**Context**

Reservoir simulation is a mathematical model used to predict the flow of fluids in a reservoir over time. It's used to predict the performance of the reservoir and optimize the extraction process. In the home task, you are provided with synthetic reservoir data containing multiple reservoir simulations with different activation times of injector well I3**.**

A reservoir is a dynamic system, meaning that its properties change over time due to fluid flow, pressure changes, and external factors like well interventions. To accurately model a reservoir, it is essential to account for these dynamic changes.

Injection wells inject water into the reservoir to maintain pressure and enhance oil recovery. The connectivity between injection and production wells is vital for understanding the impact of injected water on oil production in nearby wells.

As a reservoir is a closed system influenced by wells, events in one well will have a delayed impact on nearby wells. For instance, increasing the injection rate in an injector well may result in increased reservoir pressure, potentially leading to higher oil production rates in nearby production wells. However, this relationship may not be straightforward, and analyzing the simulation data is crucial for better understanding this interaction. In data science terms, increasing water injection in the injection well will eventually reach the production well and alter the oil flowrate with a certain time lag.

Adding more production or injection wells to a reservoir can affect flow rates and pressure distribution within the reservoir, potentially altering the behavior of existing wells. This should be considered when modeling reservoir dynamics. For example, activating an injection well and injecting water into it will increase the pressure around that well. If the pressure front propagates to a production well, it can lead to increased oil flow rates for the producing well (unless there is a water breakthrough).

Now that you have a basic understanding of these concepts, you can proceed with the home task.

## Task

Your task is **to come up with the methodology to select features and evaluate time lags which should be used to build a robust predictive model for forecasting** the flow rates of wells P1 and P2 for a 3-step horizon using provided reservoir simulation data. You may choose to forecast oil or liquid flowrates.

In the context of the home task, understanding this dynamic connectivity between injection and production wells will be essential for building a robust predictive model for forecasting the flow rates of wells P1 and P2. By analyzing the simulation data and considering factors such as distance between producers and injectors, and time of producing or injection, you can improve your model's ability to account for the impact of injection well activation on the flow rates of nearby production wells.

You may consider utilizing the following ideas to improve your solution to the problem (you do not have to implement these ideas in your solution; feel free to use whatever you think is appropriate):

* Spatial arrangement between producer and injector wells. Geostatistical methods can be used to analyze the spatial relationship between wells and improve model predictions.
* The distance between injector and producer wells, as well as the injection and production rates, can impact reservoir performance.
* Temporal patterns in production and injection data.
* Feature engineering of production and injection data.
* Change or transform the target.

## Data

The folder contains a series of computational experiments results, each consisting of 18 simulations of a synthetic oil reservoir. In the simulations, the injector well I3 is activated at different timesteps. The value in the data file's name corresponds to the simulation step when the injector well I3 was activated. Some numbers are missing because simulations with such time lags caused the hydrodynamic simulator to fail. Each simulation takes into account the underlying physics of underground flow, such as fluid flow equations, pressure gradients, and rock permeability.

## data.csv

The dataset data.csvcontains features related to production of 5 wells different wells. The features available in the dataset are as follows:

group - Type of well and can either be a producer or an injector.

cat - Well identifier.

date - Indicates the step of the simulation and can be considered as a timestamp.

water - Water flowrate produced/injected by a well. Flow rate is a measure of the volume of fluid (oil or water) passing through a well over a period of time. If the value is positive, it means that fluid (water or oil) is being injected into the well, while a negative value indicates that water is being extracted from the reservoir.

oil - Oil flowrate produced by a well. All values of oil are negative since the oil is extracted from the reservoir. An injector well cannot produce oil.

liquid - Total liquid flowrate produced by a well.

bhp - Bottomhole pressure of the well

status - This feature indicates whether the well is active or not. Even if there is a flow of liquid, if the value of status is 0, the well is not considered to be participating in the simulation.

start\_lag **-** The column represents the simulation timestep when the well I3 started operating. For instance the value 0 means that well I3 started operating immediately, while 2 introduces a lag of 2 steps.

coef - The value in this column reflects the pressure coefficient value, with a lower coefficient implying less water injection into the new well and a smaller impact on the field. This coefficient governs various forces associated with water injection in the new well I3.

is\_base - The feature signifies whether the data related to the base case simulation or not. Base case is the case in which I3 injection well was not started.

## meta.csv

The datasetmeta.csvcontains coordinates of 5 wells different wells:

cat - identifier of a well

x, y - well coordinates

**Requirements**

You are expected to address the following points in your solution:

1. Compare your model to a baseline model of your choice. Demonstrate statistically that your model outperforms the baseline.
2. Visualize the results of model predictions to provide insights into the data.
3. Propose improvements to your model. This should include any insights gained from the model and any recommendations for further enhancements.
4. Provide a concise summary of the model and the results at the end of the Jupyter notebook (1 paragraph of text in English).