Literature review

For decades, petroleum engineers and researchers are looking for a reliable and straightforward 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, to estimate the petroleum production in an oil well, the traditional methodologies include: (1) by analogy, (2) volumetric, (3) material balance, (4) decline curve fitting, and (5) reservoir simulation~\cite{thompson1984oil}. 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 useful for estimation before drilling, but lack of accuracy. \Material balance\ determines original oil-in-place which base on the law of conservation of mass. And each of those methods has limitations but can be used to cross-validation the prediction results of the prediction results generated by other means.

Curve estimation is a decline curve analysis technique based on exponential, hyperbolic, and harmonic equations. El-Banbi, A.H.et al.[1] proves that to fit production data and predict the results with a decline curve is an insufficient and unreliable way if the historical production data is inaccurate and missing. John, E.G.[2] and Li, K.et al.[3] propose several applications of fluid flow mechanism and petroleum production prediction using curve analysis.[2] and Li, K.et al.[3].

The most recent method is to estimate production values using an artificial neural network (ANN). Wong, P.M.et al.[4] proves that the Neural Network gave lower errors such as root mean square error (RMSE), and author also believes that the data preprocessing is the most critical steps in applying the ANN approach to geological problems. Gelman, A.et al.[5] discuss the data preprocessing of missing values and nan values.[5], Brownlee, J.[6] and Swalin, A.[7]. Moreover, Gharbi, R.B.,et al.[8] indicates that the Neural network model shows higher accuracy when compared to other correlation methods. ANN models are trained with more advanced, non-linear, deep & wide NN structures than the polynomial fitting equations implemented in the curve estimations methods. Instead of solving a bunch of mathematical equations to obtain the best coefficients, the neural network model updates weights to reduce the error at each step in each training epoch with objective functions and the back-propagation algorithms.

In this project, the first step of data cleaning and conditioning starts with the real meaning of features. For example, to deal with the NA's and zeros, the physical meaning plays an important role to verify if the zeros are the real data or just NA's; this is more innovative than using missing data imputation methods without considering the real meaning of features. Instead of using the curve estimation, we will focus on the artificial neural network due to the significance and potential to improve the performance of the prediction mentioned in [4].

@article{thompson1984oil,

title={Oil property evaluation},

author={Thompson, Robert S and Wright, John D},

year={1984},

publisher={Pennwell Books, Tulsa, OK}

}