Imperial College London

29/Jan/2024 – 02/Feb/2024

From Core to Computation: Deep and Machine Learning Strategies for Geological Data Interpretation in Arcadia

Applying Computational & Data Science – mini project 2

Scale Space White City

Address: 58 Wood Ln, London W12 7RZ

Imperial College London

Objectives & Learning outcomes

Objectives:

- To simulate applied computational and data science in the real world.
- To synthesise knowledge from the taught modules
- To apply techniques learned in the course to real problems
- To develop collaborative programming skills
- To reinforce best practise for software development

Learning outcomes:

- Plan and produce software collaboratively.
- Collaboratively solve problems using data and software.
- Summarise work using collaborative presentations
- Skilfully process and interpret geological datasets like core images and wireline logs.
- Construct and evaluate DL/ML models for geological feature prediction.

Imperial College London

Timetable & Spaces

	Monday	Tuesday	Wednesday	Thursday	Friday
am	Lecture/ Group assignment/ Q&A	Group working	Group working 1by1 Q&A	Group working	1 pm Code Deadline
pm	Group working	Group working 1by1 Q&A	Group working	Group working 1by1 Q&A	4:30 pm report and presentation
					Social event with pizza and drink - 17-20

Commitment:

- We estimate that each participant will dedicate approximately 45 hours to this project over the course of the week.
- This breaks down to around 9 hours per day.

Imperial College London

Timetable & Spaces

Room Allocation:

- Each group is assigned a private room from Tuesday to Friday.
- Room names are listed on your desks and in the 'groups.xlsx' file.
- The rooms are booked for your use from **8:00 AM to 8:00 PM daily**.

Key Collection & Return:

- Collect the key for your room daily.
- Return it to the reception at the end of each day.
- In each group, two members (highlighted in your group details) are responsible for this task.

Flexible Hours:

- Standard time: 9:00 AM - 6:00 PM.
- Flexible options: 10:00 AM - 7:00 PM, 8:00 AM - 5:00 PM, 11:00 AM - 8:00 PM, etc.
- Coordinate within your group for optimal timing.

Imperial College London

Important space usage Guidelines

Room Access:

- Use only the room allocated to your group.
- Do not use or enter rooms not assigned to you, as this disrupts other members' rightful usage.

Professional Conduct:

- Avoid behaviours like running in the building or blocking passages, to maintain a professional environment.

Meeting Room Access:

- Be aware of the designated persons in your group who holds the access cards.
- Avoid gathering in large groups outside rooms or lounges.

Imperial College London

Support


Cai Shengjuan


Marijan


Zayad


Parastoo


Robert

- Your primary support network is each other! Collaborate and engage within your group for solutions.
- A private Teams channel is set up for each group. Use this channel to ask us questions and for further assistance.
- We have allocated 30-minute meetings for each group. Meetings available on Tuesday, Wednesday, and Thursday for in-person or video calls. This is your chance to get direct feedback and answers.
- For quick queries, use the Teams chat. We'll respond as promptly as possible.
- GTAs will also be available daily to address your queries.
- Questions we won’t answer: Help with debugging your code!

Imperial College London

Your Feedback

From last year and mini project 1

Difficulties in working with ACSE students due to the large skill gap and the project having a heavy focus on coding

A preference for working with subsurface geoscience or energy related data

Assessment should consider the real problem context, not just model accuracy

More training on handling complex merge conflicts in GitHub

Spend less time on lecturing more on 1 by 1 meetings

We prefer smaller and private rooms

Avalon Basin Exploration

Exploration Focus:

• The Avalon Basin in the mythical Arcadia, covering the Paleocene deep-marine strata in Quadrants 204 and 205.
• Includes significant oil fields and a non-reservoir well (204/24-6).

Geological Background:

• The Avalon Basin is a NW-SE-oriented rift basin.
• Formed through rift phases from the Paleozoic to Mesozoic eras.
• The latest phase in the Neo-Paleocene epoch created a deepwater basin (over 200m depth).

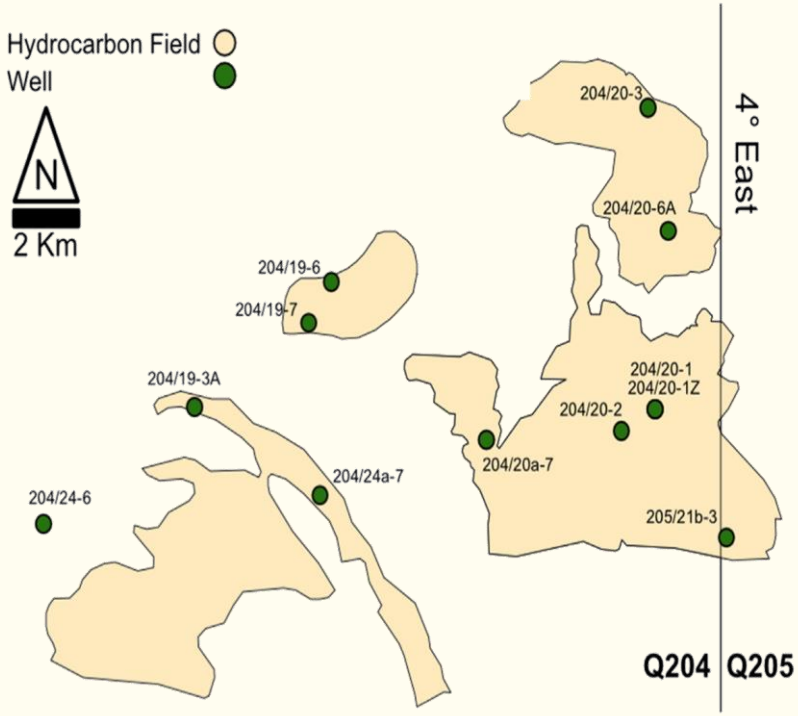

Figure 1. Map of Arcadia locating 11 wells that contain legacy wireline-log and core data.

Imperial College London

Project Overview

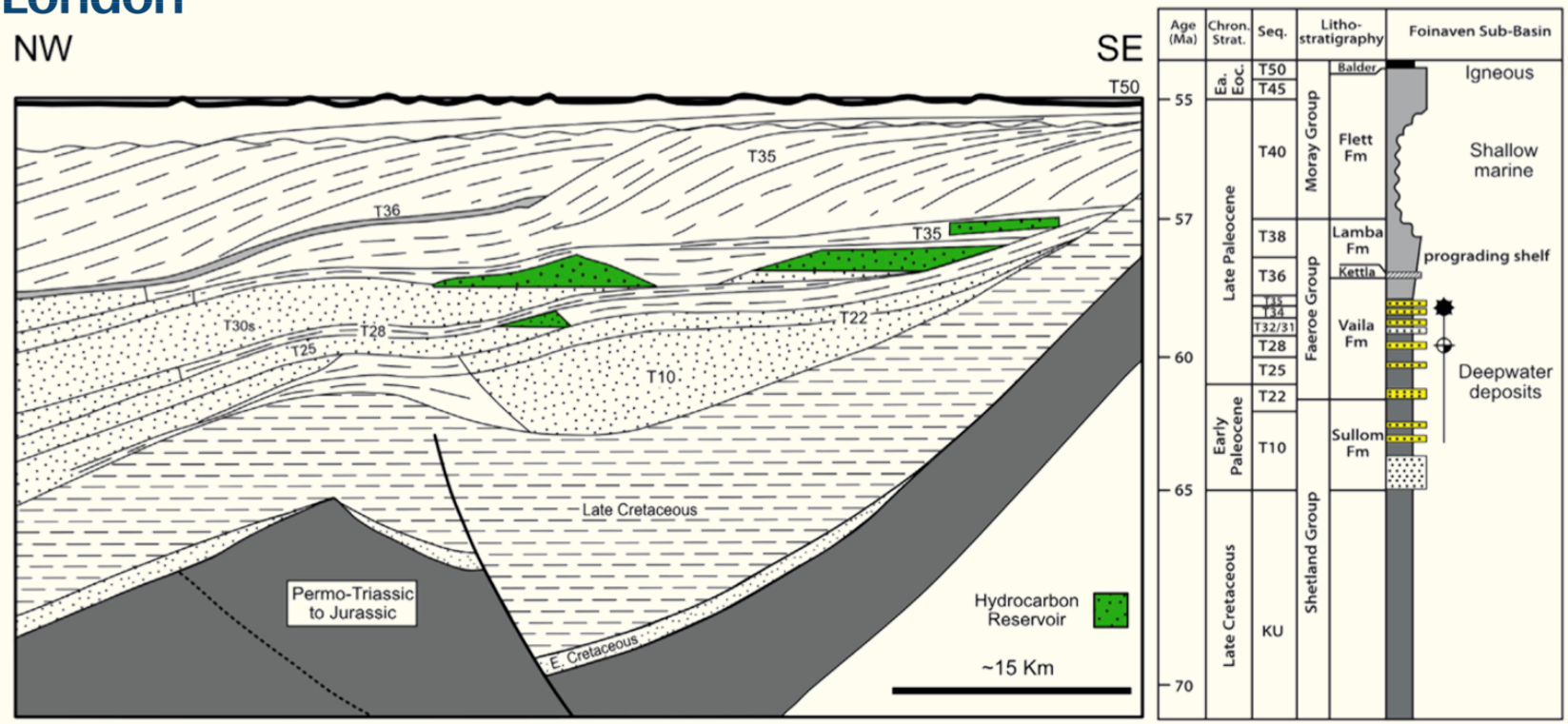

Figure 2. Diagrammatic representation of the hydrocarbon reservoirs and adjacent overburden and underburden strata

Stratigraphy and Formation:

- Comprises deep-marine siliciclastic submarine fan deposits.
- Varied facies: from sand-rich to mud-rich.
- Idealized reservoir properties for educational purposes.

Imperial College London

Materials

find more details in your repo

Geographical Context:

- Map displaying the location of eleven wells in the Avalon Basin -- Figure 1 and `*Loc_well.jpg*`.
- Diagrammatic representation of the hydrocarbon reservoirs and adjacent strata -- Figure 2 and `*Formation.jpg*`

Wireline Log Data and Core images:

- Comprehensive wireline log data covering the designated wells -- `*data/wirelineLog/nameofthewell.las*`
- Core images with their depth -- `*data/coreimage/nameofwell/nameofcore.jpg*`

And `*data/coreimage/nameofwell/depth. xlsx*`

Labels:

- Core-plug porosity and permeability data for several wells -- `*data/labels/permeability/nameofwell.csv*`
- Facies classes for each well -- `*data/labels/faciesclass/nameofwell.csv*`

Imperial College London

Deliverables

1. Report -- 1 page including text, equations, and a maximum of 5 figures.

- Outline your methodology.
- Present your findings and interpretations.

Focus on:

- Analysis of well log and core data.
- Application and results of DL/ML models.

2. Presentation -- 15-minute presentation.

- Highlight your main findings and insights.
- Use visual aids such as charts, maps, and images for clarity and impact.

Recording & Submission:

- Record your presentation using tools like MS Teams or PowerPoint.
- Upload the recording to your group channel on Teams.

Imperial College London

Deliverables

3. Python Module and Visualization Toolkit:

Module Components: (a python package)

- Classifier Model: Designed for predicting facies classes.
- Porosity Predictor Model: Focused on estimating porosity values.
- Plotting Tool: For effective display of wire log data, combined with porosity predictions and facies classifications.

Practical Demonstration: (a Jupyter Notebook)

- Submit a Jupyter Notebook that exhibits the practical use of each module component.
- Essential Inclusions:
  - Module Integration: Demonstrate how to import and utilize your custom module.
  - Model Functionality: Display how both the classifier and predictor models operate.
  - Visualization: Show the use of your visualization tool to represent:
    - Blinded well data (to be released on Friday).
    - Porosity predictions and facies classifications for each well.

Imperial College London

Deliverables

3. Python Module and Visualization Toolkit:

```
GeoPredictors
 ├── __init__.py
 ├── classifier.py
 ├── predictor.py
 ├── visualizer.py
 └── extras
      ├── __init__.py
      ├── __main__.py
      └── preprocess.py
requirements.txt
environment.yml
setup.py
README.md
```

```
GeoPredictors
 ├── __init__.py
 ├── preprocess.py
 └── models.py

requirements.txt
environment.yml
setup.py
LICENSE
README.md
```

Imperial College London

Evaluation Criteria

Software (60%):

- Skill of prediction tool;
- Quality/utility of software;
- Quality of data visualization &
- Analysis;
- Speed of operation (in training and in prediction).

Presentation/Report (20%):

- Understanding & creativity;
- Technical description of methods/ choices;
- Utility of demonstration;
- Connection to geological interpretation.

Sustainability (10%):

- Reusability & extendibility of your package;
- Documentation ;
- GitHub usage.

Teamwork (10%):

- Individual contribution;
- Peer-evaluation.

Imperial College London

Hints and helps!

1- Core Image Processing Tools:
Choose any suitable tools for exploring and preprocessing your core images.

• **Suggested Tool - rgmyr/corebreakout:**
Features: Image segmentation, depth-alignment using Mask-RCNN, CoreSegmenter API, and CoreColumn data structure.
Advantage: Pre-trained model suitable for your data

• **Alternative Tool - Segment Anything | Meta AI**

2- 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).

3- Reclassification of Facies:

- You have the liberty to reclassify facies based on domain knowledge.
- Ensure you clearly define and provide your new class labels.
- You should not have less than 4 classes

Imperial College London

Hints and helps!

IMPORTANT - Preprocessed Data Release:

Preprocessed Dataset Availability:

A preprocessed version of core images, depth, and facies classes will be released on Tuesday noon.

Dataset Details:

Includes 1 image (n, m, 3), 1 depth array (n,), and 1 label array (n,) per well.

Usage Consideration:

Opting to use this preprocessed data will result in a deduction of marks from your software component.

Imperial College London

Question

Imperial College London

Academic integrity and use of AI

• Each project will provide specific guidance about which aspects of the software must be written from scratch and which aspects can make use of packages or code written by others.

• Plagiarism must be avoided. If you use code sourced outside your group, you must include clear and proper attribution of credit.

We provide you with a references.md file in your repo. Update this file with any books, websites or AI tools you use to help you in writing this code. For tutorial sites, the main URL or landing page you used will suffice. For Q&A sites (e.g. Stack Overflow/Stack Exchange), or when interacting with ChatGPT please link the specific question/response which helped you.

• In group projects, collusion constitutes the sharing of work between groups. You are therefore strongly discouraged from discussing the group project with peers outside your project group.

• For group projects, you are actively encouraged to work collaboratively as a team sharing ideas and code within your team is exactly what you should be doing and is not collusion.

• As a guiding principle: always acknowledge the contributions of others in your work, and do not leave yourself open to allegations that you have supplied answers to enable another student or group to commit academic misconduct.

Imperial College London

Supporting materials

- Additional to the data folder that I described the details here. You are provided with:

1. Project Description (description.pdf)
2. This PowerPoint Presentation (slides.ppt)
3. GitHub Repository README
4. Group and space details (group.xlsx)

- To further support your project work, we have provided a range of additional materials.

5. ACDS Handbook 2023 (ACDS-Handbook-2023.pdf)
6. ACDS Sustainability Practices (acds-sustainability.ppt):
7. Jupyter Notebooks and Version Control (Jupyter notebooks and version control.ppt):

Imperial College London

Supporting materials

Peer evaluation:

- You'll be asked to divide a set of points between the other members of your team based on your assessment of their contribution.

- Contributions include:

- ❑ • Coding
- ❑ • Research
- ❑ • Quality Assurance (e.g. bug finding/testing/documentation)
- ❑ • Project management, & anything else which helps the project succeed

- Full details will be sent out at the end of the week

- If you don't contribute, you may have your mark reduced