A Review on Bottom-Hole Pressure Estimation in Oil Wells using Computational Intelligence Techniques

Reventh Sharma (2017A1PS0832P)

Abstract

The pressure generated by circulating fluids during the drilling process is termed as Bottom Hole Pressure. Bottom-Hole Pressure serves as an important parameter to gauge economic as well as the physical condition of well to produce. The following report provides the contribution of computational intelligence for the accurate determination of bottom hole pressure. Comparison of data-mining approaches as well as computational learning models for their accurate determination of the BHP value is conducted.

Keywords: Bottom Hole Pressure (BHP), Artificial Neural Network (ANN), Support Vector Machine (SVM), Support Vector Regression (SVR), Genetic Algorithm, Particle Swarm Optimization, Firefly Algorithm, Grey Wolves Optimization, Artificial Bee Colony, Swarm-Based Optimization.

1. Introduction:

Many engineering installation design problems are encountered with the simultaneous flow of oil, water, and gas in a vertical pipe. In an oil field, a reliable and accurate way of predicting a drop in pressure for a vertical multiphase flow is an important parameter in the design of artificial lift systems and other Well completion units. Optimization and forecast of performance of the production process are dependent on accurate bottom-hole prediction [1]. The high complexity of the multiphase flow problem provides complex theoretical models with multiple parameters to determine. Hence, empirical and semi-empirical models are highly suggestive but those are limited to experimental conditions and can't be easily extrapolated. Other methods rely upon online prediction by using ultrasound and blasting but those methods are expensive and provide time delays in the continuous process. The improvement in computational speed has led to the use of large data obtained onsite to accurately predict the bottom-hole pressure [2]. Several novel machine learning approaches have been developed and used to solve the problem of accurate BHP prediction but a review paper summarizing the current state of understanding of this topic has been lacking in the literature. Hence, in this review paper, an overview, as well as guidelines to use different computational learning model, is provided.

2. Machine Learning Models and Optimization Techniques:

Machine Learning is a union of the mathematical branch of Statistic, Numerical Theory with Computer Science. It deals with the use of currently observed data to effectively predict the output in test conditions or it deals with providing insights or patterns within the data. Several machine learning models, in recent developments used to predict bottom-hole pressure, will be introduced briefly.

2.1. Artificial Neural Networks

Artificial Neural Networks deal with interconnecting the input variable (action) to predictions (response) utilizing nodes of a network similar to neuro-connectivity in biological organisms. The layout of interconnected nodes is called the topology of the network which consists of layers of nodes where each node in $(i+1)^{th}$ layer is connected to all the nodes in the i^{th} layer as shown in fig.1. At each node, a weighted summation of values of all the nodes in the preceding layer provides a value of the node as f(W) where f is the activation function of the node. The last nodes of the neural network provide the output variable. Weights for the network are obtained by fitting the data, i.e. for each vector of input variables the output of the network and the difference in the predicted from actual value is used to adjust the weights of the network by minimizing the value of root mean squared error by numerical optimization. The data is passed to the neural network repeatedly till it fits the network such that RMSE<e, where 'e' is a predefined error value.

2.2. Least Square Support Vector Machine

Support Vector Machine is a classification algorithm that separates the data points according to their classes. Hence, when a novel input is provided the class of data point is predicted by the model.
The goal of the model is to maximize the margin-SVM as shown in fig.2. as ‘b’ for best classification, such that there is maximum classification accuracy. When a model is not linearly classifiable, the input is transferred to a higher dimensional using a kernel function space where it is assumed to behave linearly based on pre-existing knowledge of the domain.
Similarly, the input data is mapped to higher dimensional space and a linear function is obtained such that the value of output from function: ‘wTf(X)+bias’ and the data value of output variable falls within the margin of error ‘b’. The objective is to optimize the wT such that the margin is maximum. In LS-SVM the error value is (yi-ydata)^2 is also summed to the margin of error in optimization problem objective function. Hence, objective function ‘minimize z = b + sum(ei^2)’ where ‘ei = yi-ydata’ [3].

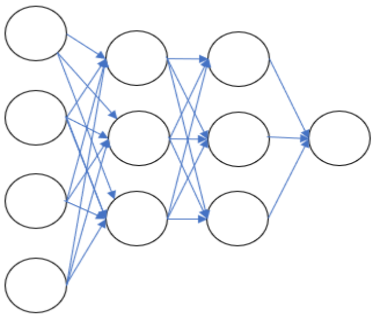

Figure 1: Artificial Neural Network (ANN)

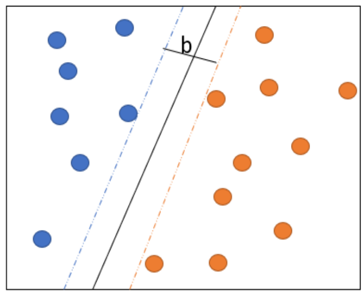

Figure 2: Support Vector Machine

2.1.3. Genetic Algorithm Optimization

The optimization technique used in numerical analysis for reaching minima/maxima follows a gradient descent approach where the parameters are tuned such that they always decrease (increase to reach maxima) along the direction of the gradient of the maximal gradient at the point.
However, gradient descent and other such numerical optimization techniques can get trapped in local minima (maxima). Hence, a population-based optimization technique is required for solving the problem more efficiently. N-values of the parameter variables are initiated from the domain of values, genotype space. The values of parameters are represented in binary form concatenating all binary parameter strings. From the genotype space, a variable set of size ‘N’ is chosen randomly with randomness probability distribution based on the values of the objective function (pmax => fmax in a maximization problem, pmax => fmin in a minimization problem). Hence, the parameter values with higher probability are represented more in this space. This is called the mating pool. The parameters are pared in the mating pool and their values are cross-over based on their binary representation by flipping bit values amongst each other. Some bits are flipped randomly in the pool which is called a mutation. The new values of parameters constitute genotype space-2. The process is iterated till the solution converges.

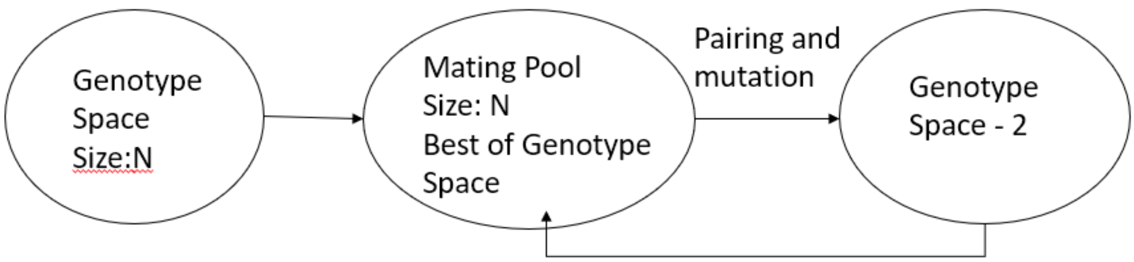

Figure 3: Algorithm for Genetic Algorithm

2.1.4. Particle Swarm Optimization

The population size of N-parameter values is chosen. The optimization technique is based on moving the parameter values towards the optimum value similar to how a swarm of birds locates their destination. The velocity of point-j in $(i+1)^{th}$ iteration is the weighted sum of velocity in i^{th} iteration, the difference from best value of j till now (where function in min/max according to the problem) and current value, the difference of best value parameters in population and j. Velocity is the difference in the value of point-j in the current iteration and previous iteration. The population values are updated till they converge.

2.1.5. Firefly Algorithm

The firefly algorithm is a swarm-based optimization technique the new position of point 'i' is chosen such that it moves closer to all the other points in the database. The amount of position change of point 'i' due to point 'j' is inversely dependent on their cartesian distance. Once all points are updated global best value is selected. The stopping criterion in the convergence of population and the value of Gbest is the optimal value of problem and value parameters is the value of point vector corresponding to Gbest. [4]

2.1.6. Grey Wolves Optimization

The population is divided into 4-parts alpha, beta, gamma, and omega. The first three are the values corresponding to the best value of objective function according to the problem. The rest omega values are updated according to influence from alpha, beta, and gamma. The procedure is repeated till convergence. [5]

2.1.7. Artificial Bee Colony

This is a swarm-based optimization model where the particles move towards the best position in their neighborhood. The population is divided into two parts onlooker and employed members. The onlooker looks for the best position within the neighborhood where function is more optimized than the current value. The employed members move toward the most optimum value in the member's neighborhood. The best solution until convergence is chosen as the optimum value. [6]

3. Data Collection and Pre-Processing

The data for the process is collected from various oil fields in the literature. Jahanandish et.al collected 413 data points from the Iranian-Oil field. The value of BHP is predicted based on 9 input parameters: flow rates of gas, water, and oil, gas to oil ration, length of pipe, the pressure at wallhead, density of oil in API, the temperature of effluents at surface and bottom of the hole. The data is partitioned as train : validate : test as 4:1:1 [1]. Ahmadi et. al used similar input parameters as Jahanandish et.al for BHP prediction. The data is partitioned as train : optimize : test as 8:1:1 [7]. Nait et.al collected data from the Algerian-Oil field with 125 points. An 8:2 train : test ratio is used to evaluate model. Nait et.al used 10 parameters as: flow rate of oil, gas, and water, densities of oil and gas, hole depth, pipe's inner diameter, gas to oil ratio, and temperature and pressure at wallhead [8, 10]. Irani et.al collected data from Parsi and Kharanj fields. 75:25 ratio was used to divide data in train : test. 7 variable measured vertical depths, flow rates of gas and liquid, length of drill pipe injected, liquid density, and temperature at the surface, and pressure on casing were used to evaluate the model [9]. Tariq et. al obtained 206 data points from published sources: Govier and Fogarasi 1975; Asheim 1986. Model inputs were the same as Jahanandish et.al [10]. Memon et.al used the data of porosity in 3-layer, permeability in all 3-directions in 3-layers, and production rate to determine BHP [12]. Pre-processing is conducted based on min-max standardization where Xstd = (X-Xmin)/(Xmax-Xmin). Pre-processing is done so that all input variables are in the same range hence, so that correlation of any one variable on output due to range factor doesn't occur.

4. Model Evaluations

Jahanandish et. al formed neural network topology as input-hidden1-hidden2-hidden3-output as 9-20-15-10-1. A non-linear activation function is used by the model. Trend as well as the statistical analysis was performed to evaluate model performance. Trend analysis predicted a stable model and statistically showed R^2 = 0.9222 [1]. Ahmadi et.al evaluated model is LS-SVM optimized by genetic algorithm. Gaussian Kernel function is used. The overall R^2 = 0.961 for the model [7]. Nait et. trained a neural network with architecture input-hidden-output as 10-12-1. The optimization was done based on traditional back-propagation, genetic algorithm, particle swarm optimization, grey wolves optimization. The R^2 values are obtained as 0.9970, 0.9983, 0.9984, 0.9980 respectively [8]. Irani et.al followed 7-4-1 Neural Network architecture. The model was trained on backpropagation (10 runs with different initialization) and artificial bee colony method. The BP-ANN provided R^2=0.9738 and ANN-ABC provided R^2=0.9999 [9]. Nait et.al performed SVR optimized by firefly algorithm, genetic algorithm, and classical method. Gaussian Kernel function is used. The R^2 values were obtained as 0.9981, 0.9981, and 0.9980 respectively [10]. Tariq et.al performed model evaluation based on architecture 9-20-1. Two types of activation functions, radial basis, and tan-sigmoidal were used. The optimization was done based on Back-Propagation and PSO. R^2 values obtained based on data are 0.879, 0.89 and 0.983 for radial basis, tan-sigmoidal and PSO-tan-sigmoidal respectively [10]. Memon builds BPNN (Back-Propagation Neural Network) and RBNN (Radial Basis Neural Network) using 5-fold cross-validation. R^2 of 0.87 and 0.96 were obtained for BPNN and RBNN respectively. For 15-testing cases 0.63 and 0.645 for BPNN and RBNN respectively [12].

5. Conclusion

The computational models have been developed using different types of input vectors to predict BHP. Since a different number of inputs are observed and the data are obtained from different sources comparison of regression models is best done by R^2 statistic. Models are also evaluated based on the input, the number of inputs required, and the feasibility of obtaining those inputs online. Models are also evaluated based on their complexity where a lesser complex model is more recommended for online operation. Occam's razor: A less complex model is more recommended than a more complex model with similar accuracy. Models are evaluated based on their fitting to test data. The model that fits more is said to have better generalization. The models evaluated based on BHP predictions from input variables as per Jahanandish et.al require bottom-hole temperature value which is infeasible to obtain in a continuous process. Input variables for models based on Nait et.al data are all feasible. Variables based on Irani et.al have infeasibility in obtaining casing pressure and true vertical depth. For Memon et.al all variables are feasible but the value of permeability and porosity aren't obtained online but in a laboratory. Hence, there is a time delay associated with the prediction. The R^2 of various models is plotted in fig.4 and table-1. R^2 values of all the models are very close to one which represents high generalization and fit to data. It is observed that the SVR-TE model is the least complex amongst all model and input to it are also feasible. Hence, the recommended model for BHP prediction is SVR-TE with 10 input variables: flow rate of oil, gas, and water, densities of oil and gas, hole depth, pipe's inner diameter, gas to oil ratio, and temperature and pressure at wallhead.

ANN	0.9222	ANN-PSO	0.9984	SVR-TE	0.9980	TS-ANN	0.8900
LS-SVM-GA	0.961	ANN-GWO	0.9980	ANN-BP	0.9738	TS-ANN-PSO	0.9830
ANN-BP	0.997	SVR-FFA	0.9981	ANN-ABC	0.9999	BP-ANN	0.6300
ANN-GA	0.9983	SVR-GA	0.9981	RB-ANN	0.8790	RBNN	0.6450

Table 1 Model vs R^2

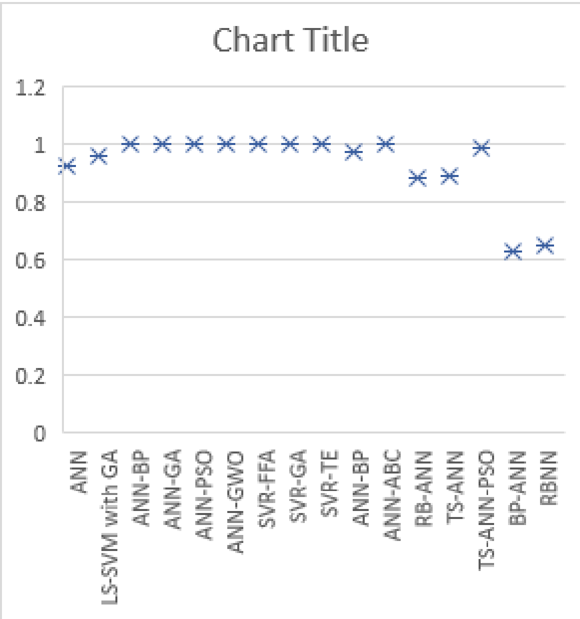

Figure 4: R^2 vs. model point plot

Reference:

1.) I. Jahanandish, B. Salimifard, H. Jalalifar, "Predicting bottomhole pressure in vertical multiphase flowing wells using artificial neural networks", Journal of Petroleum Science and Engineering, Volume 75, Issues 3–4, 2011, Pages 336-342, ISSN 0920-4105, https://doi.org/10.1016/j.petrol.2010.11.019.
2.) Rahman Ashena, Jamshid Moghadasi, "Bottom hole pressure estimation using evolved neural networks by real coded ant colony optimization and genetic algorithm", Journal of Petroleum Science and Engineering, Volume 77, Issues 3–4, 2011, Pages 375-385, ISSN 0920-4105, https://doi.org/10.1016/j.petrol.2011.04.015.
3.) X. Deng, T. Liu, Y. Luo and B. Yang, "Least squares support vector regression filter," *2010 3rd International Congress on Image and Signal Processing*, Yantai, 2010, pp. 730-733. https://doi: 10.1109/CISP.2010.5646734.
4.) Amarita Ritthipakdee, Arit Thammano, Nol Premasathian, Duangjai Jitkongchuen, "Firefly Mating Algorithm for Continuous Optimization Problems", Computational Intelligence and Neuroscience, vol. 2017, Article ID 8034573, 10 pages, 2017, https://doi.org/10.1155/2017/8034573.
5.) Seyedali Mirjalili, Seyed Mohammad Mirjalili, Andrew Lewis, "Grey Wolf Optimizer", Advances in Engineering Software, Volume 69, 2014, Pages 46-61, ISSN 0965-9978, https://doi.org/10.1016/j.advengsoft.2013.12.007.
6.) G. Pampará and A. P. Engelbrecht, "Binary artificial bee colony optimization," *2011 IEEE Symposium on Swarm Intelligence*, Paris, 2011, pp. 1-8. https://doi:10.1109/SIS.2011.5952562.
7.) Mohammad Ali Ahmadi, Morteza Galedarzadeh, Seyed Reza Shadizadeh, "Low parameter model to monitor bottom hole pressure in vertical multiphase flow in oil production wells", Petroleum, Volume 2, Issue 3, 2016, Pages 258-266, ISSN 2405-6561, https://doi.org/10.1016/j.petlm.2015.08.001.
8.) Menad Nait Amar, Nourddine Zeraibi, Kheireddine Redouane, "Bottom hole pressure estimation using hybridization neural networks and grey wolves optimization", Petroleum, Volume 4, Issue 4, 2018, Pages 419-429, ISSN 2405-6561, https://doi.org/10.1016/j.petlm.2018.03.013.
9.) Rasoul Irani, Reza Nasimi, "Application of artificial bee colony-based neural network in bottom hole pressure prediction in underbalanced drilling", Journal of Petroleum Science and Engineering, Volume 78, Issue 1, 2011, Pages 6-12, ISSN 0920-4105, https://doi.org/10.1016/j.petrol.2011.05.006.
10.) Nait Amar, M., Zeraibi, N. A combined support vector regression with firefly algorithm for prediction of bottom hole pressure. SN Appl. Sci. 2, 23 (2020). https://doi.org/10.1007/s42452-019-1835-z.
11.) Tariq, Z., Mahmoud, M. & Abdulraheem, A. Real-time prognosis of flowing bottom-hole pressure in a vertical well for a multiphase flow using computational intelligence techniques. J Petrol Explor Prod Technol 10, 1411–1428 (2020). https://doi.org/10.1007/s13202-019-0728-4.
12.) Memon, P. Q., Yong, S.-P., Pao, W., & Pau, J. S. (2015). Dynamic Well Bottom-Hole Flowing Pressure Prediction Based on Radial Basis Neural Network. Studies in Computational Intelligence, 279–292. https://doi.org/10.1007/978-3-319-14654-6_17.

Bibliography:

1.) B. Yegnanarayana, "Artificial Neural Networks," Prentice-Hall of India Pvt. Ltd., New Delhi, 1999.
2.) Chakraborty, Udit & Roy, Samir, "Introduction to Soft Computing: Neuro-Fuzzy and Genetic Algorithms", Pearson Education India, 2013.