Journal of Petroleum Science and Engineering 217 (2022) 110908

Contents lists available at ScienceDirect

Journal of Petroleum Science and Engineering

journal homepage: www.elsevier.com/locate/petrol

Capillary pressure correction of cuttings

S. Alessa [a], A. Sakhaee-Pour [a,*], F.N. Sadooni [b], H.A. Al-Kuwari [b]

[a] Department of Petroleum Engineering, University of Houston, Houston, TX, USA
[b] Environmental Science Center, Qatar University, P. O. Box, 2713, Doha, Qatar

ARTICLE INFO

Keywords:
Mercury injection capillary pressure (MICP)
K-nearest neighbor (KNN)
Locally selective combination in parallel outlier ensembles (LSCP)
Savitzky–golay filter (SG)

ABSTRACT

The accurate characterization of capillary pressure is essential in determining multiphase flow behavior in subsurface conditions. It is also essential in quantifying reservoir rock quality, reservoir fluid saturations, and the thickness of the transition zone. Mercury injection has become a routine measurement for capillary pressure characterization, but the existing technology is primitive. For samples with irregular shapes, such as cuttings, unconfined pieces are placed in an empty cell before injection. The raw capillary pressure measurements show unrealistic entry pressure corresponding to filling the empty cell and closing microcracks. This study proposes a simple relation for determining accurate entry pressure. The proposed relation is applied to the actual measurements of seven shale samples, and its performance is improved using k-nearest neighbors (KNN), locally selective combination in parallel outlier ensembles (LSCP), and Savitzky–Golay (SG) filters. The optimal solution is obtained by combining the simple relation with unsupervised machine learning and noise filtering techniques in series. The proposed relation, which is corroborated by high-resolution images, provides a new approach to determining true entry pressure and has applications in characterizing multiphase flow in unconventional formations.

1. Introduction

The U.S. Energy Information Administration estimates the proved shale gas reserve to be more than 350 trillion cubic feet (E.I.A, 2021). In addition, a U.S. geological survey reported that Texas and New Mexico's Wolfcamp Shale and Bone Spring Formation holds 46 billion barrels of oil, 281 trillion cubic feet of gas, and 20 billion barrels of natural gas liquid (Gaswirth et al., 2018). As shale's pore size is measured in nanometers, unleashing its energy potential requires a good understanding of nanofluidics, a field that studies the multiphase flow in sub-100-nm conduits (Hendraningrat et al., 2013). The distribution of the fluid in these pores may depend on fluid properties (Deinert and Parlange, 2009: Lee et al., 2020; Singh and Singh, 2011), whereas in more permeable formations, the effects of pore size are less significant.

Researchers have used mercury injection capillary pressure (MICP) measurements to estimate matrix permeability (Purcell, 1949; Swanson, 1981). For example, Tran et al. (2018) estimated shale permeability from MICP based on the acyclic pore model. The accurate characterization of capillary pressure is critical for accurate permeability estimation and characterization of the pore-throat size distribution (Peng et al., 2017).

Determining actual entry pressure from mercury injection is challenging, and researchers have proposed different models to address this problem. First, Bailey (2009) analyzed sample compression to determine entry pressure. Later, Lan et al. (2017) proposed a dual-compressibility model by dividing the pores into accessible and inaccessible. More recently, Davudov et al. (2018) studied conformance, grain compressibility, and inaccessible-pore compressibility effects to determine actual intrusion.

One of the most used methods was proposed by Comisky et al. (2011), who analyzed the variation of pore compressibility with capillary pressure and observed an exponential trend in the drainage. The trend specifies a capillary pressure range, with the upper limit corresponding to entry pressure. The lower boundary, which is not of interest here, is conformance pressure. However, this approach is challenging when there is significant noise in the data.

This study proposes a simple relation determining accurate entry pressure to correct the capillary pressure measurements obtained from mercury injection in samples with irregular shapes. The study uses two unsupervised machine learning methods and the Savitzky–Golay (SG) filter to improve the performance where there is noise. The relation is applied to the capillary pressure measurements of seven shale samples.

* Corresponding author.
E-mail address: asakhaee@central.uh.edu (A. Sakhaee-Pour).

https://doi.org/10.1016/j.petrol.2022.110908
Received 31 January 2022; Received in revised form 17 July 2022; Accepted 21 July 2022
Available online 8 August 2022

1.1. Pre-drainage stages

Often, large samples, such as core plugs, are unavailable, and cuttings or chips are the only (rock) materials available for capillary pressure characterization. Samples are usually dried and weighed before mercury injection. They are then placed inside a penetrometer bulb, as shown in Fig. 1.

Mercury injection is divided into three stages (Fig. 2). First, the bulb is filled with mercury without significant compression of the sample. This stage (I) occurs at low capillary pressure because there is no significant resistance against filling the bulb. Second, the mercury compresses the sample without invading the pores in the sample. Pore compressibility dominates the second stage because the pores are compressed more quickly than the solid domain in the matrix. The final stage is the actual entry when capillary pressure exceeds entry pressure.

1.2. Pore compressibility in the second stage

The first and second stages in Fig. 2 are similar in that the mercury does not invade the pore space in the matrix. Nevertheless, the two are distinguished because the variations of the corresponding capillary pressures with injected volume are different. In the second stage, bulk compressibility (C_b) is expressed as follows (Anderson and Jones, 1985; Jizba, 1991; Niandou et al., 1997):

$$C_b = a_1\, \sigma_{Con}^{a_2}, \tag{1}$$

where σ_{Con} is confining stress, anda_1 and a_2 are constants. Bulk compressibility represents the effects of both rock and pore compressibility (McLatchie et al., 1958; Zimmerman, 1990), and it can be expressed as follows:

$$C_b = \varnothing C_{pc} + C_m, \tag{2}$$

where C_{pc} is pore compressibility, $\varnothing$ is total porosity, and C_m is matrix compressibility.

Hawkins (1955) defined effective compressibility (C_{eff}) as follows:

$$C_{eff} = \frac{S_{oi}C_o + S_{wi}C_w + C_f}{1 - S_{wi}}, \tag{3}$$

where S_{oi} is initial oil saturation, S_{wi} is initial water saturation, C_o is oil compressibility, C_w is water compressibility, and C_f is formation compressibility.

Because the sample is dried prior to the test, the initial oil and water saturations are zero. There is also no mercury inside the sample in the first two stages; thus, effective compressibility is equal to formation compressibility:

$$C_{eff} = C_f = \varnothing C_{pc} + C_m. \tag{4}$$

Pore compressibility (C_{pc}) dominates formation compressibility because the rock compressibility (C_m) is much smaller (Ahmed, 2018). Pore compressibility is the fractional change in pore volume (PV) versus confining stress (σ_{Con}) as follows (Baker et al., 2015):

$$C_{eff} = \varnothing C_{pc} = \frac{1}{PV}\frac{dPV}{d\sigma_{Con}}. \tag{5}$$

Subsequently, we can apply the power law relation to express effective compressibility using the constants C_{po} and m:

$$C_{eff} = \frac{1}{PV}\frac{dPV}{d\sigma_{Con}} = \varnothing\left(C_{po}P_{cHg}^{m}\right). \tag{6}$$

The instrument is tuned to account for glass and mercury compressions. Thus, it is assumed that the recorded change in mercury volume (dPV_{Hg}) indicates the change in pore volume (dPV). The change in pore volume remains fairly small in the second stage. The change in the injected cumulative mercury volume (PV_{Hg}) also remains small in the second stage. The ratio of the change in mercury volume to the cumulative volume ($\Delta PV_{Hg}/PV_{Hg}$)was less than 3% in our experiments; therefore, we supposed that the cumulative pore volume could be related to the mercury volume using a constant (C_1):

$$PV = C_1\, PV_{Hg}. \tag{7}$$

We substituted Eq. (7) in Eq. (6) and set the confining stress (σ_{Con}) equal to the mercury pressure (P_{Hg}) as follows:

$$\frac{1}{(C_1)PV_{Hg}}\frac{dPV_{Hg}}{dP_{Hg}} = \varnothing\left(C_{po}P_{Hg}^{m}\right). \tag{8}$$

We simplified the relation subsequently. Our objective in this study was to find a new approach to define entry pressure using the exponential trend; hence, the actual compressibility value obtained from the mercury injection was not required.

1.3. Simple relation for determining actual entry pressure

We proposed a simple relation determining the beginning of the third stage wherein the mercury invaded the pore space, and the corresponding capillary pressure was the entry pressure. We derived an analytical relation for the second stage, as defined by Eq. (8), which was further integrated and simplified to

$$\mathrm{Ln}(P_{Hg}) = \frac{\mathrm{Ln}(\mathrm{Ln}(PV_{Hg}))}{(m+1)} + \frac{C_2 - \mathrm{Ln}(a)}{(m+1)}, \tag{9}$$

where $C_2$and a are constants. The derived relation can be expressed in a simple mathematical format:

$$Y = A\,X + B, \tag{10}$$

where Y is $\mathrm{Ln}(P_{Hg})$, X is $\mathrm{Ln}(\mathrm{Ln}(PV_{Hg}))$, A is the slope, and B is the intercept.

This simple relation identifies the beginning of the third stage. The first and second derivatives (dY/dX and d^2Y/dX^2) of this relation can help to determine where the measured data deviate from the expected trend. The entry pressure corresponds to the pressure at which the first derivative reaches a maximum, and the second derivative changes the sign. In theory, analyzing the first or second derivative is sufficient, but the simultaneous investigation of both provides further confidence,

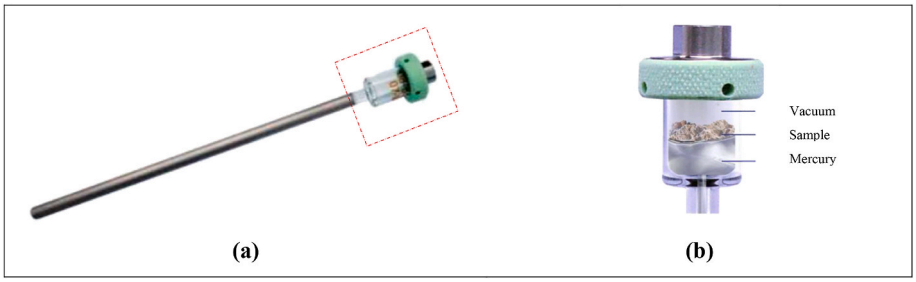

Fig. 1. (a) Empty penetrometer bulb and (b) schematic of the bulb, with crushed samples partially saturated by mercury.

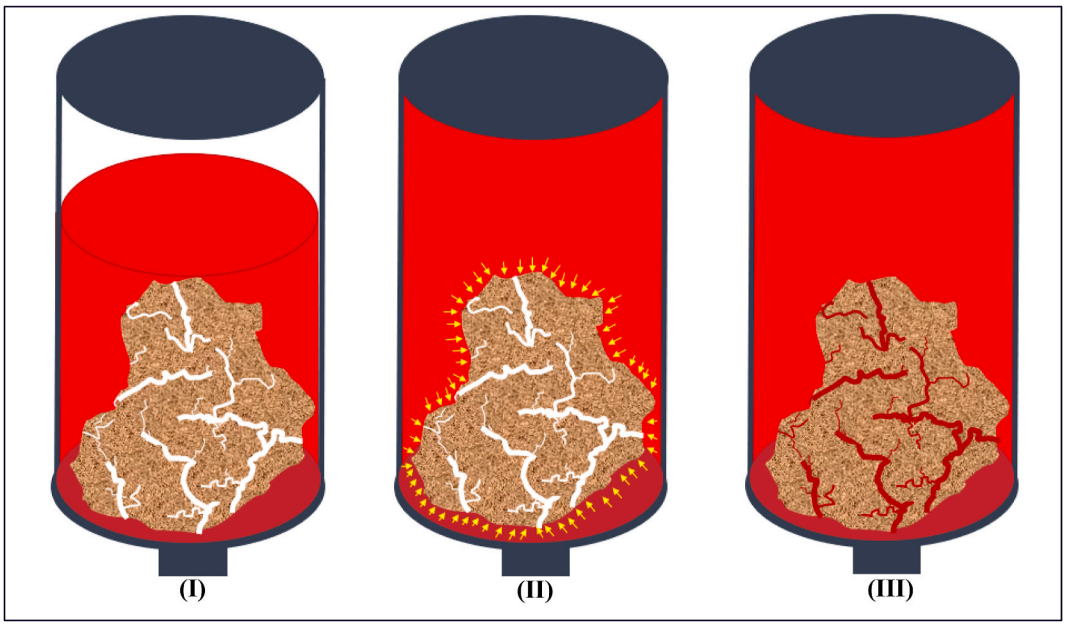

Fig. 2. Schematic illustration of mercury injection of an irregularly shaped sample: (I) filling the bulb without compressing the sample, (II) compressing the sample before intrusion, and (III) the actual intrusion.

especially when dealing with experimental data with noise.

1.4. Deviation from the simple relation

Noise in the actual measurements could be mistaken for a deviation from the simple relation (Eq. (10)). A possible source of anomalies, microfractures are openings between several nanometers and tens of micrometers in size formed by tensile stress (Slatt and O'Brien, 2011; Ougier-Simonin et al., 2016). This study used machine learning to remove the outliers and implemented filtering to discard such noise.

1.5. Machine learning algorithms for outlier detection

Two unsupervised machine learning algorithms were used to detect the outliers: k-nearest neighbors (KNN) and locally selective combination in parallel outlier ensembles (LSCP). Outlier-detection algorithms screen all data to identify which do not follow the common trend, but the absence of ground truth in unsupervised learning makes detection challenging. A relevant example is credit card fraud prevention in which the algorithm compares a behavior at a certain time with the purchasing history to determine suspicious activities (outliers). This study adopted such algorithms to determine which data points did not follow the common trend to remove them.

1.5.1. A. K-nearest neighbors (KNN)

The k-nearest neighbors algorithm (KNN) relies on distance. It is based on the notion that data are more similar where they are closer (Altman, 1992). An intuitive example of KNN is considering the nearest neighbor (1NN). For 1NN, we assign the value (or class) of a data point to its closest neighbor. This study used the Euclidean distance in KNN as follows (Deschrijver and Kerre, 2005):

$$d = \sqrt{(x_2 - x_1)^2 + (y_2 - y_1)^2}, \tag{11}$$

where d is the distance, (x_1, y_1) are the coordinates of the first point, and (x_2, y_2) are the coordinates of the second point.

The distance of KNN provides a criterion for detecting the outlier (Angiulli and Pizzuti, 2002; Ramaswamy et al., 2000; Zhao, 2018). The farthest distance of a local point to its nearest neighbors is defined as an anomaly. We examined five nearest neighbors (5NN) so that the algorithm calculated the distances of five neighbors to the local point, and the farthest neighbor was considered an outlier candidate. The average distance from all the neighbors' distances eliminated 5% of the population.

There is no established criterion for choosing the number of neighbors in KNN, but smaller numbers are not recommended for noisy data (Subramanian, 2019). KNN does not perform well where data are skewed (Zhao et al., 2019), but the LSCP framework enabled us to deal with this type of deficiency here.

1.5.2. B. locally selective combination in parallel outlier ensembles (LSCP)

LSCP is a machine learning algorithm. Similar to KNN, it is based on the distance of a data point to its closest neighbors. However, LSCP combines different detectors to improve its performance, so it is considered a parallel outlier ensemble, whereas KNN uses a single detector.

LSCP analyzes data locally to choose the best detectors, as local analysis performs better than global analysis when data are skewed (Van Stein et al., 2016). LSCP applies different models in parallel and uses the average to reduce the error (Dietterich, 2000).

First, LSCP divides the data into local regions and analyzes each region separately. It then selects the best detector(s), leading to the smallest score for each region (Zhao et al., 2019). The outlier score shows the distance of a data point from its closest neighbors. For this study, we included three detectors with 6, 12, and 18 neighbors to discard 5% of the outliers based on LSCP (Breunig et al., 2000).

1.6. Noise filtering

This study smoothed the raw data using the Savitzky–Golay (SG) filter (Savitzky and Golay, 1964). The filter has been used recently in the oil industry to preprocess data (Brenjkar and Delijani, 2022; Sabah et al., 2021). The filter uses a convolution theorem to fit subsets of adjacent data with a low-degree polynomial (Savitzky and Golay, 1964). This study used the Python toolkit (PyOD), an open-source code, to implement the data processing methods (Zhao, 2018).

2. Results

This study analyzed the capillary pressure measurements of Midra shale, which is found in Qatar. We collected outcrop samples and investigated their petrophysical properties, such as capillary pressure. Similar to other unconventional formations in the Middle East, Midra shale is less studied than most shales in the U.S. Midra shale is not extracted on an economic scale, and current investigations are at an early stage. It is a yellowish-brown and greenish-gray slate that contains fossils and changes in consistency, color, and composition throughout the region (Singh and Singh, 2011). Alessa et al. (2021) characterized the pore size distributions of Midra shale recently.

We analyzed seven shale samples, and their raw capillary pressure measurements are shown in Fig. 3. We collected 146 data points for each sample, and the capillary pressure was increased to 60,000 psi, which was the maximum pressure limitation of our machine. The raw measurements exhibited unrealistic invasion at low capillary pressures (<200 psi), corresponding to filling the empty cells and closing the microcracks. We applied the proposed relation to determine the true entry pressure. The results are discussed in detail for Samples 1 and 2, and the entry pressures of all samples are provided for completeness.

We analyzed the deviation from the simple relation. The first and second derivatives of the relation were also investigated. In theory, a deviation occurs where there is a maximum in the first derivative and the second derivative changes sign. In practice, however, choosing the corresponding capillary pressure was more complicated. Thus, we considered four scenarios to determine the actual entry pressure based on the simple relation (Table 1).

2.1. Scenario 1 (raw measurements)

The first scenario uses the raw data shown in Fig. 3. The simple relation and its first and second derivatives for Samples 1 and 2 are plotted in Fig. 4. It was difficult to determine where the first derivative reached a maximum and the second derivative changed sign corresponding to the entry pressure. The difficulty arises from the local fluctuations. We concluded that different values could be interpreted based on this scenario. Hence, we applied other scenarios to address this challenge, which are discussed subsequently.

2.2. Scenario 2 (outlier detection)

In the second scenario, we removed the outliers as potential sources of the local fluctuations observed in Fig. 4. We preprocessed the data to make outlier detection easier, as follows (Yeo and Johnson, 2000):

Table 1
Four scenarios were tested to apply the simple relation (Eq. (10)). The first scenario analyzed the raw measurements without outlier detection or noise filtering. The second scenario applied outlier detection, and the third integrated the relation with noise filtering. In the fourth scenario, outlier detection and noise filtering were incorporated.

Scenario	Outlier detection	Noise filtering	Figure
1	No	No	Fig. 4
2	Yes	No	Fig. 6
3	No	Yes	Fig. 7
4	Yes	Yes	Fig. 8

$$\psi(\lambda, y) = \begin{cases} \dfrac{(y+1)^{\lambda} - 1}{\lambda} & \text{if } \lambda \neq 0, y \geq 0 \\ \log(y+1) & \text{if } \lambda = 0, y \geq 0 \\ -\dfrac{(-y+1)^{2-\lambda} - 1}{2-\lambda} & \text{if } \lambda \neq 2, y < 0 \\ -\log(-y+1) & \text{if } \lambda = 2, y < 0 \end{cases}, \tag{12}$$

where ψ was the transformed data, y was the initial data, and λ was a power parameter. The optimum λ was obtained numerically, so the transformed data were the closest to the Gaussian distribution.

Next, we applied the KNN and LSCP algorithms to the transformed data. Fig. 5 shows the results. As discussed earlier, the KNN and LSCP algorithms analyze the overall trend of data points to determine which ones do not follow the overall behavior. The algorithms classify data into two groups: one group constitutes the main population, and the other group is the outlier. The orange area in Fig. 5 indicates the main population (95%). The outliers (5%) were removed to determine the entry pressures of the samples.

Fig. 6 exhibits the results based on the second scenario for Samples 1 and 2. The results from the first scenario are also plotted for comparison. Outlier removal enabled us to determine the entry pressure more easily in the second scenario than in the first scenario. Nevertheless, determining entry pressure was not trivial because no single value could be identified easily as a local maximum. The second derivative also showed high levels of fluctuations, especially in the second sample. The SG noise filter was applied to address this challenge in the third scenario, and the results are presented in Fig. 7.

2.3. Scenario 3 (noise filtering)

The third scenario used the SG filter to improve the performance. The filter smoothed the first derivative, and the second derivative was

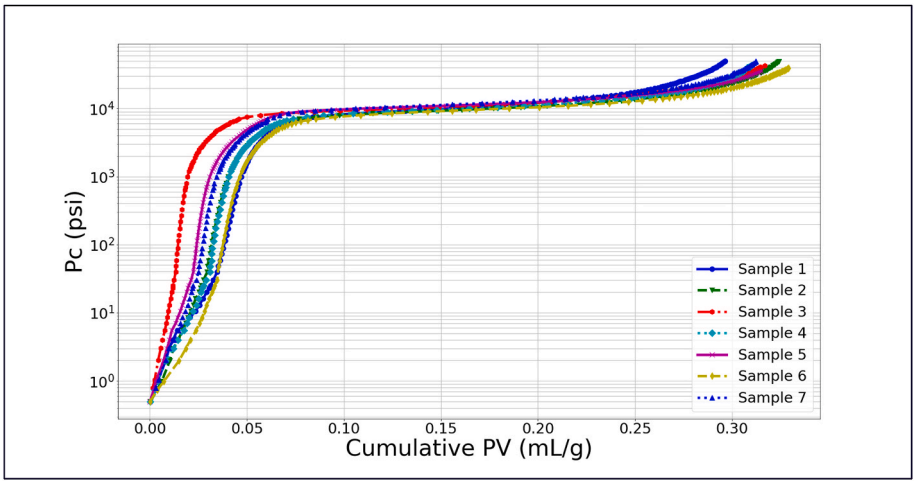

Fig. 3. Capillary pressure measurements of Midra shale samples. The raw measurements exhibited unrealistic invasion at low capillary pressures (<200 psi).

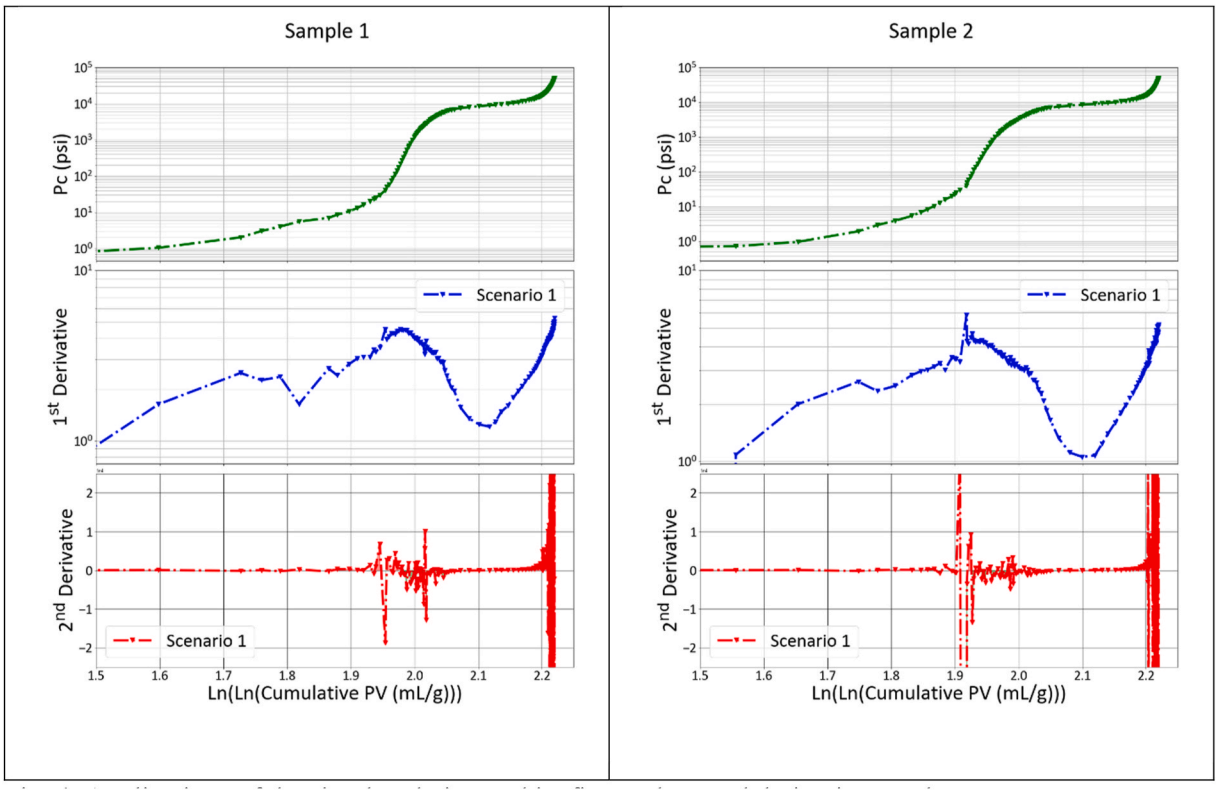

Fig. 4. Applications of the simple relation and its first and second derivatives to the raw measurements based on the first scenario for Samples 1 and 2. It was difficult to determine the entry pressure because of the local fluctuations. [1st derivative $= dY/dX$ and 2nd derivative $= d^2Y/dX^2$, where $Y = \mathrm{Ln}(P_{Hg})$ and $X = \mathrm{Ln}(\mathrm{Ln}(PV_{Hg}))$.].

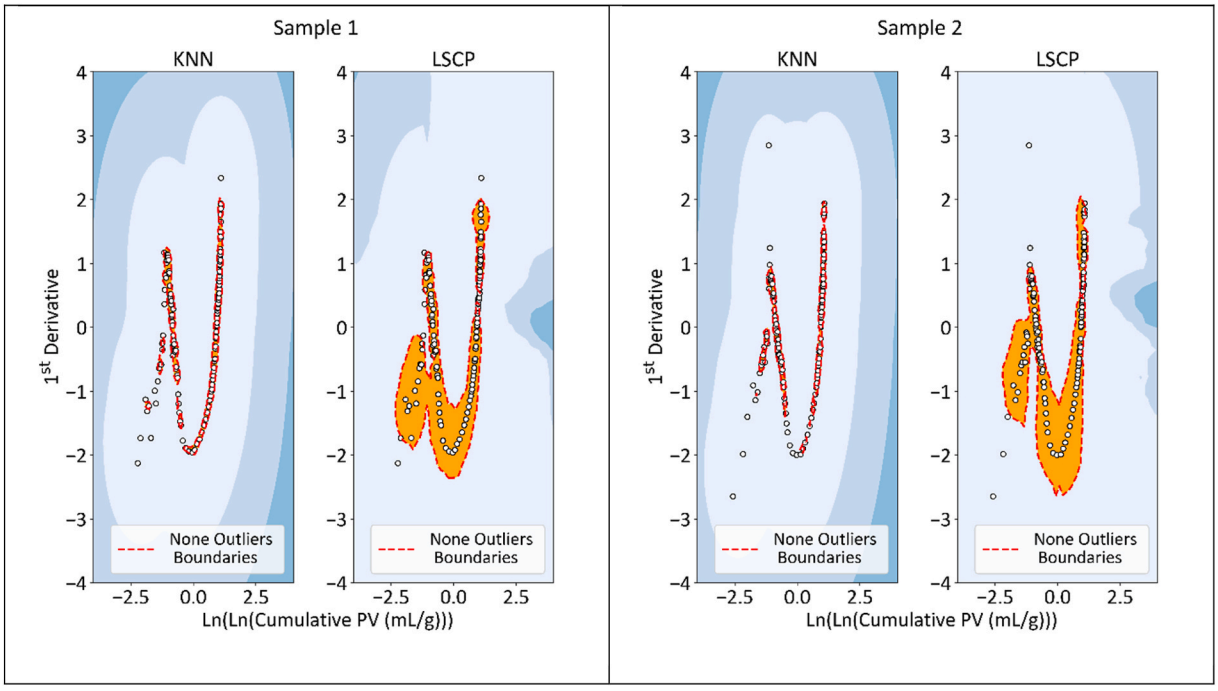

Fig. 5. The orange area shows 95% of the data kept for determining the entry pressures of Samples 1 and 2 in Figs. 6 and 8. The data points outside the orange area (5% of the population) were removed as outliers. [1st derivative $= dY/dX$, where $Y = \mathrm{Ln}(P_{Hg})$ and $X = \mathrm{Ln}(\mathrm{Ln}(PV_{Hg}))$.]

determined subsequently. There was no outlier detection in the third scenario, and all data were included (Table 1).

Fig. 7 shows the results based on the third scenario for Samples 1 and 2, with the results from the first scenario also shown for comparison. The filter was applied only to the blue region in Fig. 7, which included the entry pressure. The SG filter facilitated entry pressure determination by removing the noise (local fluctuations) from the first derivative. The fluctuations in the second derivative were also limited in the third

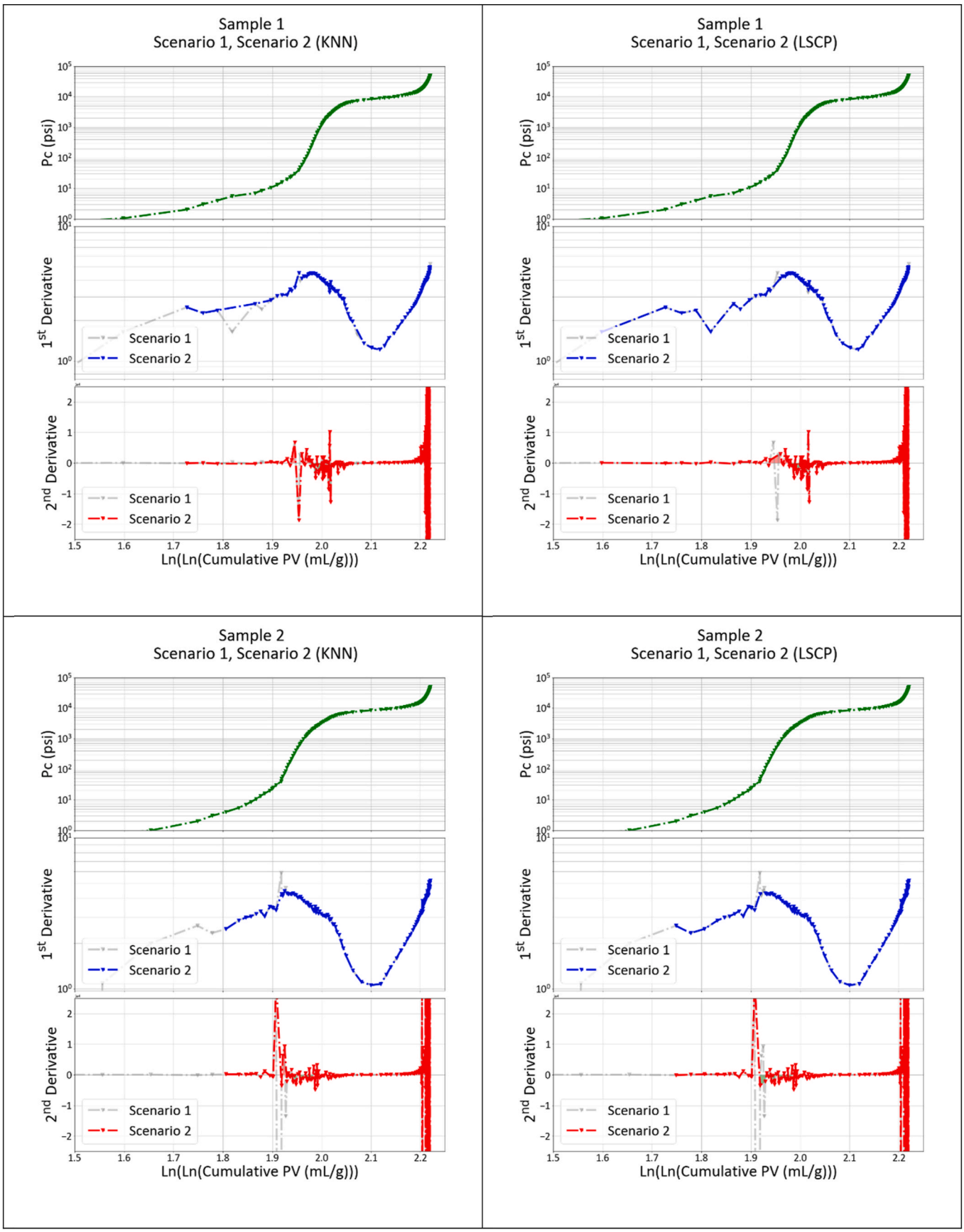

Fig. 6. Applications of the simple relation and its first and second derivatives based on the second scenario for Samples 1 and 2. The results from the first scenario are also plotted for comparison. [1st derivative $= dY/dX$ and 2nd derivative $= d^2Y/dX^2$, where $Y = \mathrm{Ln}(P_{Hg})$ and $X = \mathrm{Ln}(\mathrm{Ln}(PV_{Hg}))$.].

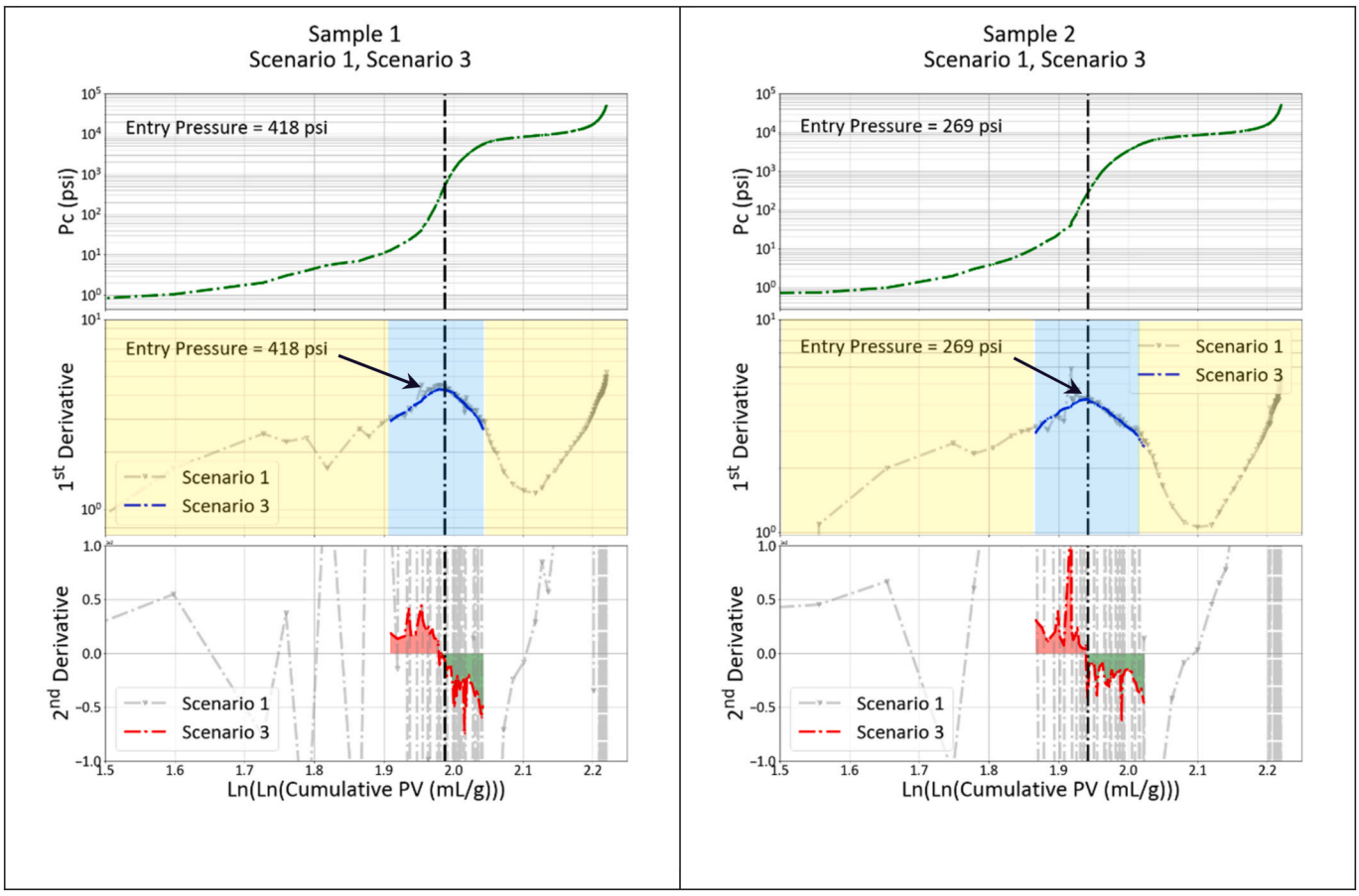

Fig. 7. Applications of the simple relation and its first and second derivatives based on the third scenario for Samples 1 and 2. The results from the first scenario are provided for comparison. [1st derivative $= dY/dX$and 2nd derivative $= d^2Y/dX^2$, where $Y = \mathrm{Ln}(P_{Hg})$ and $X = \mathrm{Ln}(\mathrm{Ln}(PV_{Hg}))$.].

scenario. One of the main features of the third scenario was that it accounted for all the data without discarding potential outliers, but it slightly altered the inputs to deal with the noise.

2.4. Scenario 4 (outlier detection and noise filtering)

The fourth scenario removed the outliers and filtered the noise to determine the entry pressure (Table 1). The outliers were discarded (5% of the population) using KNN and LSCP. The first derivative was then determined for the remainder (95% of the population). Later, the noise was filtered from the first derivative by the SG algorithm. Finally, the second derivative was calculated from the smoothed data.

Fig. 8 shows the results based on the fourth scenario, with the second and third scenarios' results also plotted for comparison. The first derivative had a smooth profile in the fourth scenario, which enabled us to easily determine its maximum value that corresponded to the entry pressure. The effects of outlier detection and the noise filter on the first derivative became clearer when we compared it with the results from the other scenarios illustrated in Fig. 8.

The second derivative sign also changed where the first derivative reached a maximum in the fourth scenario. The evident change in the behavior of the second derivative implied a different phenomenon taking place at the corresponding pressure. We interpreted this phenomenon as mercury entering the void space in the matrix.

Removing the outliers and smoothing the data helped us to easily identify the entry pressure. There was less ambiguity in the identified entry pressure based on the fourth scenario after applying the filter (SG) and outlier detection (LSCP or KNN). Comparing Scenario 4 with the other scenarios depicted in Fig. 8, we inferred that the fourth scenario was the best approach for determining the entry pressure based on the proposed relation.

Table 2 summarizes the entry pressures of the samples based on different scenarios. They are all presented for completeness, while the results from Samples 1 and 2 are discussed in detail. It was easier to identify the entry pressure for Sample 1 than for Sample 2. All scenarios led to the entry pressure of 418 psi for Sample 1, whereas the entry pressure of Sample 2 depended on the noise filter and outlier detector.

2.5. Validation

We investigated high-resolution images of the analyzed shale to test the accuracy of the proposed relation in predicting entry pressure. Table 2 lists the estimated entry pressures that are close to 300 psi. Fig. 9 shows the scanning electron microscope (SEM) images of the Midra shale. Two textures can be observed: fibers and rod-like structures, which contribute to primary and secondary porosity. These images illustrate pores with a characteristic size close to 1 μm.

The corresponding entry pressure of the shale with 1-μm conduits can be estimated with the Young–Laplace (YL) relation. The entry pressure for a single conduit is expressed as follows:

$$P_c = \frac{2\gamma\cos(\theta)}{r}, \tag{13}$$

where P_c is the capillary pressure, γ is the interfacial tension, θ is the contact angle, and r is the conduit radius. The YL relation indicated that the corresponding entry pressure was 181 psi in the mercury injection capillary pressure experiment where the contact angle was 130° and the interfacial tension was 485 dyne/cm.

The validation of the estimated entry pressure based on SEM images

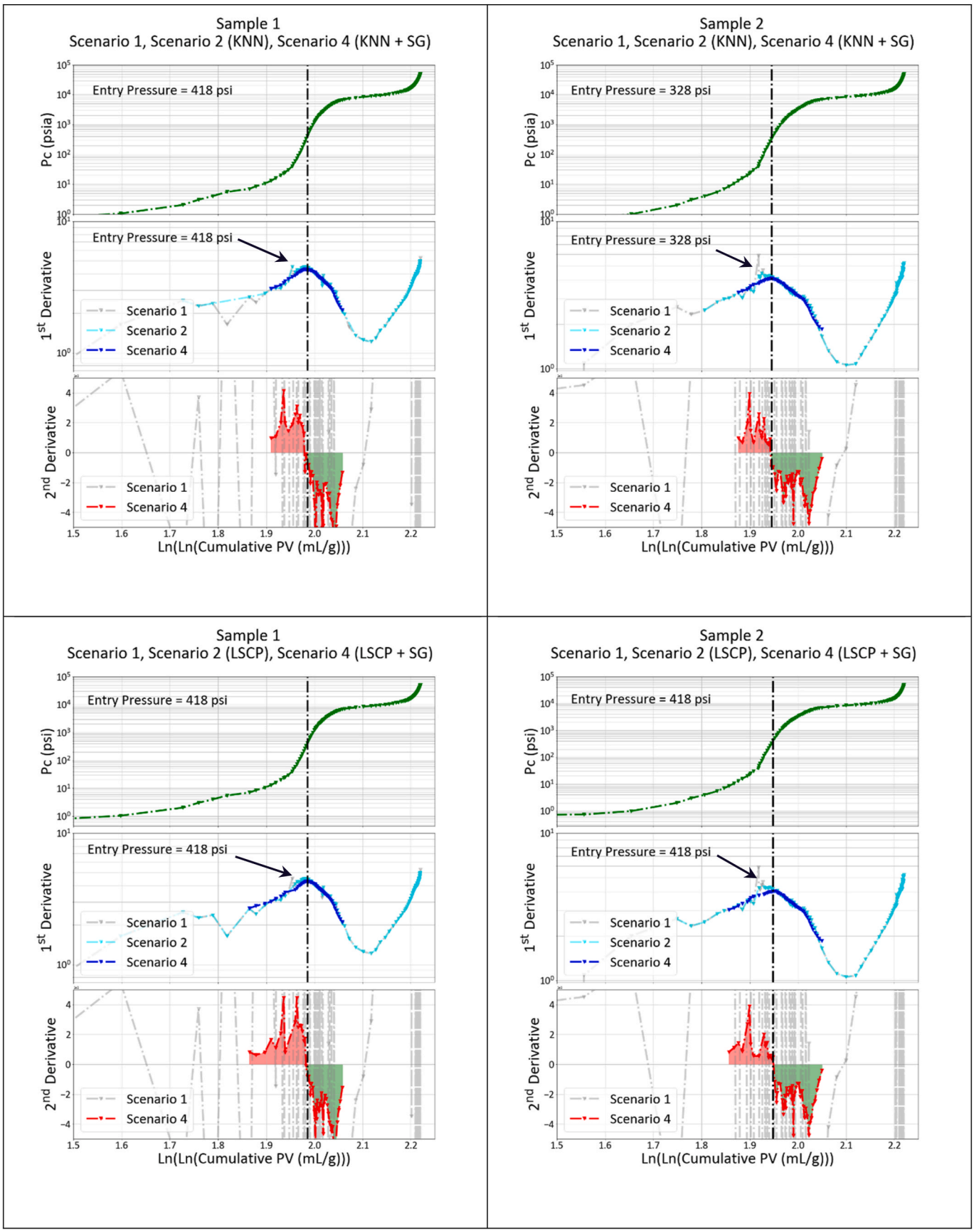

Fig. 8. Applications of the simple relation and its first and second derivatives based on the fourth scenario for Samples 1 and 2. The results from the first and second scenarios are also presented for comparison. [1st derivative $= dY/dX$ and 2nd derivative $= d^2Y/dX^2$, where $Y = \mathrm{Ln}(P_{Hg})$ and $X = \mathrm{Ln}(\mathrm{Ln}(PV_{Hg}))$.].

Table 2
The entry pressures of the shale samples, with the raw measurements shown in Fig. 3 using the proposed relation (Eq. (10)). It was challenging to determine the entry pressure based on Scenarios 1 and 2 for some samples, whereas Scenarios 3 and 4 yielded easily identifiable values for all samples.

Sample	Scenario 1 (Raw data)	Scenario 2 (KNN[a])	Scenario 2 (LSCP[b])	Scenario 3 (SG[c])	Scenario 4 (KNN, SG)	Scenario 4 (LSCP, SG)
1	418 psi	418 psi	418 psi	418 psi	418 psi	418 psi
2	Erratic	Erratic	Erratic	269 psi	328 psi	418 psi
3	Erratic	Erratic	Erratic	268 psi	268 psi	328 psi
4	Erratic	Erratic	Erratic	269 psi	328 psi	328 psi
5	Erratic	Erratic	Erratic	268 psi	268 psi	328 psi
6	328 psi	328 psi	328 psi	328 psi	328 psi	328 psi
7	Erratic	Erratic	Erratic	329 psi	329 psi	329 psi

[a] KNN: k-nearest neighbors.
[b] LSCP: locally selective combination in parallel outlier ensembles.
[c] SG: Savitzky–Golay (SG) filter.

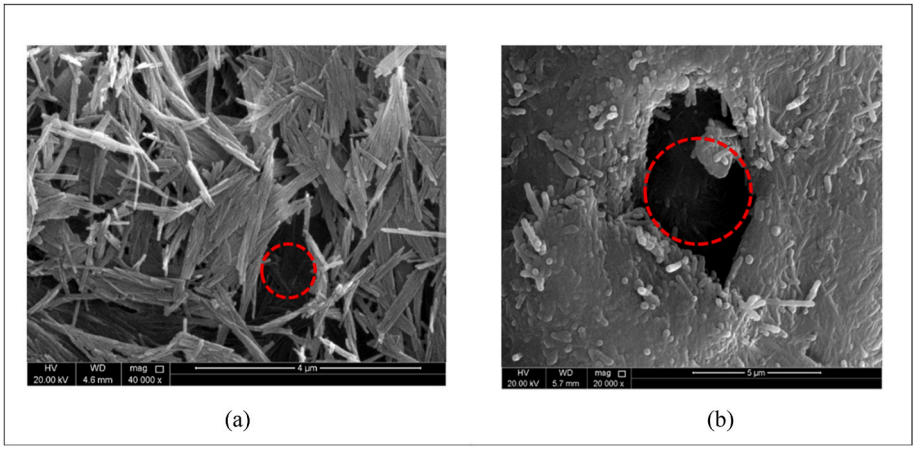

Fig. 9. SEM images of palygorskite Midra shale samples: (a) fibers and microbial-mediated palygorskite rods and (b) red circles show ~ micrometer pores, supporting the entry pressures obtained from the simple relation in this study.

provided independent verification to test the proposed relation. The validation was performed based on data from different scales because the mercury injection experiment used samples as small as a few millimeters (10^{-3} m), whereas the sizes of the images were in micrometers (10^{-6} m). Another difference was relevant to the dimension. The drainage experiments characterized pore space by probing its connectivity in a three-dimensional space, whereas the images probed a two-dimensional domain.

The estimated entry pressure based on SEM was close to the values obtained from the proposed relation, and this provided independent evidence for validation. The SEM images were obtained without confining stress, whereas the sample was compressed in MICP. This meant that the pores were wider in the images, which led to an entry pressure lower than the value obtained from the proposed relation.

3. Discussion

Our analysis showed that the SG filter preserved the main features of the data. The SG filter was biased toward noise in low-density distributions and integrated with LSCP and KNN to eliminate 5% of anomalies due to this limitation. We did not eliminate further outliers to avoid a data covariant shift. Notably, LSCP performed better than KNN in filtering noise.

Our study also showed that identifying the filter window size remains challenging. We used a relatively large window size, 13–17% of the total data, to approximate the global features. The window size's influential factor was data density at the bulb-filling and pore-compression stages (1 and 2 in Fig. 2). It was not recommended to use a large window size with our data because the profile would be biased to the high-density data points located at the later third stage of the MICP. Thus, we decreased the window size in the presence of low data density.

The window size depends on the data size, covariate shift, and quantity of noise, and it can be different for samples from the same formation. Nevertheless, we mostly used similar window sizes because the noise level was low, and the data were similar. The number of neighbors used for outlier detection also played an important role. Therefore, we were able to conclude that there is no optimal number of neighbors applicable to all samples, so the number should be determined individually.

4. Conclusions

We proposed a simple relation to correct mercury injection capillary pressure measurements for irregularly shaped samples, such as cuttings or chips. The relation detected the true entry pressure and was applied to the actual measurements of Midra shale. Four scenarios were designed to deal with the noise.

The first scenario applied the simple relation to the raw measurements, whereas the other scenarios implemented outlier detection and noise filtering techniques in machine learning. Our study has shown that it is challenging to determine actual entry pressure based on the first scenario in practice. The difficulty, which is not apparent in theory, is due to the fluctuations in the first and second derivatives.

The second scenario applied unsupervised learning outlier detection methods to eliminate spikes in the derivatives. LSCP was advantageous over KNN in filtering noise, but the second scenario was effective only where the noise level was low. The third scenario applied the SG noise filter to smooth the new relation's derivatives. However, the third scenario was not capable of determining the actual entry pressure where data were noisy. Therefore, our interpretation is that the fourth scenario,

which combines unsupervised learning outlier detection with the SG noise filter, is optimal.

The simple relation predicted that the entry pressures of the shale samples would be 328–418 psi based on different scenarios. The difference between the entry pressures of the samples was relevant to data heterogeneity. The predicted values were validated against high-resolution images of the studied shale. The proposed relation has applications in characterizing the capillary pressure measurements of samples with irregular shapes. The accurate characterization of capillary pressure is critical in modeling multiphase flow in shale formations.

The conducted study presented a systematic approach for correcting capillary pressure measurements, which was missing in the literature, by quantifying true entry pressure. It chose the entry pressure corresponding to the maximum of the first derivative of the simple relation where the second derivative turned negative. The novelty of the conducted study is that it provides a convenient, systematic tool to quantify true entry pressure and correct raw measurements.

Credit author statement

SA: Conceptualization, Methodology, Software, Validation, Formal analysis, Investigation, Writing – original draft. **ASP**: Conceptualization, Methodology, Validation, Investigation, Writing – original draft, Writing – review & editing, Supervision, Project administration, Funding acquisition. **FS**: Resources, Project administration, Funding acquisition. **HA**: Resources, Project administration, Funding acquisition.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

Data will be made available on request.

Acknowledgments

We are grateful for the constructive comments of an anonymous reviewer, a senior expert in the field, and the editor who helped us improve this work. We would like to acknowledge the support of the Qatar National Research Fund (a member of the Qatar Foundation) through Grant #NPRP12S-0305–190235. The findings achieved herein are solely the responsibility of the authors.

References

Ahmed, T., 2018. Reservoir Engineering Handbook. Gulf professional publishing.
Alessa, S., Sakhaee-Pour, A., Sadooni, F.N., Al-Kuwari, H.A., 2021. Comprehensive pore size characterization of Midra shale. J. Petrol. Sci. Eng. 203, 108576.
Altman, N.S., 1992. An introduction to kernel and nearest-neighbor nonparametric regression. Am. Statistician 46 (3), 175–185.
Anderson, M.A., Jones, F.O., 1985, June. A comparison of hydrostatic-stress and uniaxial-strain pore-volume compressibilities using nonlinear elastic theory. In: The 26th US Symposium on Rock Mechanics (USRMS). OnePetro.
Angiulli, F., Pizzuti, C., 2002, August. Fast outlier detection in high dimensional spaces. In: European Conference on Principles of Data Mining and Knowledge Discovery. Springer, Berlin, Heidelberg, pp. 15–27.
Bailey, S., 2009, October. Closure and compressibility corrections to capillary pressure data in shales. In: DWLS Workshop Presented, vol. 19.
Baker, R.O., Yarranton, H.W., Jensen, J., 2015. Practical Reservoir Engineering and Characterization. Gulf Professional Publishing.
Brenjkar, E., Delijani, E.B., 2022. Computational prediction of the drilling rate of penetration (ROP): a comparison of various machine learning approaches and traditional models. J. Petrol. Sci. Eng. 210, 110033.
Breunig, M.M., Kriegel, H.P., Ng, R.T., Sander, J., 2000, May. LOF: identifying density-based local outliers. In: Proceedings of the 2000 ACM SIGMOD International Conference on Management of Data, pp. 93–104.
Comisky, J.T., Santiago, M., McCollom, B., Buddhala, A., Newsham, K.E., 2011, November. Sample size effects on the application of mercury injection capillary pressure for determining the storage capacity of tight gas and oil shales. In: Canadian Unconventional Resources Conference. OnePetro.
Davudov, D., Moghanloo, R.G., Lan, Y., 2018. Evaluation of accessible porosity using mercury injection capillary pressure data in shale samples. Energy Fuel. 32 (4), 4682–4694.
Deinert, M.R., Parlange, J.Y., 2009. Effect of pore structure on capillary condensation in a porous medium. Phys. Rev. 79 (2), 021202.
Deschrijver, G., Kerre, E.E., 2005. Triangular norms and related operators in L*-fuzzy set theory. In: Logical, Algebraic, Analytic and Probabilistic Aspects of Triangular Norms. Elsevier Science BV, pp. 231–259.
Dietterich, T.G., 2000, June. Ensemble methods in machine learning. In: International Workshop on Multiple Classifier Systems. Springer, Berlin, Heidelberg, pp. 1–15.
Eia, U., 2021. US Crude Oil and Natural Gas Proved Reserves, Year-End 2019. US Energy Information Administration.
Gaswirth, S.B., French, K.L., Pitman, J.K., Marra, K.R., Mercier, T.J., Leathers-Miller, H. M., Schenk, C.J., Tennyson, M.E., Woodall, C.A., Brownfield, M.E., Finn, T.M., 2018. Assessment of Undiscovered Continuous Oil and Gas Resources in the Wolfcamp Shale and Bone Spring Formation of the Delaware Basin, Permian Basin Province, New Mexico and Texas, 2018 (No. 2018-3073). US Geological Survey.
Hawkins, M.F., 1955. Material balances in expansion type reservoirs above bubble point. Transactions of the AIME 204 (1), 267–271.
Hendraningrat, Luky, Li, Shidong, Torsæter, Ole, 2013. A coreflood investigation of nanofluid enhanced oil recovery. J. Petrol. Sci. Eng. 111, 128–138.
Jizba, D.L., 1991. Mechanical and Acoustical Properties of Sandstones and Shales. Doctoral dissertation. Stanford University. Retrieved from. http://search.proquest.com.ezproxy.lib.uh.edu/dissertations-theses/mechanical-acoustical-properties-sandstones/docview/303959318/se-2?accountid=7107.
Lan, Y., Moghanloo, R.G., Davudov, D., 2017. Pore compressibility of shale formations. SPE J. 22 (6), 1778–1789.
Lee, H., Shakib, F.A., Liu, K., Liu, B., Bubach, B., Varma, R.S., Jang, H.W., Shokouhimher, M., Ostadhassan, M., 2020. Adsorption based realistic molecular model of amorphous kerogen. RSC Adv. 10 (39), 23312–23320.
McLatchie, A.S., Hemstock, R.A., Young, J.W., 1958. The effective compressibility of reservoir rock and its effects on permeability. J. Petrol. Technol. 10 (6), 49–51.
Niandou, H., Shao, J.F., Henry, J.P., Fourmaintraux, D., 1997. Laboratory investigation of the mechanical behaviour of Tournemire shale. Int. J. Rock Mech. Min. Sci. 34 (1), 3–16.
Ougier-Simonin, A., Renard, F., Boehm, C., Vidal-Gilbert, S., 2016. Microfracturing and microporosity in shales. Earth Sci. Rev. 162, 198–226.
Peng, S., Zhang, T., Loucks, R.G., Shultz, J., 2017. Application of mercury injection capillary pressure to mudrocks: conformance and compression corrections. Mar. Petrol. Geol. 88, 30–40.
Purcell, W.R., 1949. Capillary pressures-their measurement using mercury and the calculation of permeability therefrom. J. Petrol. Technol. 1 (2), 39–48.
Ramaswamy, S., Rastogi, R., Shim, K., 2000, May. Efficient algorithms for mining outliers from large data sets. In: Proceedings of the 2000 ACM SIGMOD International Conference on Management of Data, pp. 427–438.
Sabah, M., Mehrad, M., Ashrafi, S.B., Wood, D.A., Fathi, S., 2021. Hybrid machine learning algorithms to enhance lost-circulation prediction and management in the Marun oil field. J. Petrol. Sci. Eng. 198, 108125.
Savitzky, A., Golay, M.J., 1964. Smoothing and differentiation of data by simplified least squares procedures. Anal. Chem. 36 (8), 1627–1639.
Singh, S.K., Singh, J.K., 2011. Effect of pore morphology on vapor–liquid phase transition and crossover behavior of critical properties from 3D to 2D. Fluid Phase Equil. 300 (1–2), 182–187.
Slatt, R.M., O'Brien, N.R., 2011. Pore types in the Barnett and Woodford gas shales: contribution to understanding gas storage and migration pathways in fine-grained rocks. AAPG Bull. 95 (12), 2017–2030.
Subramanian, D., 2019. A Simple Introduction to K-Nearest Neighbors Algorithm. Towards Data Science. https://towardsdatascience.com/a-simple-introduction-to-k-nearest-neighbors-algorithm-b3519ed98e. (Accessed 15 November 2021).
Swanson, B.F., 1981. A simple correlation between permeabilities and mercury capillary pressures. J. Petrol. Technol. 33 (12), 2498–2504.
Tran, H., Sakhaee-Pour, A., Bryant, S.L., 2018. A simple relation for estimating shale permeability. Transport Porous Media 124 (3), 883–901.
Van Stein, B., Van Leeuwen, M., Bäck, T., 2016, December. Local subspace-based outlier detection using global neighbourhoods. In: 2016 IEEE International Conference on Big Data (Big Data). IEEE, pp. 1136–1142.
Yeo, I.K., Johnson, R.A., 2000. A new family of power transformations to improve normality or symmetry. Biometrika 87 (4), 954–959.
Zhao, Y., 2018. Pyod Documentation. https://media.readthedocs.org/pdf/pyod/latest/pyod.pdf. (Accessed 15 November 2021).
Zhao, Y., Nasrullah, Z., Hryniewicki, M.K., Li, Z., 2019, May. LSCP: locally selective combination in parallel outlier ensembles. In: Proceedings of the 2019 SIAM International Conference on Data Mining. Society for Industrial and Applied Mathematics, pp. 585–593.
Zimmerman, R.W., 1990. Compressibility of Sandstones. Elsevier, Amsterdam.