Project Number: DE-FE0031557

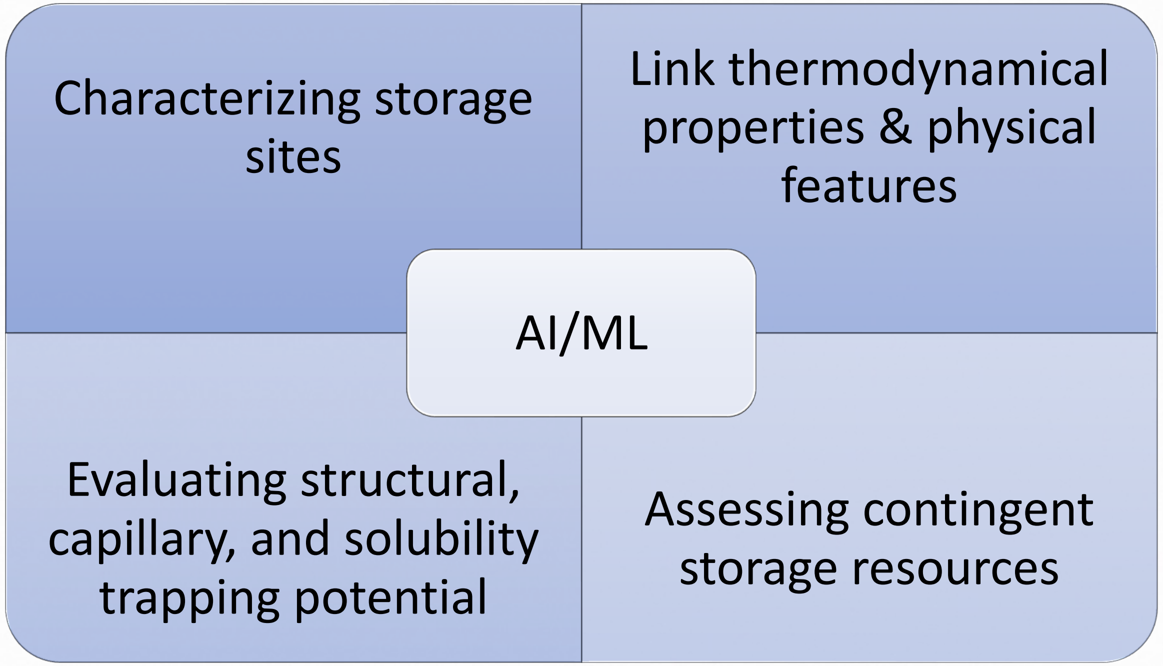
SECARB Offshore

Predicting CO_2 Solubility in Crude Oil with Machine Learning

Rupom Bhattacherjee, Kodjo Botchway, Jack C. Pashin, Goutam Chakraborty, Prem K. Bikkina

Oklahoma State University

Objective



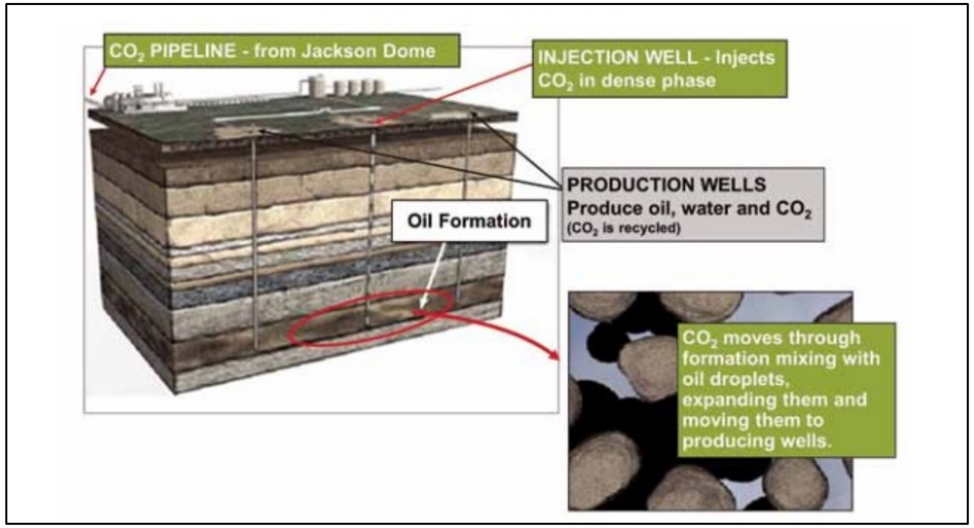
- A Decision Support System in SAS Viya is built to screen storage sites in central GOM
- Developed models to predict fluid properties (Solubility, Density, Viscosity, IFT) for CO_2-Brine system.
- Recent focus on estimating CO_2 solubility in crude oils & model oils (e.g., Decane)

Why do we want to predict solubility?

- CO_2 Sequestration
  - Sequestration involves different trapping mechanisms
  - Solubility trapping prevents plume migration and improve storage capacity

- CO_2 EOR
  - Miscible flooding prevents early CO_2 breakthrough > Better sweep efficiency
  - Oil swells > rise in oil mobility > less oil trapped

- Solubility will vary with composition of formation fluids and the reservoir condition



CO_2 EOR Mechanism (Courtesy: Denbury Resources)

Existing methods of measuring solubility

Method	Limitations
Laboratory measurements	• Costly and time-consuming.
Theoretical methods/commercial simulators	• Complicated differential equations, assumptions on equilibrium constants to exactly characterize fluid systems. • Not an ideal option for quick approximation.
Empirical correlations	• Limited applicability. • Does not have predictive power for unseen data.

What’s new with this work?

Simpler ML algorithms are employed

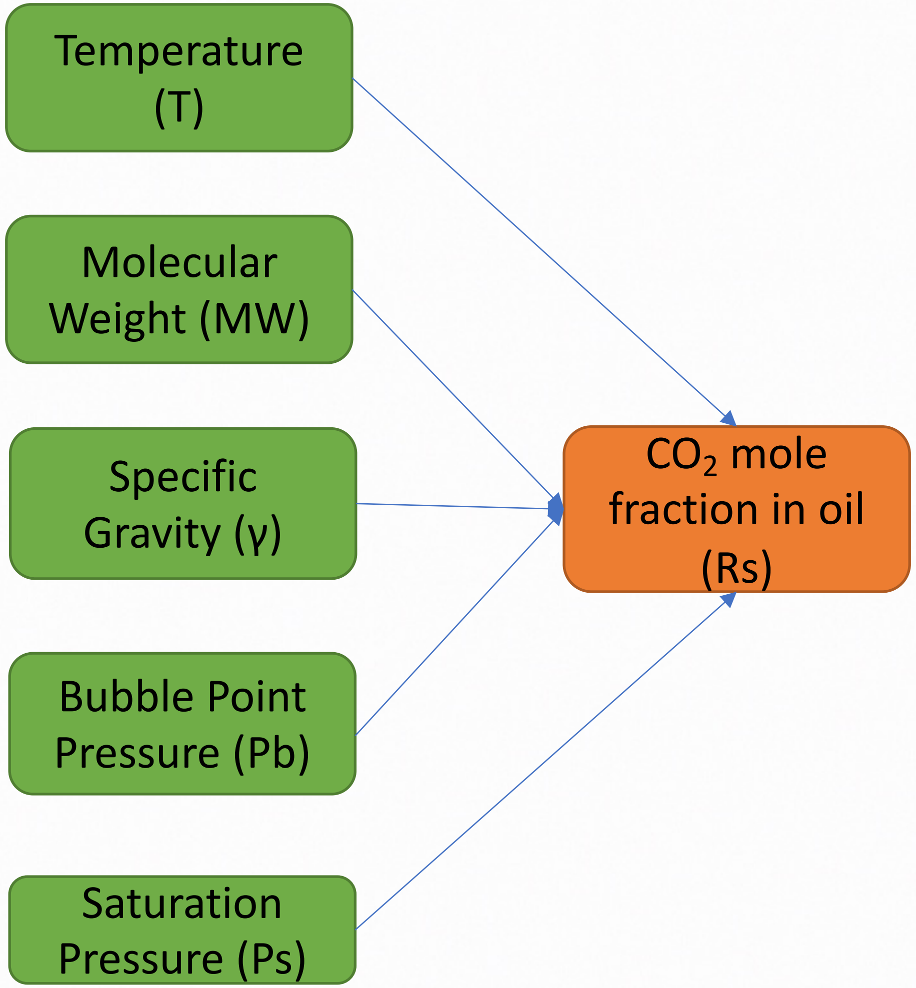
Model trained and validated with wider range of experimental data

Factors that affect solubility are taken into consideration

Interaction between parameters are considered

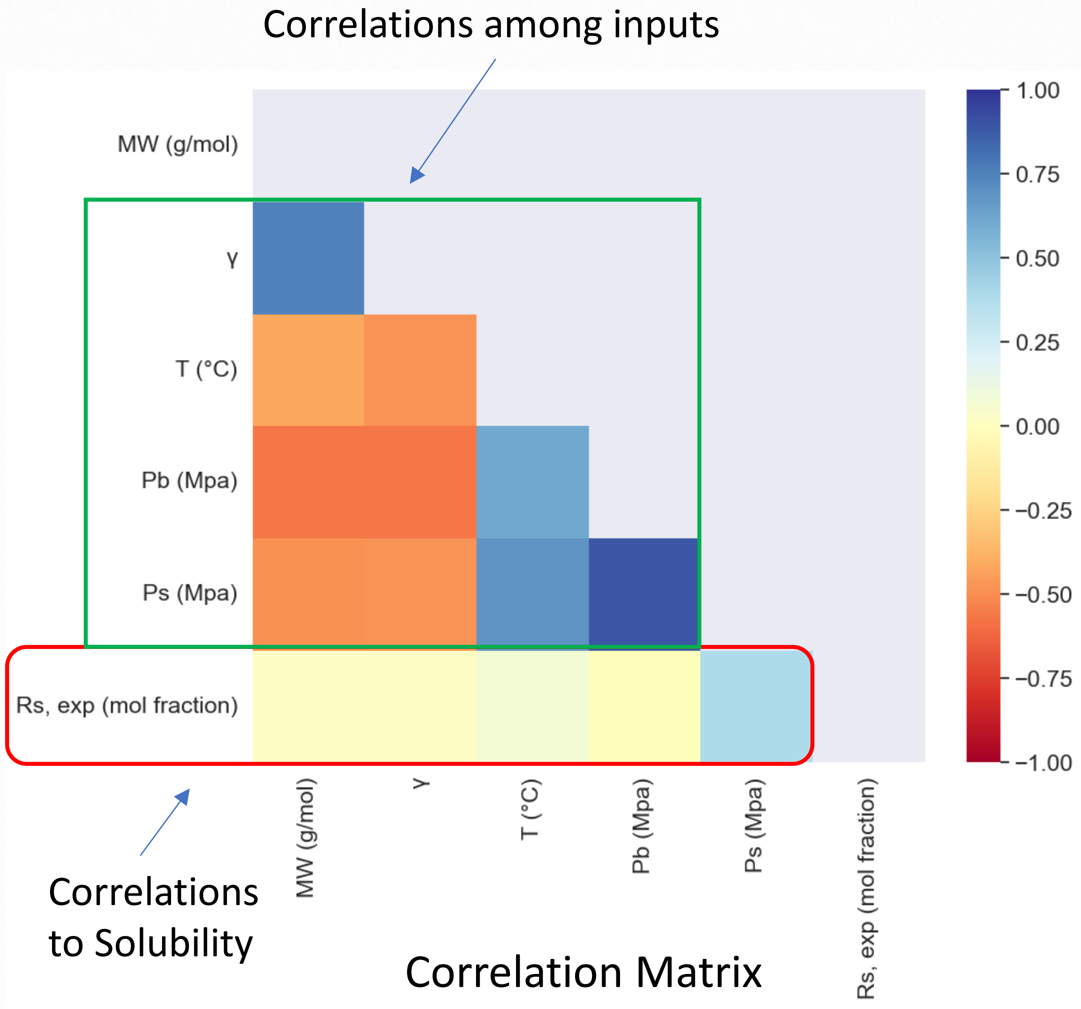
Data Overview

- 106 experimental datapoints for live oil system
- 74 datapoints for dead oil
- Temperature ranged from 28°C to 140°C
- Pressure from 3 MPa to 30 MPa



What affects solubility & how? (Live Oil)

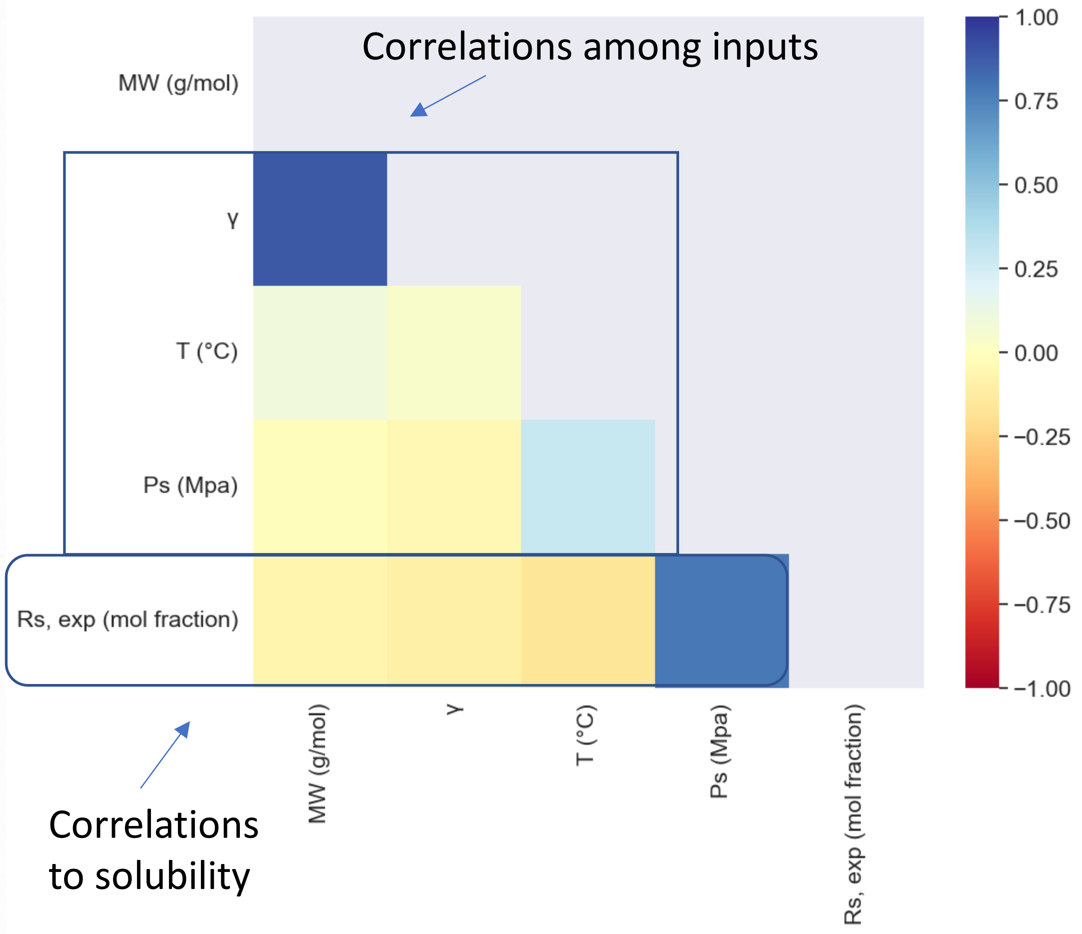
- None of the parameters have any strong linear relationship with solubility
- Input parameters are correlated with each other
  - Lighter oils also have higher bubble point pressure
- Model must consider interaction among input parameters



Correlation Matrix

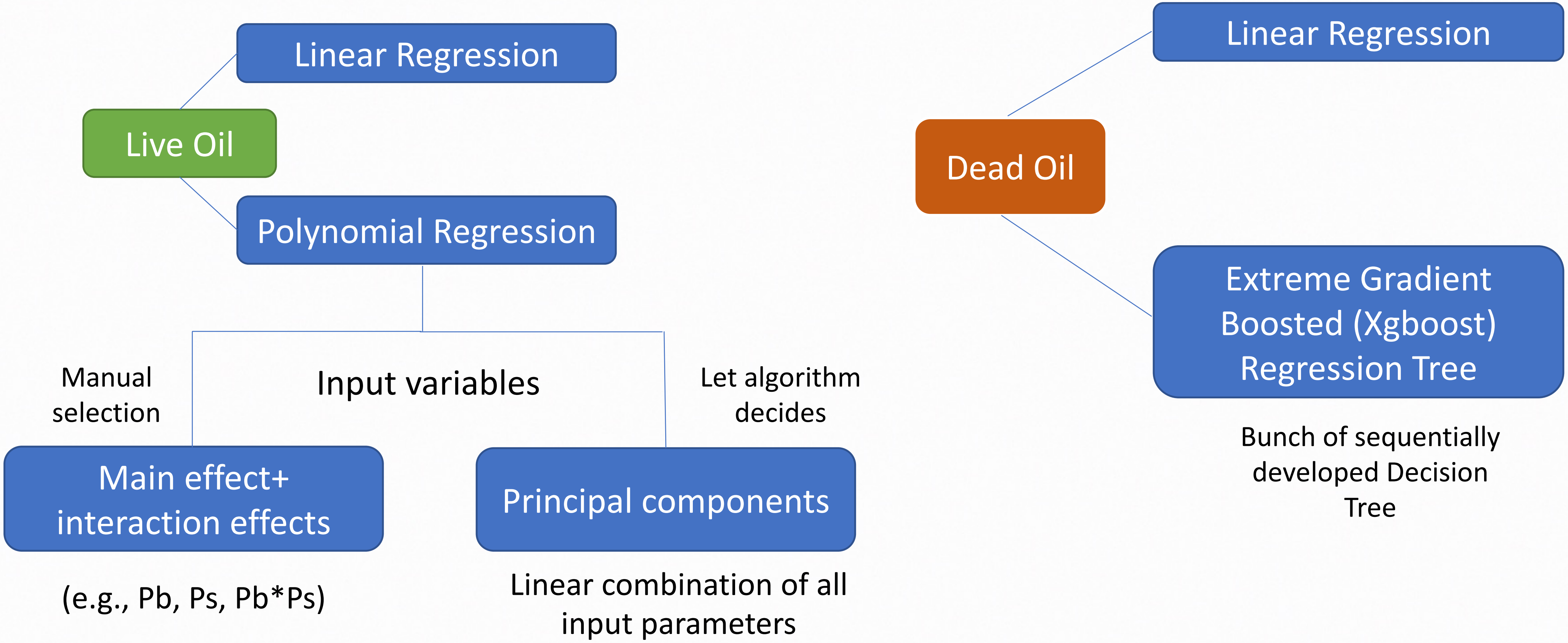
What affects solubility & How? (Dead Oil)

- Saturation pressure has stronger effect on solubility for dead oil than live oil
- No strong interaction between input parameters except for MW and specific gravity
- May need to use a non-parametric model



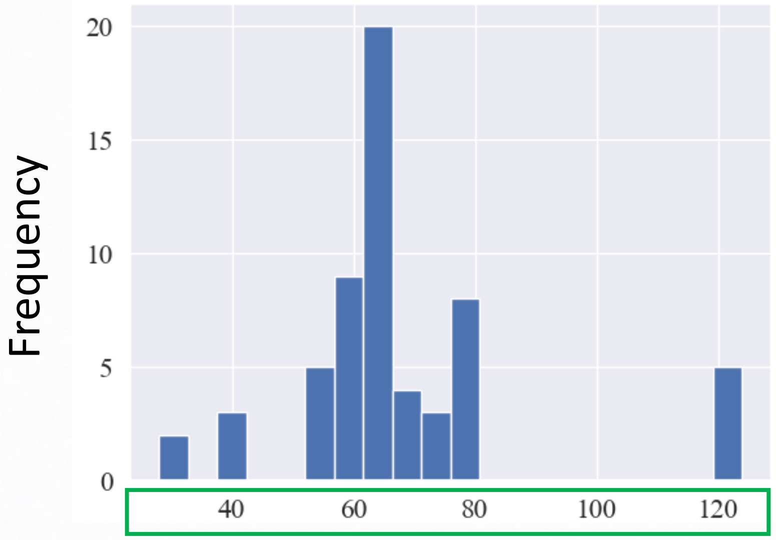
SECARB: Offshore

Model Selection

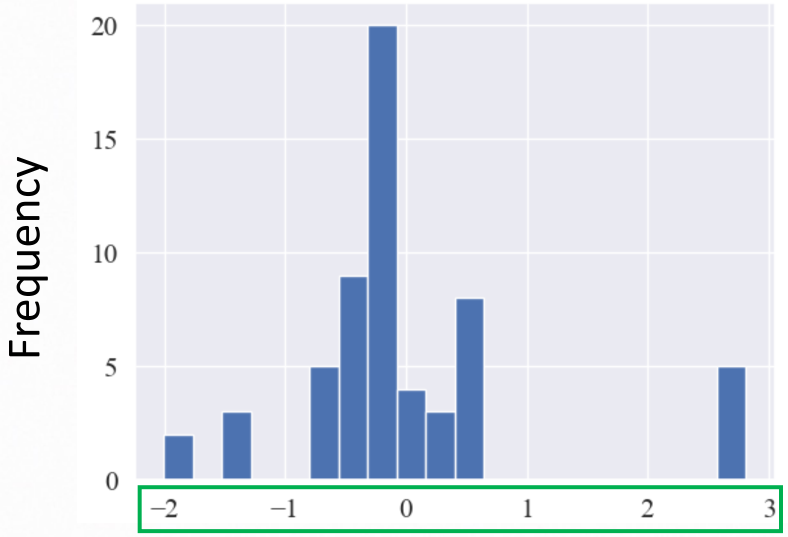


Data Preprocessing

- Scaling
  - Transforming data to have mean of zero and standard deviation of one.
  - Original features have data in different units/scales.
  - Transformed features will have comparable units of measurement.
- Creating principal components
  - Transforming input variables to their principal components.
  - Principal components are linear combination of original features.
  - Features are combined based on how well they can explain total variance in data.



Temperature (°C) before scaling

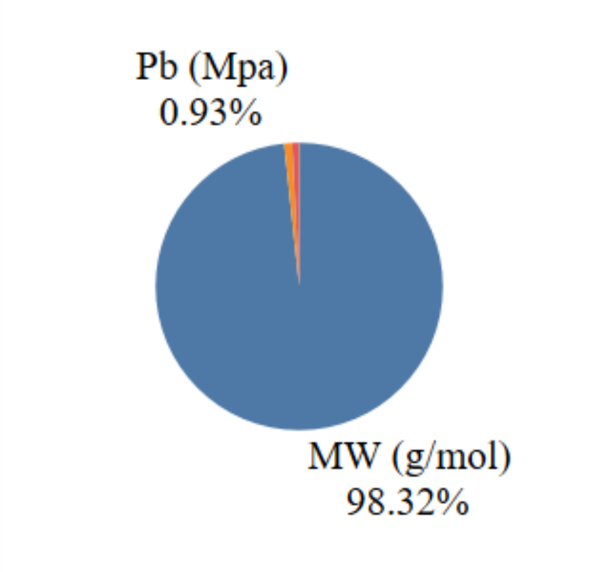


Temperature (°C) after scaling

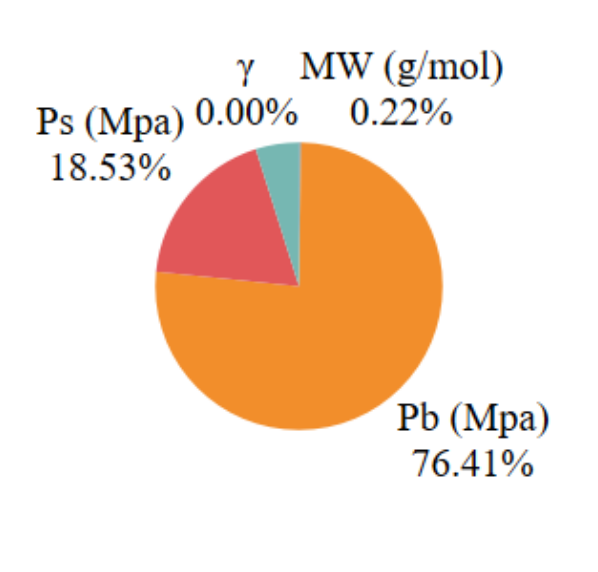
Principal Components

- Each components are linear combination of original variables.
- Four principal components explain 95% of total variance.

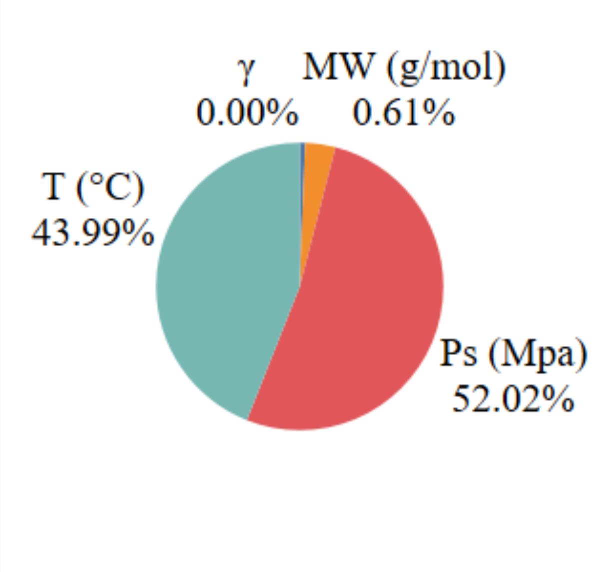
<u>How the original variables contribute to the variance explained by principal components</u>



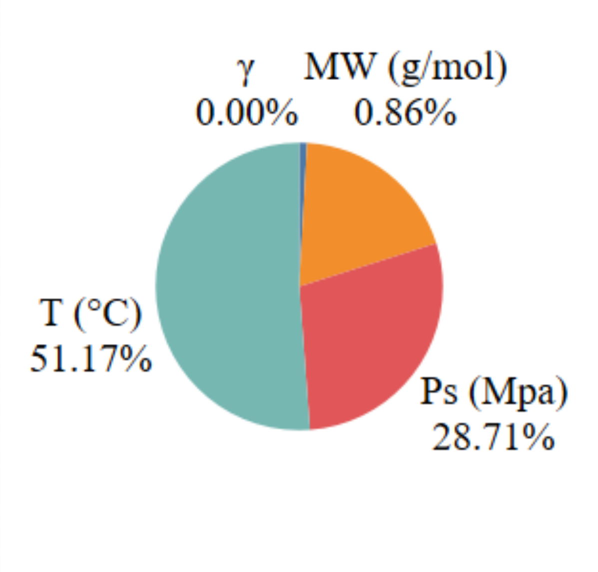
Component 1 (55%)



Component 2 (32%)

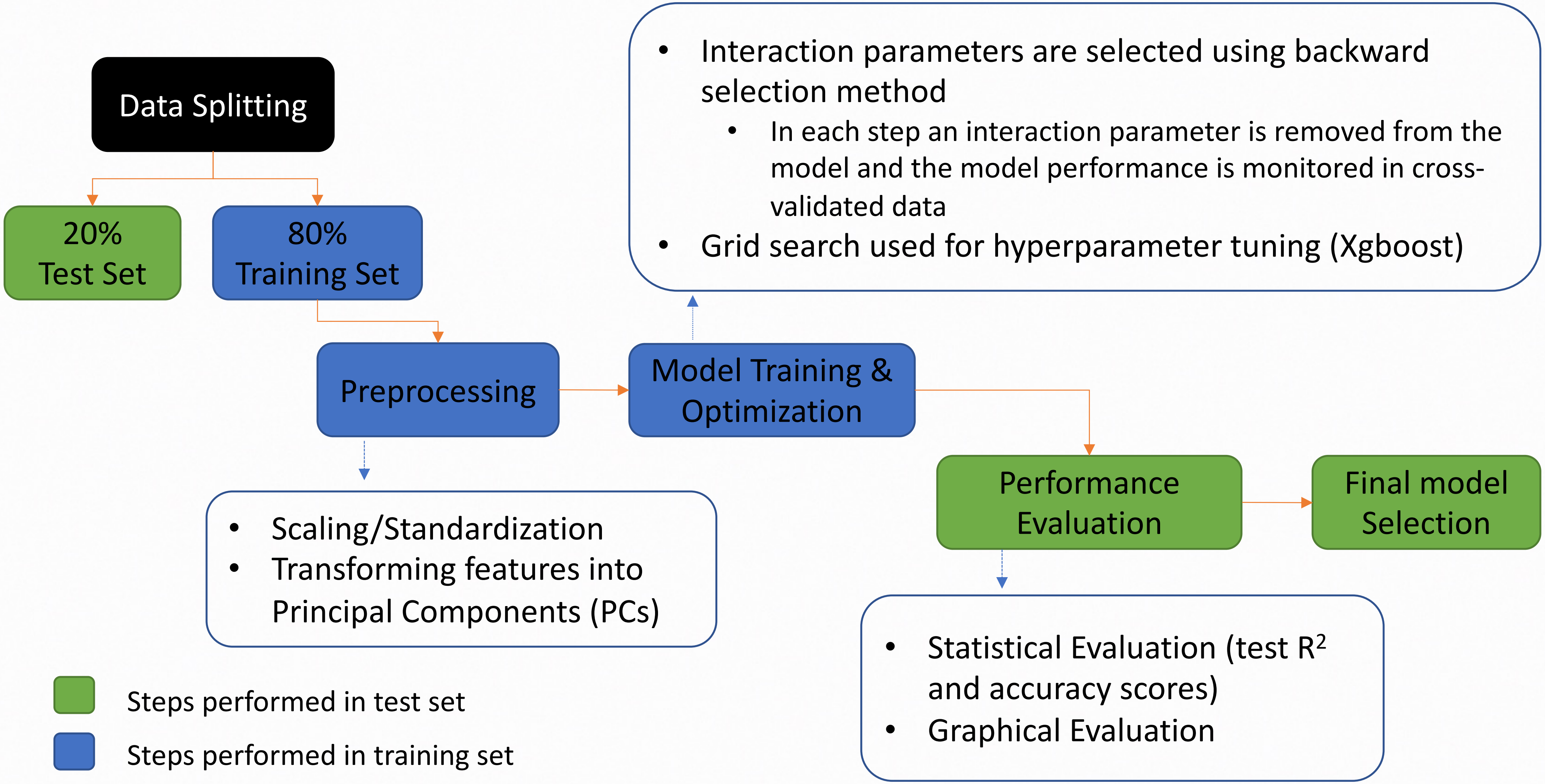


Component 3 (5%)



Component 4 (3%)

Model Construction Work-Flow

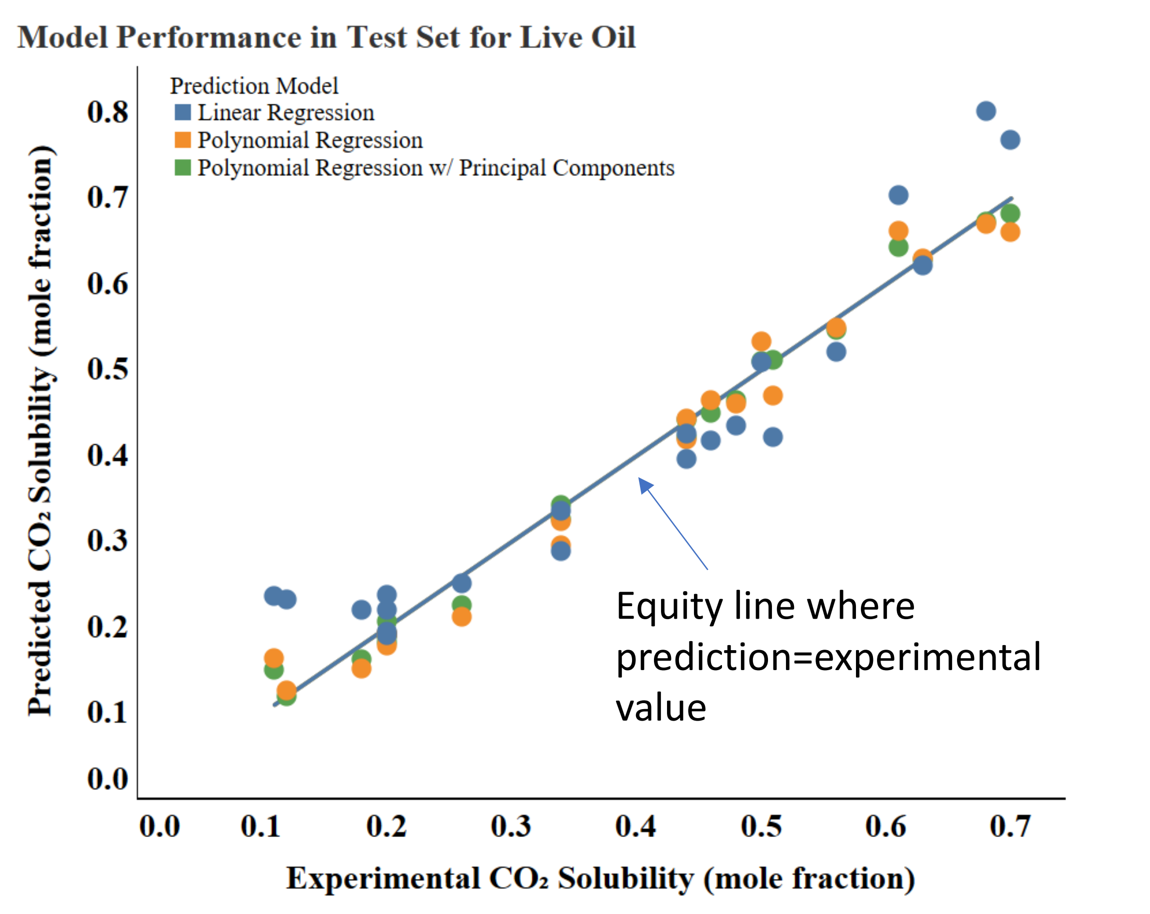


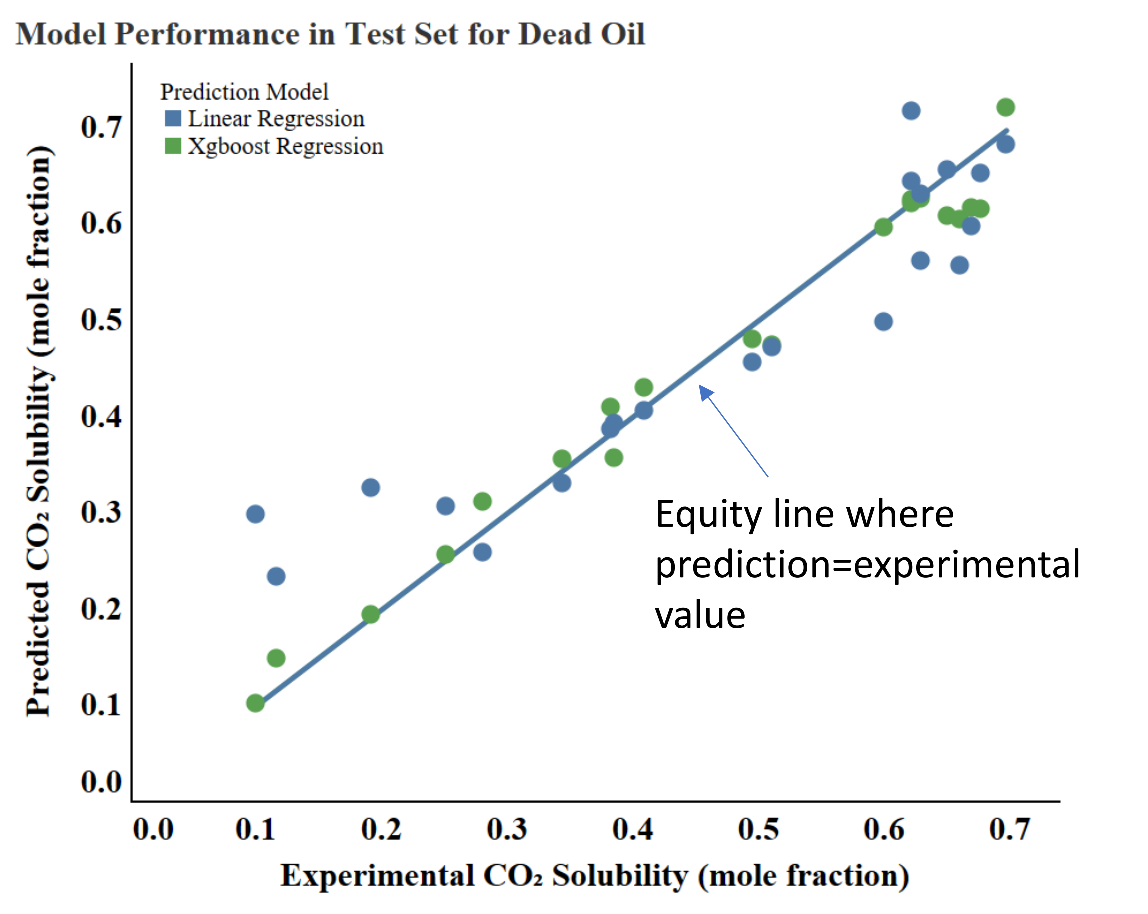
Model Evaluation: Statistical Comparison

		R^2		RMSE	
Oil Type	**Model**	Cross Validation	Test	Cross Validation	Test
Live Oil	Linear	0.84	0.81	0.04	0.06
	Polynomial with original features	0.96	0.96	0.02	0.03
	Polynomial with PCs	0.98	0.98	0.002	0.01
Dead Oil	Linear	0.76	0.75	0.08	0.08
	Xgboost	0.98	0.96	0.04	0.03

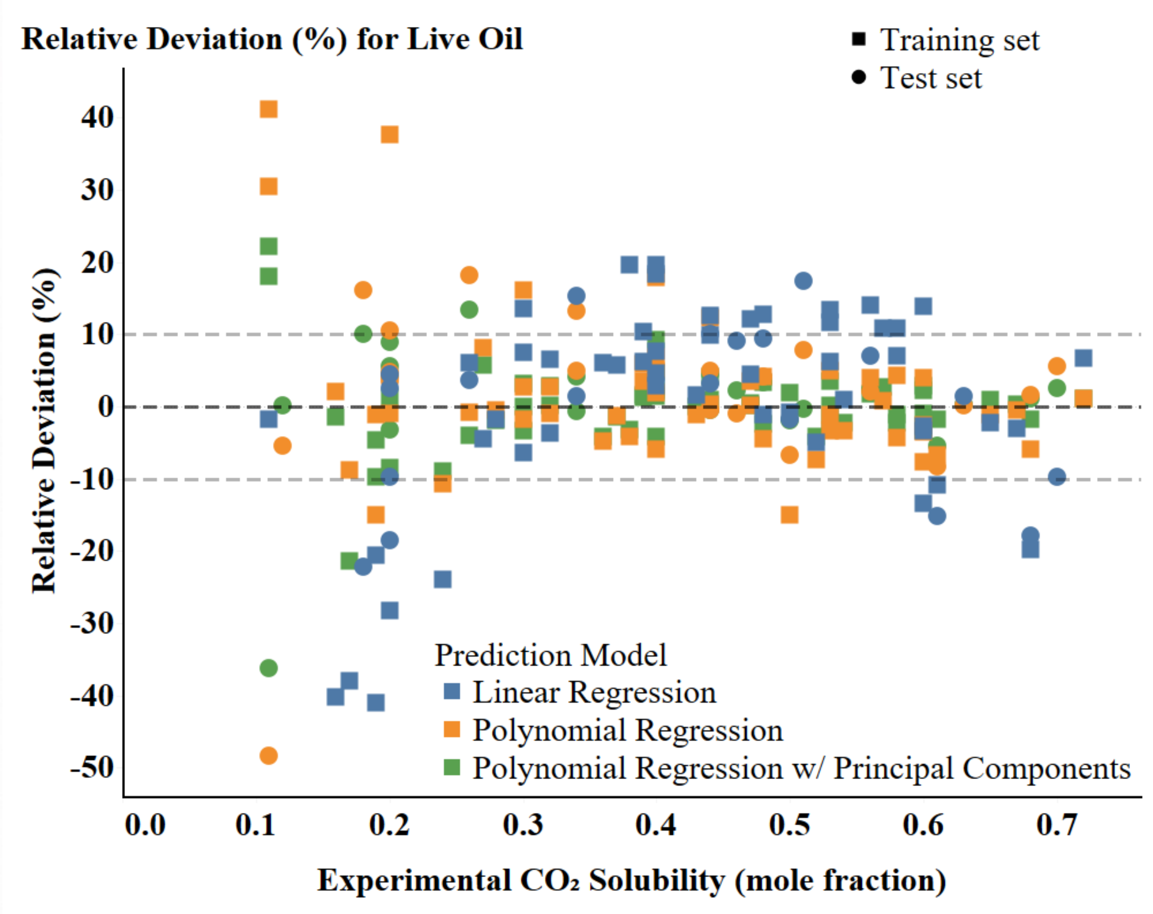
- For live oil polynomial regression with PCs gives most accurate prediction.
- Has less interpretability than polynomial model with original features.
- For dead oil Xgboost is the best performing model.

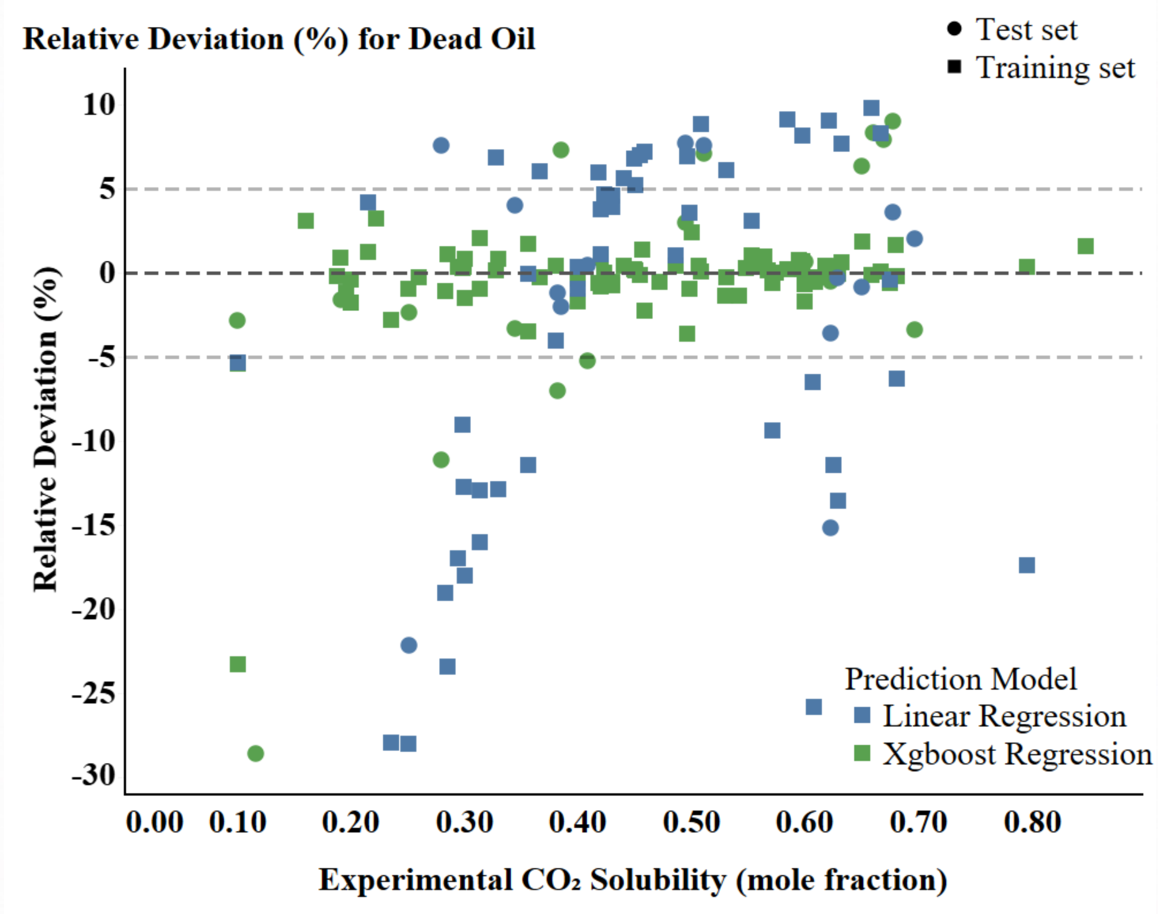
Graphical Comparison





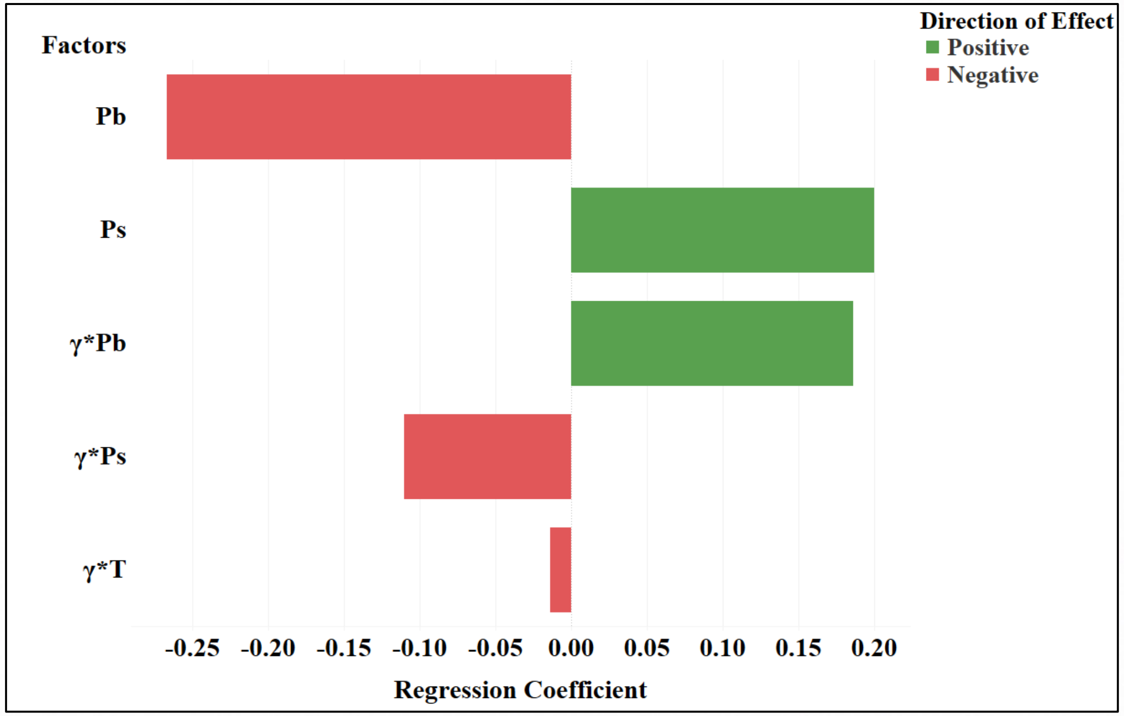
Error Analysis





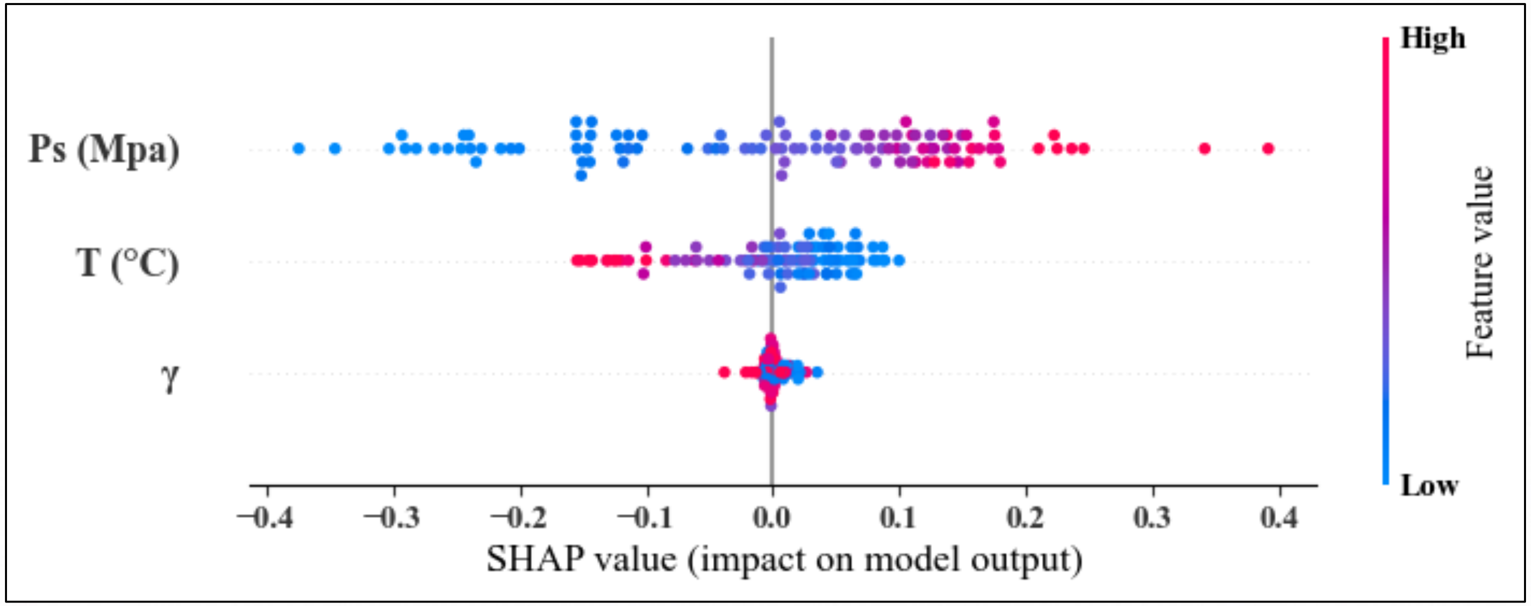
Sensitivity Analysis (Live Oil)

- Polynomial model with original features are used for interpretation
- Oil with high bubble point pressure has less solubility
- CO_2 Solubility increases with saturation pressure
- Interaction of bubble point and saturation pressure with specific gravity also affect solubility



Sensitivity analysis of the factors affecting CO_2 solubility in live oil

Sensitivity Analysis (Dead Oil)



- SHAP values are feature's contribution to individual predictions
- Higher saturation pressure leads to higher solubility
- Temperature affects solubility inversely
- Effect of specific gravity is less significant than other two parameters
  - Solubility increases as specific gravity decreases

Conclusion

- Machine learning is utilized to predict CO_2 solubility in complex multivariate system of dead oil & live oil
- Interaction effect between parameters are considered
- Principal component analysis is utilized to improve the predictivity of the polynomial model
- The developed models outperformed the benchmark model
- The predictions are in great match with experimental data
  - Within 10% (for live oil) and 5% (for dead oil) of relative deviation from experimental data

Thank You..