SPE Technical Exhibition and Networking Event - Intermediate Challenge

INTRODUCTION

Given: production.csv file of production data (gas,oil,water) for 12 wells along with time stamps for each well, blind forecast results for 6, 12 and 18 months using Arps decline curve analysis

Problem statement: to develop a data-driven time series forecasting model and compare the 6, 12 and 18 months hindcast predictions with the Arps model in terms of average RMSE.

	Well ID	DAYS	OIL	GAS	WATER
0	1	0.0	630.868000	1157.095194	66.671992
1	1	31.0	552.874349	1444.594307	33.964977
2	1	61.0	556.648235	1160.644566	0.000000
3	1	92.0	440.915721	1238.730745	0.000000
4	1	122.0	340.278752	1029.317811	0.000000

Methodology

Steps involved in the data-driven approach are as follows:

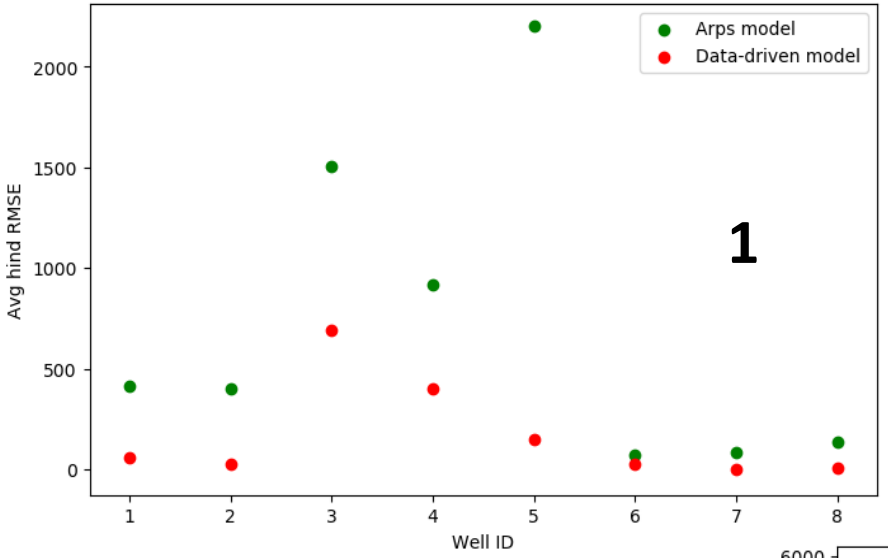
- The first 8 wells consist of monthly production data while the remaining 4 wells have the production data reported daily. Two different models are developed for these two scenarios.
- The data is restructured as a supervised learning problem where production at next time-step is predicted from production at the previous time step [‘get_history’ function]
- The data is then divided into train and test sets, the test set being the time step for which we want to predict the production. Same steps are followed for each of the two above scenarios.

Training model for first 8 wells

- Random Forest regressor is used to forecast oil, gas and water production for 6, 12 and 18 months from historical data.
- Random Forest is a tree-based method and can handle unscaled data. A total of 1000 decision trees are used to develop a robust regressor.
- The prediction production values are compared with the original data and the performance is reported in terms of RMSE.

	Well ID	Phase	hind_6_months_rmse	hind_12_months_rmse	hind_18_months_rmse	Average_hind_rmse
0	1	OIL	69.964079	89.500231	16.888770	58.784360
1	1	GAS	102.750763	229.867963	76.304394	136.307706
2	1	WATER	0.000000	11.950640	17.825878	9.925506
3	2	OIL	54.892435	17.803309	1.730484	24.808743
4	2	GAS	2.498758	9.132493	1.142792	4.258014

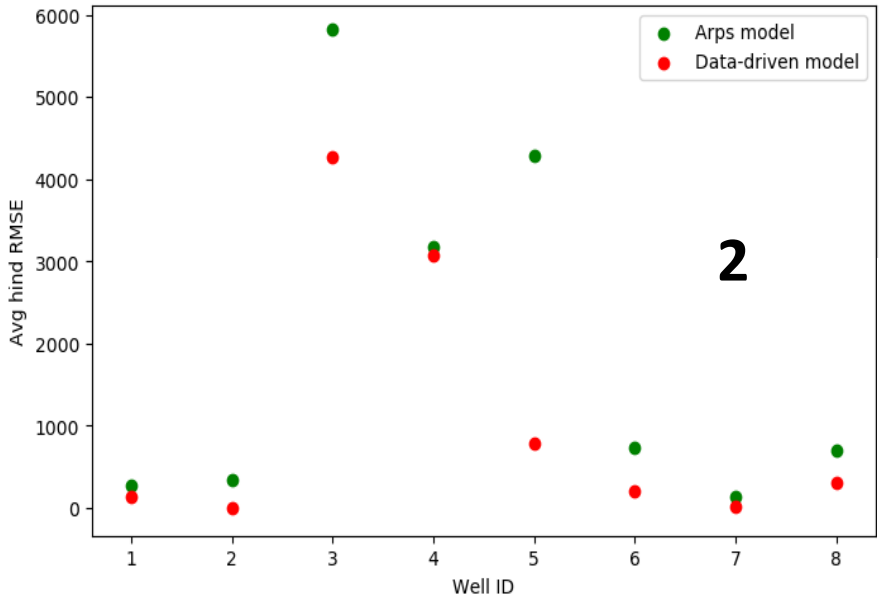
RMSE values from hindcasting 6 months, 12 months and 18 months production for the first 8 wells

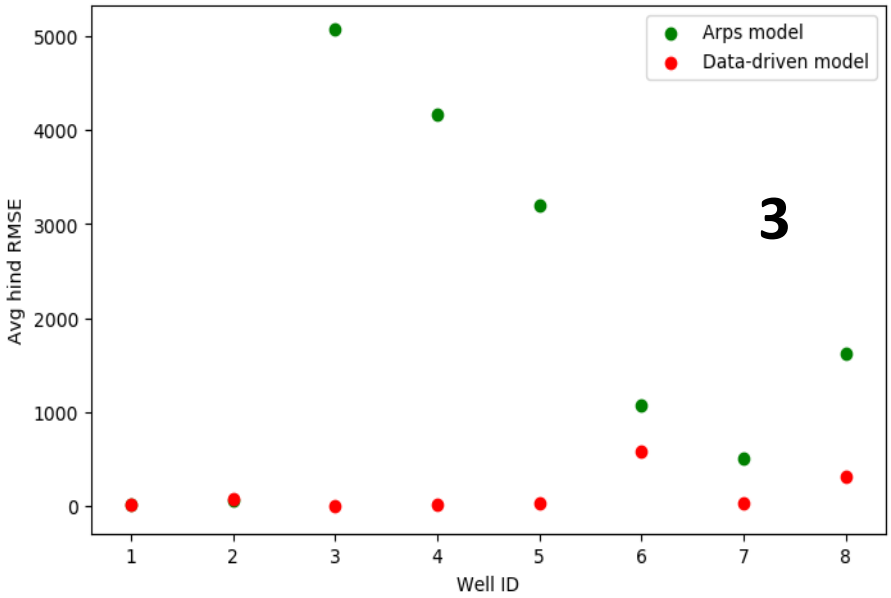


RMSE Comparison for predicted values for first 8 wells between ARPS model and Data-Driven model for:

1. Oil Phase
2. Gas Phase
3. Water Phase

For most wells, RMSE values are found to be significantly lower in the data-driven model when compared to arps model



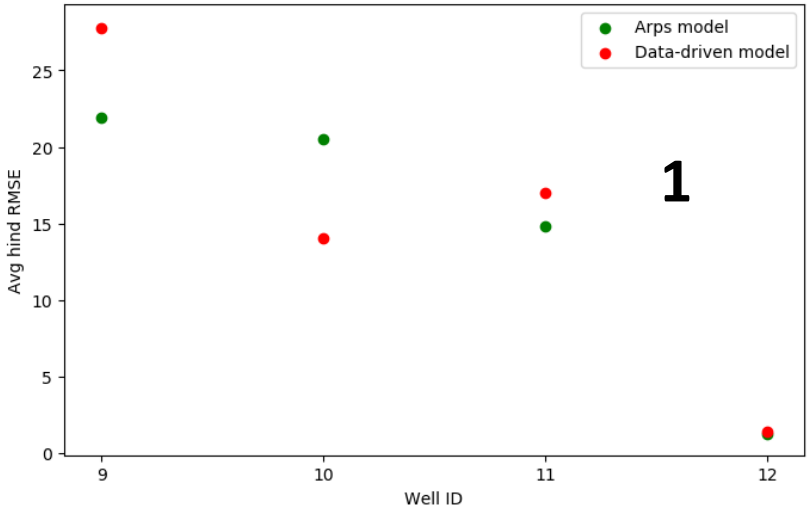


Training model for last 4 wells

- The model training follows similar steps as the first 8 wells
- 30 days are taken as time intervals for obtaining the historical data
- Daily production data from first to n-1 month is used to train the model
- 30-day production data of the nth month is predicted and RMSE is calculated with respect to the original data.

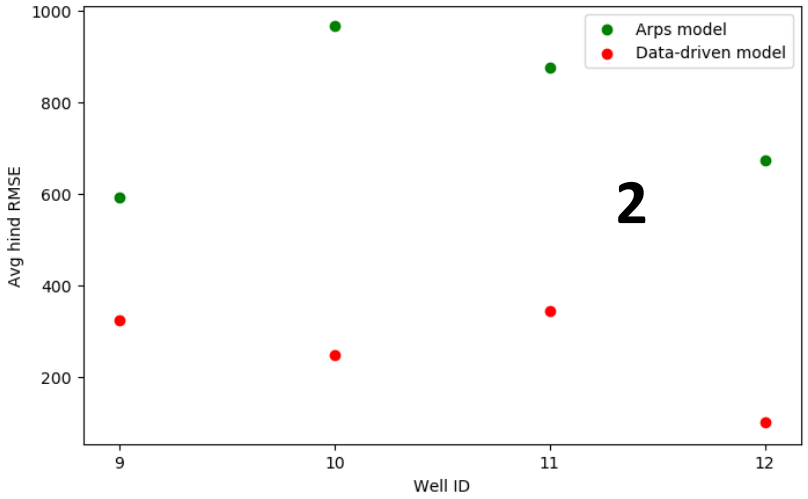
	Well ID	Phase	hind_6_months_rmse	hind_12_months_rmse	hind_18_months_rmse	Average_hind_rmse
0	9	OIL	52.045126	20.874023	10.433593	27.784248
1	9	GAS	674.159988	161.458157	136.567463	324.061869
2	9	WATER	79.930826	20.025924	23.329839	41.095530
3	10	OIL	19.659272	20.134175	2.278120	14.023856
4	10	GAS	312.935900	335.968721	89.809043	246.237888

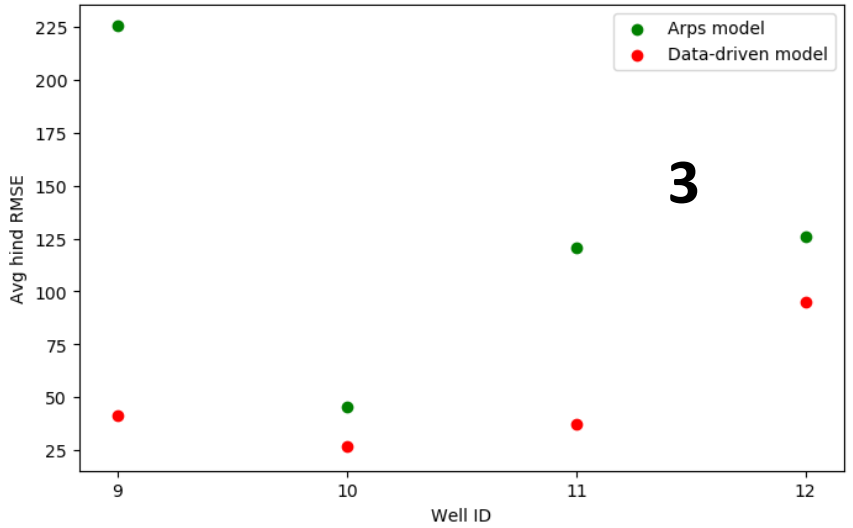
RMSE values from hindcasting 6 months, 12 months and 18 months production for the last 4 wells



RMSE Comparison for predicted values for last 4 wells between ARPS model and Data-Driven model for:

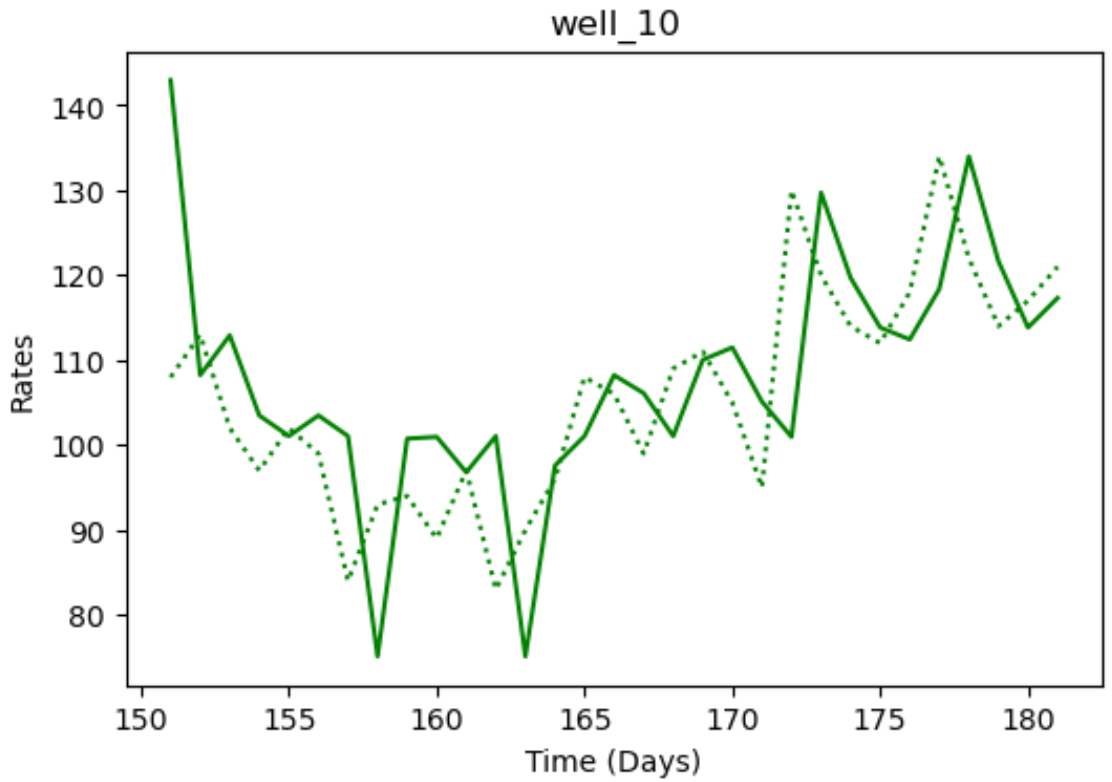
1. Oil Phase
2. Gas Phase
3. Water Phase





RMSE for the data-driven model was founds to be lower for gas and water phase and comparable for the oil phase.

- **Demo of Well 10 Oil phase production using Random Forest**



Key takeaways:

- For wells with monthly production data, the variance in the RMSE obtained from the Arps model is considerably higher than the RMSE from Random Forest.
- For most cases in this problem, the RF forecast shows significantly better results than the Arps model.
- The performance can be improved by tuning the hyper-parameters of the regressor model.
- If more data is available, deep learning techniques like RNN, LTSM can be used.