

## Shiny web applications for unsupervised learning optimization applied to diagnostic fracture injection test event detection

Lukas Sadownyk, Marcelo Guarido, Danial Zeinabady, Erfan Sarvaramini, Zhenzihao Zhang, Farshad Tabasinejad, Hashem Salari, Kristopher A. H. Innanen, and Christopher R. Clarkson

### ABSTRACT

Diagnostic Fracture Injection Tests (DFIT), are commonly used to derive key parameters and other parameters for hydraulic fracture design and modeling. Although this process can identify properties needed for well optimization, it is also time intensive and affected by human interpretation bias. In this report, we address this adversity by applying unsupervised clustering methods: *K-Means*, *DB-Scan*, *Hierarchical modeling*, and *Gaussian mixture models* to identify point density variation that correlates to key parameters on a DFIT curve. A *R-Studio Shiny Web App*<sup>®</sup> is developed to apply these methods and provide a user-friendly platform for adjusting input variables and hyperparameters. Exploring the clustering approach emphasizes the importance of different variable combinations as well as noise considerations when interpreting a DFIT curve with clustering methods. Principle Component Analysis (PCA) further demonstrates *why* clusters occur where they do along a DFIT curve. Unsupervised clustering applied to DFIT data achieves an unbiased and quick workflow for event identification that can be scalable to large datasets.

### INTRODUCTION

To ensure successful and economic development of low permeability, hydrocarbon bearing organic-rich shales, a fracture stimulation design must be implemented to effectively liberate the hydrocarbons and optimize drilling programs. The process of designing and modeling a fracture program is a computationally intensive and iterative process that requires the estimation of multiple geologic and mechanical properties. These include permeability, formation pressure, fracture half-length, minimum horizontal stress ( $Sh_{min}$ ), instantaneous shut-in pressure (ISIP), breakdown pressure, fracture extension pressure, reservoir permeability, and fluid content (Clarkson et al., 2012). Although some of these parameters can be directly measured from core samples, many studies have identified the challenges of replicating *in-situ* conditions to produce accurate results (Clarkson et al., 2012; Venieri et al., 2020). To address this challenge, innovative technologies such as DFITs have been designed to measure key parameters such as ISIP,  $Sh_{min}$ , and reservoir pressure in the borehole of a well (Jung et al., 2016). This is achieved by injecting a small volume of fluids into the target formation to create a hydraulic fracture. By measuring the downhole or surface pressure change over time (pressure-time series) and observing the pressure decline after the hydraulic fracture is created these key parameters can be derived (Figure 1).

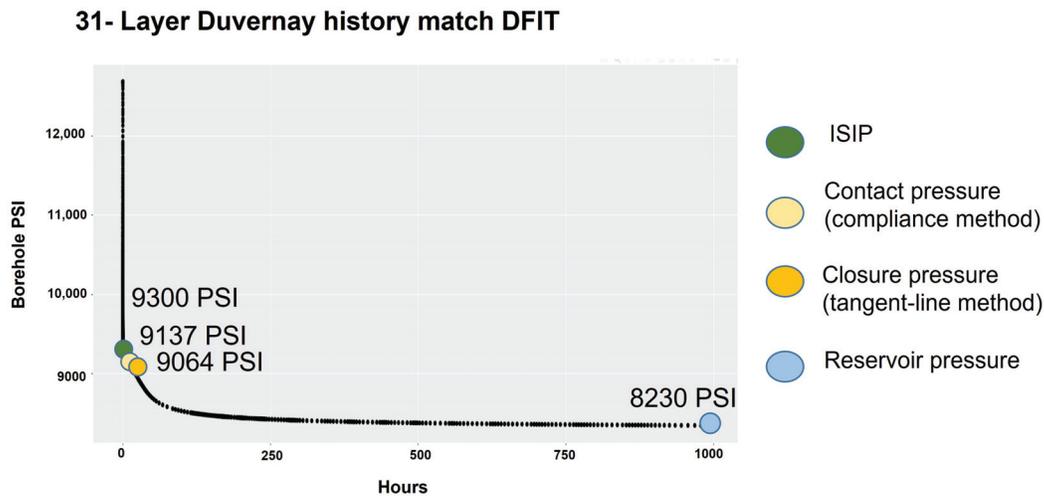


FIG. 1. ResFrac<sup>®</sup> history matched pressure falloff data obtained from simulation of a DFIT performed in the 31-layer Duvernay Formation model. Interpreted events along the curve are superimposed and color coded. Approximations of the  $Sh_{min}$  are made using both the compliance (McClure et al., 2016) and tangent line method (Barree et al., 2009).

Inspection of Figure 1 reveals that the interpretation of these key parameters directly from a pressure decline curve is non-trivial. Current methods of visualizing these hidden non-linear property relationships involve manual interpretation of derivative curves such as G-function, Bourdet derivatives, first-order derivatives, and Agarwal time (Zanganeh et al., 2018; Liu and Ehlig-Economides, 2018). The analytical nature of these methods has the unintended result of introducing human bias and error, coupled with the time-consuming prospect of defining these interpretations for multiple DFIT curves that exist for large datasets. This provides the opportunity to test and evaluate the ability of machine learning methods to resolve such adversities.

Despite the recent rise in machine learning applications to large datasets, little literature exists applying these methods for DFIT curve interpretation. Instead, current studies address interpolating missing DFIT pseudoradial flow data using Gradient Boosting (GB) and Random Forest (RF) regression methods (Mohamed et al., 2020) and the integration of real-time well stimulation datasets (injected proppant volumes, downhole pressures, and microseismic events) to identify stimulation related events using CNN, Autoencoders (AE) and Support Vector Regression (SVR) (Shen et al., 2020; Alatrach et al., 2020; Wang and Chen, 2019). This study aims to fill this gap and develop a workflow to identify reservoir parameters (*ISIP*,  $Sh_{min}$ , and *reservoir pressure*) from DFIT's with the aid of unsupervised clustering algorithms: *K-Means*, *DB-Scan*, *Hierarchical modeling*, and *Gaussian mixture models*. The application of this method intends to speed up interpretation times for datasets consisting of many DFIT curves and to eliminate human bias. Implementation and visualization of these clustering methods are complemented by the development of the *CREWES DFIT Clustering App* using *Shiny Web Apps from Rstudio*.

## Methods

This study evaluates the ability of unsupervised clustering methods *K-Means* (MacQueen, 1967), *DB-Scan* (Ester et al., 1996), *Hierarchical modeling* (Ward, 1963), and *Gaussian mixture models* (Redner and Walker, 1984) to identify key parameters: *ISIP*,  $Sh_{min}$ , and *reservoir pressure* in multivariate data. *K-Means* was selected as the baseline method for defining hyper-parameters; *DB-scan* is tested for handling noisy data; *Hierarchical modeling* offers cluster visualizations for hyper-parameter selection, and *Gaussian mixture models* offers the ability to fit model shape distributions in the form of probabilities.

The choice of applying unsupervised learning methods was influenced by its ability to eliminate bias that might otherwise exist in training datasets for supervised learning methods. Mathematically, it is hypothesized that the clustering algorithms will be purely segmenting the pressure decline curves based on the density of point distributions along original and derivative of pressure decline curves. The observations of these clusters will later allow for mathematical inferences to be made about *why* the clusters appear where they do. This approach follows a similar concept to the study by Ippolito et al. (2021) where well log facies identification is performed using unsupervised learning in conjunction with supervised learning to reduce bias. Although supervised learning will not be applied in this study, the developed method could be used to create training datasets for supervised learning applications. The application of unsupervised clustering methods also recreates a real life scenario where catalogues of events may be incomplete, or inaccurate making supervised learning unfeasible. Other studies by Li et al. (2021) use clustering more generally to identify anomalies in multivariate datasets. This idea closely parallels this study's method of identifying key-parameters ("anomalies") in a multivariate set of DFIT and pressure and pressure derivative curves.

To evaluate this method, three DFIT curves were clustered and compared to results from manual interpretation. These curves include history matched pressure decline models generated from a "simple" 3-layer Duvernay system, and a "complex" 31-layer Duvernay system. ResFrac<sup>®</sup> simulator was used to generate synthetic pump-in/shut-in response. ResFrac<sup>®</sup> is a fully coupled hydraulic fracturing, reservoir and wellbore simulator that models rigorously the key physical process involved in DFITs. The detail of ResFrac<sup>®</sup> conceptual model and numerical approach is described in McClure et al. (2021). The use of modeled curves allows for key parameters (*ISIP*,  $Sh_{min}$ , and *reservoir pressure*) to be known as simulation inputs eliminating any interpretation bias. Field data from a DFIT acquired in the Duvernay near Fox Creek, Alberta, Canada is lastly tested using the optimized hyperparameters from the model cases. In this example, key parameters are manually interpreted from the field DFIT in time (t), G-time, Agarwal time domains as well as their corresponding derivatives. Downsampling of this data was required to speed computation times in the clustering app.

Variables used for interpretation and multivariate analysis in the clustering application

are displayed in a correlation matrix in Figure 2. In this figure, there are fourteen time-dependent variables that can be interpreted to derive key parameters. To decrease noise effects, the Bourdet derivative (Duong, 1989) was applied to curves as a way of smoothing the data, this is indicated by a "B" in Figure 2. The compliance method (McClure et al., 2016) and tangent methods (Barree et al., 2009) are both used for the estimation of  $Sh_{min}$  outputting pressures of  $P_{contact}$  and  $P_{closure}$  respectively for the field DFIT. Manual interpretation workflows used in this study can be found in publications Zanganeh et al. (2018); McClure et al. (2016). Collectively,  $ISIP$ ,  $P_{contact}$ ,  $P_{closure}$ , and  $P_{reservoir}$  are the four events that were compared from model/interpretation results to unsupervised clustering results. Both  $P_{contact}$  and  $P_{closure}$  are included to determine which value is picked most consistently by cluster boundaries in the clustering app. To handle the different scales of measurement (Figure 2) scaling was applied before input into clustering algorithms to avoid any data bias.

### *CREWES DFIT Clustering App*

The developed *CREWES DFIT Clustering App* addresses the challenge of visualizing fourteen-dimensional data (Figure 2) with the application of unsupervised clustering and Principle Component Analysis (PCA). To develop the *CREWES DFIT Clustering App R-Studio*<sup>®</sup> programming software was used. *R-Studio*<sup>®</sup> offers the ability to design and create interactive web apps (*Shiny web app*) for data manipulation and visualization. The benefits of creating and using the *Shiny web app* include:

- (1) **Time saved:** Eliminate the need for a user to run multiple sections of code to generate multiple plots for multiple clustering analysis types.
- (2) **Reactive variables:** User can quickly manipulate hyperparameters for data fit.
- (3) **Intuitive display:** Back-end code runs without the user requiring extensive knowledge of programming.

Figure 3 displays a screen capture of this developed app. Additional features such as elbow plots and tree diagrams are also included as visualizations within the app to aid in hyperparameter definition (Figure 3). Evaluation of the performance of the multiple clustering outputs is achieved by defining measurements of:

- (1) Average number of events identified out of the 4 total events (varied hyperparameters).
- (2) Average number of unclassified points along curve (varied hyperparameters).
- (3) Average repeatability of classified points for varied hyper parameters (varied hyper-

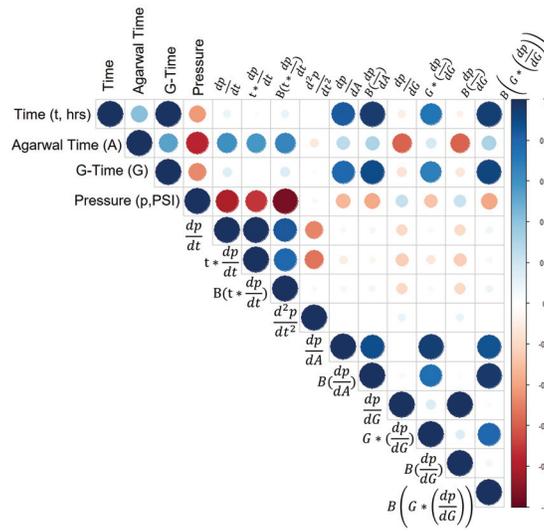


FIG. 2. Correlation matrix of all variables that can be derived from a pressure vs time DFIT measurement. The main diagonal corresponds to a variable correlated to itself, therefore, correlation is 1 (positive correlation large blue circle). Outside of the diagonal, correlations are displayed between different variables; red circles indicate negative correlation. The bottom left corner of the matrix has been eliminated due to its symmetry.

parameters).

PCA analysis is then used in the app to explore *why* cluster boundaries appear where they do (Figure 3). In this study, the cluster boundaries are defined as the event locations along the curve.

### Results

Using the "simple" and "complex" Duvernay models, optimal hyperparameters and variable combinations were tested to fit DFIT events representing key pressure data. The iterative process also allowed for inferences to be made about the relative importance of variables and their overall contribution to clustered events. This is displayed as a PCA circle in Figure 4. In this figure, DFIT pressure vs log (time) curve for the "complex" model is clustered using variables groups identified by the PCA correlation circle.

Following variable analysis, elbow plots and visual inspection is used to determine optimal hyperparameters for each clustering method. Applying the three measures of performance allows for this optimization to be quantified. Optimization results for each clustering method are displayed in Table 2. Using these clustering parameters, the three measures of performance can be further used to determine the best clustering method out of the four methods tested (Figure 5). These measures of performance are compared for the three DFIT model datasets in this study (Figure 5).

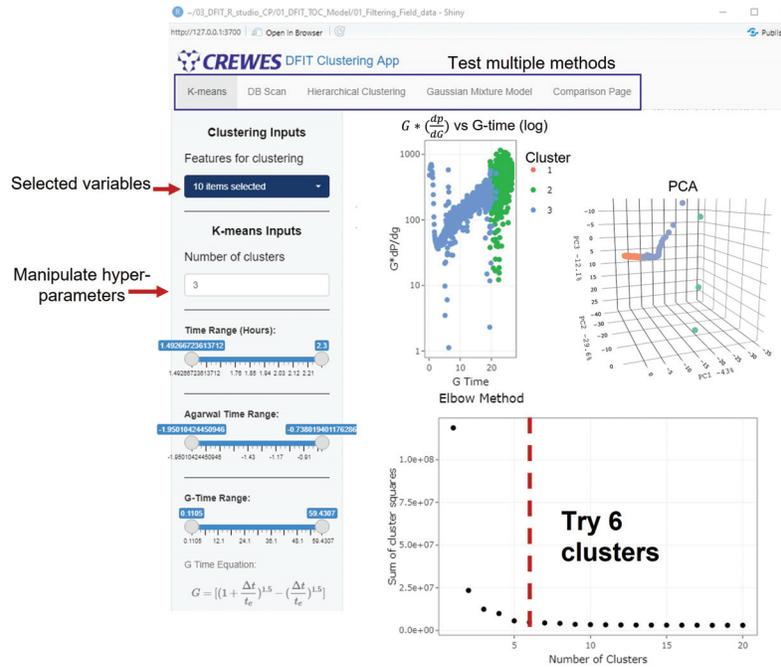


FIG. 3. Screen capture of the interactive *CREWES DFIT Clustering App* with its various analytical features highlighted.

The combination of the optimized variables, hyperparameters, and clustering method that produced cluster boundaries correlated to key parameter events displayed in Figure 6 for the "simple," "complex" and raw Field Duvernay DFIT curves respectively. On these plots, the percent difference from the true/interpreted value is also displayed.

Following the identification of optimal clustering variables, hyperparameters, and clustering methods, the mathematics of *why* clusters appear where they do is explored by examining PCA plots of fourteen-dimensions reduced to three (Figure 7).

Table 1. Optimized Clustering Hyperparameters

Clustering method	Parameters
K-means	6-clusters
DB-scan	Minimum points = 10, Search radius = 0.1
Hierarchical Clustering	Ward-D, 6-clusters
Gaussian mixture model	VEE (Ellipsoidal, equal shape and orientation) 6-clusters

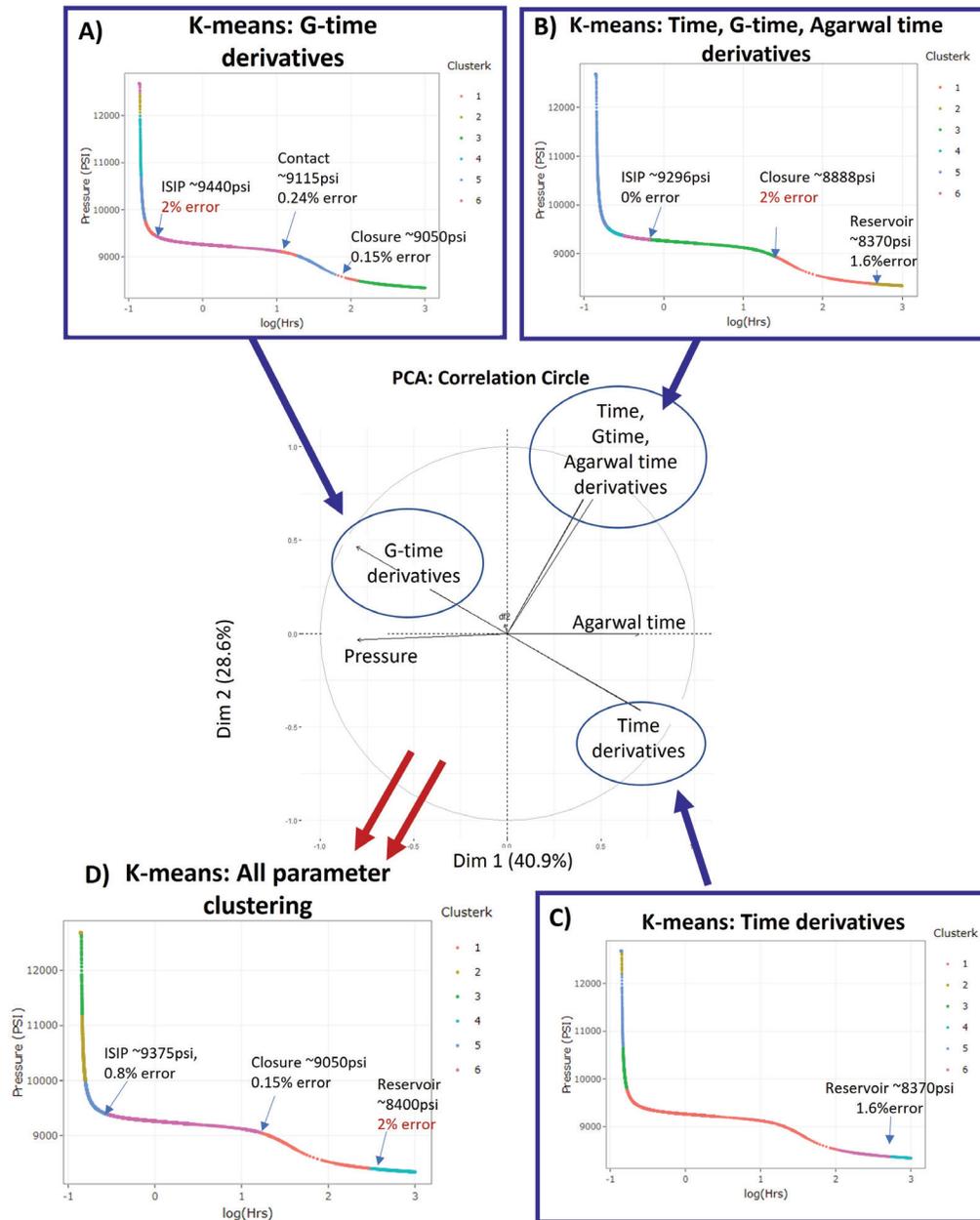


FIG. 4. Results of testing different variable clustering for K-means using the 31-layer DFIT Duvernay model. Subsets for variable testing were determined using the PCA correlation circle (center) where clusters of variables correspond to high correlation. The axes on this plot are labeled Dim 1 and Dim 2 representing dimensions that capture 41% and 28% of the variation in the data respectively. Variables outside of clusters indicate an increasing negative correlation. Using this information, plot A shows an example where only G-time derivatives are used as inputs of the clustering algorithm: the outputs of the clustering are displayed on a pressure (PSI) vs log time with key events identified and their corresponding error relative to the model inputs. Plot B shows the a similar plot using Time, G-time, and Argarwal time derivatives. Plot C illustrates the use of the with only time derivate input for this purpose. Plot D shows the result of using all variables in the clustering algorithm.

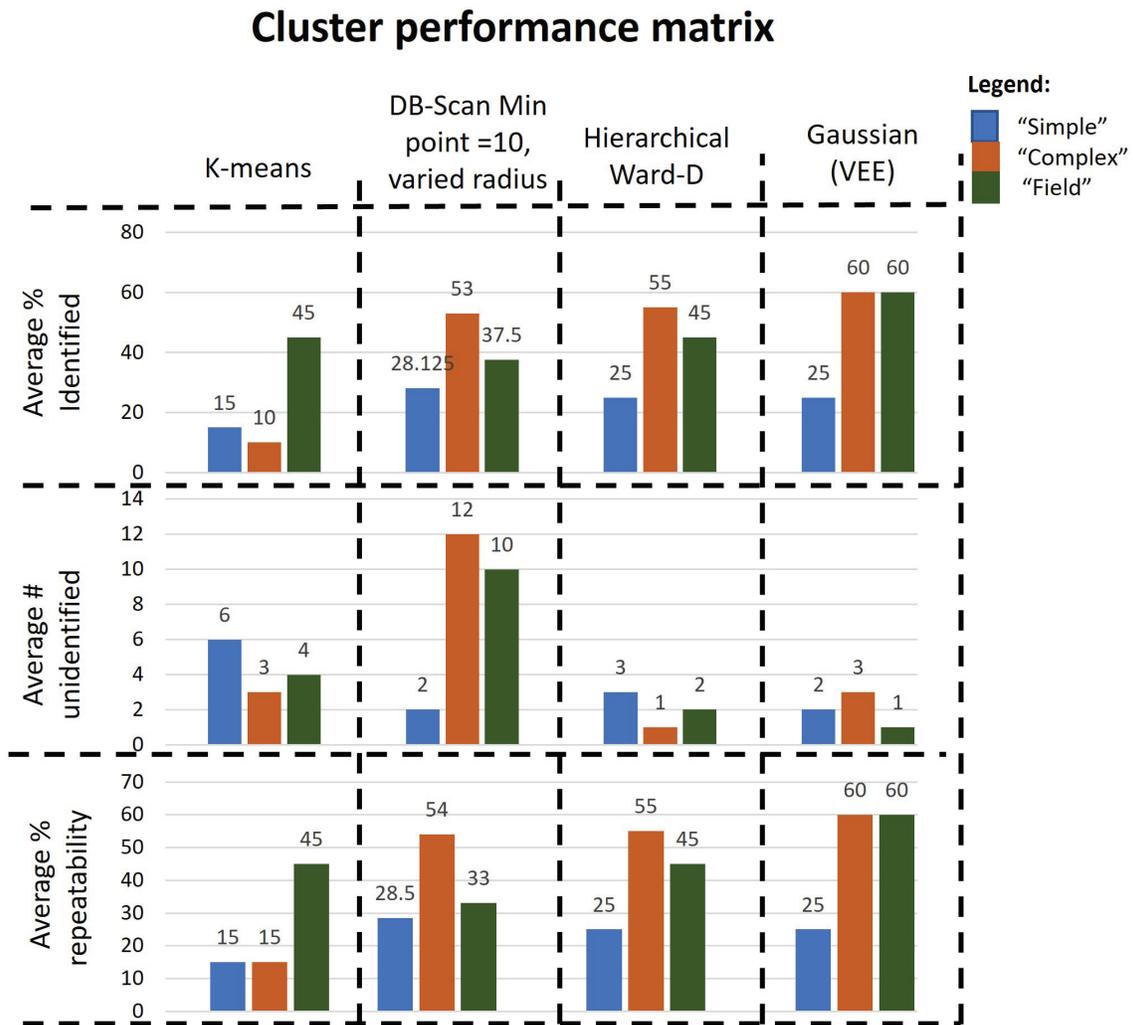
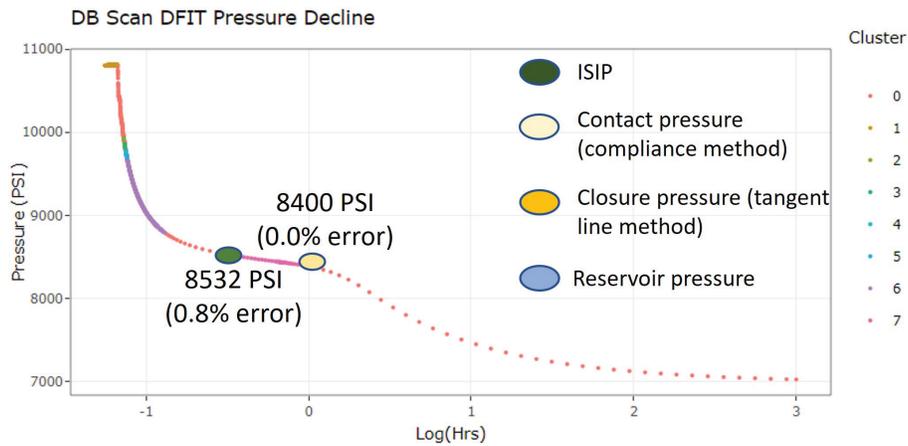
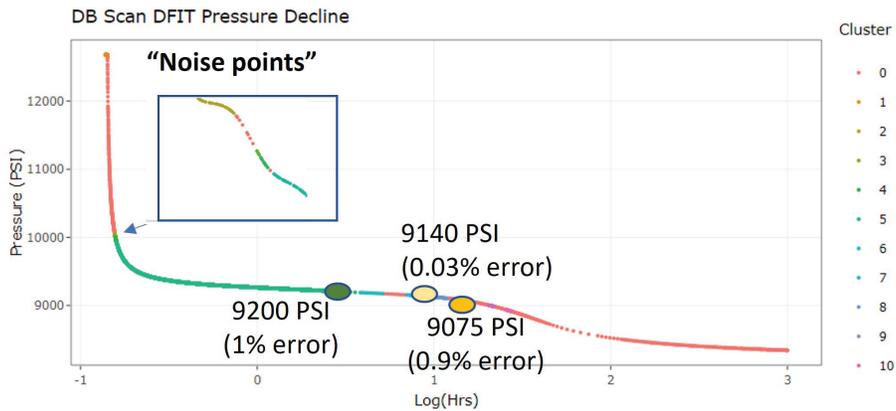


FIG. 5. The cluster performance matrix compares the three designed metrics of evaluation to selected clustering type. Bar plots for each dataset tested are colour coded, blue = "simple" model DIFT, orange = "complex" model DFIT, green = Field DFIT.

### A) 3-layer Model Dbscan



### B) 31 layer model Dbscan



### C) Field test: Gaussian mixture

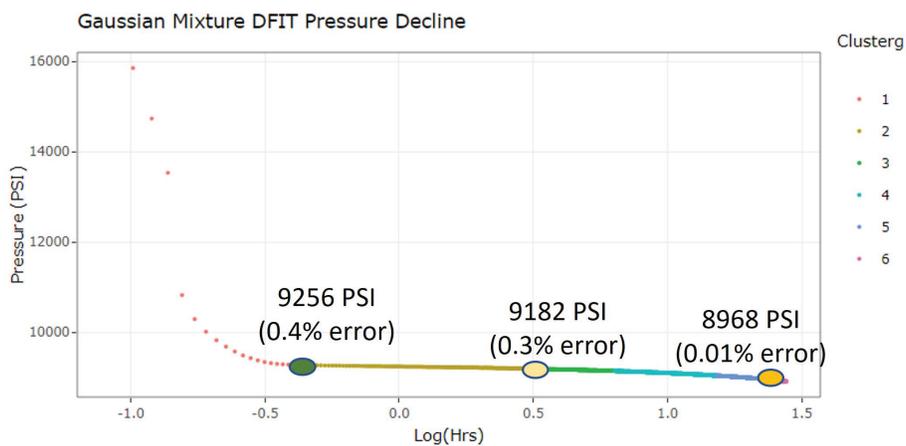


FIG. 6. Results of applying optimal variable combination, hyperparamters, and clustering method to each of the three datasets in this study. Percent error is indicated on each plot to quantify difference from true/interpreted values.

## 3D PCA: "Complex" model K-means

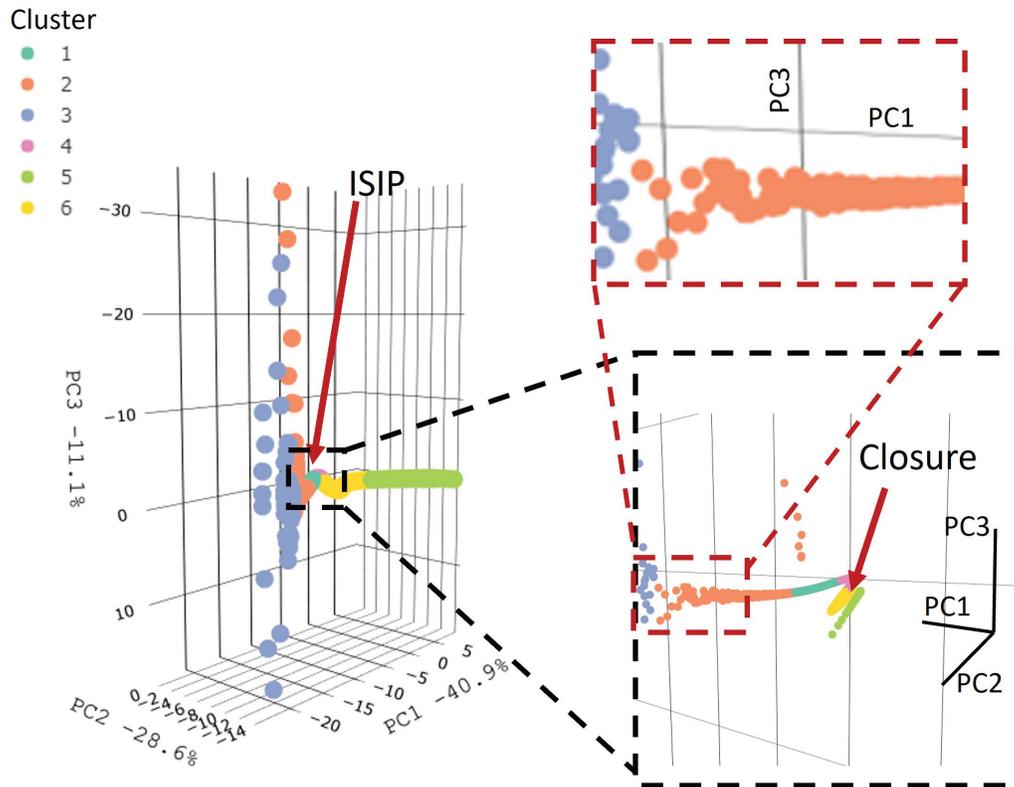


FIG. 7. Three-dimensional principle component plot used to understand distribution of clusters created from DFIT data. This case shows the "complex" DFIT model with its associated identified events using the K-means algorithm. This figure includes different perspectives of the data to understand variation.

## Discussion

This new workflow using *Rstudio Shiny Web Apps*<sup>®</sup> has helped us understand and evaluate the feasibility of using unsupervised clustering methods to interpret events on a DFIT pressure decay curve that can be used to derive key parameters. To achieve this, variable contributions from derivative curves were first explored using a PCA correlation circle. In Figure 4, it appears that reducing the original and derivative data into two dimensions (Dim 1 and Dim 2) has created natural clusters of correlation. To further explore these clusters of correlation, the clustering algorithm, K-means, was run using only the variables within each cluster. Plot A shows that clustering of only G-time derivatives has resulted in events ISIP (highest relative error), contact and closure event identification. Plot B illustrates that an accurate ISIP, and lower accuracy closure and reservoir pressure estimates, can be obtained when clustering is performed using Time, G-time, and Agarwal time derivatives.

Using time derivatives results in reservoir pressure being the only event identified in plot C. Collectively, these clusters of correlation appear to extract pressure information from different segments of the curve. The G-time events appear to give accurate contact and closure approximations (early time variables). This differs from the Time, G-time, Agarwal time correlation circle which results in accurate values for ISIP and reservoir pressure are now output. These clusters of variables will be referred to as the late/early time variables based on this distribution. Lastly, the time derivative cluster appears to correlate with only the late-time variable reservoir pressure (late-time variables).

Collectively, each variable cluster in the PCA plot appears to tell us different information for early and late time segments along a DFIT curve. Most notably, the early time variables (G-time derivatives) exist on the opposite side of the PCA correlation circle from the late time variables (Time derivatives) suggesting that these clusters of negatively correlated. This interpretation is supported by no event overlap existing in plots A and C. Plot B contains information from both the early and late time stages of the DFIT. Therefore, its cluster of variables has some correlation to the pure late and early time variables displayed in the correlation circle of Figure 4 lying orthogonal to these clusters. This may imply that depending on the events required from the clustering algorithm, different selections of variables can optimize output results for early/late time DFIT events. For the purpose of this study, a generalized approach was taken whereby all clusters of variables are merged to produce a holistic interpretation of the DFIT curve. This is displayed in plot D of Figure 4. In this plot, it appears that an averaging of the components of each variable cluster has created a holistic interpretation. This plot identifies ISIP with improved accuracy compared to plots A and C, improved closure pressure, compared to plots B and C, and a reservoir pressure that was non-existent in plot A.

The collective 14-variable approach then allows for cluster hyperparameters to be optimized. Manually testing parameters leads to the optimizations displayed in Table 2. Generally, 6-clusters for *K-means*, *Hierarchical*, and *Gaussian* clustering methods appear to fit the data best.

Following the optimization of variable inputs and hyperparameters, clustering methods are quantitatively compared using the three parameters of evaluation in the form of a cluster performance matrix (Figure 5). Using the "simple" model, DB scan appears to have identified the highest percentage of events, while having the lowest average percentage of noise points and highest repeatability with hyperparameters variation. This trend appears to shift as the model becomes more complex. For the complex Duvernay model, DB scan seems to retain high event identification and repeatability, however, the quality of value uniqueness is significantly reduced by the number of unidentified points that appear in the result (12). Compared to other methods, Hierarchical clustering has improved metrics for all three measures of performance, and the Gaussian-based method appears to have optimal performance with the highest average percentage of events identified and repeatability. Analysis of the field data has also revealed the Gaussian mixture model method as the top

performer. In the field case, the quality of the DB scan method appears to have degraded in quality while the K-means method has improved. It is important to note that Figure 5 shows the averages of multiple iterations of the clustering algorithms with varied hyperparameters to quantify the algorithm's ability to handle deviations from optimal hyperparameters (repeatability). Individual tests with optimized variable input, hyperparameters, and clustering method were found to output 75 -100% identification. The differentiator between these percentages is merely dependent on if the length of recorded data is long enough to extract reservoir pressures. This applies to the field test case.

DB-scan clustering methods appear to produce optimal results when the DFIT is "simple" and degrade as complexities are introduced. This is illustrated in Figure 6 where the application of DB-scan to the complex model (plot B) has resulted in more noise clusters being identified, degrading the uniqueness of key parameter event detection on the curve. DB-scan applied to the simple model appears to have identified accurate ISIP and contact events in Figure 6A, with little noise, while other events appear to be missed. This is explained by the lower apparent sampling of the simple model data creating sparse point density displayed in Figure 6A. Perhaps the sampling interval has affected the accuracy of the clustering methods and this may be a subject of future study.

For the field test, (Figure 6C) we see that the Gaussian method appears to have identified ISIP, contact, and closure pressures with only two unidentified cluster boundaries that exist between the contact and closure pressures. These boundaries may be indicative of a more accurate  $Sh_{min}$  estimate highlighting another area of future study. In this case, the field data were not collected for a long enough time to accurately identify a reservoir pressure, therefore, no interpretation exists in these plots. Overall we see that Gaussian clustering methods appear to handle noisy data better than DB-scan. This may be explained by the ability to fit probability distributions to the data that better handle noise. Future experiments may consist of filtering the DFIT data to see if this will improve clustering results.

There still remains the question as to *why* these clusters appear where they do. Inspection of Figure 6 does not conclusively show why these clusters occur. This question can be addressed when observing three-dimensional PCA plots for the complex model with K-means applied (Figure 7). In this figure, events appear to correlate with inflections in data trend (see ISIP and closure in original and first magnification of PCA) and frequency variation in data (see the largest magnification). This implies that a combination of dimensionality and frequency changes in the data affect point density and cluster distribution. It is hypothesized that the input of these manual interpretation derivative curves into the clustering algorithms has allowed for these changes to be identified that line up with DFIT parameter events.

## CONCLUSIONS

Successful development of unconventional hydrocarbon reservoirs is dependent on designing and modeling an effective stimulation program. This typically requires estimation of critical parameters  $ISIP$ ,  $Sh_{min}$ , and *reservoir pressure* via manual interpretation methods using DFIT pressure falloff data. Although this process may produce values for stimulation modeling, it is time-consuming and can be affected by human interpretational bias. To address this adversity, a new method of applying unsupervised clustering methods in the *CREWES DFIT Clustering App* was developed. This app allowed for quick visualization and manipulation of clustering variables and hyperparameters to find the best fit interpretation for three sets of DFIT pressure falloff data. Clustering calibration results found that different variable inputs into the clustering algorithm result in different events being identified along the DFIT curve. These were classified as early and late time variables. For a generalized interpretation of the DFIT curve, the variable clusters are merged to identify late and early events that occur along the curve.

Optimized results suggested that the *DB-scan* method can accurately define event boundaries on the "simple" DFIT models, however, the introduction of geologic complexity and noise degrades the result as more unclassified clusters appear in the interpretation. This is where the Gaussian mixture method appears to handle noise variations with improved accuracy for the "complex" and field DFIT tests. It is hypothesized that the ability to change the shape of the probability distribution fitting the data in this method has addressed any noise contamination and effects of geologic complexity. Future studies will focus on eliminating this noise with filters.

Understanding *why* clusters occur where they do is achieved by using PCA to reduce the fourteen-dimensional data down to three dimensions. This process revealed that cluster boundaries occur at inflection points (changes in dimensionality) and frequency variation in the data correlates to a variation in point density.

The *CREWES DFIT Clustering App* offers the ability to quickly interpret and reduce bias in DFIT-derived parameter estimates. Future studies will be directed toward using this method to create training data for supervised learning methods for automated event identification.

## ACKNOWLEDGEMENTS

We thank the sponsors of CREWES for continued support. This work was funded by CREWES industrial sponsors and NSERC (Natural Science and Engineering Research Council of Canada) through the grant CRDPJ 543578-19. The first author was partially supported by a scholarship from the NSERC Alexander Graham Bell Canada Graduate Scholarship (CGS-M). The first, second, fifth, and seventh authors are funded by the Canada First Research Excellence Fund, through the Global Research Initiative at the Uni-

versity of Calgary. The third and sixth authors are supported by the NSERC Alliance grant. We gratefully acknowledge Mark McClure (president of ResFrac) for providing academic licenses of ResFrac for use in this study.

## REFERENCES

- Alatrach, Y., Mata, C., Omrani, P. S., Saputelli, L., Narayanan, R., and Hamdan, M., 2020, Prediction of well production event using machine learning algorithms: Presented at the Abu Dhabi International Petroleum Exhibition and Conference.
- Barree, R. D., Barree, V. L., and Craig, D., 2009, Holistic fracture diagnostics: Consistent interpretation of prefrac injection tests using multiple analysis methods: *SPE Production and Operations*, **24**, 396–406.
- Clarkson, C. R., Jensen, J. L., and Chipperfield, S., 2012, Unconventional gas reservoir evaluation: What do we have to consider?: *Journal of Natural Gas Science and Engineering*, **8**, 9–33.
- Duong, A. N., 1989, A new set of type curves for well-test interpretation with the pressure/pressure-derivative ratio: *SPE Form Eval* **4**, **02**, 264–272.
- Ester, M., Kriegel, H. P., Sander, J., and Xu, X., 1996, A density-based algorithm for discovering clusters in large spatial databases with noise: *kdd*, **96**, 226–231.
- Ippolito, M., Ferguson, J., and Jenson, F., 2021, Improving facies prediction by combining supervised and unsupervised learning methods: *Journal of Petroleum Science and Engineering*, **200**, 108,300.
- Jung, H., Sharma, M. M., Cramer, D. D., Oakes, S., and McClure, M. W., 2016, Re-examining interpretations of non-ideal behavior during diagnostic fracture injection tests: *Journal of Petroleum Science and Engineering*, **145**, 114–136.
- Li, J., Izakian, H., Pedrycz, W., and Jamal, I., 2021, Clustering-based anomaly detection in multivariate time series data: *Applied Soft Comutung Journal*, **100**, 106,919.
- Liu, G., and Ehlig-Economides, C., 2018, Practical considerations for diagnostic fracture test (dfit) analysis: *Journal of Petroleum Science and Engineering*, **171**, 1133–1140.
- MacQueen, J., 1967, Some methods for classification and analysis of multivariate observations: *Proc. Fifth Berkeley Symp. on Math. Statist. and Prob.*, **1**, 281–297.
- McClure, M., Kang, C., Medam, S., and Hewson, C., 2021, Resfrac technical writeup: <https://arxiv.org/abs/1804.02092v9>.
- McClure, M. W., Jung, H., Cramer, D. D., and Sharma, M. M., 2016, The fracture-compliance method for picking closure pressure from diagnostic fracture-injection tests: *SPE Journal*, **04**, 1321–1339.
- Mohamed, M. I., Mehta, D., Salah, M., Ibrahim, M., and Ozkan, E., 2020, Advanced machine learning methods for prediction of fracture closure pressure, closure time, permeability, and time to late flow regimes from dfit: *SPE/AAPG/SEG Unconventional Resources Technology Conference*.
- Redner, R. A., and Walker, H. F., 1984, Mixture densities, maximum likelihood and the em algorithm: *SIAM Review*, **26**, 195–239.
- Shen, Y., Cao, D., Ruddy, K., and Moraes, L. F. T. D., 2020, Near real-time hydraulic fracturing event recognition using deep learning methods: *SPE Drill and Compl*, **35**, 478–489.

- Venieri, M., Weir, R., McKean, S. H., Pedersen, P. K., and Eaton, D. W., 2020, Determining elastic properties of organic rich shales from core, wireline logs, and 3-d seismic: *Journal of Natural Gas Science and Engineering*, **84**, 103,637.
- Wang, S., and Chen, S., 2019, Insights to fracture stimulation design in unconventional reservoirs based on machine learning and modeling: *Journal of Petroleum Science and Engineering*, **174**, 682–695.
- Ward, J. H., 1963, Hierarchical grouping to optimize an objective function: *Journal of the American Statistical Association*, **58**, 236–244.
- Zanganeh, B., Clarkson, C. R., and Jones, J. R., 2018, Reinterpretation of flow patterns during dfits based on dynamic fracture geometry, leakoff and afterflow: SPE Hydraulic Fracturing Technology Conference and Exhibition.

## APPENDIX

Table 2. List of nomenclature

<b>Parameter</b>	<b>Description</b>
$ISIP$	Instantaneous Shut In Pressure
$Sh_{min}$	Minimum horizontal stress
$P_{closure}$	Minimum horizontal stress estimation using tangent-line method (Barree et al., 2009)
$P_{contact}$	Minimum horizontal stress estimation using compliance method (McClure et al., 2016)
$P_{reservoir}$	Reservoir pressure