

# Machine Learning for Mixed Fluid Dynamics

Lara Arikan

Livia Fulchignoni

Prof. Daniel Tartakovsky

Energy Resources Engineering

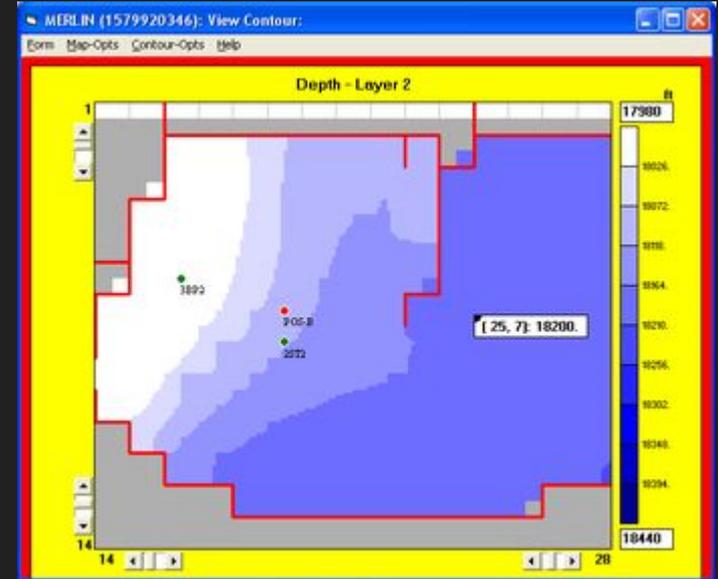
How can environmental conditions for a mixed fluid help predict its thermophysical properties?

# Outline

- Quick background
- Process:
  - Data mining
  - K-nearest neighbors classifier
- Interpretation of results
- Future plans & conclusions

# Background information

- Two properties:
  - Condensate formation volume factor ( $B_o$ )
  - Solution gas water ratio ( $R_s$ )
- Black Oil equations
  - Describe fluid flow in a petroleum reservoir
  - Needed to simulate fluid behavior
  - Used for production forecasts, transport strategies



Above: sample reservoir simulation

# Data mining

- Measurements of environmental conditions and thermophysical properties in Word documents
- No consistent format
- Solution:
- File conversion (.doc to .docx)
- Extraction only from specific tables
- Database creation

```
Subfolder: data_docs\sub1
Next: data_docs\sub1\new new
To convert: out0.docx
To convert: PVT_report_example3.docx
To convert: report_1.docx
To convert: RJS-704 CT_TR_049_13.docx
To convert: RJS-704 CT_TR_080_13.docx
Processing: out0.docx
Initialize name_to_value_dict
Processing table <docx.table.Table object at 0x0FDE9210>
Processing table <docx.table.Table object at 0x0FDE91F0>
Update name_to_value_dict
Processing table <docx.table.Table object at 0x0FDDEC90>
Update name_to_value_dict
Processing table <docx.table.Table object at 0x0FDDEFD0>
Update name_to_value_dict
Processing table <docx.table.Table object at 0x106A2A50>
Processing table <docx.table.Table object at 0x106A2E10>
Update name_to_value_dict
Processing table <docx.table.Table object at 0x106A2E30>
Update file_to_key_dict
Processing: PVT_report_example3.docx
```

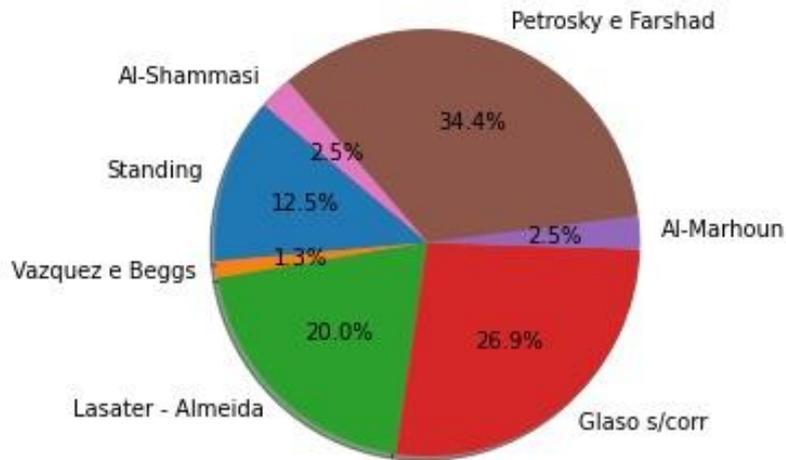
# K-nearest neighbors classifier

- Determines the class of a new data point according to its k nearest neighbors
- Data: API, P, dgs, T etc.
- 5-fold cross validation (20% of dataset tested at a time)
- For Bo: `Highest accuracy: 0.5700636942675159 for k-value 6`
- For Rs: `Highest accuracy: 0.717948717948718 for k-value 13`
- Up from 14.2% accuracy of random selection (7 models)

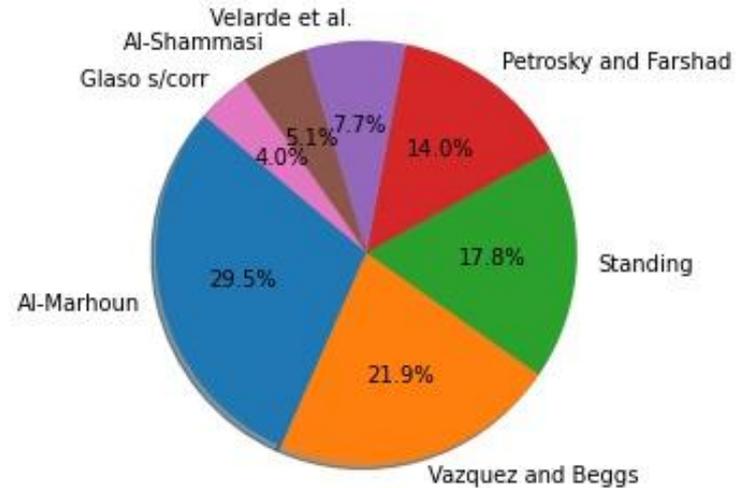
# Interpretation

Little overlap between best equations for Rs(left) and Bo(right): each separate property merits investigation

Best equations by frequency

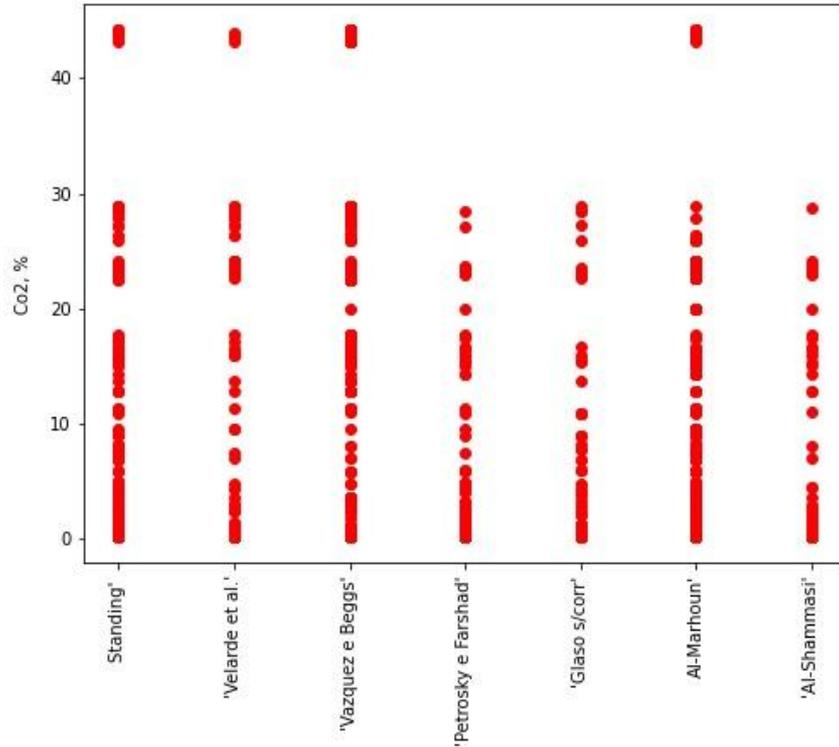


Best equations by frequency

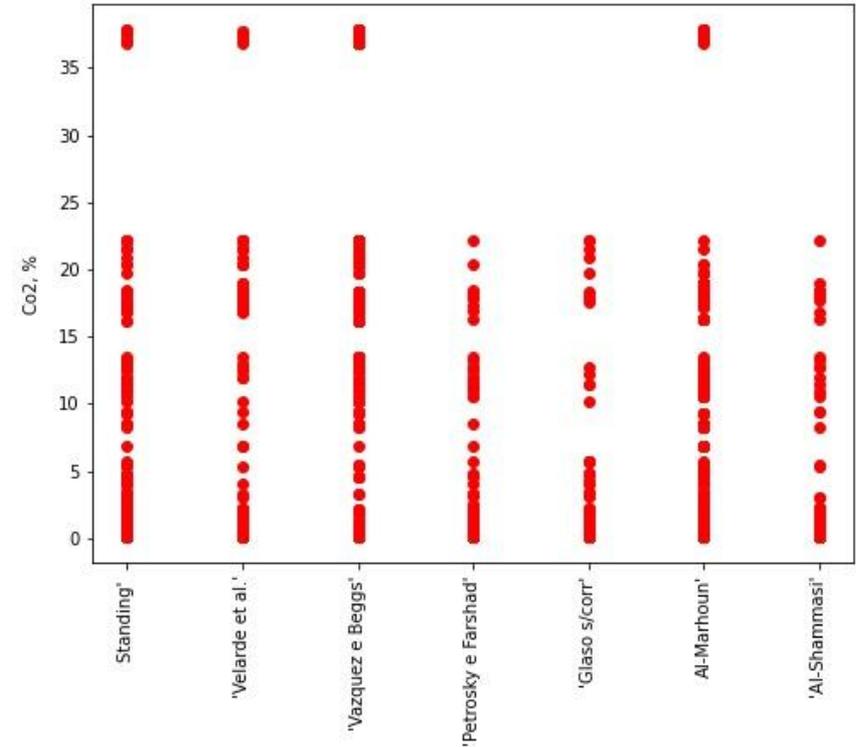


For Bo: Same four equations (Standing, Velarde, Vazquez, Al-Marhoun) best for a wider range of three parameters (especially high values)

Best equations by CO2 fraction in the gas phase

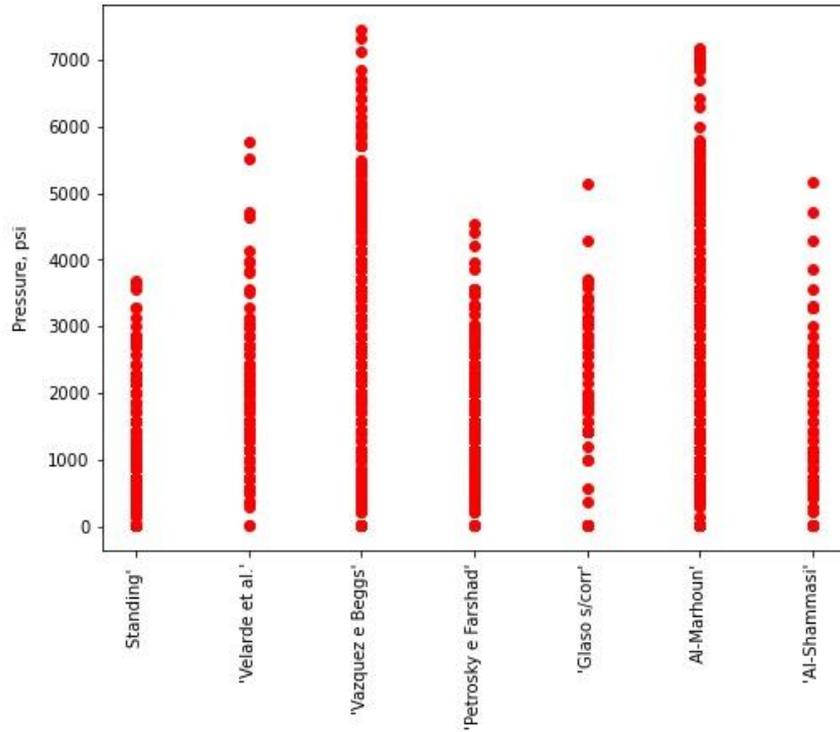


Best equations by carbon dioxide percentage in reservoir

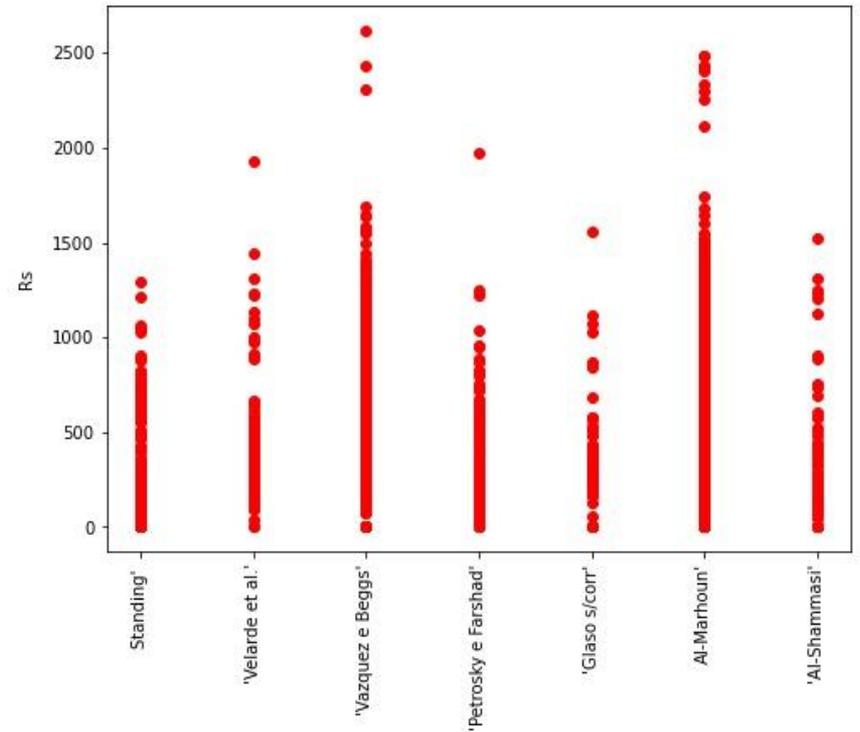


- Al-Marhoun, Vazquez uniquely good for high pressure and Rs

Best equations by pressure

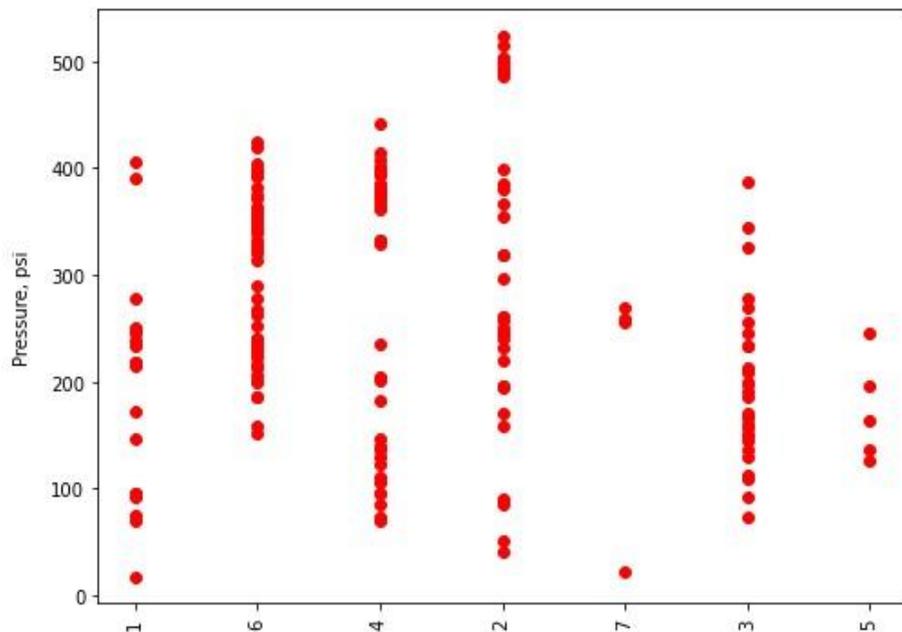


Best equations by Rs

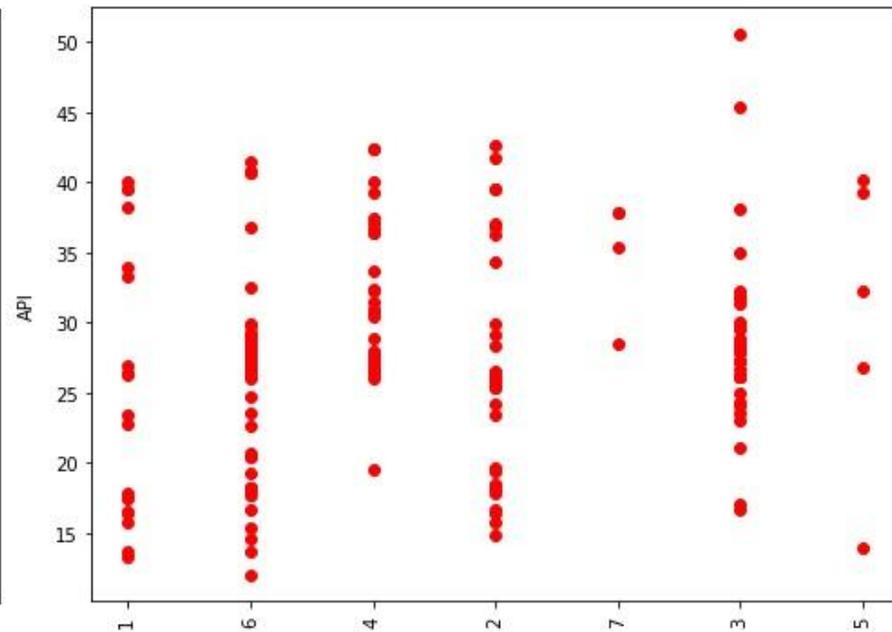


For Rs, little difference between models for each property  $\rightarrow$  multi-label classifier will improve accuracy

Best equations by pressure



Best equations by API



# Future plans

- Multi-class classification neural network
- Regression
- Consider more fluid properties (Z)
- Expand data-scraping script to review papers in a database

# Conclusions

- Important to choose best model for thermophysical properties under varying conditions:
  - Safe transport
  - Environmental and industrial applications
- Used K-nearest neighbors classifier with fair accuracy
- Other classification or regression methods may be more accurate or appropriate.