Prediction of Condensate-to-Gas Ratio for Retrograde Gas Condensate Reservoirs Using Artificial Neural Network with Particle Swarm Optimization

Sohrab Zendehboudi,§† Mohammad Ali Ahmadi,‡ Lesley James,§ and Ioannis Chatzis†

†Chemical Engineering Department, University of Waterloo, Waterloo, ON, Canada, N2L 3G1
‡Faculty of Petroleum Engineering, Petroleum University of Technology, Ahwaz, Khuzestan, Iran
§Faculty of Engineering and Applied Science, Memorial University of Newfoundland, St. John’s, NL, Canada, A1B 3X5

ABSTRACT: Condensate-to-gas ratio (CGR) plays an important role in sales potential assessment of both gas and liquid, design of required surface processing facilities, reservoir characterization, and modeling of gas condensate reservoirs. Field work and laboratory determination of CGR is both time consuming and resource intensive. Developing a rapid and inexpensive technique to accurately estimate CGR is of great interest. An intelligent model is proposed in this paper based on a feed-forward artificial neural network (ANN) optimized by particle swarm optimization (PSO) technique. The PSO-ANN model was evaluated using experimental data and some PVT data available in the literature. The model predictions were compared with field data, experimental data, and the CGR obtained from an empirical correlation. A good agreement was observed between the predicted CGR values and the experimental and field data. Results of this study indicate that mixture molecular weight among input parameters selected for PSO-ANN has the greatest impact on CGR value, and the PSO-ANN is superior over conventional neural networks and empirical correlations. The developed model has the ability to predict the CGR with high precision in a wide range of thermodynamic conditions. The proposed model can serve as a reliable tool for quick and inexpensive but effective assessment of CGR in the absence of adequate experimental or field data.

1. INTRODUCTION

Gas condensates are highly valued hydrocarbon resources. Gas condensate reservoirs are characterized by their complicated flow regimes and intricate thermodynamic behaviors.1−3 Multi-phase flow of both gas and liquid condensates is expected during the production of a gas condensate reservoir with dynamic phase behavior from the porous rock to the surface facilities. The Arun field in Indonesia, South Pars field in Iran, and Greentown field in the USA with condensate-to-gas ratio (CGR) ranging from 10 to 300 STB/MMSCF are very large gas condensate reservoirs in the world.4−6 Condensate production due to pressure drop in the reservoir improves project economics through condensate sales as produced liquid condensate is the only potential source of revenue in the severe case of a poor gas market.3−5

Gas condensate reservoirs differ from conventional gas reservoirs in their thermodynamic and flow behaviors. Prediction and optimization of hydrocarbon recovery in this type of reservoirs require careful reservoir analysis, development plan, and project management.1−4 Isothermal condensation of liquids lowers the gas deliverability in the reservoir due to condensate blockage.6,7 Low gas relative permeability in the gas−liquid region near the production well is the main reason behind reduction of gas flow rate.1−8 Proper production strategy for gas condensate reservoirs can minimize or even avoid condensate blockage, however the prediction of key parameters like CGR is necessary. High accuracy in prediction methods is needed in order to provide an overall acceptable estimate of the performance of gas condensate reserves before the implementation of a proper gas recovery plan. Numerical and analytical modeling, PVT studies, and characterizations of retrograde gas condensate reservoirs remain challenging to the reservoir engineers as a huge amount of time and energy is required to carry out such studies mentioned above, efficiently.9−10

The phase behavior of a gas condensate reservoir is strongly dependent on the P–T envelope and thermodynamic conditions of the hydrocarbon mixture. Proper thermodynamic knowledge of phase behavior in this type of reservoirs is essential in order to predict the recovery factor and CGR. Several approaches are available in the literature for performance analysis of retrograde gas condensate reservoirs.7−13 Thomas et al. (2009) proposed a systematic strategy for performance prediction of gas condensate reservoirs based on relative permeability and two-phase dynamic measurements.11

Although measuring nonlinear parameters such as CGR and dewpoint pressure is viable through some tests (e.g., Constant Composition Expansion (CCE) and Constant Volume Depletion (CVD)), however, the laboratory and field tests are quite costly and people and time intensive.16,17 As quick, inexpensive and efficient alternatives, other techniques that include theoretical or/and laboratory tests combined with artificial intelligent systems were suggested by a number of researchers.18−21 Several models have been reported in the literature to predict flow and thermodynamic behaviors of gas condensate reservoirs.12,13,15,22−26 Clearly, prediction of production rate is central in design of wellhead equipment and surface facilities. Alavi et al. (2010) used a modified Peng-Robinson equation of state in

Received: January 4, 2012
Revised: April 18, 2012
Published: April 22, 2012
their model combined with some experiments to determine physical properties of gas condensate mixtures. Their main goal was developing an appropriate design to optimize the amount of the produced condensate in a gas reservoir when pressure decreases below the dewpoint.\textsuperscript{39,40} Jokhio et al. (2002) employed the Whitson and Fevang's model to estimate the extent of produced liquid condensate.\textsuperscript{24,25} Al-Farhan et al. (2006) examined the changes in CGR at various process conditions with considering a three-stage separation process optimized using a four-layer back propagation (BP) artificial neural network (ANN).\textsuperscript{26} A number of evolutionary algorithms such as genetic algorithm (GA), pruning algorithm, and back propagation were also used to determine the network structure and its associated parameters (e.g., connection weights).\textsuperscript{26–32} England (2002) developed a global polynomial correlation for CGR in terms of \( P \) and \( T \) based on a PVT study under specific thermodynamic and geological conditions.\textsuperscript{33}

Accurate determination of PVT properties such as CGR is necessary to predict and more accurately history match production rates and recovery factors. Many correlations exist in the literature based on PVT data for various types of hydrocarbon mixtures. Most of these correlations were obtained with linear or nonlinear multivariable regression or graphical techniques where some are not very accurate.\textsuperscript{33–35} Artificial neural network (ANN) systems are potentially reliable tools for the prediction of PVT parameters such as CGR in gas condensate reservoirs if the networks are effectively trained. The enormous interconnections in the ANN offer a wider degree of freedom or a large number of fitting variables. Therefore, ANN is better able to describe the nonlinearity of the process compared to classical regression methods. Another advantage of artificial neural networks is that they are dynamically adaptive where they are able to learn and adjust to new conditions in which the efficiency of ANN is not adequate with old situations. In addition, the ANN model is capable of handling systems with multiple inputs and outputs.\textsuperscript{39,40}

This paper highlights the economic significance, production characteristics, production methods, and P-T behavior of gas condensate reservoirs with a brief description of common techniques for determining the CGR. An intelligent model is developed, based on a feed-forward artificial neural network (ANN) optimized by using particle swarm optimization (PSO), and examined to predict CGR in terms of dewpoint pressure, temperature and mixture molecular weight for retrograde gas condensate reservoirs through a rapid and accurate manner. PSO is employed to determine the initial weights of the parameters used in the ANN model. The PSO-ANN model is tested using extensive field and laboratory data. Results from this study reveal that the developed model can predict the CGR with high accuracy. The developed model can serve as a reliable tool for quick and cheap but efficient estimation of CGR in the absence of sufficient laboratory or/and field data especially during the initial stages (e.g., determination of number of wells, well placement, and estimation of optimum production rate) of development of the gas condensate reservoirs.

2. RETROGRADE GAS CONDENSATE RESERVOIRS

A typical P-T phase diagram for gas condensate reservoirs (Figure 1) has a critical temperature lower than the reservoir temperature. However, its cricondentherm is larger than the reservoir temperature. Gas condensate systems commonly appear as a single gas phase in the reservoir at the time of discovery. As the reservoir is produced, the reservoir pressure declines from the reservoir to the production well and to the surface facilities, resulting in gas condensation to produce a liquid phase. This isothermal condensation due to the reduction of pressure below the dewpoint pressure of the original hydrocarbon mixture is called retrograde condensation.\textsuperscript{1–4,11}

Gas condensate reservoirs generally produce in the range of 30–300 STB/MMSCF (standard barrels of liquid per million standard cubic feet of gas).\textsuperscript{5,12,15,23,33–38,41–47} The depth of majority of the reported retrograde gas condensate fields ranges from 4000–10,000 ft (∼1200–3000 m).\textsuperscript{5,12,15,23,33–38,41–47} Also, the ranges of pressure (\( P \)) and temperature (\( T \)) for this type of gas reservoirs are usually between 3000–8500 psi (20–55 MPa) and 150–400 °F (65–200 °C), respectively.\textsuperscript{5,12,15,23,33–38,41–47} Clearly, few gas condensate reservoirs may be found out of the ranges reported above for \( T, P \), and depth as gas condensate reservoirs may have pressures lower than 2000 psi and temperatures under 100 °F.\textsuperscript{47} The pressure and temperature values along with the wide range of compositions result in gas condensate mixtures that exhibit different and complex thermodynamic behaviors.\textsuperscript{1–4,10,11}

Since the formation has a lower permeability to liquid and the ratio of liquid viscosity to gas viscosity is fairly high in the gas-condensate reservoirs, the main portion of the condensed liquid in the reservoir is unrecoverable and considered as a condensate loss. Condensate loss is a main concern in gas reservoir engineering because the liquid condensate holds the more economically valuable intermediate and heavier components of the original hydrocarbons that are trapped in the reservoir.\textsuperscript{1–12}

Characterizing gas-condensate systems is a complex task for researchers and subject matter experts, since multiphase flow in the formation and also variations of the fluid composition significantly obscure the analysis of well tests. In the research area of gas-condensate reservoirs, the topics such as well deliverability, PVT analysis, well test interpretation, multiphase flow, rock characterization, relative permeability model, viscous stripping of condensate and interblock non-Darcy flow have been the common problems for long time.\textsuperscript{1–15}

2.1. Three Flow Regimes. According to Fevang (1995), three main different flow regimes can be identified in the hydrocarbon mixture flowing toward a production well in a gas-condensate reservoir during depletion, as depicted in Figure 2.\textsuperscript{1–25}

- Single phase gas (Region 3): A region that is far from the production well and has a pressure higher than the dewpoint value and thus only single gas phase exists in the region.
2.1.1. Condensate Blockage. Gas mobility has the maximum value further away from the wellbore where reservoir conditions dominate and gas saturation is 100% as shown in Figure 2, Region 3. Gas mobility declines significantly in Region 2 where condensate accumulates and lower gas saturation is observed. Region 1 shows an even lower gas saturation and higher liquid condensate saturation such that the liquid condensate is mobilized. Even if the condensate production is negligible, the gas mobility and consequently the well deliverability remain considerably low. This phenomenon is known as “condensate blockage”. The major resistance for gas to flow is in Region 1, and the blockage influence depends mostly on the relative permeability of the gas phase and the volume of Region 1. The gas flow resistance in Region 2 is much lower than that in Region 1.

2.1.2. Hydrocarbon Recovery and Composition Change. Composition of the gas varies as it moves toward Region 2 and the flowing gas becomes leaner of intermediate and heavy components (e.g., C4+) in the reservoir. Hence, the heavier components build up in Regions 1 and 2 as the reservoir pressure declines. Overall, the composition of the mixture everywhere in Region 1 or 2 contains heavier components when the reservoir pressure drops below the dewpoint. Region 1, near wellbore with significant liquid condensate saturation, can rapidly develop during production. This will result in producing leaner gas compared to the initial gas composition and leaving the intermediate and heavier components in Regions 1 and 2.

2.2. Production Methods and Pressure Maintenance. Production and/or recovery methods for gas condensate reservoirs usually include fast depletion, slow depletion, full dry gas cycling, partial dry gas cycling, seasonal dry gas cycling, noncondensable gas (NCG) injection, injection of tuned gas composition, and water injection. Most of these methods are employed for the purpose of pressure maintenance. In general, pressure maintenance processes are classified in four main groups: 1) an active water drive after modest decline of pressure from early production, 2) water injection techniques, 3) gas injection, or (4) combinations of these. As times goes on, some gas condensate reservoirs may experience thermodynamic conditions close to their critical points. In this case, reservoirs appear to be good candidates for some recovery methods such as injection of tuned gas compositions to offer miscibility and phase-transform processes that can enhance the recovery factor.
2.2.1. Depletion. Natural depletion in the form of fast and slow depletions involves just production. No dry or residue gas is recycled to the reservoir. Condensate recovery is effective just at the beginning of production when the produced mixture holds all original components of liquefiable hydrocarbons.\textsuperscript{3,25,41–43} The reservoir pressure is high enough (greater than the dewpoint pressure) as there is only gas phase in the formation at this stage, referred to as “fast depletion”. As the reservoir produces hydrocarbons and the pressure drops below the dewpoint, the retrograde condensation process starts. The heavier components of hydrocarbons mixture condense in the reservoir. The percentage of condensate phase is variable and depends on the composition of the mixture.\textsuperscript{3,25,41–43} Recovery of the liquid condensates at lower pressures is not guaranteed. Below the dewpoint, the relative permeability of gas phase declines causing slow depletion of the fluids. Methods have been developed in order to maintain pressure, enhance the recovery, and finally produce condensate at the surface. Advantages and disadvantages of depletion processes include the following:\textsuperscript{3,25,41–43}

**Advantages:**
- High recovery factors for gas under solution gas regimes
- Low cost for reservoir development
- Reservoir studies with fairly low cost needed

**Disadvantages:**
- Low liquid recovery efficiencies and almost irreversible loss of retrograde condensate which is left in the formation
- Reduction in well production rates when reservoir pressure declines and retrograde condensation happens in the bottom hole
- Extra compression needed as reservoir pressure is reduced
- Difficulty in managing the production process when there is an active water drive in the reservoir
- Difficulty in managing production wells operating under notably declined reservoir pressure

2.2.2. Water Injection. Only a few gas condensate reservoirs are reported to operate under natural water drive (aquifer expansion).\textsuperscript{3,25,41–43} A strong water drive is needed in order to maintain high enough pressure to reduce condensates from forming and be economical. Therefore, a careful process design and an appropriate economic analysis should be implemented if a water injection process is selected as a production method or as a technique for pressure maintenance.\textsuperscript{3,25,41–43} Other important aspects such as displacement efficiency of water pushing gas, bypassing potential, and loss of condensate phase should be considered when water breakthrough occurs via permeable zones. The mobility ratio for water flooding is satisfactorily low, leading to attaining high areal sweep efficiencies.\textsuperscript{3,25,41–43} The presence of layers with very different permeabilities in a formation may cause early water breakthrough in the production well, resulting in premature abandonment of the production plan.\textsuperscript{3,25,41–43} In general, the following concerns are present while employing water injection method for gas condensate reservoirs:\textsuperscript{3,25,41–43}

- The utilization of water injection in a massively heterogeneous gas-condensate reservoir is doubtful to be possible technically and economically due to early breakthrough and consequently loss of productivity.
- Up to 50% of the gas might be bypassed and trapped in the formation.
- It may not be possible to mobilize the gas previously trapped in the formation through a subsequent depressurization stage.
- Well lift could be a serious problem if high water cuts before and during the blowdown are experienced.

2.2.3. Gas Cycling. In dry gas injection processes, the goal is to keep the reservoir pressure over or near the dewpoint in order to minimize the amount of condensate.\textsuperscript{3,25,41–43} Dry gas cycling is considered as a particular type of miscible displacement to improve recovery.\textsuperscript{3,25,41–43} This method is very efficient in terms of displacement efficiency and having economical condensate production rates, especially when there are no markets available for the produced gas. There are two main groups for dry gas cycling, namely full cycling and partial cycling.\textsuperscript{3,25,41–43} Full gas cycling is when all the dry gas separated at the surface is recycled into the formation. Such a cycling generally leads to a material balance deficiency as pressure may slowly reduce. Some dry gas can be added as make-up gas to the injection preventing this occurrence. Partial cycling provides lower pressure maintenance over the dewpoint than the full gas cycling method.\textsuperscript{3,25,41–43} A gas sales contract usually dictates the amount of dry gas for injection into the formation where any extra gas produced is re-injected. For example, the partial cycling in the form of a seasonal gas injection is able to provide more dry gas over the off-peak summer months if the reservoir can still be operated at its maximum production rate.\textsuperscript{3,25,41–43} Besides lowering the loss of condensate in the reservoir, the pressure maintenance resulting from gas cycling hinders water influx and the entrapment of wet gas with its restrained condensate in the rear of the advancing water front.\textsuperscript{3,25,41–43} The gas cycling process ends when gas breakthrough into the production wells is observed and the rate of condensate production is not economical. Dry gas cycling methods should be considered in terms of important features such as mobility ratio, heterogeneity, and gravity which mostly affect the recovery factor.\textsuperscript{3,25,41–43} The demand for dry gas generally controls the amount of dry gas available for gas cycling. Therefore, the main disadvantage of dry gas cycling is that the sale of produced gas should be postponed until condensate production declines to its economical limit. The use of inert gas to compensate the amount of gas required for gas cycling process is an economical option for pressure maintenance and recovery enhancement.\textsuperscript{3,25,41–43} The inert gas or noncondensable gas injection generally uses nitrogen, carbon dioxide, and combustion or flue gases. Nitrogen is the most useful gas for this type of cycling since it is noncorrosive and has a high compressibility factor, leading to requiring less volume. The most important factors affecting economic viability of a gas-cycling project are liquid content of reservoir, product prices, and costs of makeup gas. The utilization of inert gas cycling has some advantages and disadvantages. The key benefit is that inert gas cycling permits early sale of gas and condensate phases, leading to higher net income.\textsuperscript{3,25,41–43} In addition, a greater recovery of total hydrocarbons is obtained since the formation holds large amounts of nitrogen rather than hydrocarbon gas when the reservoir is abandoned. Operational problems and higher operating costs are the result of corrosion if combustion or flue gas is employed for the cycling process. Extra capital costs are needed to separate the inert gas from the sales gas, to pretreat prior to compression, and to provide surface reinjection facilities. Also, premature breakthrough of inert gas due to high degrees of reservoir heterogeneity causes
additional operating expenses to find marketable gas sales. Some other shortcomings for all types of gas cycling processes include the following: 1) Gas injection is an expensive process; 2) It takes too long to reach the ultimate recovery of the field; 3) Production rate and recovery factor of gas are fairly low; therefore, gas production is less beneficial; 4) Heavy condensate fractions of hydrocarbons are precipitated in bottom hole regions; 5) Economic efficiency is poor when active water drive happens; and 6) Reservoir studies are difficult and costly.3,25,41–43

3. CGR DETERMINATION TECHNIQUES

The value of CGR in a gas-condensate reservoir depends on various factors such as API gravity, gas composition, pressure, and temperature. The preference is to determine CGR using detailed compositional tests if adequate experimental facilities and samples are available and also CGR values with high accuracy are required. However, it is usually common in PVT studies to develop CGR empirical correlations, equation of state (EOS), and numerical models due to the large number of parameters and the complexity of the experiments to obtain CGR. It should be noted here that empirical and theoretical techniques such EOS and intelligent methods usually have difficulties to predict thermodynamic properties around critical point with high precision. Therefore, the accuracy of available empirical relationships strongly depends on the quality of the samples and the proximity of reservoir conditions (e.g., P and T) to the fluid critical point.41–43 In this section, some of the techniques used for CGR determination are described, briefly.

3.1. Laboratory Methods. There are two main experimental methodologies for measuring CGR, namely constant composition expansion (CCE) and constant volume depletion (CVD). CCE is primarily used to determine the dewpoint of hydrocarbon mixtures and also CGR. In addition, the CVD method is conducted in order to simulate the behavior of reservoir production. Flash vaporization is another name for CCE and is simulated by expanding a mixture with a certain composition (z) in sequence steps of pressure reduction. The technical reader is encouraged to read refs 16 and 17 for a more detailed description of the experimental procedures.

3.2. Empirical Correlations. Several empirical correlations were proposed for CGR prediction in the literature.33–38,44 Some field data such as reservoir pressure (P), reservoir temperature (T), and fluids specific gravity are needed to develop PVT empirical correlations. For example, Cho et al. (1985) developed an empirical correlation to estimate maximum CGR value for retrograde gas-condensate reservoirs.44 Their correlation is expressed in terms of reservoir temperature and the mole fraction of heptanes-plus. Reliable experimental data are favored versus the results obtained from the correlations when assessing gas-condensate reservoirs.33–38,44 Nevertheless, accurate experimental data often are not available. One advantage of correlations is being helpful when only modest experimental information is available.

3.3. Analytical Methods. Analytical methods are usually based on EOS to evaluate CGR for gas-condensate reservoirs. This technique is very efficient to describe the phase behavior of well characterized mixtures at particular temperatures. However, the EOS should be tuned for each mixture composition and has predictive potential limitations for new compositions. Therefore, an EOS method requires a detailed description of the mixture's composition. Obtaining such quantities is expensive and time-consuming. Furthermore, there is no universal EOS available for hydrocarbon mixtures, which is able to predict CGR and other thermodynamic parameters over wide ranges of compositions, temperatures, and pressures with an acceptable accuracy. Some of the available analytical methods are including EOS models, material balance methods, and PVT delumping techniques.9,42,43,51–56

3.4. Numerical Modeling. Numerical simulations are another tool to predict CGR. An accurate numerical method to compute the CGR and well productivity is a fine-grid system. A fine-grid model has the ability to consider high-velocity effects, and most commercial simulators now incorporate options to consider non-Darcy flow and complicated phase behavior in different regions of gas-condensate reservoirs. Although, numerical simulations are proper for forecasting of reservoir behavior in detail, simpler and faster engineering calculations look more appropriate for some applications such as CGR prediction.42,43 For instance, the book titled ‘The Practice of Reservoir Engineering’ presents some simple theoretical equations and forecasting methods to compute CGR value and other important parameters (i.e., sweep efficiency, relative permeability and two-phase Z-factor) in gas condensate reservoirs when the gas cycling technique is employed.43 It should be also noted here that a strong understanding of the condensate banking behavior in gas condensate reservoirs is required for analysis of well test data and prediction of production performance using numerical methods.7,8,10,22,41–43

4. EXPERIMENTAL PROCEDURE AND DATA COLLECTION

The CGR of the hydrocarbon mixtures from the South Pars/North Dome gas field in Iran was measured by CCE and CVD methods and an average value of CGR was used for the intelligent model developed in this paper. The South Pars natural gas-condensate field is located in the Persian Gulf. It is the world’s largest single gas field, shared between Iran and Qatar. The field is composed of two independent gas-bearing formations, namely Kangan and Upper Dalan. Major gas reservoirs in this field mainly consist of limestone and dolomite. Both porosity (\(\phi\)) and permeability (\(k\)) vary widely in the field. For instance, the porosity varies from 0–35% and the permeability from zero to 1000 mD in some regions of this gigantic gas field. In addition to the data generated from the experimental part, some data from the literature were also used in developing the PSO-ANN model for CGR prediction. A part of the data used in this study is presented in Appendix A, and the rest can be tracked in the open literature.5,12,13,23,33–36,41–43

5. ARTIFICIAL NEURAL NETWORKS THEORY

Initially, McCulloch and Pitts (1943) developed the artificial neural network (ANN) theory as a mathematical technique.86 ANNs are biological inspirations on the basis of different brain functionality features.39,40,86 In recent years, ANN has been broadly employed for numerous applications such as process control, behavior prediction, model recognition, system classification, communication, and vision aspects. In this theory, ANN is generally trained to solve nonlinear and complex problems whose mathematical modeling appears to be a difficult job.39,40,86 The ANNs can take some information of a complicated system and provide an adequate model even when the data are noise infected and incomplete. The objective of an ANN model is to correlate a set of input data to a set of output patterns in a process. The network carries out this mapping scheme through training from some previous examples and also defining the input and output parameters for a particular system. ANNs contain many computational sections joined together in a fairly complicated parallel structure. In the processing units, there are a great number of simple components...
called neurons that behave like axons to obtain an empirical correlation between the inputs and outputs of a given process. 39,40,86

Figure 3 presents a classical interconnected neural network in the form of a multiple-layers structure. The network is composed of one input layer, one hidden layer, and one output layer with various duties. Each linking line holds a certain weight. ANNs are trained via adjustment of the connection weights as the desired values for output parameters are estimated with an acceptable accuracy. 30,31,86−89 The main feature of a neural network is the training stage according to a set of numerical magnitudes measured in the modeling process. The precision of the model representative is strongly dependent on the topology (structure) type of the neural network. 30,31,86−89 There are a number of different neural network structures and learning algorithms for the ANN technique. The back-propagation (BP) learning algorithm is recognized as one of the most general algorithms. This learning algorithm was selected in this study to predict the magnitude of condensate-to-gas ratio (CGR). 30,31 BP is a multiple-layer feed-forward network that includes hidden layers linking the input parameters to the output. 30,31,39,40 The simplest accomplishment of back-propagation (BP) training is the biases and neural network weights updates toward negative gradients. 30,31,39,40

The ANNs generally try to identify the data patterns throughout the training process. ANNs obtain the flexibility during the process letting the analyst operator accomplish testing the data patterns even if the work environment in the model is changed. Although ANN might take some time to learn an impulsive alteration, it is exceptional to adjust continuously varying information. 30,31,86−89

6. PARTICLE SWARM OPTIMIZATION
Particle swarm optimization (PSO) is a global optimization algorithm for the problems in which a point or surface in a multidimensional space represents the best solution. 90,91 As shown in Figure 4, PSO is initialized with a set of random particles. Then, random positions and velocities are assigned to the particles. After that, the algorithm starts searching for optima through a number of iterations where the particles are flying in a hyperspace to find the potential solutions. As time goes on, the particles learn to respond appropriately, based on their experience and also the experience of other particles in their set. 90−93 Every particle maintains the path of its best position that has obtained in the hyperspace to date. 90,91 The value of best position is named personal best. The overall best magnitude of position achieved by any particle in the population is called global best. In the period of iteration process, every particle experiences a tendency toward its own personal best and also in the direction of the global best. This is determined through calculation of a new velocity term for each particle, based on the distance from its personal best position and also its global best position. Random weights to personal best and global position velocities are then assigned to generate the new value for the particle velocity that shows effects on the next position of the particle in the next iteration. 90−93

Particle fitness mentioned in Figure 4 is a measure of how good the solution characterized by position is. For minimization problems that are the most typical categories of cases solved by PSO, lower magnitudes of the fitness field are better than higher values. However, greater values of the fitness are superior for maximization problems.

The position of a particular particle in a PSO-ANN model indicates a series of weights for the current iteration. 90,91 The dimension of each particle is equal to the number of weights corresponding to the network. Using mean squared error (MSE), the learning error of the network is calculated. The mean squared error of the network (MSE±) is defined as

\[ \text{MSE}_{\text{net}} = \frac{1}{2} \sum_{k=1}^{G} \sum_{j=1}^{m} [Y_j(k) - T_j(k)]^2 \]

where \( m \) is the number of output nodes, \( G \) is the number of training samples, \( Y_j(k) \) is the expected output, and \( T_j(k) \) is the actual output. When MSE approaches to zero, the error of our network would decrease.

The most important advantage of the PSO approach in comparison to other global minimization techniques like simulated annealing is that many members contributing in the PSO model cause the technique remarkably flexible as the PSO method prevents the neural network model from being stuck on local minima. 90−93 It should be also noted here that the particle swarm optimization (PSO) model is similar to the genetic algorithm (GA) model; however, the PSO employs a collaborative approach rather than a competitive one.

6.1. Training Using Particle Swarm Optimization. Given a fixed architecture of ANN, all weights in connections
can be coded as a genotype and applied in the PSO algorithm in order to train the network where the fitness function should be the mean squared error of the network. Using validation and testing sets, some changes can be made to obtain better fitness magnitudes with more generalization characteristics. To train the network by the particle swarm optimization, the following equations will customize the network weights until a stopping condition is satisfied. For instance, \( c_1 \) and \( c_2 \) denote "trust" parameters indicating how much confidence the current particle \( c_1 \) as a cognitive parameter has in itself and how much confidence \( c_2 \) (social parameter) has in the swarm. \( \omega \) represents the inertia weight. The equations describe how the particle velocity value and location are varying in an iteration procedure. During the training process, the above equations will customize the network weights until a stopping condition is satisfied. In this case, a lower MSE is achieved; however, a number of maximum iterations is used to end the algorithm if the examined stopping condition does not lead to termination in an appropriate time.

Generally, the inertia weight \( (\omega) \) decreases linearly with the iteration number as follows:

\[
\omega^t = \omega_{\text{max}} - \left( \frac{\omega_{\text{max}} - \omega_{\text{min}}}{t_{\text{max}}} \right) t
\]

where \( \omega_{\text{max}} \) and \( \omega_{\text{min}} \) are the initial and final values of the inertia weight, respectively, \( t \) is the current iteration number, and \( t_{\text{max}} \) is the maximum number of iterations used in PSO. The values of \( \omega_{\text{max}} = 0.9 \) and \( \omega_{\text{min}} = 0.4 \) are the proper value through empirical studies.

### 6.2. PSO-ANN Model Development

A variety of models were calibrated to find the optimum configuration of the PSO-ANN algorithm. A PSO-ANN structure with appropriate parameters can exhibit a good performance. To design a proper network, six (6) parameters were studied as given below:

- Acceleration constant for global best \( (c_1) \),
- Acceleration constant for personal best \( (c_2) \),
- Time interval,
- Number of particles,
- Number of maximum iterations corresponding to a low MSE that refers to the stopping condition, and
- Number of hidden neurons

The constant for global best \( (c_1) \), constant for personal best \( (c_2) \), time interval, and number of particles have significant effects on the optimal configuration of the PSO-ANN model. In addition, other parameters namely, maximum iteration, number of neurons in the hidden layer, and length of input and output data, are examined for calibration and optimization of the model.

The objective function employed here is mean squared error \( (\text{MSE}) \). The optimization step is vital to make sure that MSE or training error is decreasing with an increase in the number of iterations. Performance of the PSO-ANN is assessed using coefficient of determination \( (R^2) \) and Nash-Sutcliffe coefficient \( (E^2) \). If the values of \( R^2 \) and \( E^2 \) approach 1, the selected network would be a perfect model to predict the system behavior. The corresponding formulas for the above statistical parameters and other two performance measures \( (\text{e.g., mean absolute error (MEA)} \text{E} \text{A} \text{E} \text{A}) \text{nd maximum absolute error (MAAE))} \) are presented in Appendix B. It should be noted here that as MEAE and MAAE become lower, better performance for the developed neural networks is expected.

To calibrate the PSO-ANN model, three layers were used in this study. The number of input neurons is a function of the number of precursor data, and there is just one parameter (neuron) in the output layer. PSO-ANN was trained and then tested to determine the best structure using different data sets as follows:

- \( a) \), \( b) \), \( c) \), \( d) \), \( e) \)
- Maximum iterations were examined between 300 and 600.

### 6.3. Selection of Input Variables

In general, ANN works as a black box model that relates inputs variables and outputs parameters, which represents a mathematical relationship. It is crucial that key physical parameters contributing in the output of ANN models are used. Also, it is desirable to minimize the number of inputs for an ANN system. Selection of the best inputs is based on the understanding of physics of the problem. As previously mentioned, experimental works, numerical modeling, empirical equations, and PVT studies of gas condensate reservoirs have proved that CGR is a function of temperature, pressure, and composition of fluid mixture in the reservoir. To take into account the effect of mixture composition, some of the published research studies used molecular weight of \( C_{7+} \) and \( C_{30+} \) as inputs. In addition, there are some studies in the open literature that considered molecular weight of \( C_{30+} \) in their computations. In order to select the best inputs for the ANN method, the following sets of input variables were examined as shown in Table 1:

<table>
<thead>
<tr>
<th>Data Set</th>
<th>( T, P_d, M_{C_{30+}} )</th>
<th>( T, P_d, M_{C_{7+}} )</th>
<th>( T, P_d, M_{C_{30+}}, M_{C_{7+}} )</th>
<th>( T, P_d, M_{C_{30+}}, M_{C_{7+}}, M_{C_{2}} )</th>
</tr>
</thead>
<tbody>
<tr>
<td>No. 1</td>
<td>0.8995</td>
<td>0.9108</td>
<td>0.9447</td>
<td>0.9705</td>
</tr>
<tr>
<td>No. 2</td>
<td>0.8859</td>
<td>0.9036</td>
<td>0.9355</td>
<td>0.9548</td>
</tr>
<tr>
<td>No. 3</td>
<td>0.0976</td>
<td>0.0777</td>
<td>0.0541</td>
<td>0.0438</td>
</tr>
<tr>
<td>No. 4</td>
<td>12.1133</td>
<td>11.2461</td>
<td>7.7733</td>
<td>6.0004</td>
</tr>
<tr>
<td>No. 5</td>
<td>15.6432</td>
<td>13.5488</td>
<td>9.0064</td>
<td>8.1192</td>
</tr>
</tbody>
</table>

Input data set no. 3 that includes mole (or mass) fraction of all components in the mixture seems appropriate for the PSO-ANN model; however, it makes the computation very time-consuming. On the other hand, it is difficult to construct such a large data center in the neural network system, and many input data sometime cause errors in calculation of the desired parameter. However, with input data set no. 4, we consider the effect of mass or mole percentages of all components by selecting molecular weight and dewpoint pressure of the mixture. As shown in Table 1, input data set no. 4 among the selected input data sets exhibits the best performance for the PSO-ANN.
6.4. Data and Model Evaluation. Availability of reliable data is very important in the design of any ANN system. The first step which should be considered in an ANN system is finding parameters that affect the target parameter, considerably. CGR is a strong function of reservoir fluids properties. It is clear that reservoir characterizations (e.g., depth and permeability) dictate pressure distribution, local compositions, and temperature in gas condensate reservoirs that affect instantaneous values of CGR. Considering effects of reservoir characteristics, three main parameters that have direct influences on the amount of CGR include the dewpoint pressure ($P_{dp}$), the reservoir temperature ($T$), and molecular weight of mixture ($M_w$). A set of 295 experimental data was collected to develop and examine the proposed PSO-ANN model. The data series includes 45 data from South Pars gas-condensate reservoir and 250 data from the literature.\(^5,12,13,23,33,−38,41−85\) The ranges of the parameters used to develop the PSO-ANN are given in Table 2.

<table>
<thead>
<tr>
<th>property</th>
<th>min</th>
<th>max</th>
</tr>
</thead>
<tbody>
<tr>
<td>input temperature</td>
<td>152</td>
<td>355</td>
</tr>
<tr>
<td>dewpoint pressure</td>
<td>2950</td>
<td>8642</td>
</tr>
<tr>
<td>molecular weight</td>
<td>16.5</td>
<td>44</td>
</tr>
<tr>
<td>output CGR</td>
<td>4.5</td>
<td>293</td>
</tr>
</tbody>
</table>

It should be noted here that Table 2 covers wide ranges of important parameters contributing in CGR. The above ranges are applicable for the most of real gas condensate cases encountered. If a data point is selected very far from the above intervals, the methodology to implement the PSO-ANN does not change. However, some parameters of the PSO-ANN such as number of hidden neurons and number of particles should be adjusted in a way that an acceptable match is obtained. Another alternative for such a case is adding a wavelet function to the neural network structure in order to increase accuracy of the predictions.\(^91−93\)

7. RESULTS AND DISCUSSION

Precise determination of the CGR and prediction of its variations are central in developing a proper PVT model for utilization in reservoir engineering simulators. Development of an artificial neural network (ANN) is discussed here which was used to build a model to predict the amount of condensate-to-gas ratio (CGR). The ANN model trained with back-propagation network employed the Levenberg–Marquardt algorithm for prediction of the CGR. The transfer functions in hidden and output layers were sigmoid and linear, respectively. PSO was used as a neural network optimization algorithm, and the mean square error (MSE) was used as a cost function in this algorithm. The goal from implementation of the PSO was to minimize the cost function. Every weight in the network was initially set in the range of $[-1, 1]$ and every initial particle was a set of weights generated randomly in the range of $[-1, 1]$. For example, the initial and final values of the inertia ($\omega_{min}$ and $\omega_{max}$) were 0.4 and 0.5, respectively; $r_1$ and $r_2$ were also two random numbers in the range of $[0, 1]$.

The PSO-ANN model was evaluated with 3, 5, 7, 8, and 10 hidden neurons in the hidden layer. The results obtained using different numbers of hidden neurons are shown in Table 3. The accuracy of the simulations was increased with increase in the number of hidden neurons from 3 to 8. At 3, 5, and 7 hidden neurons, the network did not appear to have enough degrees of freedom to learn the process accurately. However, the performance of PSO-ANN began to decline at 10 hidden neurons. Therefore, the optimum value for hidden neurons is 8 in the PSO-ANN model, because it takes a long time for the PSO-ANN system to be trained at 10 hidden neurons and the model may over fit the data.

The acceleration constants ($c_1$ and $c_2$) correspond to the stochastic acceleration that pushes each particle toward personal best and global best positions.\(^91−93\) The $c_1$ constant determines the effect of the global best solution over the particle, while the $c_2$ constant describes how much the effect of personal best solution has over the particle. Reasonable results were obtained when the PSO-ANN was trained with acceleration constants ($c_1$ and $c_2$) in the range of 1.4–2.4. Performance of the PSO-ANN was improved as the values of $c_1$ and $c_2$ increased from 1.4 to 2.0. The main reason is that the particles are attracted toward personal best and global best with higher acceleration and abrupt movements when $c_1$ and $c_2$ values experience an increase from 1.4 to 2.0. Afterward, the PSO-ANN performance reduced as $c_1$ and $c_2$ values increased from 2.2 to 2.4. This trend indicates that the PSO-ANN is not able to converge with higher magnitudes of acceleration constants. Thus, the best $c_1$ and $c_2$ values required for an optimal structure of the PSO-ANN model were 2 where $R^2$ and $E^2$ were determined to be 0.9789 and 0.9581 for the training and 0.9087 and 0.8563 for the testing process, respectively. In addition, the minimum values of $MSE$, $MAE$, and $MMAE$ were obtained for both training and testing processes at $c_1$ and $c_2$ values equal to 2. The results for the PSO-ANN trained with different values of $c_1$ and $c_2$ are listed in Table 4.

Number of particles in the PSO-ANN dictates the quantity of the space covered in the problem. Table 5 shows the effect of number of particles on the performance of the developed intelligent model. As seen in Table 5, the PSO-ANN performance increases with an increase in the number of particles from 15 to 21. The developed neural network model was not able to describe well the data behavior at numbers of particles equal to 15 and 18 because the space covered in the problem was not adequate for these numbers. Then, a decrease in the PSO-ANN was observed when the number of particles

Table 3. Performance of PSO-ANN Based on the Number of Hidden Neurons

<table>
<thead>
<tr>
<th>number of hidden neurons</th>
<th>training</th>
<th></th>
<th>testing</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>$R^2$</td>
<td>$E^2$</td>
<td>MSE</td>
<td>MEAE (%)</td>
</tr>
<tr>
<td>3</td>
<td>0.8253</td>
<td>0.8244</td>
<td>0.0823</td>
<td>8.6542</td>
</tr>
<tr>
<td>5</td>
<td>0.8724</td>
<td>0.8609</td>
<td>0.0546</td>
<td>7.8479</td>
</tr>
<tr>
<td>7</td>
<td>0.9311</td>
<td>0.9038</td>
<td>0.0203</td>
<td>6.4563</td>
</tr>
<tr>
<td>8</td>
<td>0.9875</td>
<td>0.9563</td>
<td>0.0116</td>
<td>4.3421</td>
</tr>
<tr>
<td>10</td>
<td>0.9227</td>
<td>0.9012</td>
<td>0.0478</td>
<td>7.7534</td>
</tr>
</tbody>
</table>

dx.doi.org/10.1021/ef300443j | Energy Fuels 2012, 26, 3432–3447
was set on 22 (see Table 5). The main reason behind this decline is that the space covered in the problem appears to be too wide. The model efficiency was improved slightly when the number increased to 25. Therefore, it can be concluded that the optimum number of particles is 21. It should be also noted here that as the number of particles increases, the greater amount of space is covered in the problem, leading to a slower optimization process.

The same procedure was followed for other parameters of the network, and the optimal configuration for the PSO-ANN network was obtained as the following:

a) Number of hidden neurons = 8,
b) Number of maximum iterations = 500,
c) \(c_1\) and \(c_2\) = 2,
d) Time interval = 0.0100, and
e) Number of particles = 21.

Based on the above, the best ANN architecture is 3–8–1 (3 input units, 8 neurons in the hidden layer, and 1 output neuron).

A common question in neural networks is what is the sample size of data required to train the network? No simple procedures have been proposed to answer this question. Certainly, it requires adequately large and representative data for the neural network to learn. However, the number of samples for training depends on several parameters such as factor selection and the structure of the neural network model.88–91

Once the factor selection and neural network architecture were determined, different sampling sizes were employed to train while the testing phase of the PSO-ANN was being conducted using 45 real data of the South Pars gas field. The size of training data was selected between 50 and 500 data. Table 6 shows the results of the network performance based on the statistical investigations. Only performance of the training process in terms of \(MSE\) is shown in Figure 5. According to Table 6 and Figure 5, it can be concluded that a sample of 250 data is sufficient for training the PSO-ANN system. Therefore, 250 data points were chosen for the purpose of network training in this study. The remaining 45 samples were put aside to be employed for testing and validating the network’s integrity and robustness.

In order to evaluate the performance of the PSO-ANN algorithm, a back-propagation neural network (BP-ANN) was applied with the same data sets used in the PSO-ANN model. Predicted and measured normalized CGR values at the training and testing phases for the both BP-ANN and PSO-ANN models were compared as shown in Figures 6 and 7.

The input variables of an ANN are frequently of various types with different orders of magnitudes, such as temperature \((T)\) and dewpoint pressure \((P_d)\) in the case studied here. The same thing usually occurs for the output parameters. Thus, it is necessary to normalize the input and output parameters based on the ranges of data as they fall within a particular range. For instance, setting minimum \(-1\) and the maximum \(+1\) to have all the data within the interval \([-1, 1]\). It should be mentioned here that the legend of the vertical axis in Figures 6 and 7.
presents normalized CGR that was calculated using the following expression

\[
\text{Normalized CGR} = \frac{2(\text{CGR} - \text{CGR}_{\text{min}})}{\text{CGR}_{\text{max}} - \text{CGR}_{\text{min}}} - 1
\]  

(5)

where \( \text{CGR}_{\text{min}} \) and \( \text{CGR}_{\text{max}} \) are the minimum and maximum CGR of the data used in this study, respectively.

PSO-ANN was trained by 50 generations followed by a BP training procedure. The values of learning coefficient 0.7 and momentum correction factor 0.001 were used for the back-propagation training algorithm.

As seen in Figure 7, there is a good agreement between the model outputs and the experimental CGR, based on the testing data simulated at this phase of the study. The simulation performance of the PSO-ANN model was evaluated on the basis of mean squared error (MSE), mean absolute error (MAAE), maximum absolute error (MEAE), and efficiency coefficient \( R^2 \). The parameters \( \text{MSE} = 0.0617 \) and \( R^2 = 0.9721 \) for PSO-ANN in contrast with the \( \text{MSE} = 0.0981 \) and \( R^2 = 0.8986 \) for BP-ANN reveals that the PSO-ANN exhibits a better performance in prediction of CGR than BP-ANN (Figures 8 and 9). It should be noted here that an \( R^2 \) value greater than 0.9 generally indicates a very satisfactory model performance; while an \( R^2 \) magnitude in the range of 0.8–0.9 presents a good performance and a value less than 0.8 shows an unsatisfactory model performance, based on the statistical analysis.94,95
The extent of the match between the measured and predicted CGR values by BP-ANN and PSO-ANN networks in term of a scatter diagram are shown in Figures 8 and 9. These results convey the message that PSO-ANN has the capability to avoid being trapped in local minima. This advantage is due to the combination of global searching ability of PSO and local searching ability of BP.

The performance plots of the BP-ANN and developed PSO-ANN models are depicted in Figures 10 and 11, respectively. The curves included in the figures present the relationship between the testing, training, validation, and best models introduced for predicting CGR in terms of MSE versus number of epochs. As shown with green circles on Figures 10 and 11, the best performance for the validation phase is 0.0022 at epoch about 3 for the BP-ANN model, while the PSO-ANN model experiences the best performance ($\text{MSE} \approx 0.0015$) for the validation phase at epoch of 13.

A comparison between CGR values obtained from a correlation introduced by Fattah et al. (2009) and the proposed PSO-ANN and BP-ANN models is presented here to examine the performance of the PSO-ANN model. Table 7 shows the MSE, MEAE, MAAE, $E^2$, and $R^2$ values for these different approaches in the validation phase. It can be concluded that the performance of PSO-ANN is better than the BP-ANN model and also the empirical correlation.

An ANN is a learning structure in which the network is trained by information of a certain process or system. In general, this training stage needs considerably more processing time than what is required to evaluate the system using a single simulation run or an empirical equation. In Table 8, the training time is recorded for both the 39-bus system and the 1844-bus system, employed for the neural networks. For both systems, 250 sample data were used for training the BP-ANN and PSO-ANN. The computations were conducted using a HP laptop, Intel Pentium Dual CPU E2200 @ 2.8 GHz. Table 8 indicates that the practical system (1844-bus) needs higher training time than the 39-bus system as more input data were considered in the practical
system. Also, the table shows that the training process for the BP-ANN is faster than that for the PSO-ANN model.

The main benefit of applying computational intelligence technology to substitute the conventional experimental and numerical methods for calculation of CGR is its possibility of real-time operation. ANN has the ability to satisfactorily shrink what it learns in training process and broaden this to generate acceptable outputs for those inputs not dealt over the training stage. Once trained, the ANN system can forecast the value of CGR for a given set of input data very fast as the computation in an ANN model does not involve any iteration which is the case in numerical methods. Table 9 shows the comparison of the computational speed of the BP-ANN and PSO-ANN. Based on the results presented in the table, using the BP-ANN to estimate CGR is a little faster than the method; however, the PSO-ANN model exhibits a higher accuracy in estimating CGR.

A sensitivity analysis was conducted for the newly developed model using the analysis of variance (ANOVA) technique. The dependence of the dependent variable (CGR) on each of the independent variables was examined using the ANOVA technique. The results of the sensitivity analysis are presented in Figure 12. The higher correlation between any input variable and the output variable indicates greater significance of the variable on the magnitude of the dependent parameter. Clearly, the mixture molecular weight is the most important parameter affecting the CGR. An increase in the molecular weight leads to an increase in the CGR value. This finding is in agreement with several previously published research. In addition, impact of dewpoint pressure on the CGR supports the validity of the empirical correlations available in the literature. Increase in dewpoint pressure results in an increase in the value of CGR. It should be also noted here that as the reservoir temperature increases, the magnitude of the CGR increases.

**8. CONCLUSIONS**

An artificial neural network intelligent system optimized with particle swarm optimization was developed and tested using the experimental data from South Pars gas field and the literature to predict condensate-to-gas ratio (CGR).

Based on the results of this study, the following main conclusions can be drawn:

1. For a particular sample of South Pars field, the CGR values obtained from the experimental work and PSO-ANN are 31.8 and 31.5 STB/MMSCF, respectively. An acceptable agreement between these two determination approaches was observed.

2. Prediction of CGR was made possible using PSO-ANN, BP-ANN, and an empirical correlation. Comparing the results of this study indicates that predictive performance of the proposed PSO-ANN is better than the BP-ANN model and also the empirical correlation used in this paper.

3. For predicting the CGR, the PSO-ANN model has the capability to avoid being trapped in local optima due to the combination of global searching ability of PSO and local searching ability of BP.

4. Finding an optimum structure for the PSO-ANN model is an important feature in CGR prediction. The optimum configuration for the proposed model includes 8 hidden neurons and the acceleration constants ($c_1$ and $c_2$) equal to 2 determined by using a trial and error technique.

5. A sensitivity analysis using the ANOVA technique showed that molecular weight of mixture has the most
effect on the magnitude of the CGR predicted. Also, as the temperature increases, the CGR value increases.

6. Although the proposed ANN-PSO model works well for CGR prediction, it suffers from early convergence by the degeneracy of several dimensions; even if no local optima exist for cases considered with this model. Hence, a better evolutionary algorithm is required to be combined with PSO and ANN which is a part of our future study.

■ APPENDIX A

A part of experimental data employed in this study to develop the PSO-ANN model is listed in Table A1. To build the training network, a large volume of the CGR data available in the literature was used to obtain an appropriate neural network model.

■ APPENDIX B

$R^2$, $E^2$, MSE, MEAE, and MAAE are the statistical parameters to test the accuracy of the PSO-ANN model compared with common correlations. The equations to compute the above parameters and also the corresponding description are as follows:

$R^2$ is a statistic parameter which gives information on goodness of fit in a model. In regression analysis or fitting of a model, the $R^2$ coefficient is a statistical measure of how well the model line approximates the real data points. An $R^2$ of 1.0 indicates the model line perfectly fits the data. The following equation shows the mathematical definition of $R^2$:

$$ R^2 = \frac{\sum_{i=1}^{n} (CGR_i^P - CGR_i^M)^2}{\sum_{i=1}^{n} (CGR_i^M - \bar{CGR})^2} $$

where $CGR_i^M$ and $CGR_i^P$ are the measured CGR and predicted CGR, respectively. $\bar{CGR}$ represents the average of the measured CGR data.

The Nash–Sutcliffe model efficiency coefficient ($E^2$) can be employed to evaluate the predictive power of statistical and intelligent models. It is given as:

$$ E^2 = 1 - \frac{\sum_{i=1}^{n} (CGR_i^M - CGR_i^P)^2}{\sum_{i=1}^{n} (CGR_i^M - \bar{CGR})^2} $$

The mean squared error (MSE) is a measure of how close a fitted line or developed model is to data points. The smaller the mean squared error the closer the fit (or model) is to the actual data. The MSE has the units squared of whatever is plotted on the vertical axis. MSE is described by the following relationship:

$$ MSE = \frac{\sum_{i=1}^{n} (CGR_i^M - CGR_i^P)^2}{n} $$

where $n$ is the number of samples.

The other two performance measures that were used in this paper to assess the effectiveness of the training and testing data include mean absolute error (MEAE) and maximum absolute error (MAAE) as follows:

$$ MEAE\% = \frac{1}{n} \sum_{i=1}^{n} \left| \frac{CGR_i^P - CGR_i^M}{CGR_i^M} \right| \times 100 $$

$$ MAAE\% = \frac{\max|CGR_i^P - CGR_i^M|}{CGR_i^M} \times 100 $$

---

Table A1. Experimental Data for CGR from Various Research Studies

<table>
<thead>
<tr>
<th>run no.</th>
<th>T (°F)</th>
<th>$P_2$ (psi)</th>
<th>$M_w$</th>
<th>CGR (STB/MMSCF)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>173</td>
<td>5871</td>
<td>18.13</td>
<td>19.7</td>
</tr>
<tr>
<td>2</td>
<td>253</td>
<td>4524</td>
<td>17.44</td>
<td>8.2</td>
</tr>
<tr>
<td>3</td>
<td>215</td>
<td>5393</td>
<td>21.27</td>
<td>32.5</td>
</tr>
<tr>
<td>4</td>
<td>261</td>
<td>4345</td>
<td>23.52</td>
<td>24.8</td>
</tr>
<tr>
<td>5</td>
<td>210</td>
<td>5570</td>
<td>25.23</td>
<td>7.6</td>
</tr>
<tr>
<td>6</td>
<td>203</td>
<td>4154</td>
<td>24.80</td>
<td>33.6</td>
</tr>
<tr>
<td>7</td>
<td>281</td>
<td>5228</td>
<td>23.98</td>
<td>5.6</td>
</tr>
<tr>
<td>8</td>
<td>250</td>
<td>5962</td>
<td>21.46</td>
<td>5.8</td>
</tr>
<tr>
<td>9</td>
<td>214</td>
<td>4564</td>
<td>23.94</td>
<td>36.1</td>
</tr>
<tr>
<td>10</td>
<td>177</td>
<td>4276</td>
<td>26.56</td>
<td>26.5</td>
</tr>
<tr>
<td>11</td>
<td>252</td>
<td>3589</td>
<td>21.61</td>
<td>16.6</td>
</tr>
<tr>
<td>12</td>
<td>240</td>
<td>3948</td>
<td>19.70</td>
<td>20.6</td>
</tr>
<tr>
<td>13</td>
<td>174</td>
<td>3292</td>
<td>24.79</td>
<td>28.9</td>
</tr>
<tr>
<td>14</td>
<td>179</td>
<td>4108</td>
<td>18.98</td>
<td>6.1</td>
</tr>
<tr>
<td>15</td>
<td>281</td>
<td>3968</td>
<td>18.96</td>
<td>27.7</td>
</tr>
<tr>
<td>16</td>
<td>224</td>
<td>4005</td>
<td>21.38</td>
<td>26.9</td>
</tr>
<tr>
<td>17</td>
<td>168</td>
<td>5471</td>
<td>23.00</td>
<td>19.2</td>
</tr>
<tr>
<td>18</td>
<td>171</td>
<td>4590</td>
<td>22.96</td>
<td>23.1</td>
</tr>
<tr>
<td>19</td>
<td>182</td>
<td>4558</td>
<td>17.79</td>
<td>15.4</td>
</tr>
<tr>
<td>20</td>
<td>270</td>
<td>5339</td>
<td>17.80</td>
<td>10.0</td>
</tr>
<tr>
<td>21</td>
<td>220</td>
<td>3321</td>
<td>19.49</td>
<td>17.3</td>
</tr>
<tr>
<td>22</td>
<td>231</td>
<td>4880</td>
<td>24.23</td>
<td>33.9</td>
</tr>
<tr>
<td>23</td>
<td>223</td>
<td>4019</td>
<td>20.82</td>
<td>11.7</td>
</tr>
<tr>
<td>24</td>
<td>173</td>
<td>3982</td>
<td>25.20</td>
<td>16.1</td>
</tr>
<tr>
<td>25</td>
<td>271</td>
<td>4723</td>
<td>26.32</td>
<td>14.9</td>
</tr>
<tr>
<td>26</td>
<td>199</td>
<td>5615</td>
<td>20.69</td>
<td>7.2</td>
</tr>
<tr>
<td>27</td>
<td>197</td>
<td>3563</td>
<td>17.05</td>
<td>20.9</td>
</tr>
<tr>
<td>28</td>
<td>235</td>
<td>5025</td>
<td>22.50</td>
<td>31.2</td>
</tr>
<tr>
<td>29</td>
<td>221</td>
<td>5732</td>
<td>24.35</td>
<td>33.4</td>
</tr>
<tr>
<td>30</td>
<td>208</td>
<td>3564</td>
<td>19.86</td>
<td>27.7</td>
</tr>
<tr>
<td>31</td>
<td>242</td>
<td>3870</td>
<td>21.68</td>
<td>16.8</td>
</tr>
<tr>
<td>32</td>
<td>276</td>
<td>4482</td>
<td>18.83</td>
<td>14.0</td>
</tr>
<tr>
<td>33</td>
<td>238</td>
<td>5695</td>
<td>26.36</td>
<td>26.6</td>
</tr>
<tr>
<td>34</td>
<td>198</td>
<td>4497</td>
<td>21.56</td>
<td>32.7</td>
</tr>
<tr>
<td>35</td>
<td>189</td>
<td>3805</td>
<td>22.17</td>
<td>34.1</td>
</tr>
<tr>
<td>36</td>
<td>217</td>
<td>4597</td>
<td>18.82</td>
<td>7.1</td>
</tr>
<tr>
<td>37</td>
<td>168</td>
<td>4722</td>
<td>17.51</td>
<td>19.9</td>
</tr>
<tr>
<td>38</td>
<td>268</td>
<td>4224</td>
<td>17.13</td>
<td>8.6</td>
</tr>
<tr>
<td>39</td>
<td>237</td>
<td>2996</td>
<td>25.10</td>
<td>12.6</td>
</tr>
<tr>
<td>40</td>
<td>189</td>
<td>5118</td>
<td>24.60</td>
<td>32.8</td>
</tr>
<tr>
<td>41</td>
<td>226</td>
<td>4513</td>
<td>18.55</td>
<td>10.8</td>
</tr>
<tr>
<td>42</td>
<td>171</td>
<td>4126</td>
<td>25.90</td>
<td>8.4</td>
</tr>
<tr>
<td>43</td>
<td>267</td>
<td>3150</td>
<td>23.69</td>
<td>21.9</td>
</tr>
<tr>
<td>44</td>
<td>217</td>
<td>4056</td>
<td>23.40</td>
<td>14.6</td>
</tr>
<tr>
<td>45</td>
<td>230</td>
<td>3673</td>
<td>26.90</td>
<td>16.7</td>
</tr>
<tr>
<td>46</td>
<td>232</td>
<td>3578</td>
<td>16.85</td>
<td>29.8</td>
</tr>
<tr>
<td>47</td>
<td>245</td>
<td>5461</td>
<td>20.95</td>
<td>31.4</td>
</tr>
<tr>
<td>48</td>
<td>209</td>
<td>4164</td>
<td>21.64</td>
<td>26.3</td>
</tr>
<tr>
<td>49</td>
<td>174</td>
<td>3115</td>
<td>22.63</td>
<td>17.8</td>
</tr>
<tr>
<td>50</td>
<td>219</td>
<td>4107</td>
<td>17.37</td>
<td>25.1</td>
</tr>
<tr>
<td>51</td>
<td>171</td>
<td>5341</td>
<td>23.43</td>
<td>11.4</td>
</tr>
<tr>
<td>52</td>
<td>252</td>
<td>3460</td>
<td>17.05</td>
<td>24.3</td>
</tr>
<tr>
<td>53</td>
<td>169</td>
<td>5764</td>
<td>22.35</td>
<td>26.1</td>
</tr>
<tr>
<td>54</td>
<td>278</td>
<td>3935</td>
<td>23.98</td>
<td>23.7</td>
</tr>
<tr>
<td>55</td>
<td>253</td>
<td>3967</td>
<td>21.62</td>
<td>15.6</td>
</tr>
</tbody>
</table>
AUTHOR INFORMATION

Corresponding Author
*Phone: 519-888-4567 ext. 36157. E-mail: szendehb@uwaterloo.ca. Corresponding author address: Chemical Engineering Department, University of Waterloo, Waterloo, Ontario, Canada, N2L 3G1.

Notes
The authors declare no competing financial interest.

ACKNOWLEDGMENTS

The authors would like to thank Ali Shafei, University of Waterloo, for his help and detailed technical attention in editing the manuscript.

NOMENCLATURE

Acronyms

- AI = artificial intelligence
- ANN = artificial neural network
- BP = back propagation
- CCE = constant composition expansion
- CGR = condensate-to-gas ratio
- CVD = constant volume depletion
- EOS = equation of state
- FL = fuzzy logic
- GA = genetic algorithm
- MAAE = maximum absolute error
- MEAE = mean absolute error
- MSE = mean squared error
- Mw = molecular weight
- NN = neural network
- PSO = particle swarm optimization
- P-T = pressure-temperature
- PVT = pressure-volume-temperature

Variables

- \( T_i(k) \) = actual output
- \( P^*_i \) = best ever particle position of particle \( i \)
- \( Y_i(k) \) = expected output
- \( P^* \) = global best position in the swarm until iteration \( t \)
- \( V_t \) = velocity vector at iteration \( t \)
- \( \epsilon_1, \epsilon_2 \) = trust parameters
- \( C_j \) = hydrocarbon component (e.g., \( i = 1 \) refers to methane)
- \( E^\theta \) = Nash-Sutcliffe coefficient
- \( G \) = number of training samples
- \( k \) = permeability (mD)
- \( m \) = number of output nodes
- \( n \) = number of samples
- \( M_w \) = molecular weight
- \( P^d \) = dewpoint Pressure
- \( r_1, r_2 \) = random number
- \( R^\theta \) = coefficient of determination
- \( T \) = reservoir temperature
- \( X_i \) = position of the \( i \)-th particle
- \( z_i \) = overall composition of component \( i \)

Greek Letter

- \( \omega \) = inertia weight
- \( \phi \) = porosity

Subscripts

- \( d \) = dewpoint
- \( i \) = particle \( i \)
- \( \max \) = maximum
- \( \min \) = minimum

Superscripts

- \( M \) = measured
- \( \text{net} \) = network
- \( P \) = predicted

REFERENCES

17. Whiton, C. H.; Brule, M. Phase Behavior; Monograph Series. SPE: Richardson, TX, USA, 2000.


(66) Fang, Y.; Li, B.; Hu, Y.; Sun, Z.; Zhu, Y. Condensate Phase Behavior and Development. SPE International Oil and Gas Conference and Exhibition, Beijing, China, 2–6 November 1998.


(68) Al-Anazi, H. A.; Solares, J. R.; Al-Faiß, M. G. The Impact of Condensate Blockage and Completion Fluids on Gas Productivity in Gas-Condensate Reservoirs. SPE Asia Pacific Oil and Gas Conference and Exhibition, Jakarta, Indonesia, 5–7 April 2005.


