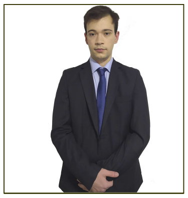
Team presentation



Juan Manuel Dri is a petroleum engineering student with a scholarship from Pampa Energía. He is doing an internship at FDC group, a consulting company in Argentina. During his last year of High school reached the final of OATEC. He really likes data science, machine learning and applying technology to solve problems in the O & G industry.

Santiago Lencinas Pupkiewicz, 23 years old advanced student of Petroleum Engineering at ITBA-Argentina, where the entire career was done with an Honor Scholarship. 2021 Petrobowl's National Champion and 2020 runner up. Passionate about Machine Learning, and the way it can improve the efficiency of the Oil and Gas operations. Currently doing an internship on FDC-Argentina, a reservoir engineering consultancy company.

Advanced Petroleum Engineer student at Instituto Tecnológico de Buenos Aires, Argentina and beneficiary of two scholarships from Chevron and Pan American Energy. Interested in state-of-the-art approaches to solve new problems and currently doing an internship in FDC a consulting company from the O&G industry.

Miranda, Lucas Lucas Miranda is a petroleum engineering student with an scholarship from Pan American Energy. Captain of the Petrobowl team of ITBA. Currently working in Tecpetrol as a Reservoir Engineering Intern. Interested in the use of Data Science to improve the Oil and Gas operations.

Agenda

- Targets & Tools
- Aproach
- Methodology
  1. Data Quality
  2. Validating sensors
  3. Correcting sensor data
- Final Remarks

Targets &Tools

The competition aims to address three tasks:

1. Detect data quality issues: In this task, the developed solution should be able detect the problems in drilling sensor data.
2. Assess data reliability: In this task, the solution should assign an uncertainty score for each sensor value. Based on the uncertainty score, ML models or field engineers can decide on whether to use such data for decision making.
3. Correct sensor data: The third task is to develop a way to correct the sensor data when data quality is poor or the uncertainty score is very low.

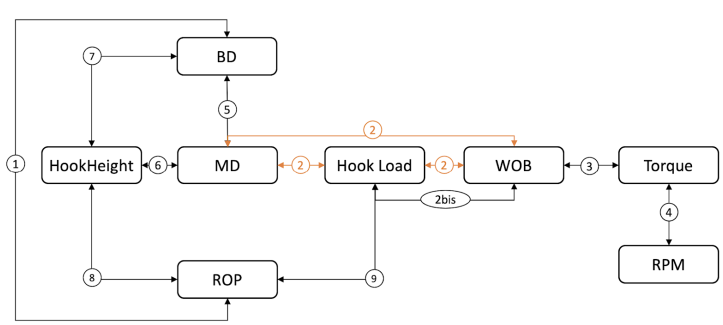
Tool used for the development:

We use python because it is a powerful, simple, and easy-to-learn tool. Besides, all the team members knew how to use it.

Approach

It was used a **bayesian network** approach to calculate all sensors using one as an and make all the same with all sensors. After the estimation of the nodes, all values are compared to make our analysis.

The network used was the one shown in the image below.



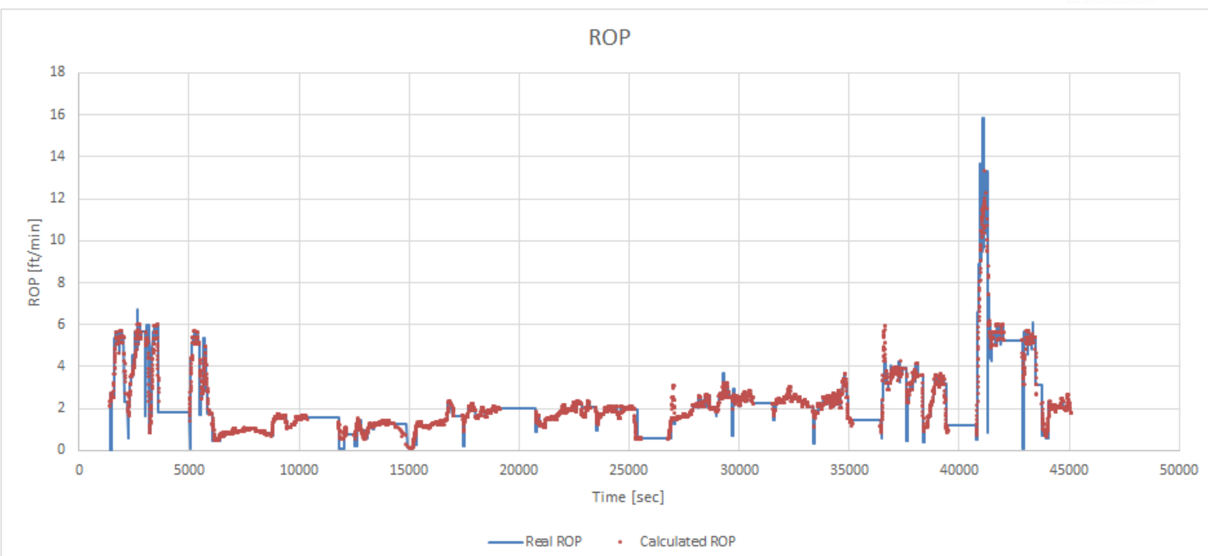
Example:

$$BD_i - BD_{i-1} = ROP_i \,.\, \Delta t \quad (Eq.\ 1)$$

Methodology: Data Quality

- Sensors may have different measure frequencies leading to some incomplete rows, so the first step is to check the incomplete rows.
- It was decided to fill the blanks with the average between the upper and lower data points in the column. Only the ROP had a particularly low frequency that makes impossible the use of an average value.
- In order to sort the ROP low frequency problem, the concept apllied was to calculate a pseudo-ROP by averaging the slope of the Bit_Depth_BD under time, i.e. the surface ROP. This pseudo-ROP is used only in the equation of the bayesian network and the REAL ROP values are used to assess the confidence of this sensor.

Pseudo-ROP



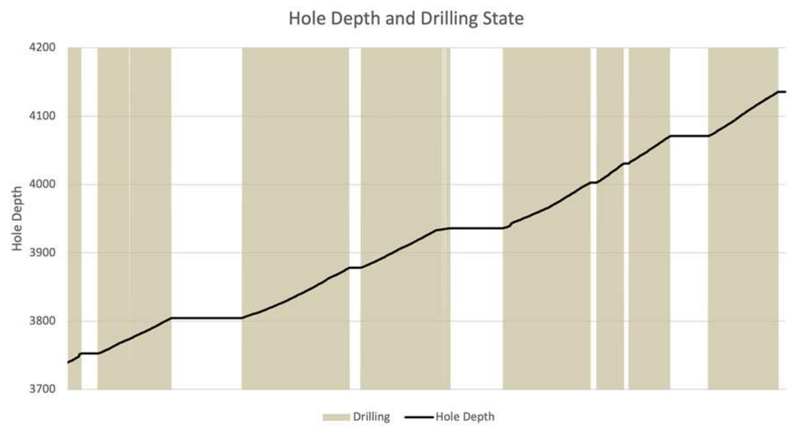
Methodology: Drilling State

Webinars

Determination of the drilling state was crucial so as to use the equation when the rig is drilling.

The drivers for these classification were:

- Hole_Depth_MD equal to Bit_Depth_BD, i.e. the bit is at the bottom of the hole
- Block_Position is lower than the previous Block_Position



SPE International

Validating sensors

Webinars

- A measured is considered as right and it's used to calculate the remaining values
- A matrix is created with every calculation

Fixed value

Calculated values

	ROP	MD	BD	HKHG	HKLD	WOB	TQ	RPM
ROP	6.012	3491.691	3491.691	77.113	236553.950	24314.032	754.109	46.933
MD	5.501	3491.855	3491.855	76.949	258530.947	2337.035	72.484	488.281
BD	5.501	3491.855	3491.855	76.949	258530.947	2337.035	72.484	488.281
HKHG	5.501	3491.888	3491.888	76.916	258530.947	2337.035	72.484	488.281
HKLD	5.661	3491.703	3491.703	77.101	251515.940	9352.042	290.057	122.019
WOB	5.661	3491.703	3491.703	77.101	251515.941	9352.041	290.057	122.019
TQ	5.759	3491.700	3491.700	77.104	255875.547	4992.434	154.842	228.572
RPM	5.658	3491.703	3491.703	77.101	251358.575	9509.407	294.938	120.000

SPE International

Validating sensors

- Once all calculation is done, a decision has to be made in order to validate or not the result.
- The algorithms compare all the values of a colum with the value of the diagonal at this given colum.
- Adaptable criteria for considering a calculation as valid

Criterial: User input

- ROP, Hook Load, Torque, RPM and WOB. In this case to validate the estimation if the **relative error** is less than 10%
- MD and BD. In this case we validate the estimation if the **absolute error** is less than 1 ft
- Hook Height. In this case we validate the estimation if the difference is less than 0.3 ft

	ROP	MD	BD	HKHG	HKLD	WOB	TQ	RPM
ROP	1.000	1.000	1.000	1.000	1.000	0.000	0.000	0.000
MD	1.000	1.000	1.000	1.000	1.000	0.000	0.000	0.000
BD	1.000	1.000	1.000	1.000	1.000	0.000	0.000	0.000
HKHG	1.000	1.000	1.000	1.000	1.000	0.000	0.000	0.000
HKLD	1.000	1.000	1.000	1.000	1.000	1.000	0.000	1.000
WOB	1.000	1.000	1.000	1.000	1.000	1.000	0.000	1.000
TQ	1.000	1.000	1.000	1.000	1.000	0.000	1.000	0.000
RPM	1.000	1.000	1.000	1.000	1.000	1.000	0.000	1.000

Validating sensors

- Measure Belief
  - Each sensor has a measure belief score between 0 and 1 that correspond to the average of all the values of the column as it is shown in the "Measure Belief" row, and the network validates the sensors that have an index greater than 0.57 (which corresponds to being validated by 4 or more of the others sensors).

	ROP	MD	BD	HKHG	HKLD	WOB	TQ	RPM	Evaluator Index
ROP	1.000	1.000	1.000	1.000	1.000	0.000	0.000	0.000	**0.57**
MD	1.000	1.000	1.000	1.000	1.000	0.000	0.000	0.000	**0.57**
BD	1.000	1.000	1.000	1.000	1.000	0.000	0.000	0.000	**0.57**
HKHG	1.000	1.000	1.000	1.000	1.000	0.000	0.000	0.000	**0.57**
HKLD	1.000	1.000	1.000	1.000	1.000	1.000	0.000	1.000	**0.86**
WOB	1.000	1.000	1.000	1.000	1.000	1.000	0.000	1.000	**0.86**
TQ	1.000	1.000	1.000	1.000	1.000	0.000	1.000	0.000	**0.71**
RPM	1.000	1.000	1.000	1.000	1.000	1.000	0.000	1.000	**0.86**
Measure Belief	**1.00**	**1.00**	**1.00**	**1.00**	**1.00**	**0.29**	**0.00**	**0.29**	-

Webinars

What do we do with the unreliable sensors?

SPE International

Correcting sensor data

Webinars

	ROP	MD	BD	HKHG	HKLD	WOB	TQ	RPM	**Evaluator Index**
ROP	1.000	1.000	1.000	1.000	1.000	0.000	0.000	0.000	**0.57**
MD	1.000	1.000	1.000	1.000	1.000	0.000	0.000	0.000	**0.57**
BD	1.000	1.000	1.000	1.000	1.000	0.000	0.000	0.000	**0.57**
HKHG	1.000	1.000	1.000	1.000	1.000	0.000	0.000	0.000	**0.57**
HKLD	1.000	1.000	1.000	1.000	1.000	1.000	0.000	1.000	**0.86**
WOB	1.000	1.000	1.000	1.000	1.000	1.000	0.000	1.000	**0.86**
TQ	1.000	1.000	1.000	1.000	1.000	0.000	1.000	0.000	**0.71**
RPM	1.000	1.000	1.000	1.000	1.000	1.000	0.000	1.000	**0.86**
Measure Belief	**1.00**	**1.00**	**1.00**	**1.00**	**1.00**	**0.29**	**0.00**	**0.29**	-

Each column in the validation matrix contains the information about how likely it is that the measure is within the acceptance criteria.
Each row have the number of times the network was able to predict the other values

The Evaluator index was constructed for each row (each fixed sensor) as the ratio of 1's over 7; and it was assumed that a sensor is a good predictor if the index is above 0.57, which implies that it is capable of predicting the real value of the vast majority of the sensors.

SPE International

Correcting sensor data

Webinars

The proposed value is calculated as the weighted average of the predicted values for the unvalidated sensor and the evaluator index for the sensors who are good predictors.

	ROP	MD	BD	HKHG	HKLD	WOB	TQ	RPM	Evaluator Index
ROP	6.012	3491.691	3491.691	77.113	236553.950	24314.032	754.109	46.933	0.57
MD	5.501	3491.855	3491.855	76.949	258530.947	2337.035	72.484	488.281	0.57
BD	5.501	3491.855	3491.855	76.949	258530.947	2337.035	72.484	488.281	0.57
HKHG	5.501	3491.888	3491.888	76.916	258530.947	2337.035	72.484	488.281	0.57
HKLD	5.661	3491.703	3491.703	77.101	251515.940	9352.042	290.057	122.019	0.86
WOB	5.661	3491.703	3491.703	77.101	251515.941	9352.041	290.057	122.019	0.86
TQ	5.759	3491.700	3491.700	77.104	255875.547	4992.434	154.842	228.572	0.71
RPM	5.658	3491.703	3491.703	77.101	251358.575	9509.407	294.938	120.000	0.86
Measure Belief	**1.00**	**1.00**	**1.00**	**1.00**	**1.00**	**0.29**	**0.00**	**0.29**	-

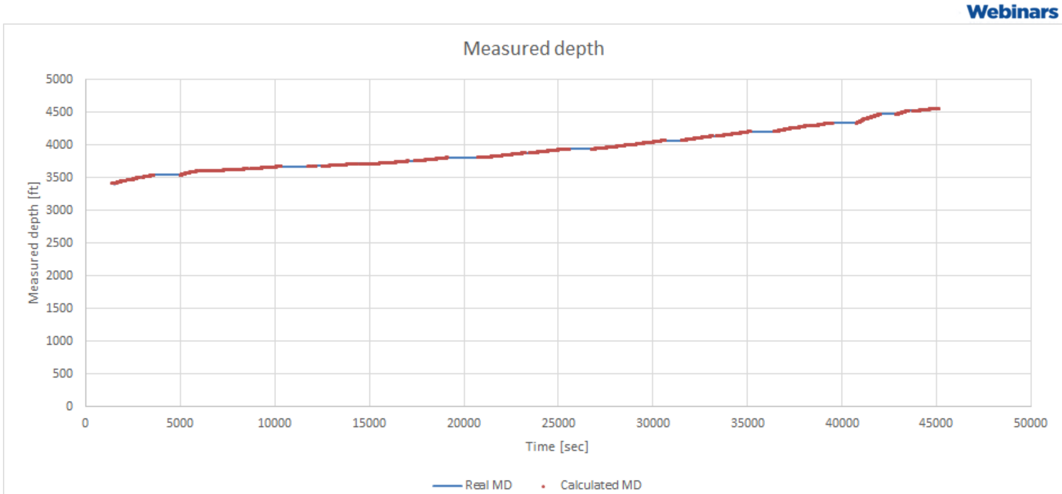
Correction
13893.73
1335.45
1335.45
1335.45
8016.04
-
3566.02
8150.92
7982.77

These calculations are done for every time step. So, the validation process for every sensor and the proposal of new values when it is required is performed for every single row of the datasheet.

SPE International

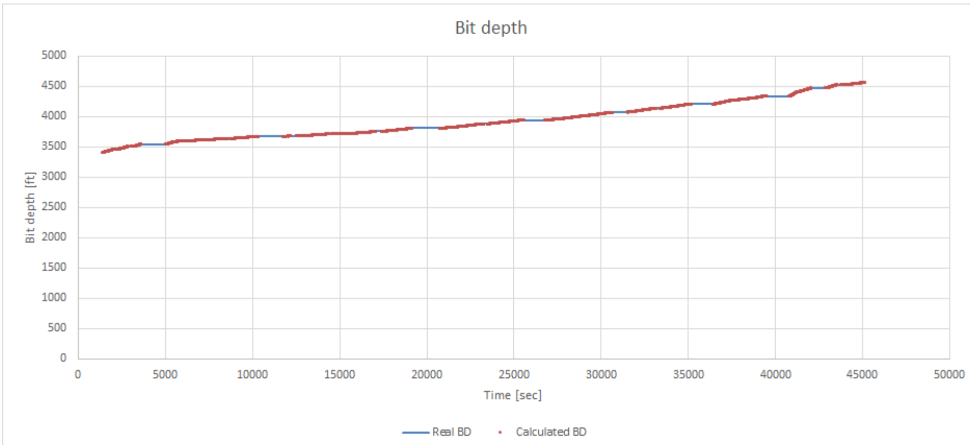
Results

Webinars

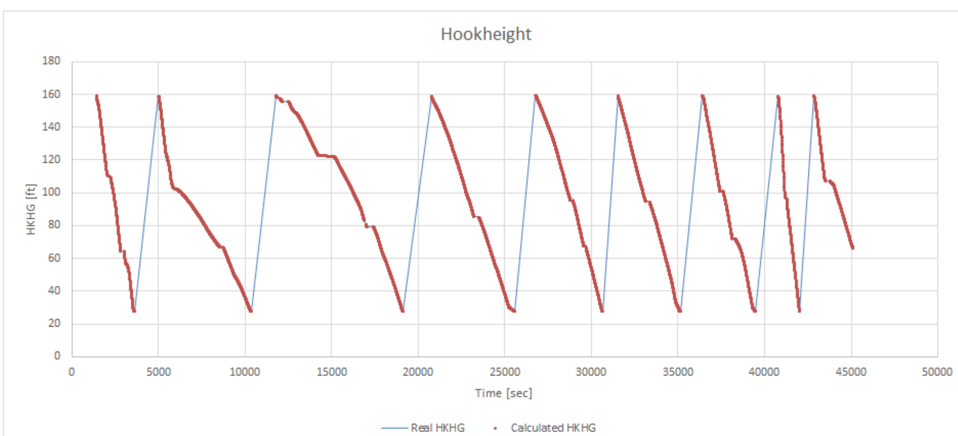


SPE International

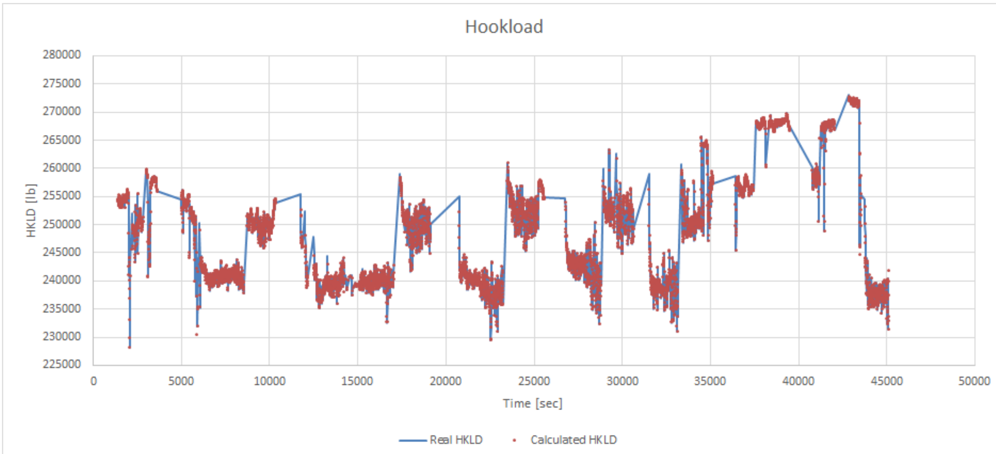
Results



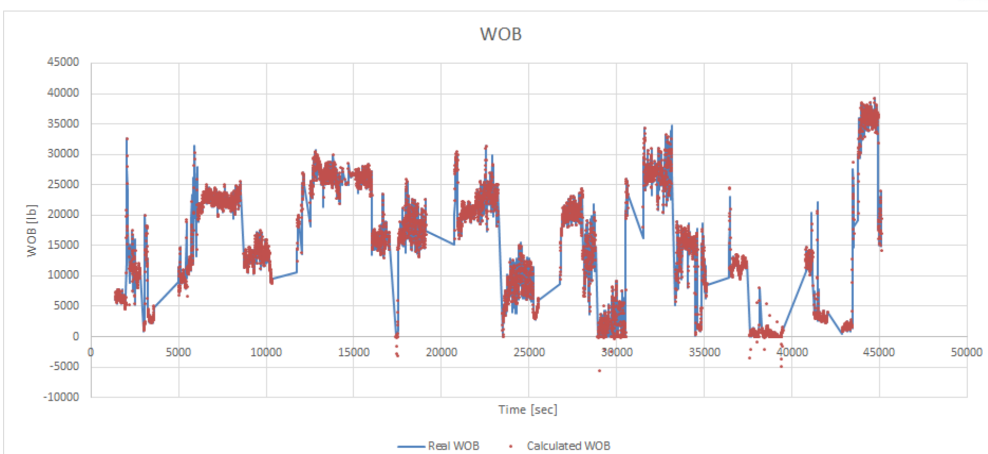
Results



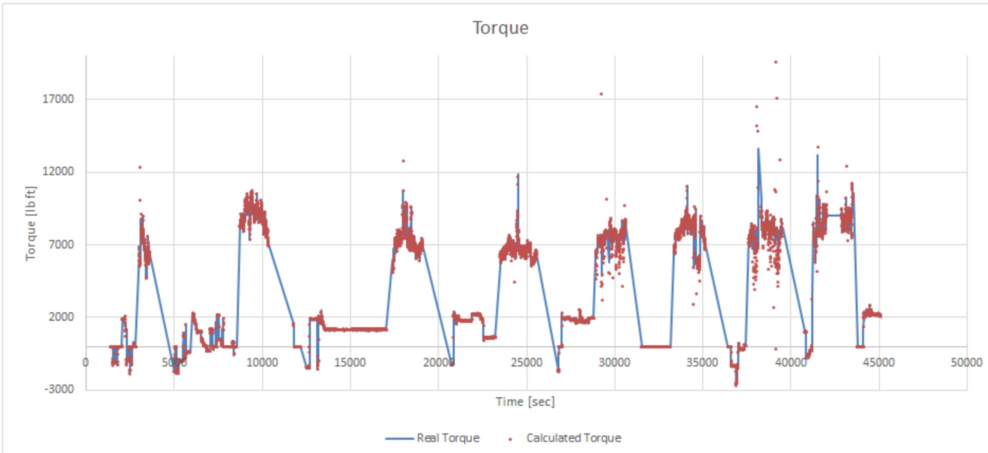
Results



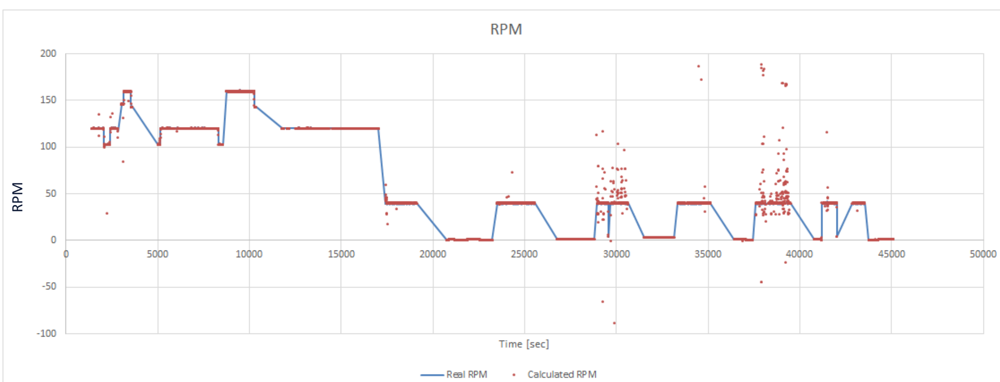
Results



Results



Results



Final Remarks

The suggested algorithm in this paper is a first approach to real time data analysis and early detection of sensor's malfunction using a pseudo-bayesian network.

Goals:

- The algorithm is easy to use and adapt to the necessity
- Development of a robust network using analytical relationships
- Scalable algorithm that allow the assessment of a great deal of sensors
- Intuitive code

Future improvements:

- Data smoothing in the pre-processing of the data set.
- Other techniques to fill missing values in the pre-processing stage.
- Further research on mathematical relations between variables to expand the network and integrate more sensors.