Literature review

For decades, petroleum engineers and researchers are looking for a simple and reliable way to predict oil production of petroleum wells. The production prediction model can help and forecast in numerical and physical ways. Technic engineers and researchers exploration mainly divided into three parts: 1. Petroleum production prediction which is the traditional method which concludes five subcategories. 2. Curve estimations, and 3. Neural networks.

For the petroleum production prediction in an oil well, the traditional methodologies include: (1) by analogy, (2) volumetric, (3) material balance, (4) decline curve fitting, and (5) reservoir simulation. Each method could be used for prediction but with different data requirements. For example, “by analogy” performs the prediction of the target well based on similar wells. This method is efficient, inexpensive, and good for estimation before drilling, but lack of accuracy. “Material balance” determines original oil-in-place which based on the law of conservation of mass. Moreover, each of those methods has limitations but can be used to cross-validation the prediction results of the prediction results generated by other methods.

Curve estimation is a decline curve analysis technique based on exponential, hyperbolic, and harmonic equations. [1] proves that fitting production data and predict the results with a decline curve is an insufficient and unreliable way if the historical production data is unreliable and missing. Several applications of fluid flow mechanism and petroleum production prediction using curve analysis are proposed such as [2] and [3].

The recent methods to estimate production values use Artificial neural network. [4] proves that the Neural Network gave a lower root mean square error (RMSE), and the authors also believe that the data pre-processing is the most important steps in applying the ANN approach to geological problems. The data pre-processing of missing values and nan values are discussed in [5], [6], [7]. Moreover, [8] Indicates that the Neural network model shows higher accuracy compared to other correlation methods. Instead of solving a bunch of mathematical equations to obtain the best coefficients in curve estimation, the neural network model updates weights iteratively and automatically, and the error reduced in each training epoch with objective functions and the back-propagation algorithms.

In our project, the first step of data cleaning and conditioning start with the physics meaning of features. For example, to deal with the zeros in the ‘AVG\_DOWNHOLE\_PRESSURE’ and ‘AVG\_DOWNHOLE\_TEMPERATURE’, the physics meanings play an important role to verify if the zeros are the type II outliers or not, which based on the traditional methods for petroleum production prediction in an oil well. Instead of using the curve estimation, we focused on the artificial neural network due to the significance and potential to improve the performance of the prediction mentioned in [4]. During the data pre-processing of missing values and zeros, we have implemented a mean value (traditional method), multilayer perceptron, and super vector regression to predict the missing value and type II outliers, then compare the performance of all three methods.

[1] “El-Banbi, A.H., Wattenbarger, R.A., 1996. Analysis of commingled tight gas reservoirs. SPE Annual Technical Conference and Exhibition, Denver, Colorado, USA, 6–9 October 1996 (SPE 36736).”

[2] “John, E.G., 1998. Simplified curve fitting using spreadsheet add-ins. International Journal of Engineering Education 14 (5), 375–380.”

[3] “Li, K., Horne, R.N., 2003. A decline curve analysis model based on Fluid flow mechanisms, SPE western regional/AAPG Pacific section joint meeting held in long beach, California, USA, 19–24 May 2003 (SPE 83470)”

[4] “Wong, P.M., Taggart, I.J., 1995. Use of neural network methods to predict porosity and permeability of a petroleum reservoir. AI Appl. 9 (2), 27–37.”

[5] Gelman, A., & Hill, J. (2006). Missing-data imputation. In Data Analysis Using Regression and Multilevel/Hierarchical Models (Analytical Methods for Social Research, pp. 529-544). Cambridge: Cambridge University Press. doi:10.1017/CBO9780511790942.031

[6] Jason Brownlee, How to Handle Missing Data with Python <https://machinelearningmastery.com/handle-missing-data-python/>

[7] Alvira Swalin, How to Handle Missing Data <https://towardsdatascience.com/how-to-handle-missing-data-8646b18db0d4>

[8] “Gharbi, R.B., Elsharkawy, A.M., Karkoub, M., 1999. Universal neural-network-based model for estimating the pressure–volume–temperature (PVT) properties of crude oil systems. Energy & Fuels 13, 454–458.”