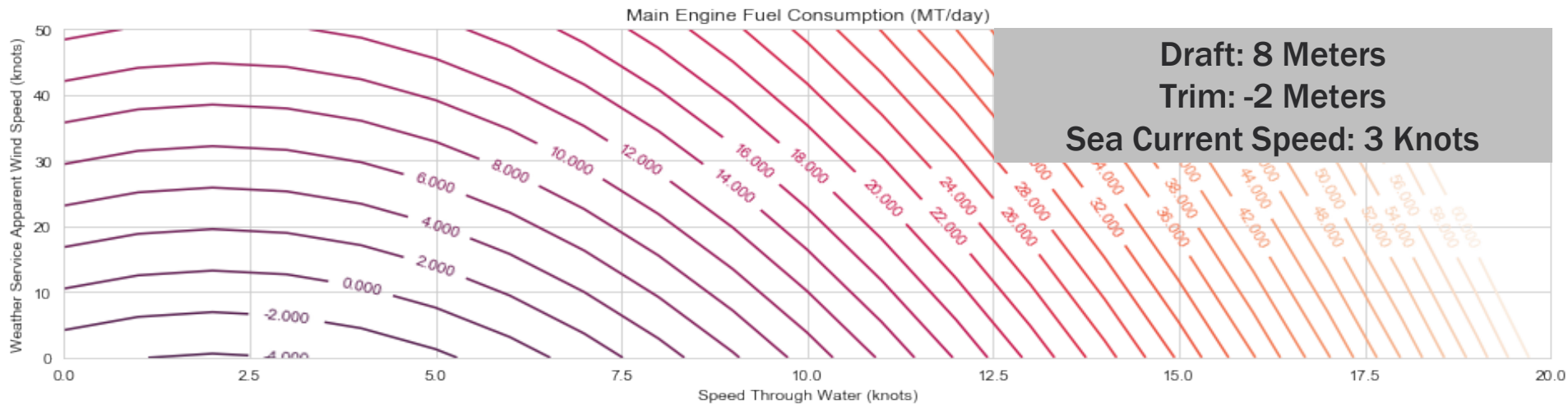
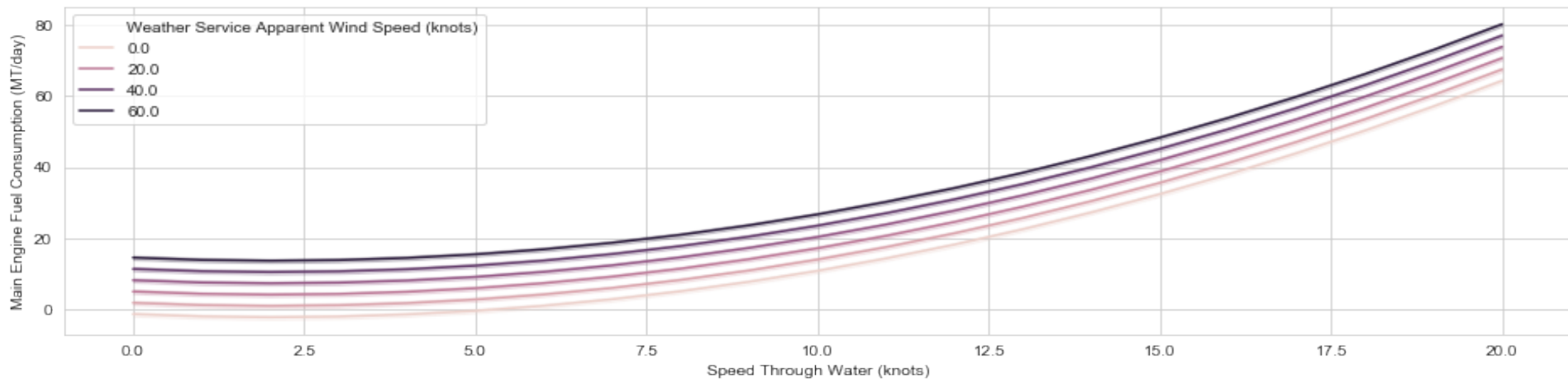


# Multi-Linear Regression Vessel Performance Curves: Sensitivity to Trim

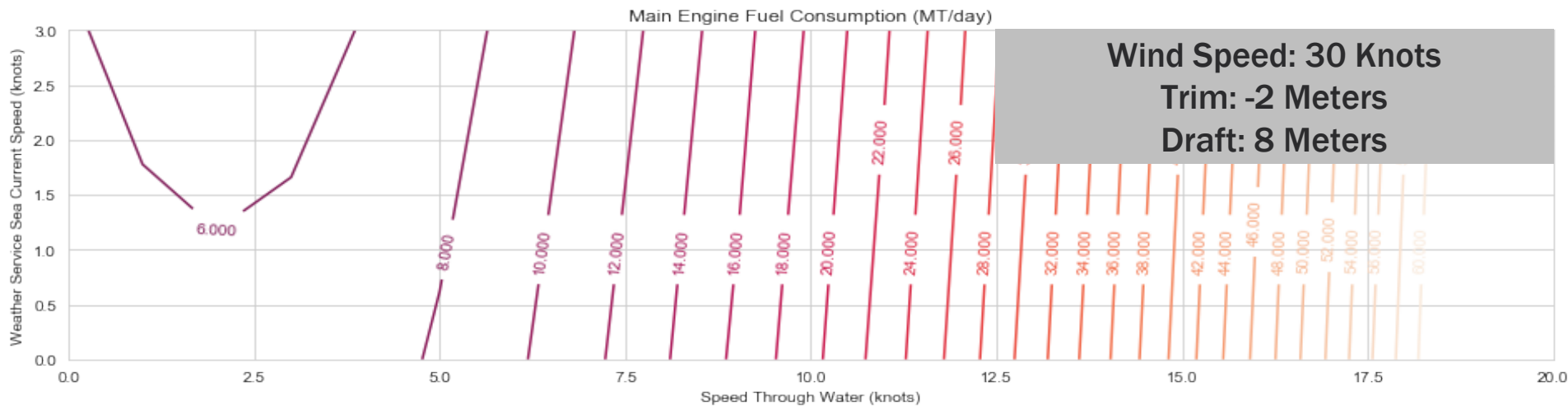
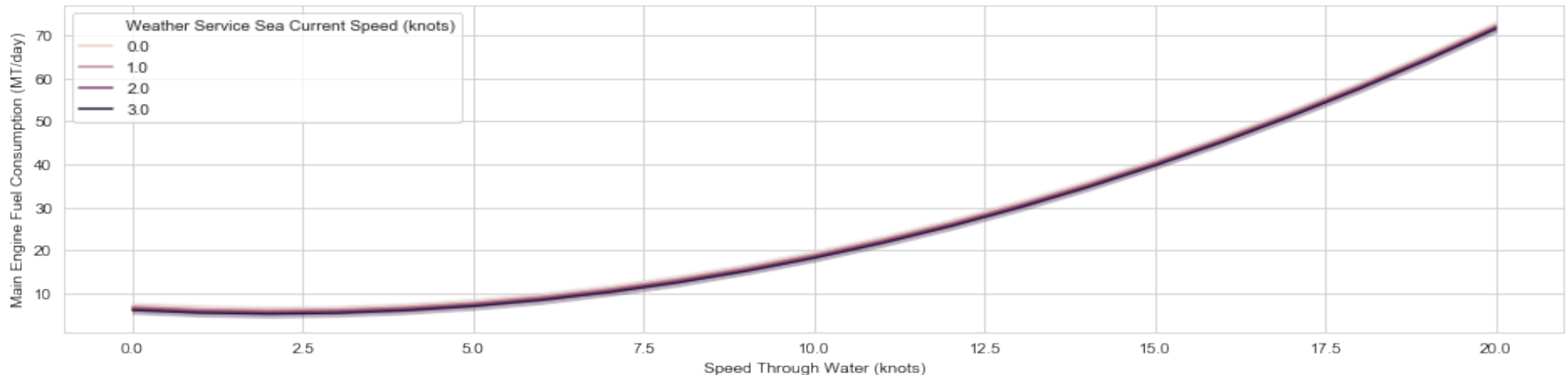




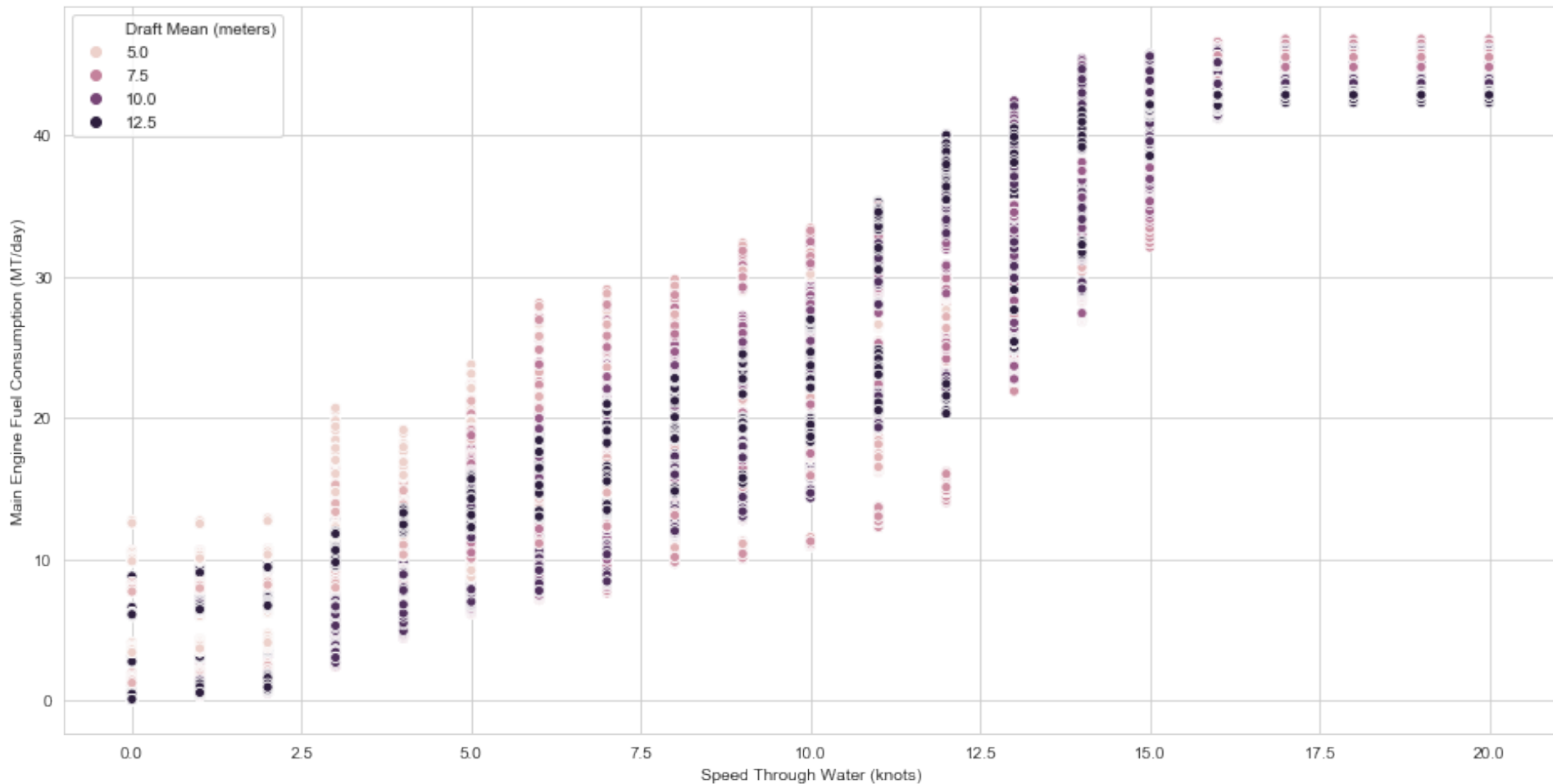
## Multi-Linear Regression Vessel Performance Curves: Sensitivity to Wind Speed



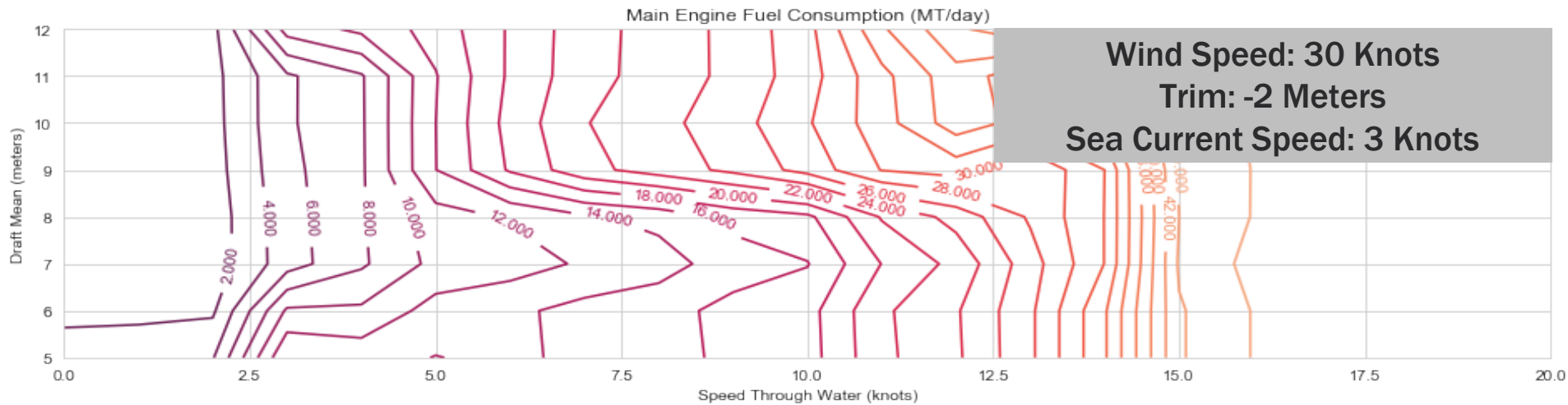
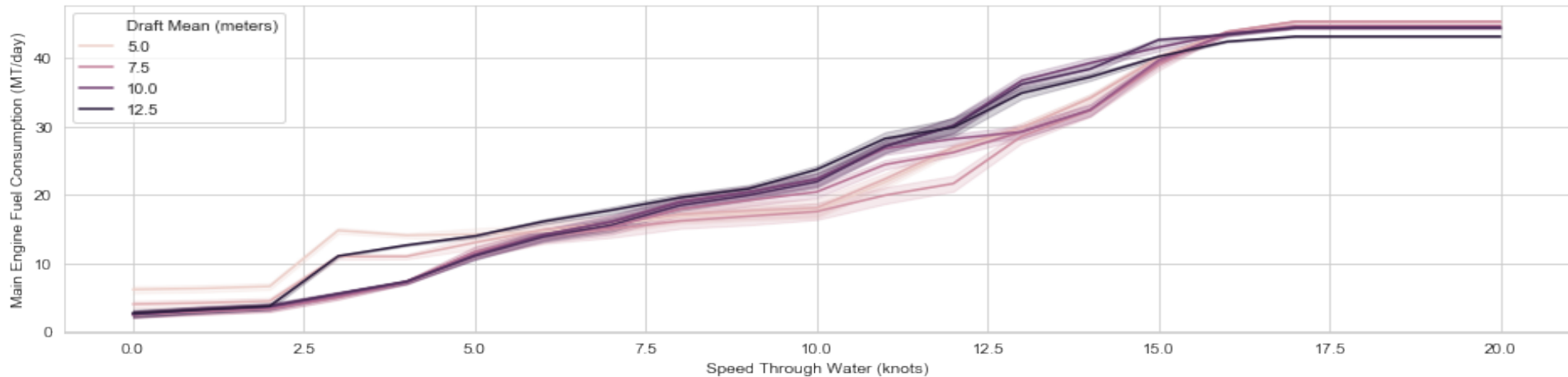
# Multi-Linear Regression Vessel Performance Curves: Sensitivity to Sea Current Speed



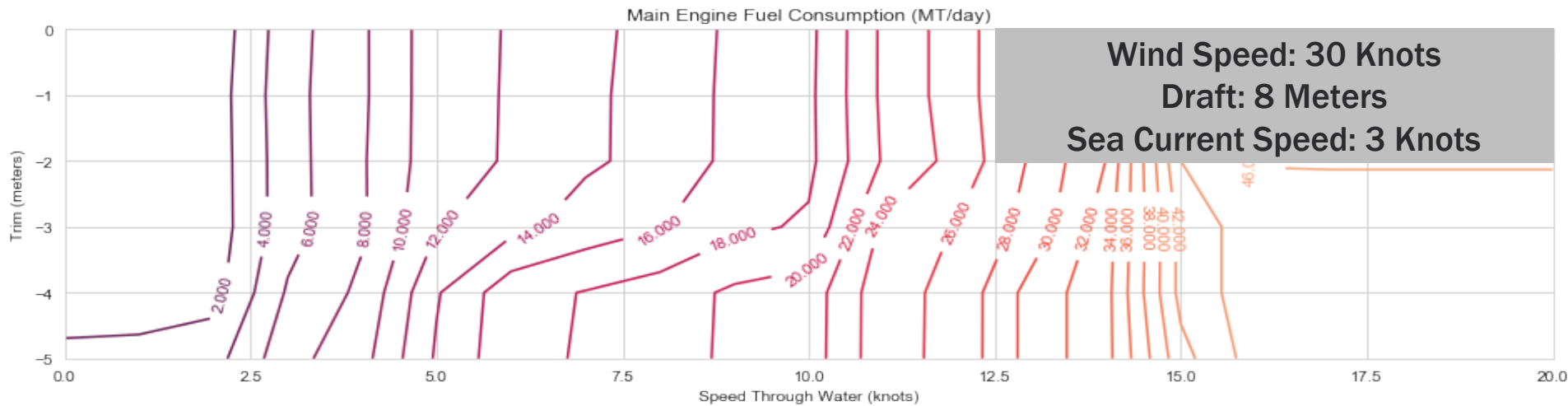
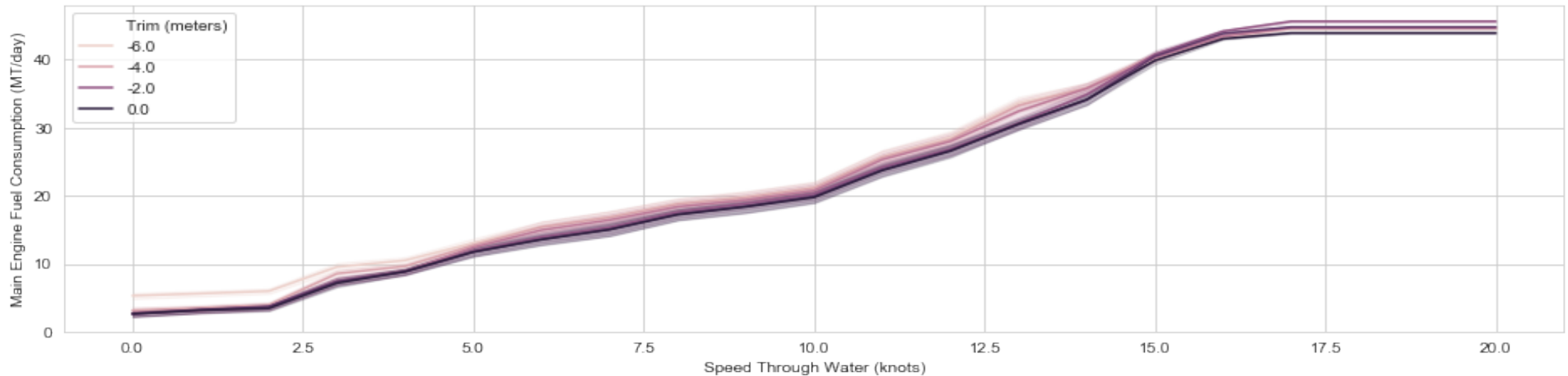
# Random Forest Regression Vessel Performance Curves: Prediction Sampling



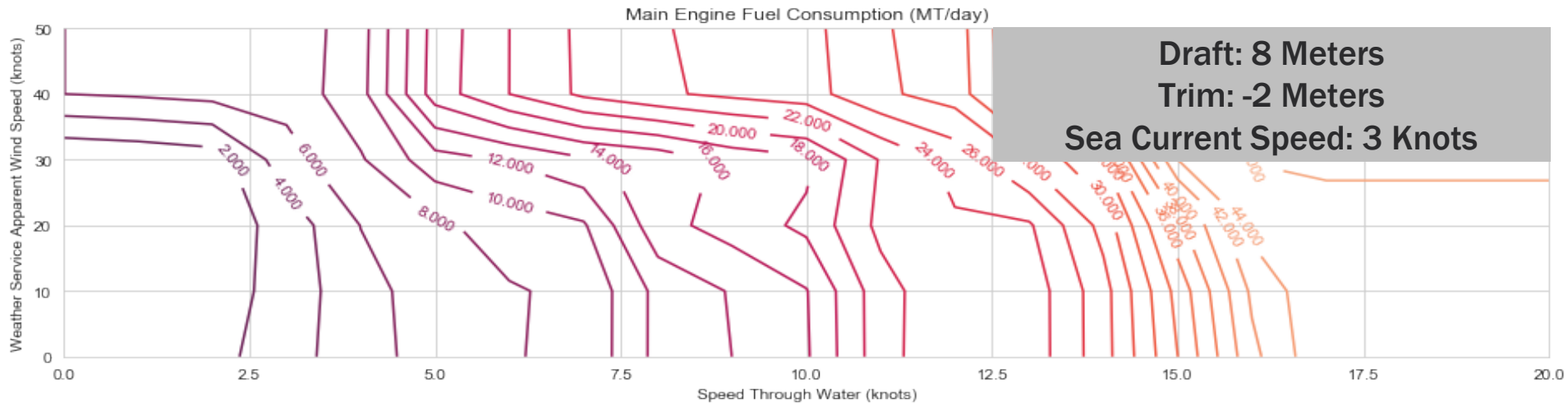
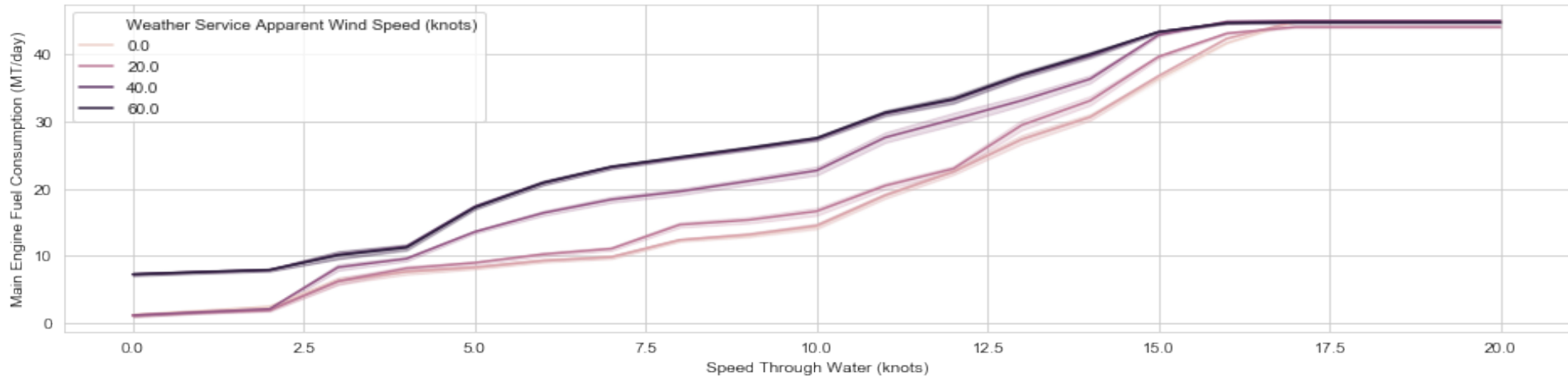
# Random Forest Regression Vessel Performance Curves: Sensitivity to Draft



# Random Forest Regression Vessel Performance Curves: Sensitivity to Trim

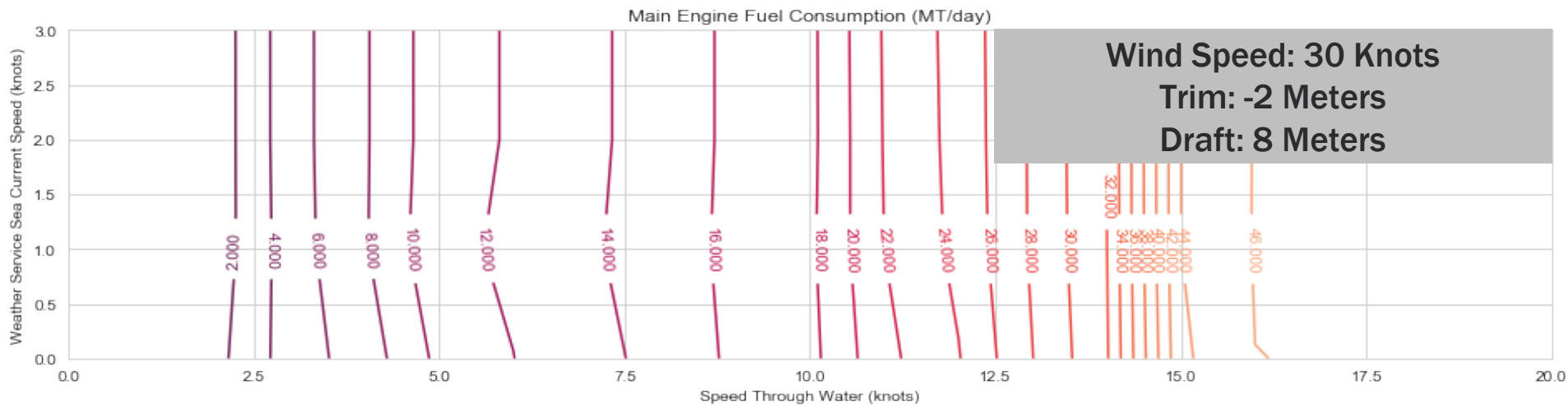
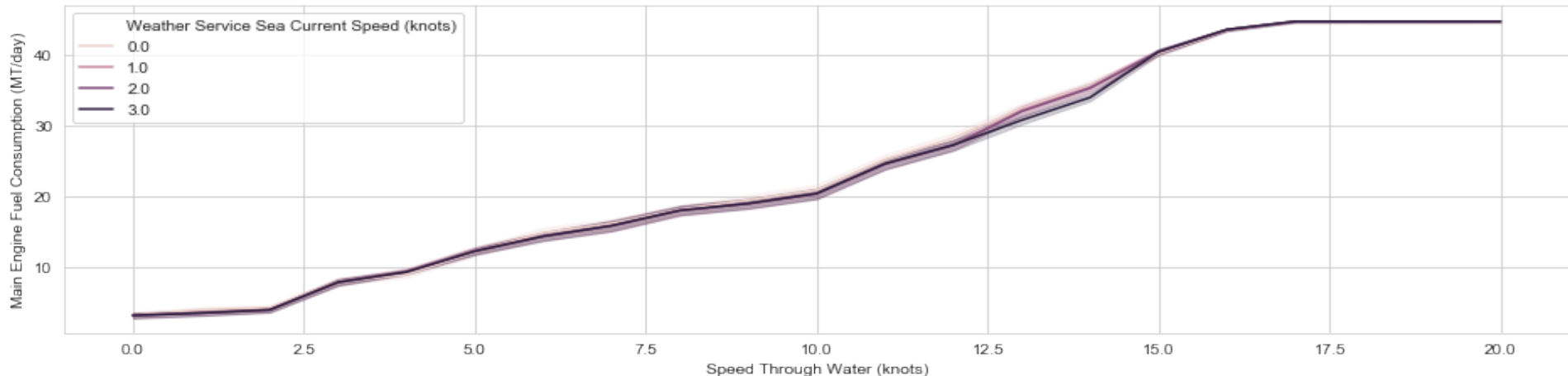


## Random Forest Regression Vessel Performance Curves: Sensitivity to Wind Speed

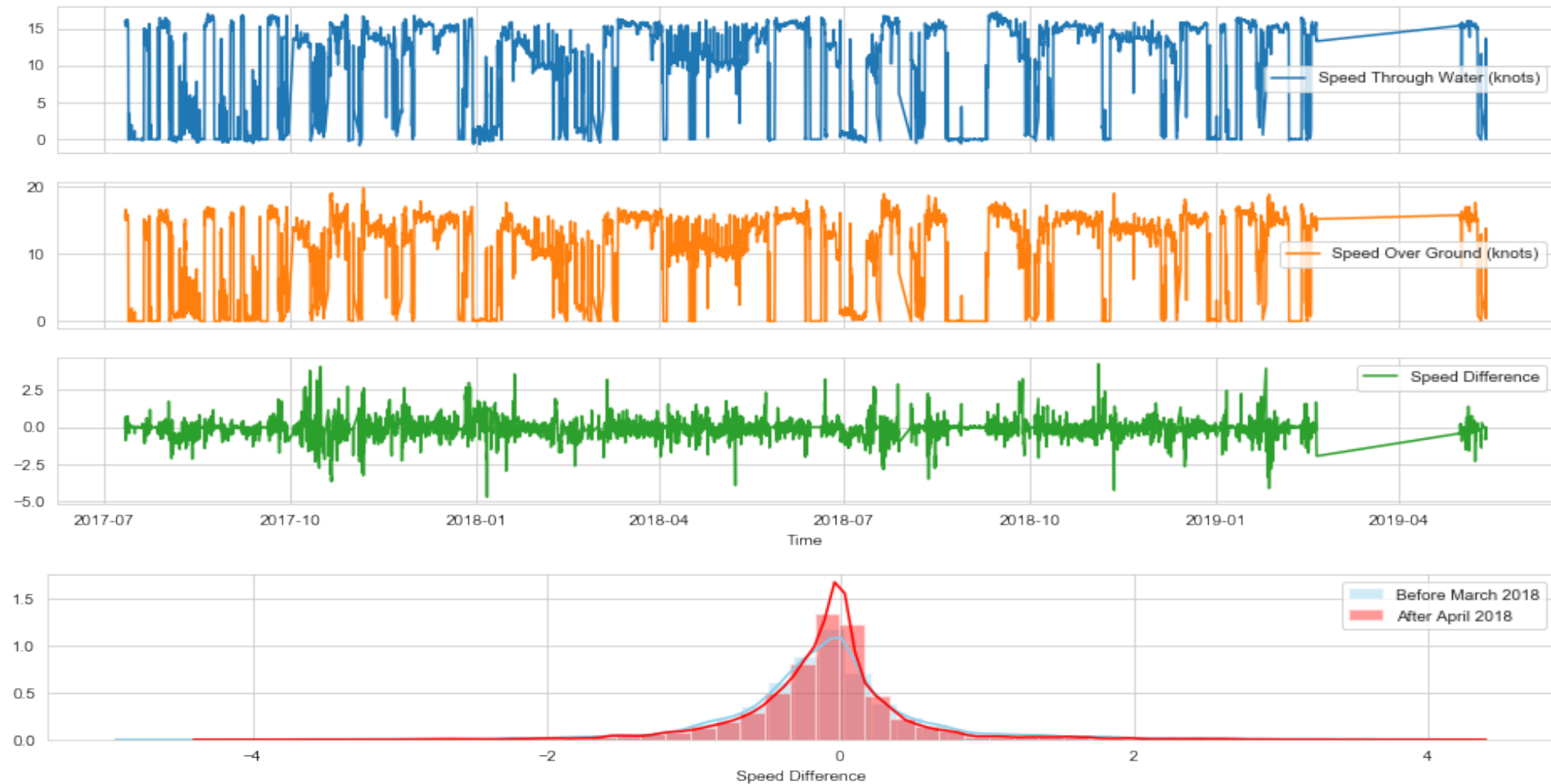




## Random Forest Regression Vessel Performance Curves: Sensitivity to Sea Current Speed



# Sensor Drift: Difference in Water and Ground Speed



# Sensor Drift: 200-Day Rolling Average



## Ship Performance

Adam C Dick

[www.linkedin.com/in/adamcdick](https://www.linkedin.com/in/adamcdick)

Project Data Stack	
Engineering Understanding	Vessel Performance IoT Sensors
Data Wrangling	Python Missingno
Feature Engineering Data Exploration	Python Pandas
Outlier Detection Feature Selection	Python Pandas
Predictive Modeling	Sci-Kit Learn Numpy / Scipy
Data Visualization	Matplotlib Seaborn