Imperial College London

Applying Computational & Data Science – mini project 2: From Core to Computation: Deep and Machine Learning Strategies for Geological Data Interpretation in Arcadia

The latest exploration in the Avalon Basin, situated in the offshore region of the mythical Arcadia, is focussed on the prolific Paleocene deep-marine strata within Quadrants 204 and 205. This study area includes significant oil-producing fields, along with an additional non-reservoir well (204/24-6), marking a key area of petroleum exploration.

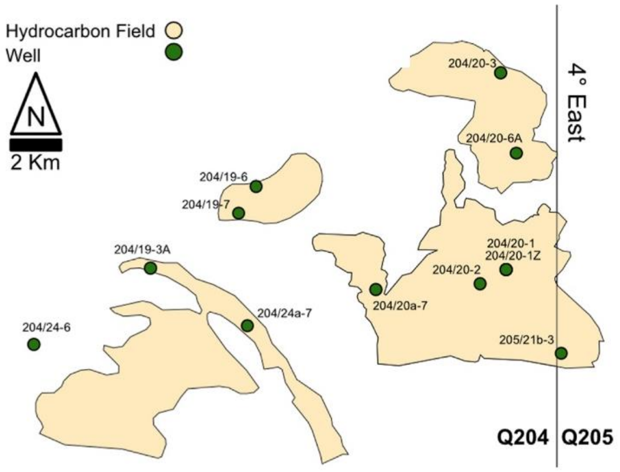

Figure 1. Map of Arcadia locating 11 wells that contain legacy wireline-log and core data.

The Avalon Basin is a NW-SE-oriented rift basin that developed through various rift phases (i.e. phases of extension) during the Paleozoic to Mesozoic eras. The latest rift phase in the Neo-Paleocene epoch led to the formation of a deepwater basin, in which water was over 200 m in depth.

The region's stratigraphy is highlighted by the mythical Elysium Formation, laid down during the Neo-Paleocene period. This formation comprises a series of deep-marine siliciclastic submarine fan deposits with varied facies, ranging from sand-rich to mud-rich. Reservoir properties in the Elysium formation are idealized for educational purposes, with approximate porosities and permeabilities conducive to various exploration scenarios.

As part of the evaluation, it is necessary to construct a consistent facies scheme that can be used to characterise the distribution of porosity and permeability using core and wireline-log data.

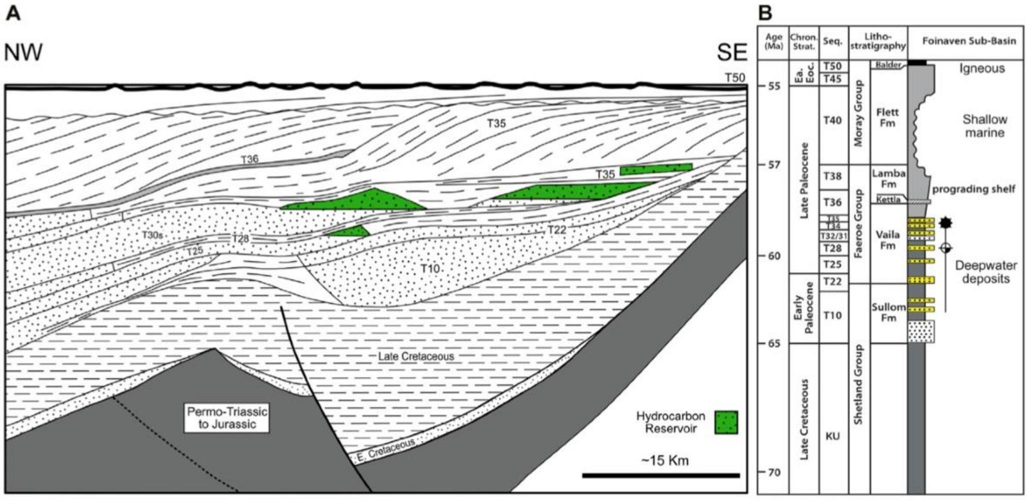

Figure 2. Diagrammatic representation of the hydrocarbon reservoirs and adjacent overburden and underburden strata

Materials

You are provided with the following materials:

1. map locating 11 wells (Figure 1);
2. wireline log data covering the area in these wells;
3. core photographs;
4. Core-plug porosity and permeability data and Facies Classifications from several wells (Table 1);
5. Diagrammatic representation of the hydrocarbon reservoirs and adjacent strata (Figure 2).

Well name	Labelled lithology	GR	PEF	SP	RESD	NPOR	DTC	RHOB
204-19-3a	78	●	●	●	●	●	●	●
204-19-6	13	●	●	●	●	●	●	●
204-19-7	72	●	●	●	●	●	●	●
204-20-1	67	●	●	●	●	●	●	●
204-20-1Z	12	●	●		●	●	●	
204-20-2	20	●	●	●	●	●	●	●
204-20-3	79	●	●	●	●	●	●	●
204-20-6a	26	●	●		●	●	●	●
204-20a-7	20	●	●	●	●	●	●	●
204-24a-6	91	●	●		●	●	●	●
204-24a-7	153	●	●		●	●	●	●

Table 1. Well-log and core dataset. Amount of labelled core in meters

Deliverables:

1. **Report – 1 page: 1 page of text/equations and up to 5 figures.**
   - A one-page report outlining your methodology, findings, and interpretations. Include:
     - Analysis of well log and core data.
     - Application and results of machine learning models.
2. **Presentation – 15 minutes:**

- A concise, engaging presentation of your key findings and insights. Include visual aids like charts, maps, and images to illustrate your points.

3. **Python Module and Visualization Toolkit**:
    - **Module Inclusions**:
        - **Classifier Model**: A model designed to predict facies classes.
        - **Porosity Predictor Model**: A dedicated model for predicting porosity values.
        - **Visualization Tool**: A utility to effectively visualize wire log data, integrating porosity predictions and facies classifications for each well.
    - **Practical Demonstration**:
        - **Jupyter Notebook Submission**: Along with the module, submit a Jupyter Notebook that demonstrates the practical application of each model component. This should include:
            - Importing your custom module.
            - Showcasing the functionality of the classifier and predictor models.
            - Illustrating how the visualization tool can be used to represent the blinded well data (that we release on Friday), porosity predictions, and facies classifications.

Evaluation Criteria:

- Thoroughness of Analysis: Depth and accuracy in analysing the provided data.
- Application of Machine Learning: Your final classification will be tested on an unseen core image.
- Clarity of Presentation and Report: How well the team communicates their findings and methodologies.
- Sustainability of your codes: Submit all code used for analysis and machine learning models in your github repo and follow the instruction for sustainability.

Hints:

- You are free to use any tools to explore and preprocess your core images.
  For example:
  [rgmyr/corebreakout: Segmentation and depth-alignment of geological core sample image columns via Mask-RCNN (github.com)]
  This is a well-documented code that provides two primary features. Firstly, it offers a deep learning process that converts raw images of core sample boxes into datasets registered by depth, using the CoreSegmenter API. Secondly, it includes a CoreColumn data structure, designed for the storage and handling of image data that is registered by depth.
  Since your core images are similar to the ones used to train this model, there is no need to retrain this model with your data and you can just use the saved model and weights to create your unlabelled images.

  Alternatively, you can use other tools for preprocessing your data like [Segment Anything | Meta AI (segment-anything.com)]

- After preprocessing your data apply a deep learning model to predict facies classes. (You are free to use wireline log data, to create new synthetic data set or just to use core images with facies classes provided for you for each well.
- To deal with imbalance target you are free to re-classify your facies based on your domain knowledge. Do not forget to provide your new class labels.
- You should not have less than 4 classes.

- Bellow you can find the definition of acronyms used for Facies classes:

- nc is no core

- s is sand

- os is oil stained

- bs is the same as nc

- is and ih are the two intermediate messy shale ones

Data Alignment

There are discrepancies between the core and wireline log data depths. To align these measurements, a certain value (in metres) must be either added to or subtracted from the core depth to match the log depth (in metres).

- 204/19-6: no shift between core and log depths
- 204/19-7: add 1.8 m to the core depths for interval 2131.00-2157.68 m (core #2), add 1.6 m to the core depths for interval 2158.00-2180.38 m (core #3), add 3.0 m to the core depths for interval 2540.00-2544.89 m (core #4)
- 204/20-1: add 3.0 m to the core depths for interval 1945.00-2116.00 m (cores #1-#6)
- 204/20-1Z: add 5.0 m to the core depths for interval 2673.00-2686.46 m (core #1)
- 204/20-2: add 2.0 m to the core depths for interval 1998.00-2018.55 m (core #1)
- 204/20-3: subtract 1.2 m from the core depths for interval 2401.00-2425.75 m (core #1), subtract 2.7 m from the core depths for interval 2427.50-2436.83 m (core #2), add 1.2 m to the core depths for interval 2658.00-2685.79 m (core #3), add 1.6 m to the core depths for interval 2958.00-2978.15 m (core #4)
- 204/20-6A: no shift between core and log depths
- 204/20-7: no shift between core and log depths
- 204/24-6: add 3.5 m to the core depths for interval 2211.00-2230.70 m (core #1), add 2.0 m to the core depths for interval 2416.00-2429.67 m (core #2), add 2.43 m to the core depths for interval 2429.67-2438.96 m (core #3), add 0.8 m to the core depths for interval 2440.20.00-2462.93 m (core #4), no shift between core and log depths for interval 2463.00-2485.60 m (core #5), add 0.4 m to the core depths for interval 2485.60-2493.50 m (core #6),
- 204/24-7: add 2.2 m to the core depths for interval 2075.00-2102.00 m (core #1), add 2.35 m to the core depths for interval 2102.00-2127.85 m (core #2), add 2.0 m to the core depths for interval 2128.50-2145.90 m (core #3), add 1.8 m to the core depths for interval 2146.00-2162.60 m (core #4), add 1.2 m to the core depths for interval 2163.00-2179.30 m (core

#5), add 1.0 m to the core depths for interval 2180.00-2188.56 m (core #6), add 0.85 m to the core depths for interval 2189.00-2206.92 m (core #7), add 0.2 m to the core depths for interval 2208.50-2240.65 m (cores #8-#9)