TOTAL
COMMITTED TO BETTER ENERGY

MEMO

PhD Thesis proposal: Merging Data-driven and physics driven modelling to optimize production of mature hydrocarbon reservoirs

Traditional tools used to make forecasts and optimize oil reservoirs production are based on complex geological modeling of the subsurface and on computationally expensive numerical solvers of the fluid flow equations in porous media. This approach enables to integrate all the available data (seismic, cores samples, wells logs, geological and physical understanding), however it is very time consuming, moreover the model obtained is usually not predictive enough for the short term production optimization particularly for large and mature fields.

More recently a different paradigm based on data-driven reservoir modeling has emerged[1] that aims at providing faster results using data analytics.

This approach has the advantage of being much faster and easy to implement, however validating and trusting the model forecasts is usually more controversial.

One of the main issues with data-driven models is to assess how much the model can be generalized to other data that has not been seen during the training. In fact in order to use these models for production optimization one needs to modify the way the wells are controlled and be able to propose new well settings that were not tested in the past.

The objective of this thesis is to investigate effective methods to combine data-driven models with simplified physics driven models with the objective of obtaining more reliable models to be used for production optimization. The combination of traditional physical models with statistical models based on machine learning and artificial intelligence is a current area of investigation in several other industries.

Main principles of data-driven the tool

The data-driven tool is based on the construction of a spatio-temporal statistical model of the water-oil-gas production and reservoir pressure for each well. The spatio-temporal model, aims at modelling the oil-water-gas evolution at two different time-steps (t, t+Δt) at each well (existing or to be drilled). We know from the laws of physics of fluid flows in porous media that if Δt is short enough the production evolution of well *l* is only affected by the values of pressure and production of very nearby wells. This assumption is very important as it will limit considerably the number of variables of our spatio-temporal function.

In order to take into account the time evolution we use recurrent neural networks and more specifically LSTM. These techniques are today among the state of art technique for time evolution statistical models, although other more classical machine learning techniques could also be tested during this work.

A data-driven tool based on LSTM has been recently developed by Total R&D and will be available for the project. One of the first objectives of this thesis will be to test the validity of the current approach in some real data. The physical laws of fluid flows in porous media are learnt by the algorithm based on historical data, however we do not have any guarantee that such physical constraints would be honoured also in the forecasts. The amount of training data depends on the field (maturity and number of wells), but it would be usually around few thousands observations. Because of this short amount of data the features used in the model needs to be selected carefully and are based on physical understanding of the input/output relationship. An effective modeling of the features to be used in the statistical model is therefore a fundamental part to improve the accuracy of the forecasts.

Recurrent Neural Networks and LSTM

We have decided to use LSTM to model the time evolution of the system. LSTM are more suitable to classical neural networks used in[1] for modeling time series: the LSTM state vector and weights are modified at each time steps to take into account some possible evolutions of the input-output relation occurring through time.

Uncertainty

In some cases the amount of available data may not be sufficient for the neural network model to learn the behavior of the system in order to obtain the required accuracy in the forecasts. Moreover the available data is of very different type and resolution (3D data (seismic, logs), production data at different locations, operational constraints… Interpreted/synthetized data coming from geoscientist synthesis could then be used to build the neural network model (drainage areas and average properties in these areas such as remaining oil in place and permeabilities/porosities). This choice will however necessitate to test different set of input data resulting in different possible NN models.

Other approaches such as data-augmentation (generating noisy data) or using dropout could be used to define a more robust model.

The objective is to be able to propose forecasts envelope based on different possible models. Each model may have a different weight according to its likelihood (accuracy in honouring the training data) similarly to a Bayesian approach.

New wells prediction

A problem strictly related to the previous problem of taking into account uncertainty is how to provide the missing data for new wells (the wells not available in the training data). A similar approach to the one considered above could be adopted for missing wells, however the construction of a more systematic technique to define missing value properties for such new wells should be defined in this case. Some classical approaches such as kriging or random forests will be tested. In fact we should also try to use production data to better evaluate such missing value problem and not only spatial correlation.

Including Physical constraints

The idea is to investigate the possibility of including physical constraints such as material balance directly into the neural networks models as for example an additional term in the objective function to minimize. Normally the training data will respect all the physical laws (unless there are major errors in this data), however the considered neural network model may not be able to reproduce exactly all the training data. In this case adding an additional term in the learning function could be useful to push the NN to honour more particulalry such constraints (the method used is similar to the well known technique of Lagrange multipliers).

Another interesting approach should be to test this constraint in the forecasts obtained by the different possible neural networks models that will be generated and to use it as an additional criteria to perform model selection.

Defining real experiments to improve the data-driven model forecasts

In order to increase the reliability of the data-driven model some real experiments could be planned: such experiments should short in order to not affect considerably the reservoir production but should be long enough to let the statistical model learn the reservoir behavior. For example in order to train the model to learn the effect of changing the well controls several variations of such controls should be performed. This is not always the case in real reservoir as operators tends to not perform too many changes of well controls to avoid loosing production.

Combining data-driven models with simplified physics driven models

The classical reservoir modeling workflow involves the construction of a very complex geological model of the subsurface where the petrophysical properties (permeabilities, porosities, …) are defined statistically (due to the lack of information) in every small region of the reservoir. More precisely a reservoir grid is built in which all the different properties are defined. This same grid discretization or an upscaled one is then used to solve the numerical equations of fluid flow in porous media. In order to represent the real heterogeneity of the reservoir typically millions of grid cells are used, therefore the simulation of the model to produce forecasts can be very long (several hours on large parallel computers).

Moreover in order to fill the properties of each grid cell of the model very complex workflows are used combining different data sources at different scales (seismic, core samples and well logs) and involving many different domain experts (geologists, geophysicists, but also experts on fluids, geostatistics, geomechanics, …). In this work we would like to investigate much more simple approaches to fill the physical model parameters based on pure data-driven approaches. To this end the resulting physical model should contain a number of macroscopic parameters not observable directly from the data but that could be derived by matching historical production data. One of the main challenges would then be to obtain the possible values of these parameters by assimilating production data but also other "static" information such as seismic data or well logs. This results in having to solve an inverse problem. Stochastic optimization techniques and machine learning methods could then be used to solve this problem.

Such simplified data and physics driven models will be fast to run and could be used as such to guide the optimization of the reservoir by testing many different scenarios (changing the injection, the operating well pressures, but also shutting wells, converting producers into injectors or drilling new infill wells). Methods to combine such models with purely data-driven models as described above will be investigated with the objective of obtaining more reliable forecasts and a more robust production optimization.

References (short list)

1. Shahab D. Mohaghegh "Data-Driven Reservoir Modeling" SPE 2017
2. Shahab D. Mohaghegh « Converting detail reservoir simulation models into effective reservoir management tools using SRMs; case study – three green fields in Saudi Arabia » Int. J. Oil, Gas and Coal Technology, Vol. 7, No. 2, 2014 115.
3. Hochreiter, S. & Schmidhuber, J. « Long short-term memory ». Neural Comput. 9, 1735–1780 (1997).
4. ElHihi, S. & Bengio, Y. Hierarchical recurrent neural networks for long-term dependencies. In Proc. Advances in Neural Information Processing Systems 8 http://papers.nips.cc/paper/1102-hierarchical-recurrent-neural-networks-for-long-term-dependencies (1995).
5. Sutskever, I. Training Recurrent Neural Networks. PhD thesis, Univ. Toronto (2012).
6. Pascanu, R., Mikolov, T. & Bengio, Y. On the difficulty of training recurrent neural networks. In Proc. 30th International Conference on Machine Learning 1310– 1318 (2013).
7. Weston, J. Chopra, S. & Bordes, A. Memory networks. http://arxiv.org/ abs/1410.3916 (2014).
8. Singh, N., Javeed, A., Chhabra, S., & Kumar, P. (2015). Missing Value Imputation with Unsupervised Kohonen Self Organizing Map. In Emerging Research in Computing, Information, Communication and Applications (pp. 61-76). Springer India.
9. Iyer, A., & Petkovic, D. (2016). Practical Tool to Understand Structure of Machine Learning Training Databases with Missing Data.