A System of Key Performance Indicators for Validating Matched Reservoir Analogues through Case-based Reasoning Algorithms*

Cenk Temizel[1], Suryansh Purwar[2], Amit Agarwal[2], Azer Abdullayev[2], Pedro Mora[2], Alejandro Chacon[2], and Alkan Gulses[3]

Search and Discovery Article #41311 (2014)
Posted March 31, 2014

*Adapted from extended abstract prepared in conjunction with poster presentation at GEO-2014, 11th Middle East Geosciences Conference and Exhibition, 10-12 March 2014, GEO-2014©2014

[1]Aera Energy, Bakersfield, CA, USA
[2]Halliburton, Bakersfield, CA, USA
[3]University of Southern California, Los Angeles, CA, USA

Abstract

Advances in the Digital Oilfield Concept in the last decade have not only increased the amount of data captured but have also improved the quality and precision. Furthermore, initiation of the robust application of relatively-new technologies in petroleum engineering including computer science lets us turn data into knowledge in a more efficient and consistent way. This paradigm change has great potential for use yet is not being applied in many areas that use more 'traditional' or 'primitive' ways. The application of case-based reasoning techniques in reservoir analogues is one of such areas where these relatively new techniques are not studied and documented in a proper and efficient manner.

Temizel and Dursun (2013) listed different metrics and methods of k-means clustering algorithms that can be harnessed within the scope of effectively finding reservoir analogues. A second aspect of this approach involves the validation of the reservoir analogues using generalized dimensionless forms of reservoir performance and development parameters systematically. This is a crucial step in finding and employing analogue techniques that can be used to make decisions about large investments in new discoveries as well as field development plans. Although analogues have been used in day-to-day activities in the industry, the industry still lacks a systematic way of validating matches. This will be critical for developing automated data-mining-based digital oilfields. In our study, we extend the first step of matching reservoir analogues with cased-based reasoning algorithms to the key parameters and their use in validating matched analogues in a dimensionless, scalable domain. This study outlines the significant parameters in reservoir analogue matching and will serve as a guide for engineers in the industry in decision-making processes in field development by shedding light on most of the uncertainties that naturally exist in critical aspects of development and design, and most importantly economics.

Introduction

The many uncertainties regarding the potential performance of a field particularly in complex and harsh environments, such as deep-water, limit the number of wells being drilled due to high drilling costs. Growing exploration in deeply buried reservoirs and in reservoirs with increased complexity that have relatively scarce information available has necessitated the use of reservoir analogues techniques. A good geological description of the reservoir's complexity is basic for working towards optimum field development. However, the limitations/constraints associated with describing a complex reservoir could be overcome by using reservoir analogues. Using analogues provides a better understanding and a more effective way of dealing with reservoir problems. It can help in recognizing potential problems early in the development phase and so that mitigation strategies can be planned. In addition, it contributes to the E&P best practices of overall knowledge capture, retention and reuse.

Unfortunately, there is an inherent problem with using reservoir analogues—a perfect analogue cannot be produced and the use of an incorrect analogue can result in erroneous decisions. It is therefore important to analyze the usability of an analogue based on reasoned and critical assessment of the problems and principal objectives involved. Case-based reasoning (CBR) is the process of finding solutions to new problems using the solutions of previously solved problems. Cases describe specific situations and are indexed by relevant features. All cases contain data about:

- Identification such as owner, place, date and time, geology, well depth etc.
- Specific issues and failures
- Recommended procedures to solve the problem
- Success ratio of solved cases and lessons learned

The methodology of CBR is based on the past decisions made about the same type of problems with the assumptions that similar problems have similar solutions (Avramenko 2006). CBR is typically described as a cyclic process, comprising the following four stages (Temizel and Dursun 2013):

1. RETRIEVE the most similar case(s).
2. REUSE the case(s) to attempt to solve the problem.
3. REVISE the proposed solution if necessary.
4. RETAIN the new solution as part of a new case.

Abel et al. (1996) used CBR in a system to support petrographic analysis. Irrgang et al. (1999) used CBR to cut drilling costs by deriving alternative plans using information from previously drilled wells. Skalle et al., (1998, 2000) applied CBR to develop best practices on tackling stuck drill string issues and for improving the efficiency of oil well drilling. Popa et al. (2008) demonstrated how CBR could be applied for well failure diagnostics and planning. Abdollahi et al. (2008) used a case-based approach for diagnosis of well integrity problems. Shokouhi et al., (2009) used CBR for determining root causes of drilling problems by using knowledge from previous cases. Shokouhi et al. (2010) also reviewed application of CBR in different fields of petroleum engineering. In drilling operations for example, CBR could be applied for planning, problem solving, optimization, decision-making, well integrity, and pattern recognition. This could help to reduce costs up to 30% and address knowledge problems due to limited access to human experts and the age gap problem.

Table 1 shows the list of parameters defined by SPE and SEC to find the reservoir analogue. The highlighted parameters are used as the reservoir parameters in our simulations.

Using Distance Metrics to Find Similarities between Analogues and the Reference Model

Table 2 shows the distance metrics used for the numerical attributes.

Explanation

We are using certain distance metrics between the reference model and five models/analogues based on seven attributes. These features are Average Porosity, Average Permeability, Average Thickness, Average NTG, Average Viscosity, Average API gravity, and Average Initial Pressure. Note that the primary objective here is to illustrate the method, the detailed examination or use/selection of parameters is of secondary importance.

In our formulation below, distance metrics will be shown as d_n, with n denoting the sample model number (n = 1, 2...5). In addition, data points in model vectors will be written as x_n^m, with subscript n being the model number as before and superscript m indicating the attribute number (m = 1, 2 … 7). Thus, using all seven attributes, four distance metrics will be calculated based on different distance metric definitions. These values are shown in Table 3 following the equations used. Table 4 shows the values of attributes for the analogue models.

Cityblock Distance

In this metric, we calculate distances in the analogue by moving through blocks; no diagonal move is allowed. Thus, we just calculate the absolute value of the distances between the reference and the sample vectors for different features directly as:

$$d_n = \left|x_1^1 - x_n^1\right| + \left|x_1^2 - x_n^2\right| \ldots + \left|x_1^m - x_n^m\right| \qquad (m = 1, 2 \ldots 7)$$

By using this formula for every sample model, five of the distance metrics can be found. Note that the cityblock metric is a basic measure of how far apart two quantities are. Therefore, it is a simple and useful tool in analyzing any kind of relationship between two sets of data.

Euclidean Distance

This metric tells us the shortest distance between two vectors using the well-known Pythagorean formula. For *m*-dimensions it is defined as:

$$d_n = \sqrt{\left|x_1^1 - x_n^1\right|^2 + \left|x_1^2 - x_n^2\right|^2 \ldots + \left|x_1^m - x_n^m\right|^2} \qquad (m = 1, 2 \ldots 7)$$

Euclidean distance is the main measurement used to compare two quantities. It calculates the root mean square of the distance. Its importance comes from the fact that the shortest geometrical path between two points is given based on it.

Chebyshev Distance

In mathematics, Chebyshev distance is a metric that gives the maximum of the difference of two vectors by comparing their entries one by one.

$$d_n = \max\left\{\left|x_1^1 - x_n^1\right|, \left|x_1^2 - x_n^2\right|, \ldots, \left|x_1^m - x_n^m\right|\right\} \quad (m = 1, 2 \ldots 7)$$

This metric is important because rather than giving the cumulative similarity of data, extreme points are taken into account and compared. Therefore, even if all points are close except for one extreme point, the Chebyshev distance becomes that point's distance emphasizing efficiency in grouping. Thus, it is useful especially in clustering. Therefore, the three metrics given above are sub-classes of the Minkowski metric. An additional metric used is cosine distance.

Cosine Distance

The cosine distance metric is a measure of similarity between two vectors defined in an inner product space. It basically gives the '1-cosine' between these two vectors. When similarity increases, vectors tend to be aligned making the angle smaller, giving rise to higher values of cosine. Thus '1-cosine' is to decrease, yielding smaller distance metrics for closely related vectors, just like before. Cosine similarity is generally used in the positive domain.

$$d_n = 1 - \frac{x_1^1 x_n^1 + x_1^2 x_n^2 + x_1^3 x_n^3 \ldots + x_1^m x_n^m}{\sqrt{x_1^1 x_1^1 + x_1^2 x_1^2 + x_1^3 x_1^3 \ldots + x_1^m x_1^m}\ \sqrt{x_n^1 x_n^1 + x_n^2 x_n^2 + x_n^3 x_n^3 \ldots + x_n^m x_n^m}} \quad (m = 1, 2 \ldots 7)$$

In the expression above, the numerator calculates the inner product while the denominator is finding the absolute values of the vectors under consideration. Cosine values are supposed to be limited to the range of 0 to 1. Table 3 and Table 4 show the values of attributes for reference model 1 and the analogue models, respectively.

Analogue-Matching Results Using Distance Metrics

Based on these distance metrics and related formulae, the results are ranked and shown in Table 5 and Figure 1. The distance values for our five trial models are d_2, d_3,...d_6. Apparently, based on the results of different type of metrics, all trends look similar so that model 2 fits the best and model 6 fits the worst. Note that in the calculations, normalized values with respect to reference data are used instead of the pure data given in Table 2, to prevent biasing due to differences among the magnitudes of the different physical quantities.

Reservoir Simulation

Due to difficulties associated with obtaining and publishing confidential data while keeping in mind the scope of this study, a standard SPE simulation model was used. As mentioned, concept is of primary concern here rather than the values or results obtained throughout the calculation, matching of the models, and the validation of the matched-analogues. The model used for simulation was the SPE Comparative Solution Project 9 model. It is a dipping black-oil reservoir model with 25 oil producers and 1 injector. Reservoir lifetime of this synthetic case is 01/01/1980 – 06/19/1982. The reservoir simulator used for this study was the Nexus® simulator. Values for porosity, permeability, viscosity, average thickness, net-to-gross ratio, API gravity, and initial pressure were taken from the reference and analogue model values listed in Table 1 and Table 2.

Table 6 shows the simulation output for Oil in Place, Cumulative Oil Production, Cumulative Gas Production, and Cumulative Water Production for reference model 1 and five sample models. In addition to Table 6, Figure 2, Figure 3, and Figure 4 show the plots of the simulation output for Cumulative Oil, Cumulative Gas and Cumulative Water Production for the reference model and five sample analogue models.

Reservoir production performance and recovery is of utmost importance among the objectives of finding the ‘right’ analogues. As performance is a key unknown in the beginning while being the ultimate crucial parameter, we must validate the consistency and strength of the methods and algorithms used for matching analogues. Some performance parameters include cumulative oil production, cumulative water production, cumulative gas production and recovery. Other parameters are used in calculating the distance between the reference and the analogue models. Each parameter/attribute was normalized with respect to the respective parameter of the reference model to eliminate the effect of the difference in order of magnitudes between attributes in terms of numerical values. Engineering judgment should still be used to check and fine-tune the decisions and the way the methods are used. The operational strategy for the fields might differ and affect parameters such as the cumulative production of fluids. Thus, there is no generic solution, it is case-specific and adjustments by reservoir engineers should be one of the required inputs of this process. Results of different simulation models are outlined in Table 7.

Results and Discussion

With the advance in technology and so-called ‘smart fields’ concept in the oil industry, it is crucial not only to set standards for each and every process but also to establish algorithms that will enable the automation of processes in state-of-the-art workflows to process large amounts of data where workflows turn data into knowledge. Similarity between the reference model and the analogue models has been illustrated using different numerical distance metrics in MATLAB. The similarity to the reference model was first calculated using the model attributes and then the performance attributes were used to obtain the distances or the similarity indices between the reference and the analogue models. The main objective was to compare the similarity indices obtained using model features and performance measures that are of utmost importance to the operators. The main objective here is to provide a means of comparison between the aforementioned parameters in calculating the distance measures and thus to understand the validity of the initial analogue matches done with model features and the distance metrics (Table 8).

As seen in Figure 5, cityblock distance experienced more difference between the values calculated using the model and the performance values. Differences between values of cosine distances are the lowest. This might be due to this metric's smoothening/dampening effect. Overall, although not perfect, parallel trends between the model and the performance values are important in terms of the proportionality and trend of these two approaches while putting more trust on the initial model values' validity in using them in matching the analogues. This study can be extended in a more detailed manner. Methods and concepts illustrated are far more important than the numerical values used here.

Acknowledgments

We would like to thank Aera Energy (a Shell-ExxonMobil affiliate), Halliburton, and the University of Southern California for using the computational resources. The MATLAB code used in this study is provided below.

MATLAB CODE

```
% GEO PAPER ATTRIBUTES (BASED ON MATLAB PAGE)

% Reference values
M1=[0.33 2000 100 0.8 100000 9 3500];

% Model values
M2=[0.30 1000 60 0.7 10000 13 4200];
M3=[0.24 350 55 0.8 400 22 5000];
M4=[0.20 100 40 0.7 100 26 5350];
M5=[0.25 10 30 0.6 50 32 6200];
M6=[0.15 0.1 25 0.6 100 36 5600];
```

MATLAB CODE used

```
% Model Matrix
M0 =[0.33 2000 100 0.8 100000 9 3500
  0.30 1000 60 0.7 10000 13 4200
  0.24 350 55 0.8 400 22 5000
  0.20 100 40 0.7 100 26 5350
  0.25 10 30 0.6 50 32 6200
  0.15 0.1 25 0.6 100 36 5600];

% Normalized model matrix
M=zeros(6,7);
M(1,:)=M1./M1;
```

```
M(2,:)=M2./M1;
M(3,:)=M3./M1;
M(4,:)=M4./M1;
M(5,:)=M5./M1;
M(6,:)=M6./M1;

% Function: Use M for normalized, M0 for unnormalized form
% Models: 'euclidean', 'seuclidean', 'cityblock', 'minkowski',
% 'chebychev', 'mahalanobis', 'cosine', 'correlation',
% 'spearman', 'hamming', 'jaccard'

X=zeros(2,7);
X(1,:)=M(1,:);
 for i=1:1:5
  for j=1:1:7
   X(2,j)=M(i+1,j);
  end
 pdist(X, 'chebychev')
end
```

References Cited

Abdollahi, J., I.M. Carlsen, P. Randhol, E. Tenold, H. Haga, and J. Jakobsen, 2008, A case-based approach to understand the complexity of causal connections related to well integrity problems: OnePetro, SPE #111129, 14 p.

Abel, M., E.B. Reategui, and J.M.V. Castilho, 1996, "Using case-based reasoning in a system that supports petrographic analysis, *in* Artificial Intelligence in the Petroleum Industry: Symbolic and Computational Applications II, Chapter 7, p. 159-172.

Avramenko, Y. 2006, Similarity Concepts for Case-Based Design in Process Engineering: Computers and Chemical Engineering, v. 30/3, p. 547-557.

Hodgin, J.E., and D.R. Harrell, 2006, The selection, application, and misapplication of reservoir analogs for the estimation of petroleum reserves: OnePetro, SPE #102505, 15 p.

Irrgang, R., C. Damski, S. Kravis, E. Maidla, and K. Millheim, 1999a, A case-based system to cut drilling costs: OnePetro, #SPE 56504, 17 p.

Popa, A.S., C.G. Popa, M. Malamma, and J. Hicks, 2008, Case-based reasoning approach for well failure diagnostics and planning: OnePetro, SPE #114229, 7 p.

Shokouhi, S.V., A. Aamodt, and P. Skalle, 2010, Applications of CBR in Oil Well Drilling: A General Overview: Intelligent Information Processing V. Springer Berlin Heidelberg, p. 102-111.

Shokouhi, S.V., A. Aamodt, P. Skalle, and F. Sormo, 2009, Determining root causes of drilling problems by combining cases and general knowledge: Case-Based Reasoning Research and Development: Springer Berlin Heidelberg, p. 509-523.

Skalle, P., A. Aamodt, and J. Sveen, 1998, Case-Based Reasoning, a method for gaining experience and giving advise on how to avoid and how to free stuck drill strings: IADC Middle East Drilling Conference, Dubai, Proceedings of IADC Middle East Drilling Conference, v. 1, 12 p.

Skalle, P.A., J. Sveen, and A. Aamodt, 2000, Improved efficiency of oil well drilling through case based reasoning: PRICAI 2000 Topics in Artificial Intelligence. Springer Berlin Heidelberg, p. 712-722.

Temizel, C., and S. Dursun, 2013, Efficient Use of Methods, Attributes, and Case-Based Reasoning Algorithms in Reservoir Analogue Techniques in Field Development: OnePetro, SPE #163700, 19 p.

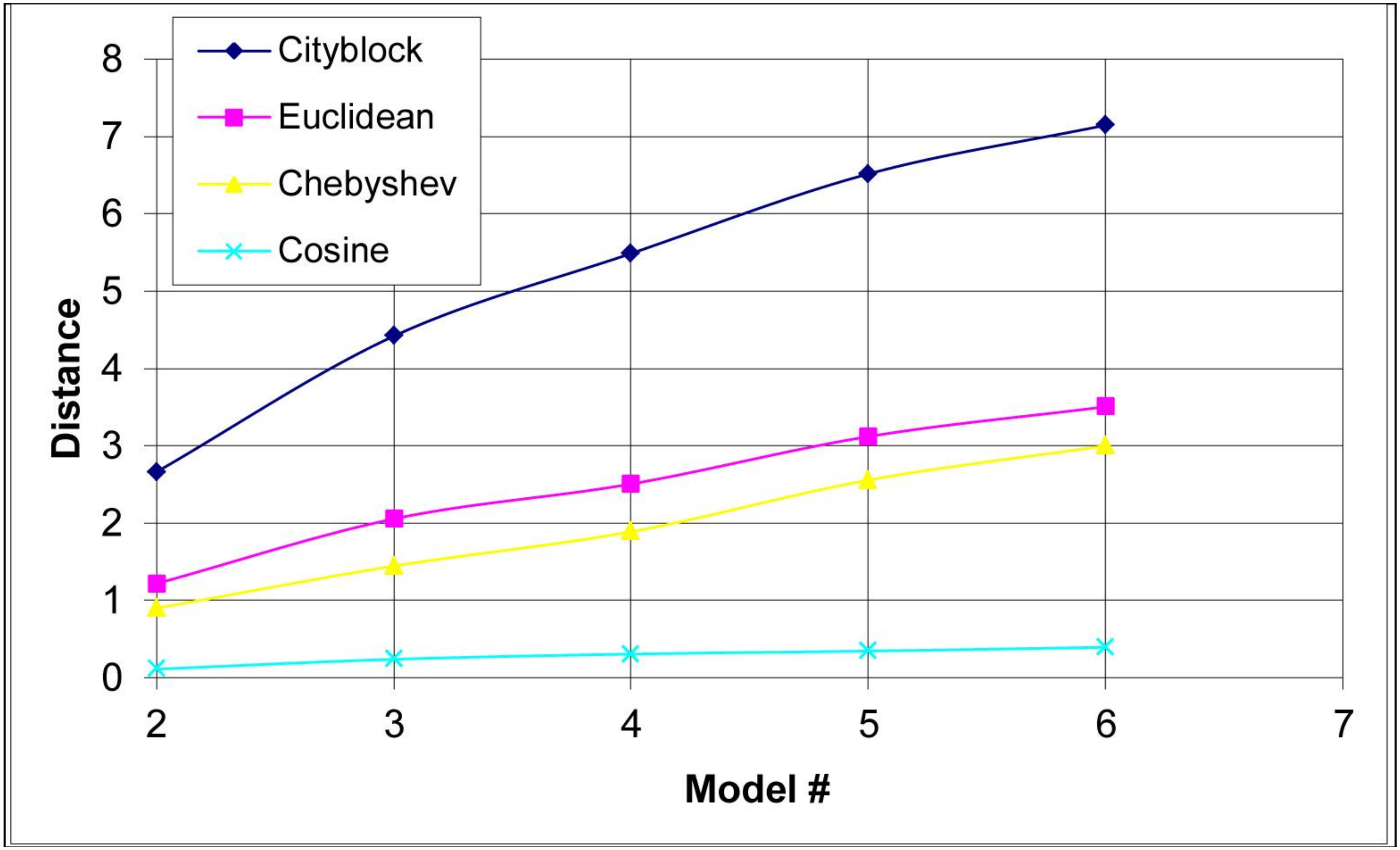

Figure 1. Distance vs. analogue models using different distance metrics.

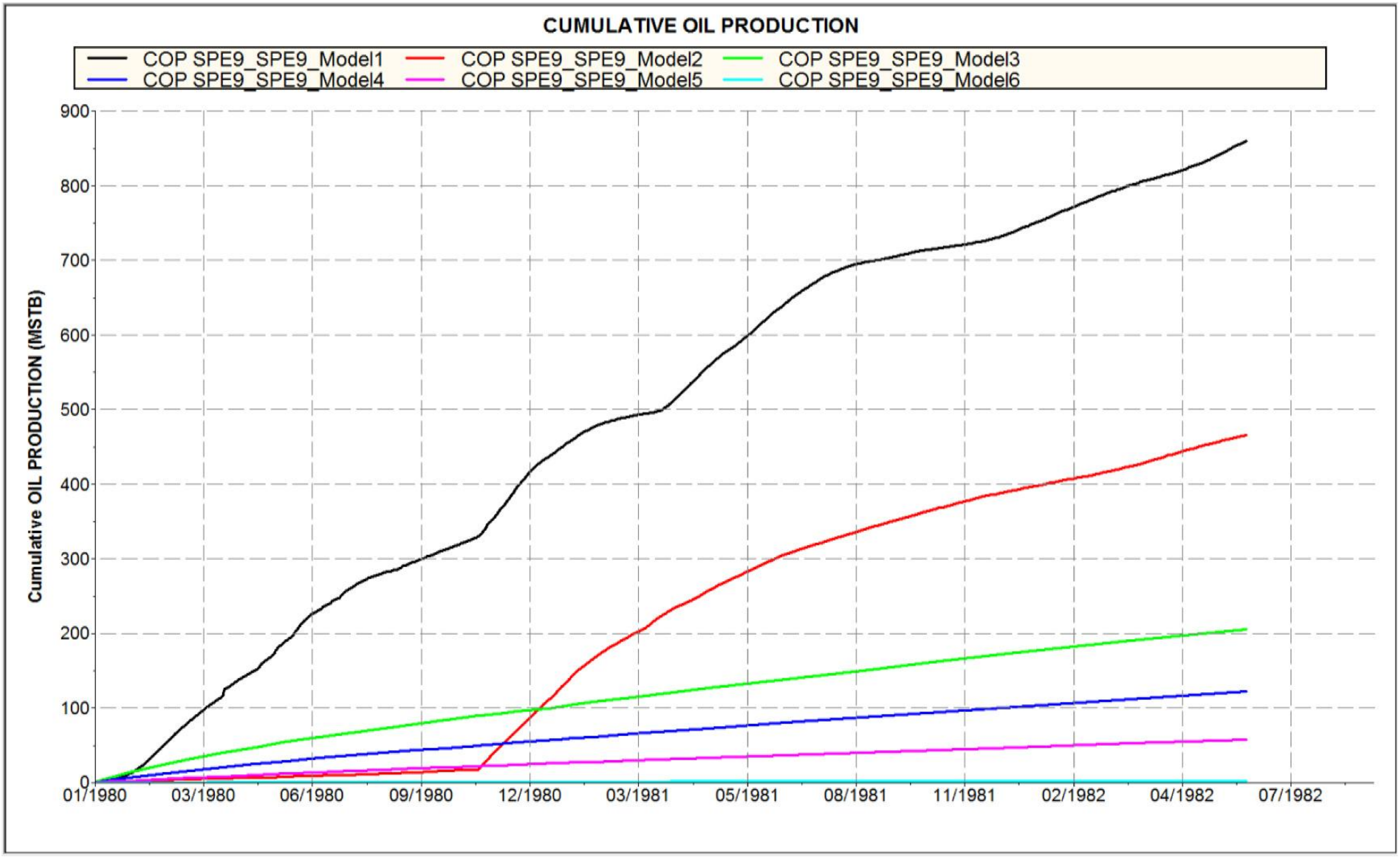

Figure 2. Cumulative oil production.

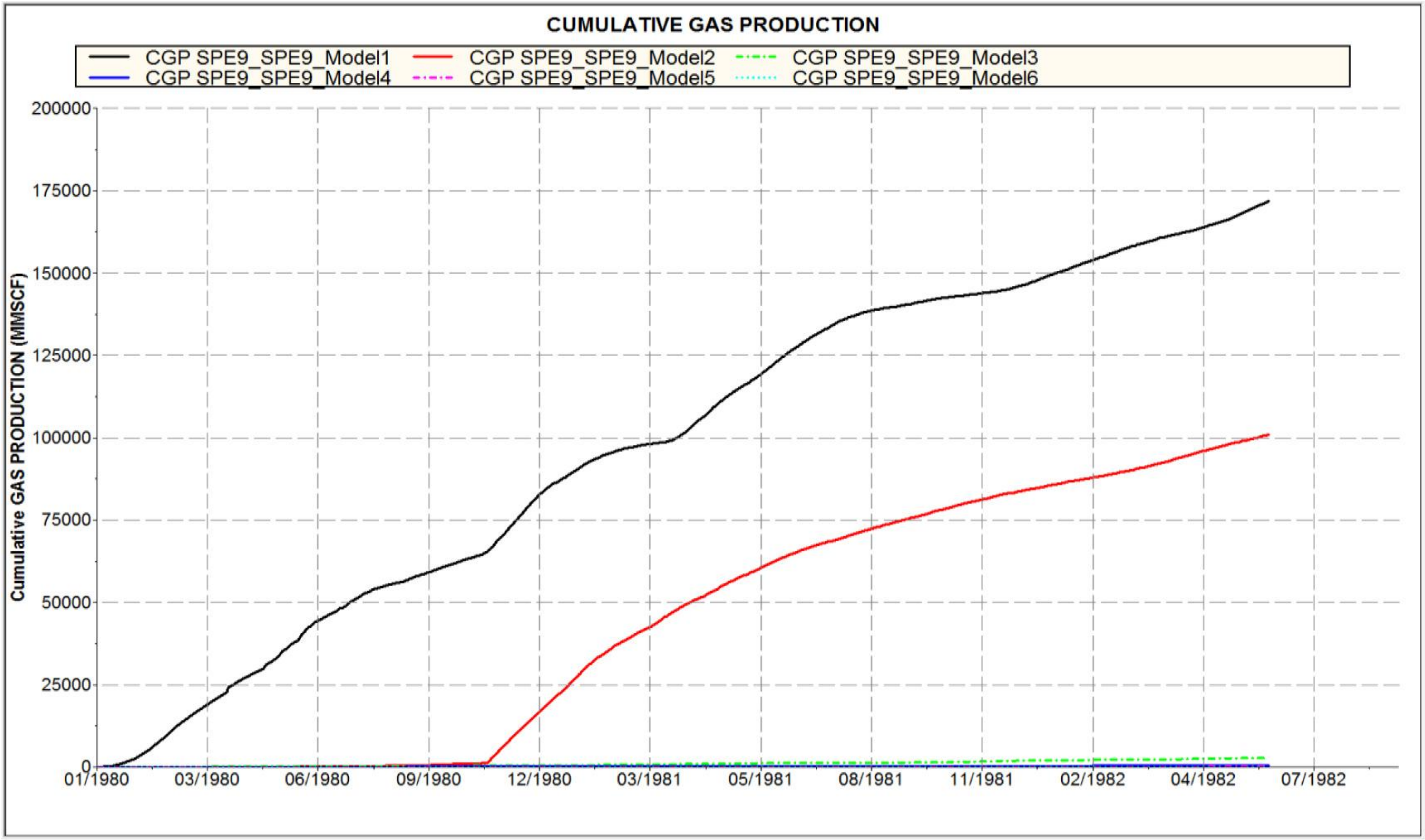

Figure 3. Cumulative gas production.

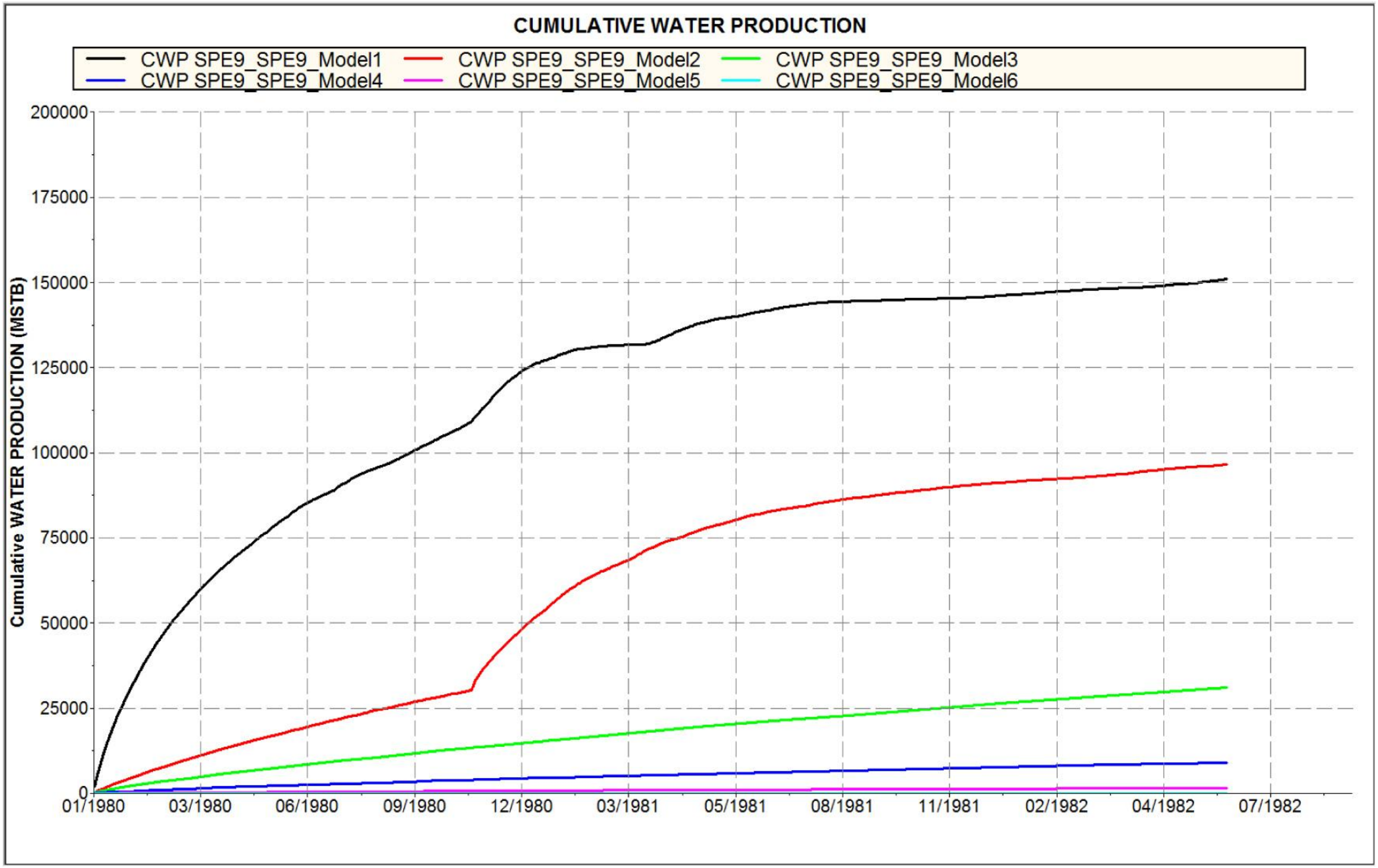

Figure 4. Cumulative water production.

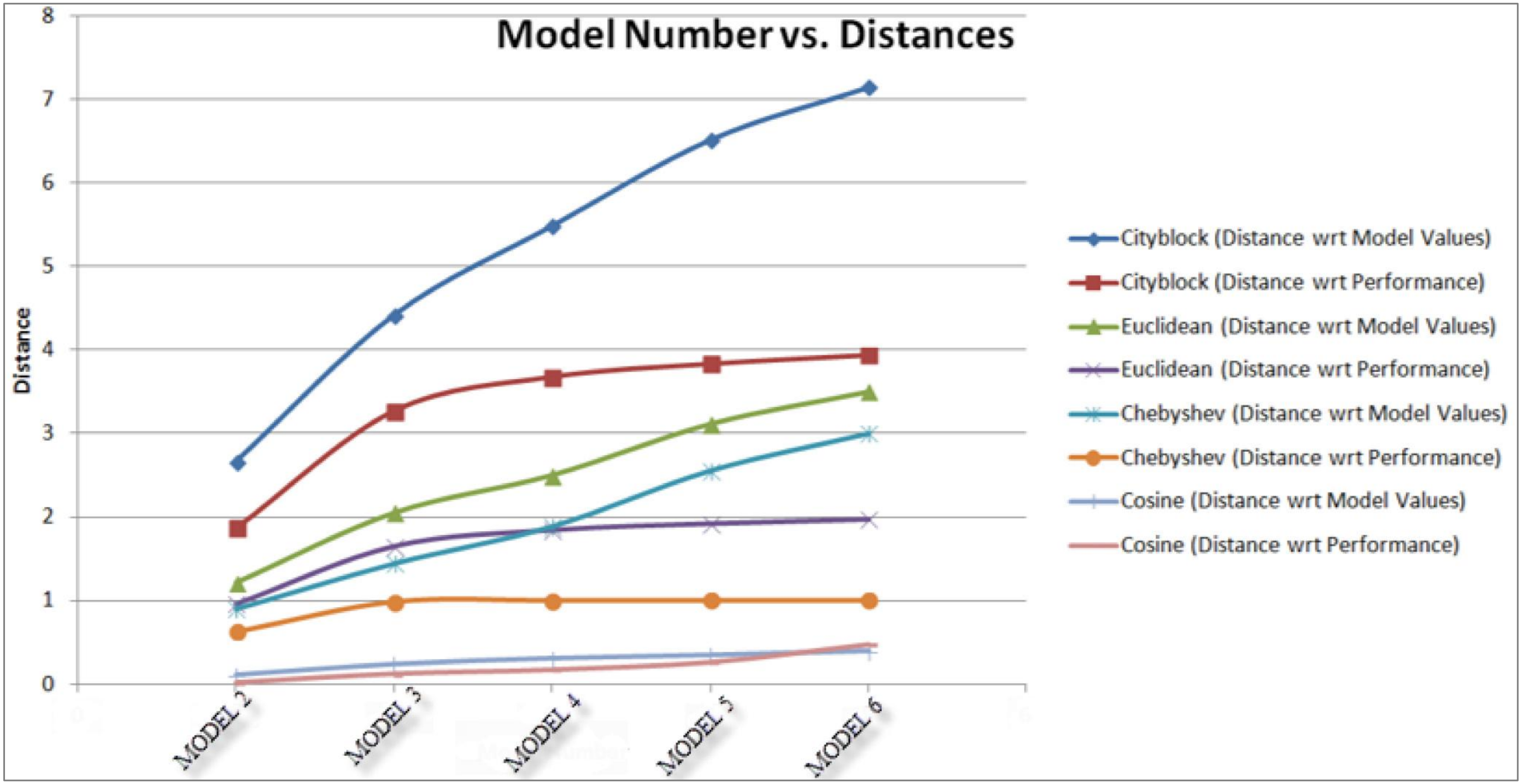

Figure 5. Comparison of results of distances evaluated using model and performance values.

Reservoir Matching Parameters	
Proximity	**Engineering Properties**
*Lateral Distance Between Subject and Analog	Average Depth (Ft)
Rock Properties	Original Bottom Hole Pressure (psia)
*Porosity (%)	Original Bottom Hole Temperature (Deg F)
*Permeability (md)	*Availability of MDT Data (Number of Data Points)
Permeability Distribution (Dykstra Parsons or Kv/Kh Ratio)	*Fluid Samples Obtained from MDT Data (Yes/No)
Gross Thickness (Ft)	Average Initial Well Producing Rate (BBL/Day or MCF/Day)
*Net Pay Thickness (Ft)	Average Drainage Area/Well (Acres)
Net-to-Gross Ratio	Average Well Spacing (Acres/Well)
*Hydrocarbon Saturation (1-Sw %)	Drive Mechanism
*Type/Quantity of Core Data	Production Mechanism (Development Scheme)
*Type of Open-Hole Evaluation Logs Available	Primary Product (Oil/Gas)
Geological Properties	Original In-Place Volume (Primary Product Oil/Gas)
Predominate Lithology	Cumulative Production of Primary Product/Date
*Geological Age	Recovery Factor To Date (%)
Depositional Environment	Estimated Primary Ultimate Recovery Factor (%)
*Same correlative stratigraphic interval or reservoir (Yes or No)	Method Determining Primary Recovery Factor (Performance/Simulation)
*Continuity (Comment on compartmentalization)	Total Recovery Factor Including Secondary Recovery (%)
Reservoir Area (Ac)	Ratio Producers to Injectors
Type of Original Fluid Contacts (LKH, O/W, G/W, GOC)	Flood Pattern Type
Productive Column Height (Ft)	**Fluid Properties**
*Availability of Seismic Data (2-D/3-D)	Oil Gravity (Deg API)
	Initial Solution GOR (CF/BBL)
	Oil Viscosity (cp)
	Mobility Ratio
	Gas Gravity
	Initial Condensate Yield (BBL/MMCF)
	Inert Gases (Mol % Each)

Table 1. List of reservoir parameters defined by SPE and SEC (Hodgin et al. 2006).

Metric	Description
Euclidean	Euclidean distance
Secludiean	Standardized Euclidean distance. Each coordinate difference between rows in X is scaled by dividing by the corresponding element of the standard deviation
Cityblock	City block metric
Minkowski	Minkowski distance, the default exponent is 2.
Chebychev	Chebychev distance (maximum coordinate difference)
Mahalanobis	Mahalanobis distance, using the sample covariance of input X
Cosine	One minus the cosine of the included angle between points (treated as vectors)
Correlation	One minus the sample correlation between points (treated as sequences of values)
Spearman	One minus the sample Spearman's rank correlation between observations (treated as sequences of values)
Hamming	Hamming distance, which is the percentage of coordinates that differ.
Jaccard	One minus the Jaccard coefficient, which is the percentage of non-zero coordinates that differ.

Table 2. Distance metrics for numerical attributes.

Reference Model Values	
Feature (Unit) \ Reference Model	***Model 1 (Reference)***
Average Porosity (fraction)	0.33
Average Permeability (md)	2000
Average Thickness (ft)	100
Average NTG (fraction)	0.8
Average Viscosity (cp)	100000
Average API gravity (API degrees)	9
Average Initial Pressure (psi)	3500

Table 3. Values of attributes for reference model 1.

Analogue Model Values					
Sample Model / Feature (Unit)	***Model 2***	***Model 3***	***Model 4***	***Model 5***	***Model 6***
Average Porosity (fraction)	0.3	0.24	0.2	0.25	0.15
Average Permeability (md)	1000	350	100	10	0,1
Average Thickness (ft)	60	55	40	30	25
Average NTG (fraction)	0.7	0.8	0.7	0.6	0.6
Average Viscosity (cp)	10000	400	100	50	100
Average API gravity (API degrees)	13	22	26	32	36
Average Initial Pressure (psi)	4200	5000	5350	6200	5600

Table 4. Values of attributes for analogue models.

Distance Between Reference (Model 1) and Other Models Based on Seven Attributes					
Sample Model Distance / Metric Type	Model 2 d_2	Model 3 d_3	Model 4 d_4	Model 5 d_5	Model 6 d_6
Cityblock	2.6604	4.4167	5.4854	6.5139	7.1444
Euclidean	1.2171	2.0542	2.5057	3.1187	3.5043
Chebyshev	0.9000	1.4444	1.8889	2.5556	3.0000
Cosine	0.1119	0.2389	0.3066	0.3456	0.3944

Table 5. Distance metric values between the reference model and the sample models.

	Model 1 (reference)	Model 2	Model 3	Model 4	Model 5	Model 6
OIL IN PLACE (MSTB)	3074930	1138440	837452	374190	289138	195933
CUMULATIVE OIL PRODUCTION (MSTB)	860.229	465.914	206.114	121.881	57.591	2.526
CUMULATIVE GAS PRODUCTION (MMSCF)	171698.2	100867.4	2749.963	428.88	66.391	2.984
CUMULATIVE WATER PRODUCTION (MSTB)	150968.3	96421.9	30916.1	9056.202	1487.865	43.416

Table 6. Simulation output.

Distance between Reference (Model 1) and Other Analogue Models based on Performance Parameters					
Sample Model Distance / Metric Type	***Model 2*** d_2	***Model 3*** d_3	***Model 4*** d_4	***Model 5*** d_5	***Model 6*** d_6
Cityblock	1.8620	3.2672	3.6741	3.8288	3.9330
Euclidean	0.9526	1.6457	1.8403	1.9160	1.9673
Chebyshev	0.6298	0.9840	0.9975	0.9996	1.0000
Cosine	0.0173	0.1211	0.1695	0.2610	0.4751

Table 7. Results based on performance parameters.

Attributes	Model 2	Model 3	Model 4	Model 5	Model 6
Cityblock (Distance wrt Model Values)	2.66	4.417	5.485	6.514	7.144
Cityblock (Distance wrt Performance)	1.862	3.267	3.674	3.829	3.933
Euclidean (Distance wrt Model Values)	1.217	2.054	2.506	3.119	3.504
Euclidean (Distance wrt Performance)	0.953	1.646	1.84	1.916	1.967
Chebyshev (Distance wrt Model Values)	0.9	1.444	1.889	2.556	3
Chebyshev (Distance wrt Performance)	0.63	0.984	0.998	1	1
Cosine (Distance wrt Model Values)	0.112	0.239	0.307	0.346	0.394
Cosine (Distance wrt Performance)	0.017	0.121	0.17	0.261	0.475

Table 8. Distances with respect to model and performance values.