Advances in Geo-Energy Research

Vol. 2, No. 2, p.148-162, 2018
www.astp-agr.com

Original article

A transparent Open-Box learning network provides insight to complex systems and a performance benchmark for more-opaque machine learning algorithms

David A. Wood*

DWA Energy Limited, Lincoln, United Kingdom

(Received February 28, 2018; revised March 15, 2018; accepted March 16, 2018; available online March 20, 2018)

Citation:
Wood, D.A. A transparent Open-Box learning network provides insight to complex systems and a performance benchmark for more-opaque machine learning algorithms. *Advances in Geo-Energy Research*, 2018, 2(2): 148-162, doi: 10.26804/ager.2018.02.04.

Corresponding author:
*E-mail: dw@dwasolutions.com

Keywords:
Learning networks
transparency of variable relationships
benchmarking machine learning performance
prediction of complex petroleum systems
soft-computing solutions
under-fitting/over-fitting

Abstract:
It is now commonplace to deploy neural networks and machine-learning algorithms to provide predictions derived from complex systems with multiple underlying variables. This is particularly useful where direct measurements for the key variables are limited in number and/or difficult to obtain. There are many petroleum systems that fit this description. Whereas artificial intelligence algorithms offer effective solutions to predicting these difficult-to-measure dependent variables they often fail to provide insight to the underlying systems and the relationships between the variables used to derive their predictions are obscure. To the user such systems often represent "black boxe". The novel transparent open box (TOB) learning network algorithm described here overcomes these limitations by clearly revealing its intermediate calculations and the weightings applied to its independent variables in deriving its predictions. The steps involved in building and optimizing the TOB network are described in detail. For small to mid-sized datasets the TOB network can be deployed using spreadsheet formulas and standard optimizers; for larger systems coded algorithms and customised optimizers are easy to configure. Hybrid applications combining spreadsheet benefits (e.g., Microsoft Excel Solver) with algorithm code are also effective. The TOB learning network is applied to three petroleum datasets and demonstrates both its learning capabilities and the insight to the modelled systems that it is able to provide. TOB is not proposed as a replacement for neural networks and machine learning algorithms, but as a complementary tool; it can serve as a performance benchmark for some of the less transparent algorithms.

1. Introduction

The potential of neural networks and machine learning has been recognised since the 1940s (McCulloch and Pitts, 1943; Hebb, 1949; Farley and Clark, 1954; Ince, 1992 (with reprint of Turing, 1948)). There have been many steps forward in their development and application, notably the perceptron algorithm (Rosenblatt, 1958), the backpropagation algorithm (Werbos, 1975), multi-layer perceptrons (Cybenko, 1989), feedforward networks (Scarselli and Tsoi, 1998), variously driven by gradient descent methods (Behnke, 2003), radial basis function networks (RBFN) applying various radially symmetrical functions applied to three network layers (Santos et al., 2013). These advances have gradually transformed artificial neural networks (ANN), and other neural networks, into popular and widely applied system learning tools with a range of complex deep learning capabilities (Schmidhuber, 2015). The number of neurons employed influences issues associated with under-fitting and over-fitting of the systems modelled (Aalst et al., 2010). However, as such networks become more complex it is more and more difficult to reveal their inner calculations, and they become black boxes to many of the users that employ them (Heinert, 2008). It is difficult and time consuming to extract useful information and functional relationships about how these complex algorithms are making their predictions and the relative significance of the input variables to those predictions. It is possible to extract functional relationships from some such networks, but this transforms them into whit boxes (Elkatatny et al., 2016), not fully transparent boxes.

Other machine learning techniques including support vector machine (SVM) (Cortes and Vapnik, 1995), least squares support vector machine (LSSVM) (Suykens and Vandewalle, 1999), and fuzzy logic combinations with ANN such as Adaptive Neuro-Fuzzy Inference Systems (ANFIS) (Sugeno and Kang, 1988; Jang, 1993) suffer from the same transparency limitations as neural networks. Nevertheless, this lack

https://doi.org/10.26804/ager.2018.02.04.

of transparency has not inhibited these algorithms from being widely and successfully applied, often hybridized with various optimization algorithms, to improve predictions from various oil, gas and other industrial systems (Li et al., 2013; Meng and Zhao, 2015; Zamani et al., 2015; Choubineh et al., 2017; Yavari et al., 2018). Indeed, there is a growing body of research in the petroleum sector that enters dataset records with multiple variables into opaquely coded soft-computing machine learning and neural network functions (MatLab, in particular), blindly accepting the results in many cases without even checking the veracity of the input data values. If these results demonstrate reduced prediction errors versus traditional formulaic relationships, then such blind machine learning is frequently claimed to provide a superior prediction tool. Although the prediction results may be impressive from such an approach, without understanding how those prediction processes work in detail, or how the underlying variables contribute to the predictions in relative terms, leaves significant uncertainty about the underlying system. This leads to doubts about how apparent performance improvements can be confidently applied to and relied upon related systems or datasets.

The learning network methodology and algorithm described here focuses on transparency of all calculations and providing clarity regarding the relative contribution that each variable makes to the prediction of the dependent variable /objective function (OF). This transparent open box (TOB) network makes no claims here in terms of improved precision of prediction compared to other learning network methodologies, but this learning network does provide fundamental insight to the factors determining the level of precision it is able to achieve. This makes it a useful complementary methodology with which to benchmark the prediction performance of more complex and opaque learning network algorithms. It also helps to decide whether it is actually necessary and/or appropriate to run more opaque learning algorithms, if a system can be adequately modelled by a transparent alternative.

This paper is organized as follows: section 2 describes, step-by-step, how the TOB learning network is constructed, and tuned; section 3 describes how the TOB is optimized and further tuned with the aid of sensitivity analysis; section 4 describes the application of the TOB algorithm to three quite distinct oil field data sets; section 5 presents a discussion of the potential benefits of the TOB algorithm and the type of oil and gas datasets to which it could be beneficially applied; section 6 draws conclusions for the study.

2. Description of transparent open box (TOB) learning network methodology

The TOB learning network methodology involves a set of simple, systematic and rigorous steps that enable its development and progress in predicting the objective function of the system being modelled to be clearly followed and interrogated. There are fourteen steps divided into two stages:

1) building and tuning the network (steps 1 to 10 described in Fig. 1);

2) optimizing the network and sensitivity analysis (steps 11 to 14, described in Fig. 2).

Step 1: setup a 2D array containing each of the data records of the dataset in rows, each independent variable defining the system to be predicted in columns, and the dependent variable/objective function (OF) in the final column.

Step 2: sort and rank the records into ascending (or descending) order of OF values. If there are many data records with the same OF values, then one independent variable should be selected as a secondary sorting criterion.

Step 3: calculate basic statistical metrics for each variable covering the entire data set. These should include, minimum, maximum, mean, variance (used for normalization) and a range of percentiles available for use in clustering.

Step 4: Normalize all the data variables in the entire dataset. Working with normalized data (i.e., between 0 and 1 or -1 and +1) removes scaling biases associated with the different units associated with the variables.

X'=(X- Xmin)/(Xmax-Xmin)provides normalized values (X') in the range 0 to 1;

X'= 2*[(X- Xmin)/(Xmax-Xmin)]-1 provides normalized values (X') in the range -1 to +1.

Either of these normalization methods can be used, but once selected the same method should be used consistently.

Step 5: repeat statistical analysis to calculate basic statistical variables on the normalized scale for each variable covering the entire data set. Use these statistics to assign each independent variable to numbered clusters. For example, by establishing the minimum, 20-percentile, 40-percentile, 60-percentile, 80-percentile and maximum values, those values can be used as thresholds between five different cluster numbers. Normalized variable values that lie between the minimum and 20-percentile are assigned to cluster#1, normalized variable values that lie between the 20-percentile and 20-percentile are assigned to cluster#2, etc. For the objective function a larger number of clusters is determined by narrowing the interval between the percentiles. By using every tenth-percentile (and minimum (P0) and maximum (P100)) as cluster thresholds, then ten numbered clusters can be distinguished; by using every fifth-percentile (and minimum (P0) and maximum (P100)) as cluster thresholds, then twenty numbered clusters can be distinguished. These clusters are useful in ensuring that data subsets sample a comprehensive range of the objective function (see Step 6).

Step 5a (optional): a simple quick-look learning network can be established by matching the cluster number allocations of the independent variables to test data records and selecting the best matches to place that test record in a predicted cluster of the OF. The precision of predictions made in this way depend on the width of the percentile clusters used for the dependent variable. For some applications, where lower level of precision is acceptable, this approach can provide a rapid and simple prediction.

Step 5b (optional): refines the cluster approach by calculating the squared errors of each variable in the best matching records to further refine the prediction. This provides an alternative learning network approach incorporating elements of steps 6 to 10 with the step 5a cluster analysis. This approach is not developed further here, as it is less precise than steps 6 to 10.

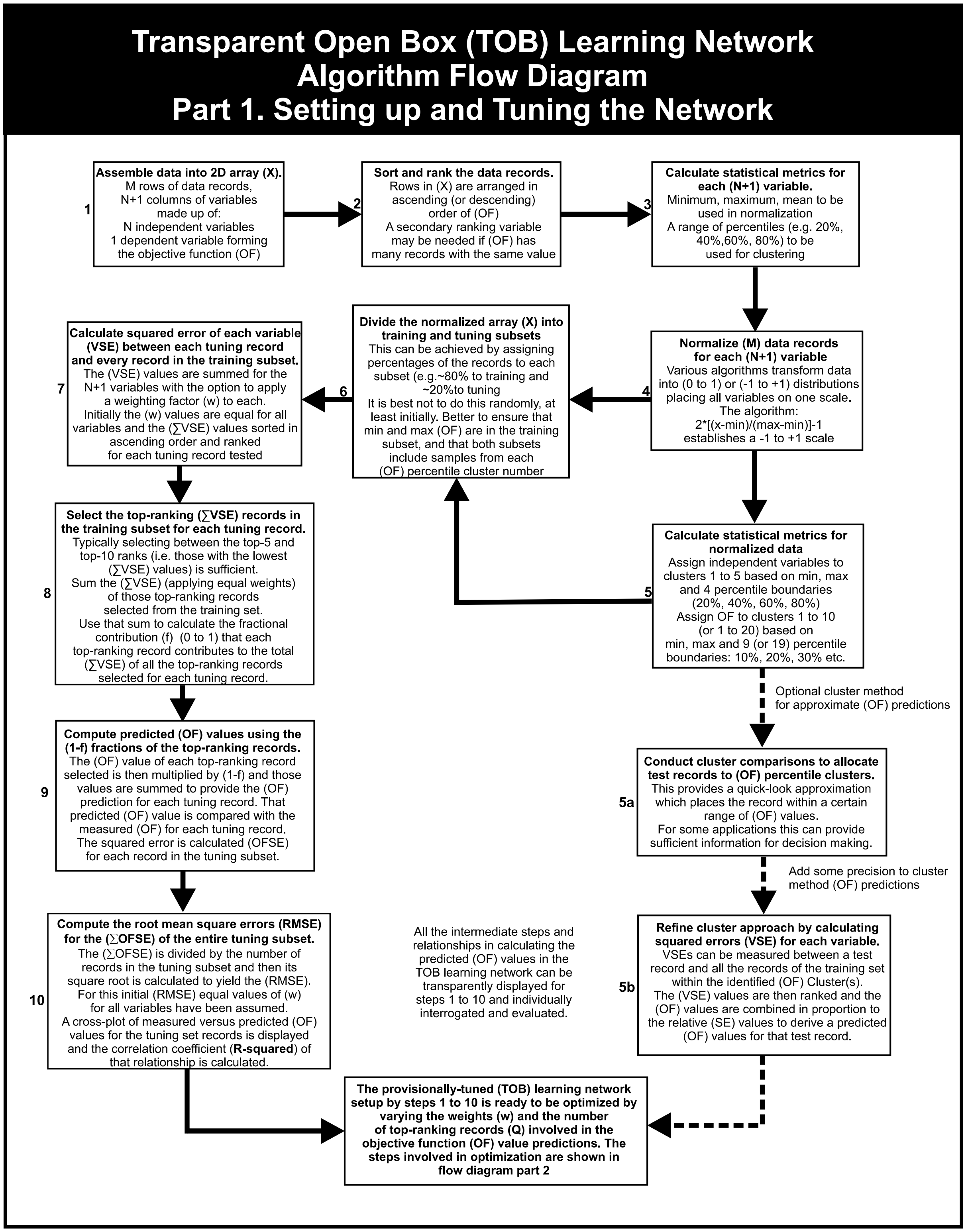


Fig. 1. Flow diagram summarizing ten steps involved in constructing and tuning a transparent open box (TOB) learning network. Steps 1 to 10 are described in more detail in the text.

Step 6: dividing the data set into a training subset, a tuning subset and a testing subset is an important selection. In many learning network methodologies, a random approach is applied. However, random allocation does not always lead to representative selections. It is important that the training subset includes data records at or close to the minimum and maximum values for the OF in the full dataset. If this does not happen in a random selection, then it is likely that systematic errors will occur in the prediction made at either end of the OF range sampled. Consequently, the methodology here advocates forcing the data records with the minimum and maximum OF values into the training subset. Also, it is important that the tuning set includes records that sample each percentile cluster used to define the OF range. A random approach is unlikely to achieve that. How many records to place in the training, tuning and testing subsets depends upon the size and complexity of the system being modelled. This requires some trial and error and sensitivity analysis (see step 13). Typically, the training set should be as large as possible (with multiple records included in each of the OF percentile cluster), e.g., an initial starting point could be between 70% to 90% of the dataset allocated to the training subset, if there are sufficient remaining records to provide adequate coverage of the OF clusters in the tuning and testing subsets (particularly the tuning subset, as this will influence how comprehensively the network can be improved through learning).

Step 7: A key step in the methodology, this tests each of the data records in the tuning subset for differences between each of its variable values and ***every*** data record in the training set. These differences for each variable are measured and expressed as variable squared errors (VSE). Squaring the differences removes the influences of negative signs in the difference calculations and enables those VSE to be summed for each data record comparison to provide summed error value ($\sum$VSE) that can be used to sort and rank the matches between the tuning data record and each record of the training subset. Weighting factors (w_1 to w_{N+1}) are applied to the error differences for each of the $N+1$ variables. In the initial tuning process, the same weighting value is applied to all the variables (e.g., w_1 = w_2, w_{N+1} = 1 or 0.5), so no preference is applied to any of the variables in the initial sorting and ranking process. By ranking the matched records in ascending order of their $\sum$VSE values, the top-ranking records (i.e., those with the lowest $\sum$VSE values) for matches to the tuned subset record being assessed can be identified and selected for detailed OF prediction (steps 8 and 9).

Step 8: A number (Q) of the top-ranking matched training subset records for each tuning subset data record are selected. The integer value of Q typically varies between 2 and 10 and is later used as a variable in the optimization process. These Q records in the training subset are identified for each tuning record in the subset. The sum of the $\sum$VSE values for the Q records (i.e., $\sum_1^Q \sum$VSE) for each tuning set data record (applying equal weighting (w) values for w_1 to w_N, with w_{N+1} = 0, so the value of the dependent variable does not influence the detailed tuning calculations) is used to assess the relative proportion of the($\sum_1^Q \sum$VSE) error that is contributed by each record making up the Q set of top-ranking matches. By dividing each $\sum$VSE of the Q set of records by ($\sum_1^Q \sum$VSE) the fractional contribution, f, (where, f = 0 to 1 and $\sum f$ = 1) to the ($\sum_1^Q \sum$VSE) error is established. The matched data record with the highest value of f is the one contributing the most to the ($\sum_1^Q \sum$VSE) error; whereas, the matched data record with the lowest value of f is the one contributing the least to the ($\sum_1^Q \sum$VSE) error.

Step 9: The objective functions (OF) of the top-ranking matching records for each record of the tuning subset contribute to the prediction of the OF value for that tuning data record in proportion to their $(1-f)$ values. The matching record with the highest (1-f) value is the one that contributes least to the ($\sum_1^Q \sum$VSE) error. If Q equals 2 and one has an f value of 0.8 then the other has an f value of 0.2. In that case, as just the top-two ranking matches are used in the prediction calculation, the OF value of the record with $f=0.2$ contributes 80% $((1-f)*100)$ to the predicted OF value, and the other record (with $1-f$ = 0.8) contributes just 20% to the predicted OF value for that tuning subset record. The predicted OF value calculated by this method is then compared to the measured OF value for that record, by taking the difference between them and squaring that difference to yield a OFSE prediction error measure for that tuning subset data record. The OFSE can be calculated using normalised values for the OF, but it is more meaningful to use actual values for the OF to visualise the significance of the prediction errors involved.

Step 10: The sum of the OFSE (squared errors of the predicted versus measured objective function values) for all the tuning subset records ($\sum$OFSE) is calculated. The contribution of each record to the total prediction error is available for display and analysis. This sum-of-the- squared-errors value ($\sum$OFSE) is then divided by the number of records in the tuning subset and its square root calculated to provide a root mean squared error (RMSE). The RMSE becomes the objective function for the optimization process (steps 11 to 14). An additional key metric calculated is the correlation coefficient (R^2) between the measured and predicted OF values. Clearly, the closer the R^2 value is to 1 the better the prediction performance of the unweighted TOB learning network developed for the system dataset modelled. Cross-plotting the measured versus predicted values in an X-Y graphic helps to visualise the prediction performance and identify potential anomalous data records (i.e., outliers) worthy of closer scrutiny. Other statistics worth calculating, that provide useful insight to the prediction performance, are the standard deviation (SD) and the average absolute relative deviation (AARD%) for the prediction errors of the actual values (not normalized values) of the OF for the entire tuning subset.

Having completed steps 1 to 10 the TOB learning network is established and ready to be optimized, with all the intermediate steps and calculated values available for scrutiny and analysis for a prediction with no weighting yet applied. In the next section the optimization of that TOB learning network is described.

3. Key role for optimization and sensitivity analysis in the learning process of the TOB network

The RMSE and R^2 values established for the tuned and unweighted TOB learning network are used as benchmarks for the improvements in prediction that can be achieved by optimization and sensitivity analysis applied to the network. The steps 11 to 14 described here (Fig. 2) outline a sequence that can use most optimization algorithms.

Step 11: the objective function of the optimization process is to minimize the RMSE value calculated in step 10 of the TOB learning network setup. This objective function is optimized by varying the weights (the values for w_1 to w_N, with $w_{N+1} = 0$) between the constraint limits 0 and 1, and Q, the number of top-ranking records included in the prediction calculations, between the integer limits of 2 to up to 10. Three distinct approaches can be adopted for the optimization and detailed analysis of the TOB learning network, identified as A, B, and C in Fig. 2.

A. 100% Excel.

All calculations are conducted on a Microsoft Excel spreadsheet with cell formula, using Excels Solver optimizer, which includes a generalized reduction gradient (GRG) non-linear algorithm and an evolutionary algorithm (EA). Both of these optimization algorithms are powerful fast, capable of handling quite large datasets and flexible in the sense that they offer various tuning options (population size, multi-start and alternative seeding options). This makes the formulas involved in all intermediate calculations readily visible and auditable. For tuning subsets larger than about thirty records and for large training subsets this makes the spreadsheet very large and cumbersome to manipulate. The approach is therefore more suited to small and mid-sized data sets.

B. Hybrid Program Code Plus Excel Solver.

The setup of the learning network (step 1 to the first part of step 8) are conducted using any programming language (VBA, R, Octave, MatLab, Python, etc.) for the calculations with the output placed in a spreadsheet. Steps 8 to 10 are set up as cell formulas in the spreadsheet and the Solver optimizers are deployed for the optimization process. The Solver optimizers can easily be driven by VBA code, but they need to optimize objective functions that are related to cell formulas on the spreadsheet. Hence, the need to set up steps 8 to 10 with cell formulas. It is therefore possible with one coded macro in VBA to setup the TOB learning network and then run the Solver optimizers on it. If other programming languages are used then two distinct sets of code are required, one with VBA and the setup code with the other language. In practice it is therefore typically more convenient for this hybrid approach to use VBA.

C. 100% Program Code Not Using Excel's Solver Optimizer.

Conduct all the TOB learning network set up and optimization in a programming language that does not employ Excel's Solver optimizer. This can be readily achieved in VBA (using a customized optimizer not Excel's solver), Octave, MatLab or Python. This approach makes sense for large datasets as it avoids large cumbersome spreadsheets with cell formulas to handle and adjust. An Excel spreadsheet can still be used to display and analyse the dataset and results, but few cell formulas would be involved. This alternative has the advantage of flexibility to adjust to different sized data sets quickly, but the disadvantage of the intermediate calculations only being auditable in the software code.

There are pros and cons to all three alternatives, with selection depending upon the dataset dimensions and the manner in which the application is to be deployed. For many field applications operators may prefer one or other of these alternatives to fit with software availability.

Step 12: Compare the optimized (minimum RMSE value and its associated R^2 value) weighted solution with the unweighted solution establish by Step 10. As well as verifying that a significant improvement has been achieved by the optimizer, the key information to review is the value of the weights (w_1 to w_N) that are associated with the optimum solution. Typically, the weights will be high for some independent variables (the variables having significant influence on the prediction accuracy of the TOB learning network) and zero, or close to zero, for others (the variables having no, or insignificant, influence on the prediction accuracy of the TOB learning network). Graphical analysis of the highly significant independent variables and their relationship with the dependent variable needs to be carefully assessed at this point.

It is also important to conduct sensitivity analysis at this stage to establish how robust the optimized solution is with respect to: 1) different numbers of records in the training and tuning subsets; 2) the influence of different values of Q (one will be associated with the minimized RMSE case) on the optimum solution selected and the RMSE and R^2 values of those solutions; 3) the impact of varying the optimizer's tuning parameters, or the optimizer algorithm itself (e.g., Solver's GRG versus Solver's EA).

The information gained from this sensitivity analysis and closer inspection of the relationships between the significant independent variables (those with high w values) and the dependent variable will help to decide whether the optimized learning network has been improved and modified in such a way that its predictions of the dependent variable can be relied upon.

Step 13: Evaluate the prediction performance of the selected weightings for the optimized TOB learning network for the testing subset. The testing subset includes data records representative of the systems range of dependent variable functions that have not been involved in either the training subset or the tuning subset, but for which accurate measured values of the dependent variable (OF) are available. If the TOB network is effective then the RMSE and R^2 values achieved for the testing subset should be close to that of the tuning subset, but probably displaying slightly higher RMSE and slightly lower R^2 values. Slightly poorer performance should be expected as the data records of the testing subset have not been able to influence the tuning of the TOB learning network.

Step 14: If the RMSE and R^2 values for the tuning and testing data sets achieve levels of accuracy that are sufficient to generate confidence in their prediction accuracy then deploy the TOB learning network for practical applications. If the

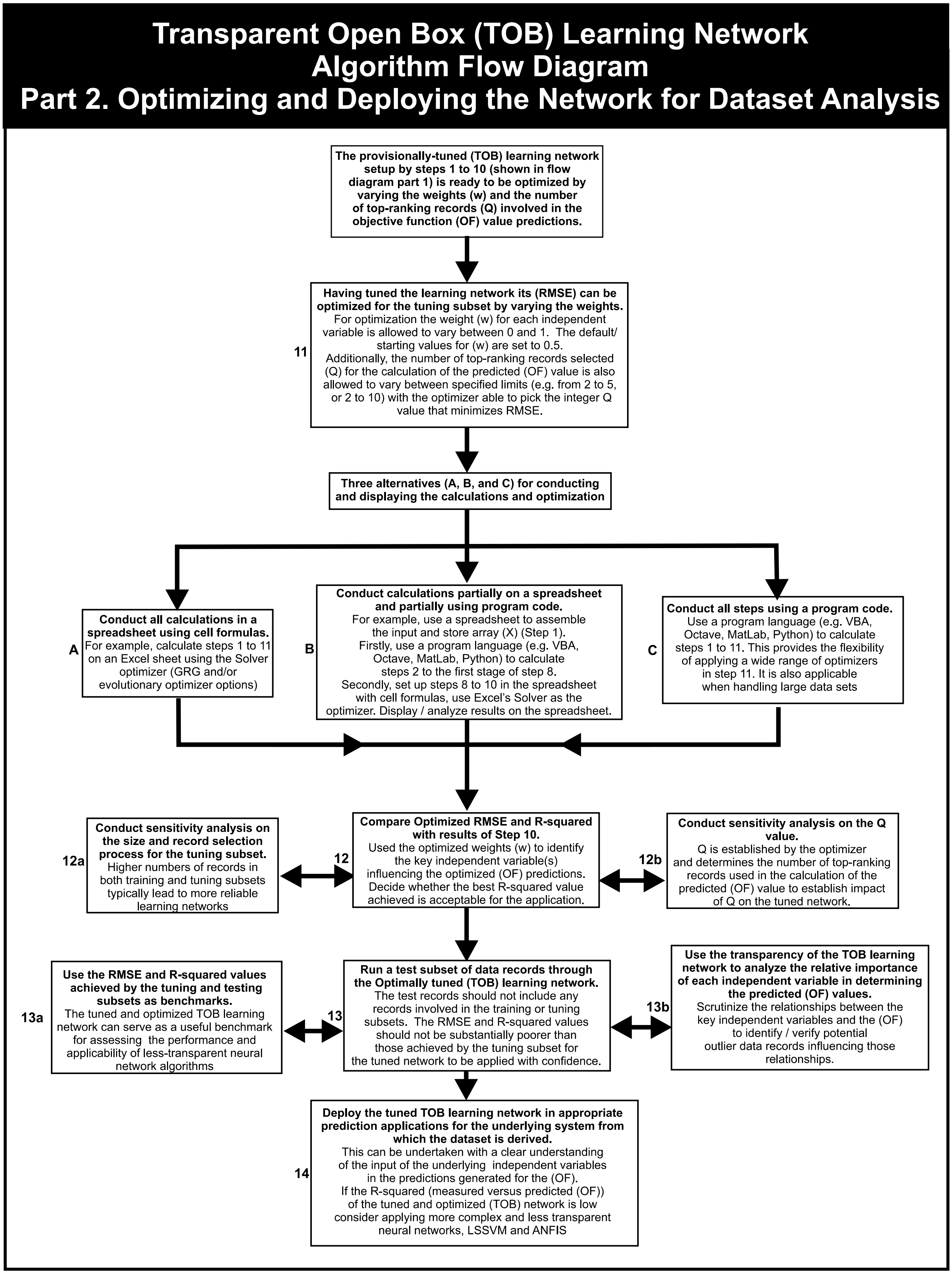


Fig. 2. Flow diagram summarizing four steps involved in optimizing and fine-tuning through sensitivity analysis a transparent open box (TOB) learning network. Steps 11 to 14 are described in more detail in the text.

level of uncertainty remains unacceptably high (i.e., RMSE too high and R^2 too low) then consider applying more complex and less transparent learning networks and machine learning algorithms (e.g., ANN, RBF, ANFIS, LSSVM) using the accuracy achieved by TOB to benchmark their performance.

It is worth noting that for some complex systems with large numbers of independent variables displaying very poor correlations with the dependent variable achieving a high R^2 value is not very likely with a TOB learning network or other less transparent networks. In such cases the question to be addressed is does a learning network (TOB or other) provide a better solution than the alternative prediction measures available. If the answer to that question is yes, then it may still be worthwhile deploying the optimized network with a low (but better-than-otherwise-achievable) R^2 prediction performance.

4. Example analysis and insight provided by the TOB learning network

Here, datasets for three petroleum-related systems are evaluated with TOB networks to predict their objective functions. Only portions of the evaluations are described to illustrate the type of insight and prediction improvements that TOB learning networks can achieve and to highlight the type of petroleum activities that can benefit from this methodology. Detailed analysis of each system is not provided, as this will be provided in papers focused specifically on each data set. The hybrid program code plus Excel Solver approach (alternative B in step 11) was used for each evaluation presented.

4.1 Dataset 1 (prediction of loss circulation while drilling)

Lost circulation is a significant issue and risk when drilling oil and gas wells and a number of machine learning algorithms have been applied in attempts to provide reliable predictions of loss severity (Sheremetov et al., 2008; Moazzeni et al., 2010). The example considered here involves datasets of drilling metrics recorded while penetrating two formations: zone A above the main reservoir; and, zone B the main reservoir. The dependent variable is the occurrence and quantity of circulation loss (i.e., loss severity in barrels/hour). There are sixteen independent variables (N = 16) defining the system, with the dependent variable being the seventeenth variable.

1. length of section drilled (feet)
2. borehole size (inches)
3. weight on bit (WOB) (tons)
4. pump rate (gallons/minute)
5. pump pressure (pounds per square inch)
6. drilling fluid viscosity (centipoise)
7. drilling fluid shear stress at shear rate 600 rpm (lb/100ft^2)
8. drilling fluid shear stress at shear rate 300 rpm (lb/100ft^2)
9. drilling fluid gel strength (shear stress in quiescent state) (lb/100ft^2)
10. drilling time (hours)

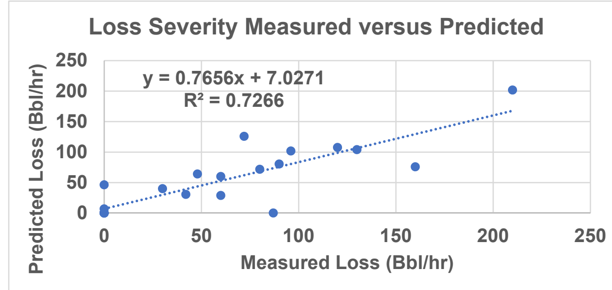


Fig. 3. Predicted versus measured loss severity for tuning data set (23 records) for Dataset 1-Formation A.

11. drilling fluid velocity (feet/second)
12. drilling fluid solids (percent)
13. drilling bit rotation speed (revolutions per minute)
14. pore pressure (pounds per square inch)
15. drilling fluid pressure (pounds per square inch)
16. formation fracture Pressure (pounds per square inch)
17. loss severity (barrels/hour)-objective function

For formation A there are 93 (M = 93) data records; for formation B there are 289 (M = 289) data records. For both formations the analysis of the TOB through to step 12 are presented using tuning subsets of 23 records in both cases; meaning that the training subset consists of 70 records for formation A and 266 records in the case of formation B. Tables 1 and 2 summarizes the prediction performance of the TOB learning networks constructed and optimized for these two formations.

For formation A the evenly weighted TOB network (involving errors treated evenly for all sixteen independent variables, and for Q = 3) achieves RMSE of 52.6 bbls/hour and R^2 of 0.3129 (predicted versus measured loss severity). This is not that impressive as the range of loss severity in the entire data set is *min* = 0 barrels/hour and *max* = 270 barrels/hour (with 35 records displaying 0 barrels/hour loss severity, i.e., no loss of circulation). Optimization of the TOB network significantly improves its prediction performance, with the minimum error achieved by the Solver GRG optimizer applying the multi-start option and a population of 150 (RMSE of 31.1 bbls/hour and R^2 of 0.7266). That optimum solution involves Q = 5 and applies non-zero weights to only the following variables (variable with highest weight listed first):

drilling fluid gel strength (variable 9) w = 1.00
drilling fluid solids (variable 12) w = 0.43
pump pressure (variable 5) w = 0.0027
pore pressure (variable 14) w = 0.0017
drilling fluid pressure (variable 15) w = 0.0017
drilling time (variable 10) w = 0.00099.

All other independent variables involve w = 0, so contribute nothing to the loss severity prediction. The two key variables influencing the optimized TOB learning prediction, drilling fluid gel strength and drilling fluid solids show poor correlations with loss severity for the 23 records of the tuning subset (R^2 = 0.1085 and 0.1912, respectively). Fig. 3 shows a cross plot of predicted versus measured loss severity for the optimized tuning subset.

Table 1. Loss severity prediction performance of TOB learning network applied to Dataset 1-Formation A.

Transparent Open Box (TOB) Learning Network Results and Variable Weightings in the Prediction of Loss Severity for Formation A (Dataset 1)									
Variable Description	Variable Number	Pre-optimization Even Weightings	Best Solution Solver GRG Multi-start	Best Solution Solver Evolutionary Algorithm	Solver GRG with no Multi-start	GRG Multi-start with Q constrained	GRG Multi-start with Q constrained	GRG Multi-start with Q constrained	GRG Multistart with Q constrained
Q Constained to	Integer #	3	2 to 10	2 to 10	2 to 10	6	4	3	2
Q selected for solution	Integer #	3	5	5	5	6	4	3	2
Prediction Performance of Optimum Solution									
RMSE	barrels/hour	52.61	31.13	31.15	32.70	31.40	33.32	43.29	46.69
R^2	%	0.3129	0.7266	0.7266	0.7102	0.7273	0.6976	0.5213	0.4628
Weightings ($0 \leq w \leq 1$) Applied to solutions									
Length drilled	1	0.5	0	<0.0001	0	<0.0001	0	0	0
Hole size	2	0.5	0	<0.0001	0.79712	0.24942	0	0	0
wieght-on-bit	3	0.5	0	<0.0001	0	0	0	0	0
Pump rate	4	0.5	0	<0.0001	0	0	0	0	0
Pump pressure	5	0.5	0.00271	0.00340	<0.0001	<0.0001	0	0.00177	0
Viscosity	6	0.5	0	<0.0001	<0.0001	0	0.00825	0.03992	0.21492
Shear stress at 600 rpm	7	0.5	0	<0.0001	0	0	0	0	0
Shear stress at 300 rpm	8	0.5	0	<0.0001	0	0	0	0	0
Gel strength	9	0.5	1.00000	0.99544	1	1	1	1	1
Drilling time	10	0.5	0.00099	0.00135	<0.0001	<0.0001	0.00265	0.01721	0.06840
Mud velocity	11	0.5	0	<0.0001	0.65324	0.00294	0.18591	0	0
Solid percent	12	0.5	0.43280	0.56492	0.01028	0.01024	0.24159	0	0
Bit rotation speed	13	0.5	0	<0.0001	0	0	0	0.00306	0
Pore pressure	14	0.5	0.00166	0.00366	0.000600339	0	0.00818	0.14226	0.02024
Mud pressure	15	0.5	0.00173	0.00064	0	0	0	0	0
Fracture pressure	16	0.5	0	<0.0001	0	0	0	0	0
Loss Severity	Obj Fn	0	0	0	0	0	0	0	0

Constraining the value of Q to values other than 5 during optimization leads to higher RMSE and lower R^2 values for the minimum solutions found. For values of Q = 6 the minimum solution is close in performance to the optimum solution with Q = 5. However, as Q is reduced progressively from 4 to 2 the prediction performance of the TOB network deteriorates significantly.

For formation B (main reservoir) the evenly weighted TOB network (involving errors treated evenly for all sixteen independent variables, and for Q = 3) achieves RMSE of 15.8 bbls/hour and R^2 of 0.1401 (predicted versus measured loss severity). This is not that impressive as the range of loss severity in the entire data set is *min* = 0 barrels/hour and *max* = 850 barrels/hour (with 91 records displaying 0 barrels/hour loss severity, i.e., no loss of circulation; and the second highest loss severity being 65 barrels/hour). Optimization of the TOB network does improve its prediction performance (Table 2), with the minimum error achieved by the Solver GRG optimizer applying the multi-start option and a population of 150 (RMSE of 13.21 bbls/hour and R^2 of 0.3073). However, a relatively low level of confidence remains in the prediction performance of the optimized TOB network.

The best optimum solution found involves Q = 6 and applies non-zero weights to only the following variables (variable with highest weight listed first):

borehole hole size (variable 2) w = 1.00

drilling fluid shear stress at shear rate 600 rpm (variable 7) w = 0.1608

drilling fluid gel strength (variable 9) w = 0.0847

length of drilled section (variable 1) w = 0.0129

drilling time (variable 10) w = 0.00032

All other independent variables involve w = 0, so contribute nothing to the loss severity prediction for the optimized solution. Note the difference in the variables selected by the formation B TOB network compared to the formation A TOB network. Only gel strength and drilling time are selected by both networks. The two key variables influencing the optimized TOB learning prediction for formation B, i.e., borehole hole size and drilling fluid shear stress at shear rate 600 rpm, show very poor correlations with loss severity for

Table 2. Loss severity prediction performance of TOB learning network applied to Dataset 1-Formation B.

Transparent Open Box (TOB) Learning Network Results and Variable Weightings in the Prediction of Loss Severity for Formation B (Dataset 1)									
Variable Description	Variable Number	Pre-optimization Even Weightings	Best Solution Solver GRG Multi-start	Best Solution Solver Evolutionary Algorithm	Solver GRG with no Multi-start	GRG Multi-start with Q constrained	GRG Multi-start with Q constrained	GRG Multi-start with Q constrained	GRG Multistart with Q constrained
Q Constained to	Integer #	3	2 to 10	2 to 10	2 to 10	5	4	3	2
Q selected for solution	Integer #	3	6	6	6	5	4	3	2
Prediction Performance of Optimum Solution									
RMSE	barrels/hour	15.76	13.21	13.22	13.44	13.61	14.41	14.82	16.53
R^2	%	0.1401	0.3073	0.3046	0.2821	0.2661	0.2076	0.1817	0.1235
Weightings ($0 \leq w \leq 1$) Applied to solutions									
Length drilled	1	0.5	0.01286	0.00814	0.01801	0.00022	1.00000	0	0
Hole size	2	0.5	1.00000	0	1.00000	1.00000	1.00000	0	0
wieght-on-bit	3	0.5	0	0	0	0	0	1.00000	0
Pump rate	4	0.5	0	0	0	0.00727	0	0	0
Pump pressure	5	0.5	0	0	0	0	0	0	0
Viscosity	6	0.5	0	0	0	0	0	0	0
Shear stress at 600 rpm	7	0.5	0.16084	1.00000	1.00000	1.00000	0	0	0
Shear stress at 300 rpm	8	0.5	0	0	1.00000	0	0	0	0
Gel strength	9	0.5	0.08469	0.93542	0.35598	0.26806	0.50069	0.00259	0
Drilling time	10	0.5	0.00032	0.00052	0.00477	0.00059	0.24511	0.00037	0
Mud velocity	11	0.5	0	0	0.00005	0.00001	0	0	0
Solid percent	12	0.5	0	0	0.36081	0.00414	0.42393	0	0
Bit rotation speed	13	0.5	0	0	0	0	0	0	0
Pore pressure	14	0.5	0	0	0	0	0.38808	0.01042	0
Mud pressure	15	0.5	0.	0	0.00482	0	1.00000	0	0.76363
Fracture pressure	16	0.5	0	0	0	0	0.17086	0	0
Loss Severity	Obj Fn	0	0	0	0	0	0	0	0

the 23 records of the data set (R^2 = 0.0394 and 0.0194, respectively). Fig. 4 shows a cross plot of predicted versus measured loss severity for the optimized tuning subset, with significant dispersion and uncertainty around a linear correlation line fitted to that data.

Constraining the value of Q to values other than 6 for the formation B TOB network during optimization leads to higher RMSE and lower R^2 values for the minimum solutions found. As Q is reduced progressively from 5 to 2 the prediction performance of the TOB network deteriorates significantly (for Q = 2, R^2 = 0.1235).

Loss of circulation is a notoriously difficult variable to predict during drilling, because it depends on a large number of uncorrelated variables and varies significantly from formation to formation (as shown for formations A and B) and from location to location in the same field. Consequently, it is not surprizing that the prediction performances of the TOB learning networks are not spectacular. However, the TOB network for formation A is potentially useable, whereas that for formation B is unlikely to help much in the prediction of loss severity. This suggests that there is scope to apply more complex neural network methodologies (using the performances of the TOB network as benchmarks to assess their performances) to try and improve prediction of loss severity in formations A and B. Also, there is scope for sensitivity analysis to provide more insight to the relationships between the key variables impacting loss severity in formations A and

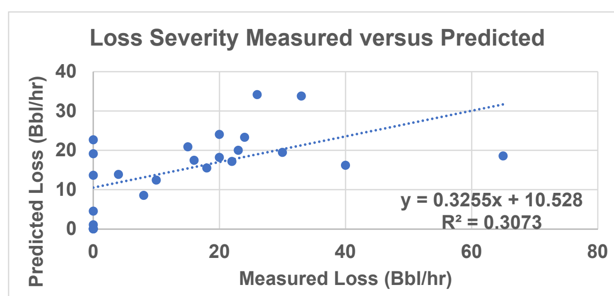


Fig. 4. Predicted versus measured loss severity for tuning data set (23 records) for Dataset 1-Formation B.

Table 3. Table 3 Bubble point pressure (Pb) prediction performance of TOB learning network applied to Dataset 2.

Transparent Open Box (TOB) Learning Network Results and Variable Weightings in the Prediction of Bubble Point Pressure (Dataset 2)									
Variable Description	Variable Number	Pre-optimization Even Weightings	Best Solution Solver GRG Multi-start	Best Solution Solver Evolutionary Algorithm	Solver GRG with no Multi-start	GRG Multi-start with Q constrained	GRG Multi-start with Q constrained	GRG Multi-start with Q constrained	GRG Multistart with Q constrained
Q Constained to	Integer #	3	2 to 10	2 to 10	2 to 10	6	5	3	2
Q selected for solution	Integer #	3	4	4	3	6	5	3	2
Prediction Performance of Optimum Solution									
RMSE	psi	76.99	44.41	44.41	75.86	50.29	47.20	75.86	73.32
R^2	%	0.9703	0.9895	0.9895	0.9711	0.9877	0.9882	0.9711	0.9745
Weightings ($0 \leq w \leq 1$) Applied to solutions									
Gas specific gravity	1	0.5	0.000002	0.000002	0.55690	0.85176	0.000001	0.189633	0
Oil specific gravity	2	0.5	0.000011	0.000011	0.60129	0.43459	0	0.204750	0
Oil gravity API	3	0.5	0	0	0	0	0.000001	0	0
Gas-to-oil ratio (GOR)	4	0.5	0	0	0.85277	1.00000	0.017804	0.290383	1.000000
Oil temperature	5	0.5	1.000000	0.999731	0.88380	0	1.000000	0.300946	0
Oil formation volume factor	6	0.5	0	0	0.07326	0.00863	0	0.024948	0.000504
Bubble point pressure	Obj Fn	0	0	0	0	0	0	0	0

B. Moreover, further analysis using different sizes of tuning data sets could help to improve the prediction performance of these TOB networks. This additional analysis is not presented here.

4.2 Dataset 2 (prediction of bubble point pressure P_b of oil produced from multiple wells drilled into a single oil field based on a pressure-temperature-volume (PVT) data analyses).

The objective function is the bubble point pressure (P_b) measured in pounds per square inch units, an important and widely predicted oil field metric (Gomaa, 2016). For the dataset evaluated here, there are 75 (M = 75) data records and six variables (N = 6) defining the system, with the dependent variable being the seventh variable.

1. specific gravity of gas (relative to air)
2. specific gravity of oil (relative to water)
3. oil gravity (measured in degrees API units)
4. gas to oil ratio (GOR) (standard cubic feet/stock tank barrel)
5. oil temperature (degrees Fahrenheit)
6. oil formation volume factor (FVF)

This set of variables is not totally independent as specific gravity of oil (γg) and oil gravity API have a perfect negative correlation with each other as they are linked through the formulaic relationship: γg = 141.5/(API+131.5). Also, there is a strong positive correlation between P_b and GOR: R^2 = 0.8874, increasing to 0.9176 if three outlier data records are ignored. It is interesting to see how these relationships influence the behaviour of the TOB learning network.

Table 3 summarizes the prediction performance of the TOB learning networks constructed and optimized for Dataset 2. This presents the analysis of the TOB through to step 12 with the dataset divided into a training set with 55 records and a tuning set of 20 records.

The evenly weighted TOB network (involving errors treated evenly for all six "independent" variables, and for Q = 3) achieves RMSE of 76.99 psi and R^2 of 0.9703 (predicted versus measured Bp). This is quite impressive prediction accuracy as the range of B_p in the entire data set is min = 2,150 psi and max = 4,250 psi. Optimization of the TOB network does improve its prediction performance, with the minimum error achieved by the Solver GRG optimizer applying the multi-start option and a population of 150 (RMSE of 44.41 psi and R^2 of 0.9895). That optimum solution involves Q = 4 and applies non-zero weights to only the following variables (variable with highest weight listed first):

oil temperature (variable 5) w = 1.0
oil gravity (variable 2) w = 1.08E-05
gas gravity (variable 1) w = 2.179E-06

All other independent variables involve w = 0, so contribute nothing to the loss severity prediction for the optimized solution. This is interesting as at first sight it implies that GOR (which is highly correlated with measured B_p in the dataset) has contributed nothing to the TOB network's prediction of B_p. While it is certainly true that in the optimization sequence (steps 11 to 12) GOR has been eliminated in the B_p prediction process for the tuning subset, it has played a significant role in steps 7 and 8 during the development of the TOB learning network. The selection of the top-ranking matches in the training set for each record in the tuning set is made by ident-

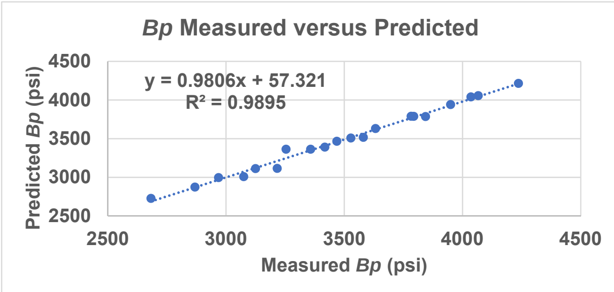


Fig. 5. Predicted versus measured bubble point pressure (B_p) for tuning data set (20 records) for Dataset 2.

ifying those with the lowest squared error. A close match between the GOR of the record being tested with those top-ranking matches is where GOR has had its influence. What is noteworthy is that once those top-ranking records are selected for further optimization the differences between their GOR values is small, making them a less sensitive metric for the optimizer to focus upon. The optimizer finds oil temperature, which displays a relatively poor positive correlation with B_p (R^2 = 0.1218) a more sensitive discriminator in those top-ranked matching records selected for optimizing the B prediction performance of the tuning subset. These subtleties of the optimization process are only possible because of the transparent nature of the TOB learning network and its display of the intermediate calculations.

Fig. 5 shows a cross plot of predicted versus measured loss severity for the optimized tuning subset, with very little dispersion and uncertainty around the linear correlation line fitted to that data. This optimized network could clearly be deployed with confidence to predict B_p in this oil field, subject to the further sensitivity analysis described in step 13.

Constraining the integer value of Q to values other than the optimizer selection of 4 for the Dataset 2 TOB network during optimization leads to higher RMSE and lower R^2 values for the minimum solutions found (Table 3). For a Q value of 5 the optimized result is only slightly worse than for the optimum Q = 4, and the solution is still dominated by oil temperature, although GOR also contributes as the second-most-important variable. When Q is constrained to a value of 6 the optimized prediction depends on GOR, gas gravity and oil gravity as highly weighted variables to derive its predictions (with no contribution from oil temperature). When Q is constrained to a value of 3 the optimized prediction depends on oil temperature and GOR (with almost equal weightings) and gas gravity, oil gravity and FVF all contributing as weighted variables to derive its predictions. When Q is constrained to a value of 2, there is little improvement in the optimized solution (*RMSE* = 73.32 psi, R^2 = 0.9745 compared to the evenly-weighted TOB networks setup solution. For Q constrained to a value of 2 the optimized prediction depends almost exclusively on GOR with a minor weighted contribution from FVF (with no contribution from oil temperature). These results highlight the value of conducting sensitivity analysis on the Q value applied to the TOB learning network.

It can be concluded that there is relatively limited scope for the less transparent learning networks and machine-learning algorithms to improve upon the prediction process of the TOB learning network for dataset 2. Indeed, if such networks are able to make prediction improvements for this dataset, it will not be possible to interrogate the relative contributions of the independent variables to their optimum solutions in the way described here for the TOB learning network.

4.3 Dataset 3 (prediction of liquid flow rate from well test data from a producing oil field)

The objective function is liquid flow rate (Q_L) measured in units of stock tank barrels per day (where liquid involves oil plus produced water). Artificial intelligence algorithms have been extensively applied for predicting flow rates through wellhead chokes (Elhaj et al., 2015; Choubineh et al., 2017). For the dataset evaluated here, there are 180 (M = 180) data records and four variables (N = 4) defining the system, with the dependent variable being the fifth variable.

1. wellhead choke size (in one-sixty-fourths of an inch)
2. wellhead pressure (pounds per square inch)
3. base solids and water (BS&W) (as a percentage of produced fluids)
4. gas to liquid ratio (GLR) (standard cubic feet/stock tank barrel)

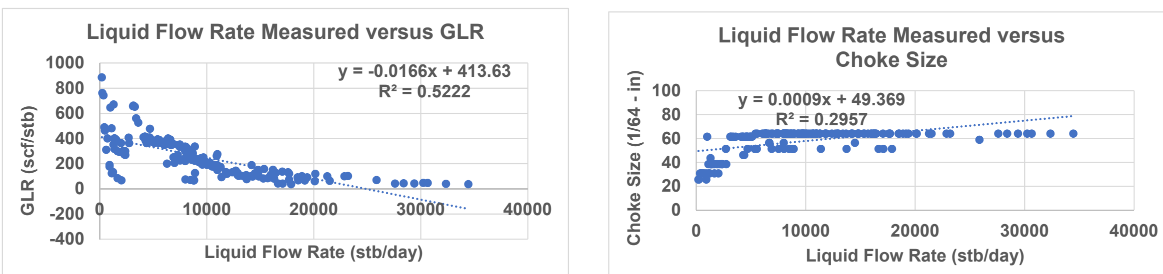
The independent variables display distinctive individual relationships with Q_L for dataset 3: gas to liquid ratio shows a moderate negative correlation with Q_L (R^2 = -0.5222); BS&W shows a poor negative correlation with Q_L (R^2 = -0.0656); Wellhead pressure shows a poor positive correlation with Q_L (R^2 = 0.0731); and, choke size shows a moderate positive correlation with Q_L (R^2 = 0.2957). In fact, the correlation between choke size and Q_L would be higher were it not for the fact that the maximum choke size is capped at 64/64 inches for the well tests (112 of the 180 data records involve a choke size of 64/64 inches). Again, it is interesting to see how these relationships influence the behaviour of the TOB learning network.

Table 4 summarizes the prediction performance of the TOB learning networks constructed and optimized for Dataset 3. This presents the analysis of the TOB through to step 12 with the dataset divided into a training set with 180 records and a tuning set of 23 records.

The evenly weighted TOB network (involving errors treated evenly for all four independent variables, and for Q = 3) achieves RMSE of 752.61 psi and R^2 of 0.9812 (predicted versus measured Q_L). This is an impressive prediction accuracy as the range of Q_L in the entire data set is *min* = 200 stb/day and *max* = 35,000 stb/day. Optimization of the TOB network does improve its prediction performance, with the minimum error achieved by the Solver GRG optimizer applying the multi-start option and a population of 150 (RMSE of 434.97 stb/day and R^2 of 0.9936). That optimum is also achieved using Solver's evolutionary optimization algorithm. The optimum solution involves Q = 4 and applies non-zero weights to only the following variables (variable with highest weight listed first):

Table 4. Oil flow rate (QL) prediction performance of TOB learning network applied to Dataset 3.

Transparent Open Box (TOB) Learning Network Results and Variable Weightings in the Prediction of Liquid Flow Rate (Dataset 3)									
Variable Description	Variable Number	Pre-optimization Even Weightings	Best Solution Solver GRG Multi-start	Best Solution Solver Evolutionary Algorithm	Solver GRG with no Multi-start	GRG Multi-start with Q constrained	GRG Multi-start with Q constrained	GRG Multi-start with Q constrained	GRG Multistart with Q constrained
Q Constained to	Integer #	3	2 to 10	2 to 10	2 to 10	6	5	3	2
Q selected for solution	Integer #	3	4	4	4 6	5	3	2	
Prediction Performance of Optimum Solution									
RMSE	stb/day	752.61	434.97	434.97	721.01	508.71	499.57	664.39	721.01
R^2	%	0.9812	0.9936	0.9936	0.9835	0.9931	0.9926	0.9854	0.9835
Weightings ($0 \leq w \leq 1$) Applied to solutions									
Wellhead choke size	1	0.5	0.94926	0.96947	1.00000	1.00000	1.00000	1.0000000	0.74428
Wellhead pressure	2	0.5	0	0	0.00316	0	0	0	0.00235
BS&W	3	0.5	0	0	0.00063	0	0	0	0.00047
Gas-to-liquid ratio (GLR)	4	0.5	0.00145	0.00147	0.07705	0.00020	0.00056	0.0002869	0.05735
Liquid flow rate	Obj Fn	0	0	0	0	0	0	0	0



(a) (b)

Fig. 6. Relationships between the two key independent variables and measured liquid flow rate (Q_L) (the objective function) for all records (180 records) in Dataset 3.

wellhead choke size (variable 1) w = 0.9493

gas to liquid ratio (variable 4) w = 0.0015

The other independent variables involve w = 0, so contribute nothing to the loss severity prediction for the optimized solution. This is logical as choke size and GLR display moderate correlations with Q_L that complement each other. GLR shows a wide dispersion at low Q_L values (Fig. 6(a)), so is not much use as a discriminator in that region of the data set. On the other hand, Choke size being capped at 64/64 inches shows little variation across the higher flow rate data records in the data set (Fig. 6(b)). The fact that the optimum TOB learning network solution uses both of these variables highlights its ability to select and identify the most useful variables in optimizing its prediction of liquid flow rate. These subtleties of the optimization process are revealed due to the transparent nature of the TOB learning network and its display of the intermediate calculations.

Fig. 7 shows a cross plot of predicted versus measured liquid flow rate for the optimized tuning subset, with almost no dispersion and uncertainty around the linear correlation line fitted to that data. This optimized network could clearly be deployed with confidence to predict liquid flow rate (Q_L) in this oil field, subject to the further sensitivity analysis described in step 13.

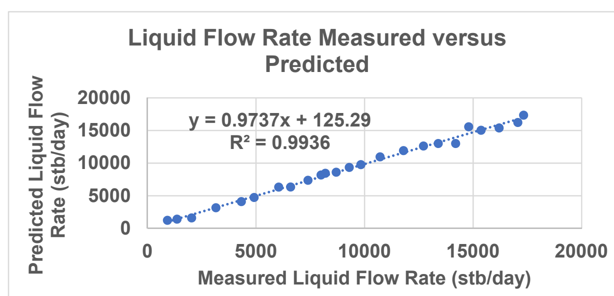


Fig. 7. Predicted versus measured liquid flow rate (Q_L) for tuning data set (23 records) selected for Dataset 3.

Constraining the integer value of Q to values other than the optimizer selection of 4 for the Dataset 3 TOB network during optimization leads to higher RMSE and lower R^2 values for the minimum solutions found (Table 4). For a Q value of 5 the optimized result is only slightly worse ($RMSE$ = 499.57 stb/day; R^2 = 0.9926) than for the optimum Q = 4, and the solution remains controlled by choke size and a minor contribution from GLR (w_{GLR} = 0.0006). When Q is constrained to a value of 6 the optimized Q_L prediction is slightly worse than for Q = 5 ($RMSE$ = 508.71 stb/day; R^2 = 0.9931) with GLR making an even smaller contribution (w_{GLR} = 0.0002). When Q is constrained to a value of 3 the optimized prediction deteriorates further ($RMSE$ = 664.39 stb/day; R^2 = 0.9854), but with choke size and GLR still the only variables selected to contribute to the optimized prediction. When Q is constrained to a value of 2, there is little improvement in the optimized solution ($RMSE$ = 721.00 stb/day; R^2 = 0.9835) compared to the evenly-weighted TOB network's setup solution. For Q constrained to a value of 2 the optimized prediction remains dominated by choke size followed by GLR, but with contributions from wellhead pressure and BS&W. These results highlight the value of conducting sensitivity analysis on the Q value applied to the TOB learning network to provide insight to how the variables are contributing to the optimum solutions.

It can be concluded that there is relatively limited scope for the less transparent learning networks and machine-learning algorithms to improve upon the prediction process of the TOB learning network for dataset 3. Indeed, if such networks are able to make prediction improvements for this dataset, it will not be possible to interrogate the relative contributions of the independent variables to their optimum solutions in the way described here for the TOB learning network. Equations/formulas linking the independent variables to liquid flow rate that have been established using well tests from many oil fields or across regions are traditionally used for predicting Q_L. The problem with the formulaic approach is that in detail the relationships between independent variables and the dependent variable vary for each field/reservoir. Hence, trying to apply a single formulaic relationship universally is likely to lead to errors and sub-optimal R^2 (prediction versus measured Q_L). The TOB learning network offers a middle ground between formulaic solutions and opaque networks and machine learning algorithms; it can provide high degrees of precision, but it can also identify the relative contributions of the independent variables in determining those predictions.

5. Discussion

The detailed description provided here of the novel TOB learning network demonstrates that it can be constructed in a systematic and simple sequence of steps, and it can be applied easily to wide range of small and mid-sized datasets using spreadsheets, coded algorithms or a hybridized combination of the two. The TOB networks applied to example data sets to derive useful predictions illustrate that not only can the TOB prediction capability be significantly improved using optimizers, it can also provide significant and useful insight to the key variables contributing to the optimized solutions. In this sense the TOB learning networks can provide more useful information about the datasets they are applied to than many of the less transparent neural networks and machine learning algorithms (ANN, RBF, ANFIS, LSSVM) that are now commonly deployed to perform such predictions.

The main improvements and/or benefits of the TOB algorithm proposed are that it provides transparency and quantifies the contributions made by the underlying variables to the predictions that it generates. Neural networks and other machine learning algorithms typically fail to do this or must be coupled with complex simulation algorithms to do so. Also, by monitoring through sensitivity analysis the prediction performance of TOB networks in terms of the Q-factor during optimization, insight to the under-fitting and over-fitting tendencies of specific solutions can be gained.

Future work is planned to apply the TOB learning network to various datasets and compare its prediction performance with that of ANN, RBF, ANFIS and LSSVM. It is not the intention here to propose that the TOB network should be used instead of these less transparent networks, but rather that it should be used as a benchmark to measure the relative prediction improvements and performance that such learning networks can provide.

Petroleum-industry-related datasets that are potentially attractive to work on with the TOB network approach include: 1) potentially improving the prediction of source rock characteristics (quality, quantity and maturity, identification of sweet spots) from wireline and measurement while drilling data through specific formations, for which there is some geochemistry and thermal maturity measured data to work with; 2) predicting other shale characteristics from logged information; and 3) predicting wellbore cleaning performance in complex wellbore designs from drilling datasets. In addition, to many other potential surface and sub-surface oil and gas industry applications, the TOB learning network methodology also has many potential applications in other industrial sectors that need to predict behaviours of variables dependant on a range of poorly correlated variables and non-linear relationships.

6. Conclusions

The transparent open box (TOB) learning network as described and implemented offers a novel tool with several advantages over established machine-learning approaches to predicting system behaviours that are dependent on a number of variables related in a non-linear and/or a poorly correlated manner involving much uncertainty. These include its ability to:

- readily reveal the intermediate calculations involved in its predictions;
- clearly display the relative contributions of independent variables in predicting dependent variable values;
- achieve significant learning and prediction improvements through the application of standard optimizers, gradient descent and/or evolutionary (including Excel's Solver options);
- provide insight to under-fitting and over-fitting tendencies based on sensitivity analysis of its Q factor;
- be effectively implemented on spreadsheets, by stand-alone code or a combination of the two;
- act as an intermediate step towards deploying more com-

plex machine learning and soft computing algorithms to predict outcomes of more complex systems.

The information gained, and performance observed, from optimized TOB learning networks can act as a useful benchmark for less transparent and more complex neural networks and machine learning algorithms applied to the same datasets.

There are many surface and sub-surface datasets in the petroleum industry that lend themselves to the benefits provided by learning networks. Systems in which a range of variables can be measured quickly and relatively cheaply (e.g., from wireline and other down-hole logs), but for which certain key associated variables are expensive and time-consuming to measure directly, requiring laboratory analysis (e.g., PVT analysis; geochemistry/thermal maturity; several geomechanical attributes of prospective formations; wellbore fluid performance), if they can be measured reliably at all (e.g., drilling fluid loss of circulation severity). The TOB learning network offers a powerful and auditable soft computing algorithm that can be deployed at the field/reservoir level to provide useful prediction and insight to such systems. The TOB methodology also offers potential benefits that can be deployed in many industries and data analysis/prediction applications.

Acknowledgments

This work is influenced by my interactions with many researchers around the world on machine learning algorithms over many years. That has stimulated my desire to develop more transparent learning-based algorithms.

References

Aalst, W.M.P., Rubin, V., Verbeek, H.M.W., et al. Process mining: A two-step approach to balance between underfitting and overfitting. Softw. Syst. Model. 2010, 9(1): 87-111.

Behnke, S. Hierarchical neural networks for image interpretation. Lect. Notes Comput. Sci. 2003, 2766(3): 1345-1346

Choubineh, A., Ghorbani, H., Wood, D.A., et al. Improved predictions of wellhead choke liquid critical-flow rates: modelling based on hybrid neural network training learning based optimization. Fuel 2017, 207: 547-560.

Cortes, C., Vapnik, V.N. Support-vector networks. Mach. Learn 1995, 20(3): 273-297.

Cybenko, G. Approximation by superpositions of a sigmoidal function. Math. Control Signal. 1989, 2(4): 303-314.

Elhaj, M.A., Anifowose, F., Abdulraheem, A. Single gas ow prediction through chokes using articial intelligence techniques. Paper SPE 177991 Presented at Society of Petroleum Engineers Saudi Arabia Section Annual Technical Symposium and Exhibition, Al-Khobar, Saudi Arabia, 21-23 April, 2015.

Elkatatny, S., Tariq, Z., Mahmoud, M. Real time prediction of drilling fluid rheological properties using Artificial Neural Networks visible mathematical model (white box). J. Pet. Sci. Eng. 2016, 146: 1202-1210.

Farley, B., Clark, W.A. Simulation of self-organizing systems by digital computer. IEEE Trans. Inf. Theory 1954, 4(4): 76-84.

Gomaa, S. New bubble point pressure correlation for middle east crude oils. Int. Adv. Res. J. Sci. Eng. Technol. 2016, 3(12): 1-9.

Hebb, D. The organization of behavior. Science 1961, 133(3466): 1749-1757.

Heinert, M. Artificial neural networks-how to open the black boxes. Application of Artificial Intelligence in Engineering Geodesy, 2008.

Ince, D.C. Collected Works of AM Turing-Mechanical Intelligence. Mechanical intelligence, 1992.

Jang, J.S. ANFIS: Adaptive-network-based fuzzy inference system. IEEE Trans. Syst. Man. Cybern. Syst. 1993, 23(3): 665-685.

Li, Y., Xie, R., Yu, L. Flow pattern identification of oil-gas-water three-phase flow based on NPSO-LSSVM algorithm. J. Theor. Appl. Inf. Technol. 2013, 48(2): 933-938.

McCulloch, W., Pitts, W. A logical calculus of ideas immanent in nervous activity. Bull. Math. Biol. 1943, 5(4): 115-133.

Meng, M., Zhao, C. Application of support vector machines to a small-sample prediction. Adv. Pet. Explor. Dev. 2015, 10(2): 72-75.

Moazzeni, A.R., Nabaei, M., Ghadami Jegarluei, S. Prediction of lost circulation using virtual intelligence in one of the Iranian oil fields. Paper SPE 136992 Presented at SPE Nigeria Annual International Conference and Exhibition, Tinapa-Calabar, Nigeria, 31 July-7 August, 2010.

Rosenblatt, F. The Perceptron: A probabilistic model for information storage and organization in the brain. Psychol. Rev. 1958, 65(6): 386-408.

Santos, R.B., Rupp, M., Bonzi, S.J., et al. Comparison between multilayer feedforward neural networks and a radial basis function network to detect and locate leaks in pipelines transporting gas. Chem. Eng. Trans. 2013, 32: 1375-1380.

Scarselli, F., Tsoil, A.C. Universal approximation using feedforward neural networks: A survey of some existing methods, and some new results. Neural Netw. 1998, 11(1): 15-37.

Schmidhuber, J. Learning complex, extended sequences using the principle of history compression. Neural Comput. 1992, 4(2): 234-242.

Schmidhuber, J. Deep learning in neural networks: An overview. Neural Netw. 2015, 61: 85-117.

Sheremetov, L., Batyrsgin, I., Filatov, D., et al. Fuzzy expert sSystem for solving lost circulation problem. Appl. Soft Comput. 2008, 8(1): 14-29.

Sugeno, M., Kang, G. Structure identification of fuzzy model. Fuzzy Sets. Syst. 1988, 28(1): 15-33.

Suykens, J.A.K., Vandewalle, J. Least squares support vector machine classifiers. Neural Process. Lett. 1999, 9(3): 293-300.

Werbos, P.J. Beyond regression: New tools for prediction and analysis in the behavioral sciences. Boston, Harvard

University, 1975.
Yavari, H., Sabah, M., Khosravanian, R., et al. Application of adaptive neuro-fuzzy inference system and mathematical ROP models for prediction of drilling rate. J. Oil Gas Sci. Technol. 2018.
Zamani, H.A., Rafiee Taghanaki, S., Karimi, M., et al. Implementing ANFIS for prediction of reservoir oil solution gas-oil ratio. J. Nat. Gas Sci. Eng. 2015, 25: 325-334.