ABSTRACT
Fracture modeling plays a valuable role to understand the fluid flow in carbonate reservoirs. For this, the fracture characterization to generate Discrete Fracture Networks (DFNs) can take advantage of analogue outcrops through Virtual Outcrop Models (VOMs), acquired by Unmanned Aerial Vehicles (UAV) and digital photogrammetry. The stochastic DFN generation is an important step in reservoir modeling as it brings more representative data to the process and has long been studied. However, optimizations concerning automatizing some of the steps necessary to its generation like data clustering are still open to advancements. In this sense, this work aims the fracture data clustering and the definition of the number of clusters when gathering data for the stochastic process, developing an Elbow method for spherical data and a balanced K-means, both based on Fisher statistics. For this, we interpreted fracture planes in a VOM that recreates a carbonate reservoir analogue from the Jandaíra Formation, in the Northeast, Brazil. As result, we show a workflow for immersive fracture interpretation alongside a 3D stochastic DFN model with fracture intensity of $22.57 \text{m}^{-1}$ for cell sizes of $1\text{m}^3$. Regarding the clustering balance, our method achieved a lower standard deviation between sets while maintaining the Fisher values greater to obtain fracture sets with lower dispersion. Additionally, the Elbow method implementation proved a beneficial step to the workflow as it reduced the interpretation bias of family clusters. These results alongside the proposed workflow bring a better understanding of the outcrop geometry while offering data scalability for reservoir modeling.

INDEX TERMS
3D stochastic DFN modeling, Elbow method, Fisher statistics distribution, K-means optimization, Spherical data clustering

I. INTRODUCTION
Carbonate reservoirs (e.g., Pre-salt of the South Atlantic Margin offshore Brazil) host important hydrocarbon reserves. The carbonates present intrinsic characteristics of porosity and permeability that regulate fluid storage and fluid flow in reservoirs [1]–[3]. These characteristics are impacted by diagenetic processes, dissolution, and tectonic events that affect the reservoirs, generating fractures and discontinuities [4]. Thus, the fracture attributes contribute to the estimation of the geomechanical behavior of carbonate rocks necessary for hydrocarbon extraction.

Fractures can be defined as a result of a rupture or crack in the rock body that generates a physical discontinuity [5]. They often appear as a connected network of fractures, usually represented in numeric terms as Discrete Fractures Networks (DFNs) [6].

Identifying this kind of features is also an important task when analyzing impact flow on geothermal reservoirs [7], hydrocarbon-bearing formations [8], groundwater aquifers [9], leakage of $\text{CO}_2$ storage [10] and nuclear waste [11].
broadening the knowledge over these formations allowing proper exploration and preemptive actions.

Specifically, deep reservoirs, like the Brazil’s offshore pre-salt reservoirs (6km below sea level), suffer from the limitations in spatial resolution of the seismic assessment and with the sparsity of well log data [12]. These limitations in reservoir assessment lead us to the study of analogue outcrops, which are exposed rock formations with characteristics of composition and geometry similar to the reservoirs studied [13].

The fracture network investigation in analogue outcrops is carried out by direct measurements like the scanline method, where information such as length, strike and dip, and aperture are acquired [14], [15], usually by an analyst in situ. Due to access restrictions and limited spatial coverage, remote sensing techniques, to obtain the Virtual Outcrop models (VOMs), are widely used to obtain a larger set of geometric information about fractures on the outcrop. In this case, the interpretation of fractures can be made directly on the computer screen.

Either in situ or using VOMs, the fracture network information also integrate roughness, connectivity, and spacing, which are is modeled in deterministic and stochastic DFN models [6]. In the deterministic DFN model, attributes such as density and intensity are generally estimated by the $P_{xy}$ system [16]. The fracture intensity is obtained by the number of fractures that cross a scan line ($P_{10}$), by the sum of the fracture lengths, divided by the area measure of the sample region ($P_{21}$), and by the sum of the areas of fractures contained in a volume ($P_{32}$) expressed in meters ($m^{-3}$). A linear correlation can be estimated observing the indexes $P_{10}$, $P_{21}$, and $P_{32}$, allowing the estimation of the volumetric $P_{32}$ from 1D and 2D data [12], [17]. Such correlations are supported by a constant $C$, based on the parameters of fracture orientation, radius size distribution and orientation of the sampling window or scan line [18].

To estimate the $P_{32}$ index, the dimension of the sample to be measured must be three-dimensional, in which orientation measurements (strike/dip) are performed. This information can be incorporated into the stochastic DFN modeling and scaled to the reservoir models. In these models, the quality of the simulation is directly related to the collection of information from the fracture network [17]. The orientation parameters, needed to create the DFN models, must consider their mean values and dispersion values obtained with Fisher distribution statistics [19], [20], and the quantity of sets decision will influence how the fracture are clustered and how the Fisher statistics will be propagated to the stochastic model.

Works that have generated 3D stochastic DFN models from strike/dip data relied on visual interpretation of fracture sets in diagrams and statistics generated by 2D software like the FracPaQ [21], [22] and FRACMAN [23] to identify the best quantity of sets. With some works [24], [25] reducing the quantity of sets arbitrarily applying metrics to identify and agglutinate co-planar sets, also using the Fisher concentration to highlight the possible number of clusters in the stereonet. Additionally, some works have proposed modifications to the clusterization method based on the K-means algorithm, employing Particle Swarm Optimization and Fuzzy clustering [25], [26], also optimizing the cluster initial positions. However, the quantity of cluster is still set arbitrarily.

In this regard, we propose two optimizations to the 3D stochastic DFN generation workflow by: adapting the Elbow method to the definition of fracture sets using the spherical K-means and Fisher statistics; and making a balance check to the clusters obtained from the K-means algorithm to verify if the process should be restarted with new center positions or not. These strategies aims to obtain better input data to the Fisher statistics definitions necessary to the stochastic generation modeling.

For this, we use the software Mosis XP [27], [28], to interpret fracture planes in a VOM obtained of an outcrop analogue located in the Jandaira Formation, in Brazilian Northeast. From this data, our proposed workflow is applied and validated generating a 3D stochastic DFN and estimating the volumetric fracture intensity for the model cells. This region is well studied in geological terms [29], [30], however no study regarding outcrop geometry and generation of 3D stochastic was carried. This work is presented as follows: the proposed method, statistics and computations to extract 3D data are presented in Section 2; the interpretation scenario and experimental setting are presented in Section 3; followed by the results and comparisons in Section 4; accompanied by the discussion and finding conclusion.

II. 3D STOCHASTIC DFN MODELING

The proposed workflow to determine the volumetric intensity integrates the 3D stochastic DFN generation from VOM interpretation of fracture planes, as illustrated in Figure 1.

The workflow starts with the fracture plane points acquisition from VOM. The rest of the workflow describes the data clustering and statistics for each possible set, and finishes with the 3D fracture generation and the volumetric 3D DFN. This workflow is better described in the following sections.

A. PROCESSING FRACURE PLANE DATA

The interpreted plane data extracted from the VOM is composed of 3D point coordinates $(x,y,z)$, in which the positional and directional information will be computed. Directional and positional information from points placed in 3D space necessary for fracture characterization can be obtained by basic geometry operations. As the first plane information, the centroid is calculated to obtain the central position of a group of points, given by

$$
\text{centroid} = \frac{\sum_{i=1}^{N} x_i}{N}, \frac{\sum_{i=1}^{N} y_i}{N}, \frac{\sum_{i=1}^{N} z_i}{N}
$$

where $N$ is the quantity of points that compose the plane and $x_i \in \{x_0, x_1, ..., x_N\}$, $y_i \in \{y_0, y_1, ..., y_N\}$, and $z_i \in \{z_0, z_1, ..., z_N\}$.
\{z_0, z_1, ... z_N\} are the points coordinates.

The next plane information is the plane equation and its components that are given by

\[Ax + By + Cz + D = 0\]  \hspace{1cm} (2)

where A, B, C, and D are the equation terms (with D representing a positional term) [25]. These terms are computed with the principal component analysis (PCA) technique [31] that fits the point data to its eigenvalues and eigenvectors, where the third eigenvector gives the plane normal vector \(N(A, B, C)\) usually normalized in the form \(n(a, b, c)\). The positional term \(D\) in the plane equation is given by

\[D = \text{centroid} \cdot n(a, b, c).\]  \hspace{1cm} (3)

After obtaining the plane equation terms, the strike and dip angle are calculated in terms of \(\alpha\) and \(\beta\) angles given by

\[\alpha = \arccos\left(\frac{c}{\sqrt{a^2 + b^2 + c^2}}\right)\]  \hspace{1cm} (4)

and

\[\beta = q \pm \arccos\left(\frac{a}{\sqrt{a^2 + b^2}}\right)\]  \hspace{1cm} (5)

where \(q \in 0^\circ, 180^\circ\) or \(360^\circ\) according to the quadrant that the resultant vector of \(a\) and \(b\) points [32].

These equations give us the strike and dip in radians, that are later converted to degrees for visualization in the stereonet and other statistics. Depending on the quadrant, we must correct the strike direction considering the \(a\) and \(b\) values, where, if \(a\) and \(b\) are greater than zero the strike is \(360 - \text{deg}(\beta)\) and if \(a\) is greater than zero and \(b\) is lesser than zero the strike is \(180 + \text{deg}(\beta)\). Also, considering the "right-hand rule" [33] we consider opposite strike directions as the same angle, computing directions for the first and the second quadrant only.

\textbf{B. FISHER STATISTICS}

The strike and dip data obtained from the 3D interpretations are the main data for the stochastic plane generation. However, this data is reduced to spherical statistics information computed with Von Mises and Fisher [34] statistics for each fracture set. The Fisher statistics [35] are spherical statistics that indicate the dispersion and distribution of point positions in a sphere. For the Fisher statistics, the mean strike and dip for a set are observed in terms of the normal vector of each plane represented in polar coordinates \(\phi\) and \(\theta\). Where \(\phi\) (latitude) is given by \(90 - \text{dip}\), and \(\theta\) (longitude) is given by the angle rotated from the prime meridian using the bearing angle.

The mean direction of Fisher is given by the sum of the direction cosines [35]

\[\overline{l} = \sum_{i=1}^{N} l_i\quad \overline{m} = \sum_{i=1}^{N} m_i\quad \overline{n} = \sum_{i=1}^{N} n_i\]  \hspace{1cm} (6)
where \( l_i = \cos(\phi_i)\cos(\theta_i) \), \( m_i = \cos(\phi_i)\sin(\theta_i) \), \( n_i = \sin(\phi_i) \), and \( i \) is an individual plane of the \( N \) observations.

After computing the mean direction cosines, they are converted in unit vectors dividing by the length of the dip mean vector \( R \) given by

\[
R = \sqrt{(\bar{l})^2 + (\bar{m})^2 + (\bar{n})^2}
\]

and the unit vectors are given by

\[
\hat{l} = \frac{\bar{l}}{R}, \quad \hat{m} = \frac{\bar{m}}{R}, \quad \hat{n} = \frac{\bar{n}}{R}.
\]

The mean directional unit vectors are converted to mean strike and dip are then given by

\[
\delta = \arcsin(\hat{n})
\]

for dip angle, and

\[
\text{strike} = \begin{cases} 
\arccos\left(\frac{l}{\cos(\delta)}\right) & \text{if } \hat{m} \geq 0 \\
360 - \arccos\left(\frac{l}{\cos(\delta)}\right) & \text{otherwise}
\end{cases}.
\]

The index of dispersion of the Fisher distribution \( k \) is given by

\[
k = (N - 1)/(N - R)
\]

where \( k \) values close to 0 indicate not coplanar planes and positive values to \( \infty \) indicate strongly coplanar planes. In terms of distribution, the Fisher statistic [36] is given by

\[
f(\phi, \theta) = \frac{k}{4\pi \sinh(k)} \exp^{k[\cos(\theta_0)\cos(\theta) + \sin(\theta_0)\cos(\phi - \phi_0)]}.
\]

\[
C. SPHERICAL DATA CLUSTERING
\]

The basic K-means algorithm clusters objects in sets of \( K \) defined clusters, with random centroid clusters initially set in the first iteration. The objects are then associated with the closest cluster center given a distance metric. After each iteration, the centroid position of the cluster is recalculated. The algorithm converges when no significant change in centroid positions is observed.

To obtain the optimal number of clusters for the K-means clustering the most common algorithms used are the Elbow method and the Silhouette method [37]. For the Elbow method algorithm, the optimal number of sets is given by the Elbow method in which a set of different numbers of clusters are tested evaluating the squared sum of distances of samples and clusters centers. Using a line plot as a reference, the desired number of clusters is identified by the inflection point in the graph. Our implementations modify the K-means and the Elbow method to use the Fisher \( k \) statistics to obtain optimized clusters sets.

For the K-means algorithm we consider the spherical K-means which uses the angular distance between two strike/dip positions as polar coordinates \((\phi/\theta)\) to measure the clusters and centers proximity. We use an additional step that verifies if the generated clusters have optimal Fisher \( k \) values, considering that are no set with Fisher \( k \) values below the average Fisher \( k \) minus the standard deviation of all sets. This approach guarantees that we have balanced clusters (Figure 2b), instead of some clusters with low dispersion and others with high dispersion (Figure 2a). If this condition is not met, the K-means algorithm is restarted with new random centers. These steps are detailed in Algorithm 1.

\[
SP_{12} = \frac{|d_1 - d_2|}{\sqrt{a^2 + b^2 + c^2}}
\]

where \( d_1 \) and \( d_2 \) are two parallel discontinuities (planes). For non-parallel planes, a virtual scanline can be used to estimate the distance between planes that cross the designated line, where the average, minimum and maximum spacing can be computed [40].
Algorithm 1 Spherical K-means with Fisher k groups check
1: Convert Strike/dip data to lat/lon
2: while optimal_Fisher_k ≠ True do
3:    Get random K lat/lon positions as initial set centers
4:    while moving_centers ≠ False do
5:      Compute distances between centers and measures
6:      Associate measures to closest centers
7:      Change center positions to new centers position in each group
8:      if distance(centers, new_centers) ≤ tolerance then
9:         moving_centers ← False
10:        centers ← new_centers
11:     end if
12:    end while
13:    Compute Fisher_k_values for each group
14:    if Fisher_k_values ≤ (max(Fisher_k) * 0.25) then
15:        optimal_Fisher_k ← False
16:    else
17:        optimal_Fisher_k ← True
18:    end if
19: end while
20: Output: centers

Algorithm 2 Elbow method for spherical data clustering
1: Input: strike_data, dip_data, max_clusters, max_iterations
2: initialize mean_fisher_k array
3:   ▶ Compute mean_fisher_k for max_iterations runs for clusters from 1 to max_clusters
4: for i_cluster = 1, i_cluster++ do
5:    sum_fisher_k = 0
6:    for i = 1, i <= max_iterations, i++ do
7:        mean_fisher_k[i_cluster] = (sum_fisher_k/max_iterations)²;
8:    end for
9:   ▶ Triangulate the squared mean_fisher_k as positions in a graph and get the point with longest distance to the reference line
10: x_i, x_j ← first and last values from mean_fisher_k
11: y_i, y_j ← first and last index positions from mean_fisher_k
12: Initiate distances array
13: for i_cluster = 1, i < max_clusters, i++ do
14:    x_i = i
15:    y_i = mean_fisher_k[i]
16:    numerator ← abs((y_j - y_i) × x_i - (x_j - x_i) × y_i)
17:    denominator ← \sqrt{(y_j - y_i)² + (x_j - x_i)²}
18:    distances[i] ← numerator/denominator
19: end for
20: ideal_n_clusters ← position of the highest value on distances
21: Output: ideal_n_clusters

E. 3D FRACTURE GENERATION AND STOCHASTIC DFN MODELING

With Fisher statistic parameters and spacing for each set, we can generate the stochastic DFN by estimating random planes that tend to follow the same Fisher distribution of the original sets when applying stochastic techniques.

This process starts by establishing the quantity of fracture planes to be generated for each fracture set. Our implementation considers the mean plane sizes to determine the quantity of fracture layers to be generated, as the height (z axis) of the model is divided by the mean fracture size.

In each layer, the quantity (n) of fracture planes to be generated in each set is given by the width (x axis) times the depth (y axis) of the model divided by the set spacing.

After getting the quantity of fractures for each layer (n), we generate two lists of random uniform values R₁ and R₂ of size n. To generate the new fracture data we apply these random values in the Fisher distribution sampling function given by

\[ \theta = 2 \arcsin \sqrt{-\log[R_1 (1 - \lambda) + \lambda]/2k} \]  

(14)

and

\[ \phi = 2\pi R_2 \]  

(15)

where k is the Fisher dispersion, and \( \lambda = \exp(-2k) \). Which are rotated using the rotation matrix given a initial \( \theta_0 \) and \( \phi_0 \).

The rotation matrix used to obtain the final sample values \( \theta' \) and \( \phi' \) is given by

\[
A(\theta_0, \phi_0, 0) = \begin{pmatrix}
\cos \theta_0 & \sin \theta_0 & 0 \\
\cos \phi_0 & \cos \theta_0 \sin \phi_0 & -\sin \phi_0 \\
\sin \phi_0 & \cos \theta_0 \cos \phi_0 & \cos \phi_0 \\
\end{pmatrix}
\]  

(16)

The random generated plane directions are then converted to their normal vector information to be placed in the stochastic DFN following a uniform random positions for each layer.

To estimate the volumetric fracture intensity \( P_{32} \) we consider a sample volume with the same size and position to the stochastic DFN model, while this volume is subdivided in cells of equal sizes. In each cell, we compute the fracture plane intersections with the cell faces to obtain the vertices that make the fracture portion polygon inside the cell to compute its area.
These intersections are obtained using the Möller-Trumbore ray-triangle intersection algorithm [41] considering the cell faces and fracture planes as triangle pairs and the rays as triangle vertices. Fracture plane vertices inside the cells are identified using the Delaunay triangulation using the cell edges as reference.

The polygon areas using the extracted vertices are computed subdividing the polygon using the Delaunay triangulation and estimating the area of triangles that fit the polygon. All polygon areas inside a sample cell are summed and then divided by the volume to compute $P_{32}$ for each cell.

The generated algorithm and scripts presented are packaged in a application named vizfrac3D which will generate the following products as presented in Figure 3. This chart illustrates what a user could expect in terms of input and output to our script in an easier way, complementing the main workflow presented in Figure 1.

![Application for 3D stochastic DFN generation](image)

**FIGURE 3:** Application for 3D stochastic DFN generation vizfrac3D.

### III. METHOD VALIDATION

#### A. STUDY AREA AND GEOLOGICAL SETTINGS

For this study we used a ravine within the Lajedo Soledade (Figure 4B), located in Apodi municipality, in the Rio Grande do Norte state, Brazil. This outcrop presents rocks within the southeast portion of the Potiguar basin [42], [43], precisely within the Jandaíra formation (Figure 4A), deposited between Turonian and Campanian stages (ca. 93.9-83.6 Ma) [29], [44].

This formation is characterized by the presence of a monotonous and homogeneous succession of carbonate rocks predominantly classified as mudstones, grainstones, and packstones, composed of grains of primary carbonate minerals (calcite) with strong posterior transformation of primary minerals for dolomite [30], [46], [47]. Grainstones are grain-supported rocks and have a relatively small fraction (<1%) of clay-like grain material (<20µm), while packstone is a supported rock by grains, but with a clay fraction ratio >1%, and mudstones are not supported by grains and are predominantly formed by fragments of grain size <20µm [48]. The rocks have a minority fossiliferous content of bivalves, mollusks, and algae remain [46].

The Jandaíra Formation is locally preserved in an expansion that exceeds 600m in thickness with a slight bedding dip (< 5 degrees) to NNE at the regional level [47], [49]. This late Cretaceous formation comprises a fracture network that is predominantly in the N-S; E-W directions, and secondarily NE-SW and NW-SE directions [44], [47], in both cases with sub-vertical dips associated to the tectonic evolution of the Potiguar basin. This fractured system is characterized by vertical or sub-vertical joints and veins. Horizontal fractures and fracture parallel to the Jandaíra formation bedding also occur. These structures are possibly related to the lithostatic pressure relief process in the Potiguar Basin region [44], [50].

Given the geological characterization and the extensive areas of naturally fractured carbonate outcrops present in the Jandaíra Formation, this region is of great interest for geological studies focused on hydrocarbon reservoirs, like the pre-salt reservoirs, in the Brazilian offshore. The Lajedo Soledade outcrop, in particular, has NE-SW faults and E-W fractures, which control its drainage system [29]. This outcrop presents an important process of karstification resulting from the dissolution of carbonate rocks by meteoric waters, as described in [29], which led to the creation of cavities such as the ravine used for this study. The karst dissolution process in the ravine, as well as the characterization of stylolites and primary discontinuities are best described in [30], based on field and laboratory analyses. However, a more comprehensible geometric analysis of these features is yet to be seen, and this is one key motivation for this work.

#### B. VOM ACQUISITION

For the UAV image acquisition in the ravine, we used a DJI Mavic 2 Pro embedded with a Hasselblad 20 Megapixels camera. The ravine represents 7,000m² of the outcrop, in which 376 photos were taken with a camera tilt of 45°, in a flight lasting 14 minutes and height of 3m for subsequent photogrammetric reconstruction and generation of the textured 3D model.

The photogrammetric reconstruction follows the flow (Figure 5) proposed by the software Agisoft Metashape [51], version 1.6.2, which employs algorithms of SfM [51], [52] identifying key points (Scale-invariant Feature Transform - SIFT algorithm) in image pairs, and through bundle adjustment estimate camera location to generate the point cloud, whereafter the Multi-View Stereo algorithm [53], [54] is used to create the dense point cloud. The dense point cloud

1 http://www.agisoft.com/downloads/user-manuals/
is triangulated generating a 3D mesh that will be texturized generating the VOM.

FIGURE 5: Common SfM/MVS workflow to generate 3D texturized models from UAV image acquisition.

C. FRACTURE INTERPRETATION IN IMMERSIVE ENVIRONMENT

We used the software Mosis XP designed for geological interpretation in an immersive virtual reality (iVR) system with a head-mounted display (HMD) to extract the fracture planes. This software integrates the Mosis suite, developed by the Vizlab | X-Reality and GeoInformatics Lab. With the iVR system set, the VOM is loaded on Mosis XP project ready for immersive geological interpretation where in this work we use 3D data from the plane measurement tool exported as xyz data.

The HMD used is an HTC Vive with a pair of controllers (also visible within software) tracked by two infrared towers. The computer specs required for the iVR system is relatively high due to the dual visual output of the HMD (one for each eye) at a high refresh rate (90Hz) to avoid motion sickness. The computer setup is an Intel Core i7-6820HK 2.70GHz (3.60GHz) CPU, 64GB RAM, an Nvidia GTX 1080 8GB RAM graphic card, and Windows 10, 64 bit operational system.

D. VALIDATION

For validation of the proposed method to generate the 3D stochastic DFN we evaluate the Fisher statistics of both stochastic and deterministic models. Each generated and interpreted set will have a mean strike and dip, that considering the polar coordinates, is possible to determine the angle distance between the stochastic and deterministic fracture sets. This validation is done considering 30 stochastic generation for each fracture set where we compute the mean and standard deviation of the angle distance.

To validate the proposed method to balance the K-means clusters after the elbow method definition, we run the original spherical K-means clustering and the modified proposed method 10 times for each set considering a small subset. The dispersion Fisher $k$ value is used in terms of mean, standard deviation, min and max values to show the impact of the proposed method.

E. EXPERIMENTAL SETTINGS

The fracture plane data from Mosis XP consists of x,y, and z coordinates which are converted to strike, dip and centroid data following the steps presented in the prior section, while the algorithm to obtain the optimal number of clusters or fracture families is presented in Algorithm 1. We use the average of 50 runs as max iterations considering 1 to 10 clusters.

After computing Fisher and spacing statistics the fracture planes for the 3D DFN model are generated. The volumetric
sampling in this case will give the $P_{32}$ areal fracture intensity per volume with mean and standard values.

The 3D stochastic model to be generated will have a size of 20m$^3$ with individual cube samples of 1m$^3$, while fracture planes will have 4m$^2$.

**IV. RESULTS**

The UAV-SfM-MVS photogrammetric processing (Figure 6) resulted in a textured 3D model with 899,055 triangles faces and texture resolution of 6 times 8k pixels showed in Figure 7a. The iVR interpretation is illustrated in Figure 7b, where we captured 638 planes using the bedding plane tool (Figure 7c). Lastly, the measured plane locations are shown in Figure 7d.

![FIGURE 6: Point cloud generated from photo alignment in the Agisoft Metashape software.](Image)

Applying the Elbow method in the fracture plane data from the iVR interpretation, we identified the optimal number of clusters as shown in Figure 8, highlighted by the inflection point with 7 clusters.

The Table 1 shows the results of the clustering balancing method in comparison to the original spherical clustering, where in 10 runs for each of the selected number of clusters considering a subset of 10% of the data, the proposed method showed a lower standard deviation. Also, the unmodified method achieved Fisher $k$ values lower as 1 indicating groups with no directional trend.

**TABLE 1: Result validation for the proposed clustering balancing considering Fisher $k$ values not lower than 25% range of the max value against the unmodified random spherical K-means considering 10 runs (70 random groups for each method).**

| Fisher $k$-values | mean | std  | min  | max  |
|-------------------|------|------|------|------|
| Original          | 94.15| 79.74| 1.02 | 496.07 |
| Balanced          | 70.90| 31.36| 90.43| 151.20 |

Within each of the seven clusters shown in Figure 9 we computed the Fisher statistics as shown in Table 2, while a visual representation of the Fisher dispersion is shown in Figure 10.

![Figure 10](Image)

**TABLE 2: Directional statistics for each set showing mean direction/mean dip and computed dispersion values and Fisher angles.**

| Set | Mean Strike/Dip | Fisher Angle | R-Value | $k$-Value |
|-----|-----------------|--------------|--------|----------|
| 1   | 64°/77°         | 2.69°        | 0.97   | 31.08    |
| 2   | 91°/58°         | 3.70°        | 0.96   | 24.38    |
| 3   | 135°/76°        | 3.00°        | 0.98   | 44.53    |
| 4   | 165°/79°        | 3.49°        | 0.94   | 17.55    |
| 5   | 217°/11°        | 4.25°        | 0.95   | 19.27    |
| 6   | 104°/80°        | 4.12°        | 0.98   | 59.80    |
| 7   | 155°/51°        | 4.70°        | 0.96   | 23.00    |

The mean and standard deviation of the spacing in each set were also calculated as shown in Table 3.

**TABLE 3: Spacing information regarding fracture planes for each set. Set families with high mean and standard deviation spacing are guaranteed to have at least one fracture represented in the final model.**

| Set 1 | Set 2 | Set 3 | Set 4 | Set 5 | Set 6 | Set 7 |
|-------|-------|-------|-------|-------|-------|-------|
| $n$   | 92    | 63    | 114   | 99    | 61    | 167   |
| $\mu$ | 21.34 | 4.78  | 7.55  | 3.85  | 2.31  | 14.27 |
| $\sigma$ | 30.40 | 7.78  | 14.68 | 5.03  | 3.75  | 19.58 |

In Figure 11 the stochastic model is shown with the fracture generated for each set, alongside the volumetric model with 8,000 sample blocks colored according to the volumetric fracture intensity $P_{32}$, which presented a mean intensity of 22.57m$^{-1}$, and standard deviation of 7.92.

The Figure 12 shows the stereonets with fracture planes generated for each clusters, matching the Figure 10, while the number of fracture planes generated are shown in Table 4.

**TABLE 4: Quantity of fracture planes generates for each set.**

| Set 1 | Set 2 | Set 3 | Set 4 | Set 5 | Set 6 | Set 7 |
|-------|-------|-------|-------|-------|-------|-------|
| $n$   | 1360  | 90    | 415   | 260   | 520   | 860   | 140   |

For the stochastic process validation considering 30 stochastic runs we achieve low deviation on the distance angle values indicating the statistical stability, while the mean angle distances is minimal indicating that the stochastic generations replicated the same behavior found in the deterministic model as shown in Table 5.

**TABLE 5: Deterministic and stochastic mean strike/dip from each set and mean angular distance computed for method validation considering 30 stochastic generations.**

| Set | Deterministic | Stochastic (Mean) | $\mu$ distances | $\sigma$ distances |
|-----|---------------|-------------------|-----------------|-------------------|
| 1   | 64°/77°       | 64°/77°           | 0.0012°         | 0.0008            |
| 2   | 91°/58°       | 92°/47°           | 0.1914°         | 0.0518            |
| 3   | 135°/76°      | 135°/76°          | 0.0017°         | 0.0010            |
| 4   | 165°/79°      | 165°/79°          | 0.0649°         | 0.0026            |
| 5   | 217°/11°      | 22°/11°           | 0.0056°         | 0.0024            |
| 6   | 104°/80°      | 104°/80°          | 0.0011°         | 0.0007            |
| 7   | 155°/51°      | 155°/51°          | 0.0016°         | 0.0015             |

**V. DISCUSSION**

The first of the proposed improvements, namely the modified Elbow method to identify the optimal number of clusters...
FIGURE 7: Fracture interpretation in Mosis XP in immersive mode. The red marked area in (a) is detailed in (b) showing the measured planes and scale. In (c) the points marked to capture fracture planes and in (d), illustration of the planes captured in the outcrop.

FIGURE 8: Elbow method with 10 different cluster sizes.

(Figure 8), proved an important step to the stochastic DFN generation workflow as it reduce the bias when selecting the number of fracture sets from the gathered data. It also contributed to generate fracture data respecting a lower direction dispersion used for the Fisher generation. This is a crucial factor as a reduced number of set can generate sets that not follow a visible trend, and prior works like [21]–[25] have generally ignored these characteristics of groups’ dispersion.

Additionally, the proposed clustering balancing using the Fisher $k$ values of dispersion guarantee that the selected clusters will follow a more defined trend. This is shown in our validation step as without our method, the K-means spherical clustering generated groups with values as low as 1.02 for the Fisher $k$ value of dispersion, while the balanced results presented a minimum values of 30.43 and lower standard deviation between groups (Table 1). Our approach tackle a concern also find by [25], [26], however they only adjust the ideal position of clusters centers and not the final clusters.

With cluster balancing, the data distributed in the sphere
in their respective sets shown in Figure 9 and the results presented in Table 2, prove the validity of the proposed method. Additionally, from the proposed optimizations, our stochastic model is able to reproduce the input data (10 and 11) with average strike/dip differences equal to 0.03°, having a set with a difference of 0.19° and all others with values less than 0.006° (Table 5).

Regarding the DFN modeling, an important aspect to be set is the plane form and dimensions that represent the fracture, while the most common forms are ellipsoidal, rectangular, or with the planes extending to infinity. This will impact how the area of the plane will be measured to estimate the fracture intensity in the volume $P_{32}$ or how the topology will be constructed as suggested by [55]. However, the squared shape represented was the most appropriate to our fracture model. Besides that, the sample volume size needs to be further explored, as prior works have employed sizes varying from centimeters to tens of meters [21], [56]–[58].

As for the VOM interpretation, the constant evolution and improvement of UAV-SfM-MVS techniques for generating VOMs have resulted in increasingly lighter meshes [59], which can increase the accessibility for such immersive interpretations, as performance capabilities of today’s computers, graphic cards, and HMDs enable more people to experience virtual reality.

The studied area presents steep and narrow sinkholes and canyons from karstification and subsequent erosion of these faults [29], [30], which makes access to these places difficult and dangerous. The adequate collection of fracture planes was possible due to the particularity of the immersive virtual environment that allows navigating the outcrop safely and reaching places that would not be possible in a real environment. Thus, the proposed framework showed a viable alternative as a mean to safely provide three-dimensional fracture data in adequate quantity and quality for the generation of numerical models from realistic fracture networks, which reinforces previous work on these themes.
VI. CONCLUSION

In this work we presented a workflow to generate 3D stochastic DFN models, with algorithms optimization for fracture plane clusterization. To obtain three-dimensional data, we interpreted a virtual outcrop model in an immersive environment. The mains conclusions are listed as follows:

- The decision regarding the number of fracture sets often disregarded can be automatized adapting the Elbow method to work with spherical data, reducing the user bias in the process;
- Clusterization data can be improved by the analysis of spherical statistics;
- The automatic cluster balancing brings more data quality for the stochastic processes generating fractures with similar behavior to the input data.

Additional future works aim to integrate DFN model with multi-scale data from regional observations and porosity data to generate a representative geo-cellular model of the region. Also, further validation of the statistical models is to be carried out incorporating machine learning techniques. Further analysis need to be made concerning the number of interpretations to verify how representative to the DFN the acquired interpretations are.

CODE AVAILABILITY

The code script developed in Python and the related data is available at the Github repository (https://bit.ly/3vtOGRx), under the BSD-3 license.

CONTRIBUTORS

G.R. and A.M.J. conception and software development, analysis and interpretation of the data and drafting the article. G.R., A.M.J., and L.T. analyzed the results. G.R. and A.M.J. wrote the text. All authors reviewed the manuscript.

REFERENCES

[1] Herson Oliveira da Rocha, Jéssica Lia Santos da Costa, Antonio Abel González Carrasquilla, and Alfredo Moisés Vallejos Carrasco. Permeability estimation and analysis of fracture networks using resistivity logs in an offshore aptian carbonate reservoir pre-salt, in the southeastern santos basin. Journal of Applied Geophysics, 184:104241, 2021.

[2] Juliette Lamarche, Arthur P.C. Lavenu, Bertrand D.M. Gauthier, Yves Guglielmi, and Océane Jayet. Relationships between fracture patterns, geodynamics and mechanical stratigraphy in carbonates (south-east basin, france). Tectonophysics, 581:231–245, 2012. Crustal Stresses, Fractures, and Fault Zones: The Legacy of Jacques Angelier.

[3] Arthur P.C. Lavenu and Juliette Lamarche. What controls diffuse fractures in platform carbonates? insights from provence (france) and apulia (italy). Journal of Structural Geology, 108:94–107, 2018. Spatial arrangement of fractures and faults.

[4] Knut Bjorlykke. Petroleum geoscience: From sedimentary environments to rock physics. Springer Science & Business Media, 2a edition, 2010.

[5] DCP Peacock, CW Nixon, A Rotevatn, DJ Sanderson, and LF Zuluaga. Glossary of fault and other fracture networks. Journal of Structural Geology, 92:12–29, 2016.

[6] Qinghua Lei, John Paul Latham, and Chin Fu Tsang. The use of discrete fracture networks for modelling coupled geomechanical and hydrological behaviour of fractured rocks. Computers and Geotechnics, 85:105438, 2020.

[7] Alireza Shahbazi, Ali Saeidi, and Romain Chesnaux. A review of existing methods used to evaluate the hydraulic conductivity of a fractured rock mass. Engineering Geology, 265(November 2019):105438, 2020.

DECLARATION OF INTERESTS

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.

ACKNOWLEDGEMENTS

We would like to thank the funding provided by Petróleo Brasileiro S.A. (PETROBRAS), Agência Nacional do Petróleo, Gás Natural e Biocombustíveis (ANP) grant Numbers 4600556376 and 4600583791, and the Coordenação de Aperfeiçoamento de Pessoal de Nível Superior – Brasil (CAPES) – Finance Code 001, that made this work possible.

CODE AVAILABILITY

The code script developed in Python and the related data is available at the Github repository (https://bit.ly/3vtOGRx), under the BSD-3 license.

CONTRIBUTORS

G.R. and A.M.J. conception and software development, analysis and interpretation of the data and drafting the article. G.R., A.M.J., and L.T. analyzed the results. G.R. and A.M.J. wrote the text. All authors reviewed the manuscript.
[9] Michail Pavlis and Enda Cummins. Assessing the vulnerability of groundwater to pollution in Ireland based on the COST-620 Pan-European approach. Journal of Environmental Management, 133:162–173, 2014.

[10] S. Bigi, M. Bani, A. Alemanni, S. Lombardi, A. Campagna, E. Borisova, and M. Loizzo. CO2 flow through a fractured rock volume: Insights from field data, 3D fractures representation and fluid flow modeling. International Journal of Greenhouse Gas Control, 18:183–199, 2013.

[11] L. Guo, X. Li, Y. Zhou, and Y. Zhang. Generation and verification of three-dimensional network of fractured rock masses stochastic discontinuities based on digitalization. Environmental Earth Sciences, 73(11):7075–7088, 2015.

[12] Alessandro Giuffrida, Fabrizio Agosta, Andrea Rustichelli, Elisa Panza, Vincenzo La Bruna, Martin Eriksson, Stefano Torrieri, and Maurizio Giorgioni. Fracture stratigraphy and dfn modelling of tight carbonates, the case study of the lower cretaceous carbonates exposed at the monte alpi (basilicata, italy). Marine and Petroleum Geology, 112(September 2019), 2020.

[13] Baptiste Lepillier, Pierre Olivier Bruna, David Bruhn, Eivind Bastensen, Alexandros Danilidis, Oscar Garcia, Ana Torabi, and Walter Wheeler. From outcrop scanslines to discrete fracture networks, an integrative workflow. Journal of Structural Geology, 133, apr 2020.

[14] M Mauldson, WM Dunne, and MB Rohrbaugh Jr. Circular scanlines and circular windowing tools for characterizing the geometry of fracture traces. Journal of Structural Geology, 23(2-3):247–258, 2001.

[15] Hannah Watkins, Clare E Bond, Dave Healy, and Robert WH Butler. Appraisal of fracture sampling methods and a new workflow to characterise heterogeneous fracture networks at outcrop. Journal of Structural Geology, 72:67–82, 2015.

[16] William S. Dershowitz and Hans H. Herda. Interpretation of fracture spacing and intensity. volume All Days of U.S. Rock Mechanics/Geomechanics Symposium, 06 1992. ARMA-92-0757.

[17] Takako Miyoshi, Davide Elmo, and Steve Rogers. Influence of data analysis when exploiting DFN model representation in the application of rock mass classification systems. Journal of Rock Mechanics and Geotechnical Engineering, 10(6):1046–1062, 2018.

[18] Xiaohai Wang. Stereological interpretation of rock fracture traces on borehole walls and other cylindrical surfaces. PhD thesis, Virginia Tech, 2005.

[19] Stephen Donald Priest. Discontinuity Analysis for Rock Engineering. Springer Science & Business Media, 1993.

[20] Lanru Jing and Ove Stephansson. The basics of fracture system characterization–field mapping and stochastic simulations. In Developments in Geotechnical Engineering, volume 85, pages 147–177. Elsevier, 2007.

[21] A Giuffrida, V La Bruna, P Castelluccio, E Panza, A Rustichelli, E Tondi, M Giorgioni, and F Agosta. Fracture simulation parameters of fractured reservoirs: Analogy with outcropping carbonates of the inner apulian platform, southern italy. Journal of Structural Geology, 123:18–41, 2019.

[22] Luca Smargiassi, Marco Mercuri, Stefano Tavani, Antonio Pignolasha, Michael Kettermann, Andrea Billi, and Eugenio Carmignani. D Discrete Fracture Network (DFN) models of damage zone fluid corridors within a reservoir-scale normal fault in carbonates: Multiscale approach using field data and UAV imagery. Marine and Petroleum Geology, 126(January), 2021.

[23] N Parrino, F Agosta, P Di Stefano, G Napoli, F Pepe, and P Renda. Fluid storage and migration properties of sheared Neptunian dykes. Marine and Petroleum Geology, 102:521–534, 2019.

[24] Dongdong Pan, Shucai Li, Zhenhao Xu, Yichi Zhang, Peng Lin, and Haiyan Li. A deterministic-stochastic identification and modelling method of discrete fracture networks using laser scanning: Development and case study. Engineering Geology, 262(April):105310, 2019.

[25] Wentao Xu, Yangsong Zhang, Xiaozi Li, Xinyong Wang, Fei Ma, Jiuhin Zhao, and Yue Zhang. Extraction and statistics of discontinuity orientation and trace length from typical fractured rock mass: A case study of the Xinchang underground research laboratory site, China. Engineering Geology, 269(February):105553, 2020.

[26] Yanyan Li, Q. Feng, P. Sjögren, O. Stephansson, and L. Jing. Measuring fracture orientation at exposure by using a non-reflector total station. Engineering Geology, 59(1-2):133–146, 2001.

[27] Richard W Allmendinger, Nestor Cardozo, and Donald M Fisher. Structural geology algorithms: Vectors and tensors. Cambridge University Press, 2011.

[28] Nicholas I Fisher, Toby Lewis, and Brian JJ Embleton. Statistical analysis of spherical data. Cambridge university press, 1993.

[29] Vincent S. Cronin. Finding the mean and 95 percent confidence interval of a set of strike-and-dip or lineation data. Environmental and Engineering Geoscience, 14(2):113–119, 2008.

[30] Kanti V Mardia and Peter E Jupp. Directional statistics, volume 494. John Wiley & Sons, 2009.

[31] Trupti M Kodiniraya and Prashant R Makwana. Review on determining number of cluster in k-means clustering. International Journal, 1(6):90–95, 2013.

[32] Adrián J. Riquelme, Antonio Abellán, and Roberto Tomás. Discontinuity spacing analysis in rock masses using 3D point clouds. Engineering Geology, 195:185–195, 2015.

[33] Deheng Kong, Charalampos Saroglou, Faqun Wu, Peng Sha, and Bo Li. Development and application of k-n-fm photogrammetry for quantitative characterization of rock mass discontinuities. 141(Febuary), 2021.

[34] Andreas Buyer and Wulf Schubert. Calculation the spacing of discontinuities from 3D point clouds. Procedia Engineering, 191:270–278, jan 2017.

[35] Tomas Möller and Ben Trumbore. Fast, minimum storage ray-triangle intersection. Journal of graphics tools, 2(1):21–28, 1997.

[36] O da C Pessoa Neto, UM Soares, JG Da Silva, E.H Roesser, CP Florencio, and ACV De Souza potiguar. Boletim de Geociências da PETROBRAS, 15(2):357–369, 2007.

[37] A K Alves de Medeiros and Valéria Centurion Córdoba. Stratigraphic analysis of ponta do mel carbonate platform, albain-cenomanian in potiguar basin, ne of brazil. Journal of South American Earth Sciences, 102:102675, 2020.

[38] Juliana A G Lopes, David L De Castro, and Giovanni Bertotti. Quantitative analysis of the tectonic subsidence in the Potiguar Basin (NE Brazil). Journal of Geodynamics, 117:60–74, jun 2018.

[39] RCT Cassab. Paleontologia da formação jandaíra, cretáceo superior da bacia potiguar, com ênfase na paleobiologia dos gastrópodos. Rio de Janeiro, 2003.

[40] Stefano de Graaf, John J.G. Reijmer, Giovanni V. Bertotti, Francisco H.R. Bezerra, Caroline L. Cazarin, Kevin Bisdom, and Hubert B. Vonhof. Fracturing and calcite cementation controlling fluid flow in the shallow-water carbonates of the Jandaira Formation, Brazil. Marine and Petroleum Geology, 80:382–393, 2017.

[41] Daniel F. Menezes, Francisco H. Bezerra, Fabrizio Balsamo, Andrea Arcari, Rubson P. Maia, and Caroline L. Cazarin. Subsidence rings and fracture pattern around dolines in carbonate platforms – Implications for evolution and petrophysical properties of collapse structures. Marine and Petroleum Geology, 113(October 2019):104113, 2020.

[42] Robert J Dunham. Classification of carbonate rocks according to depositional textures, 1962.
Racolte et al.: Spherical K-means and Elbow method optimizations for 3D stochastic DFN

[49] Giovanni Bertotti, Stefan de Graaf, Kevin Bisdom, Brigit Oskam, Hubert B. Vonhof, Francisco H. R. Bezerra, John J. G. Reijmer, and Caroline L. Cazarin. Fracturing and fluid-flow during post-rift subsidence in carbonates of the Jandaíra Formation, Potiguar Basin, NE Brazil. Basin Research, 29(6):836–853, 2017.

[50] Juliana G. Rabelo, Rubson P. Maia, Francisco H.R. Bezerra, and Carlos C. Nascimento Silva. Karstification and fluid flow in carbonate units controlled by propagation and linkage of mesoscale fractures, Jandaíra Formation, Brazil. Geomorphology, 357, 2020.

[51] Matthew J. Westoby, James Brasington, Niel F. Glasser, Michael J. Hambrey, and Jennifer M. Reynolds. ‘structure-from-motion’ photogrammetry: A low-cost, effective tool for geoscience applications. Geomorphology, 179:300–314, 2012.

[52] A Corradetti, S Tavani, M Parente, A Iannace, F Vinci, C Pirmez, S Torrieri, M Giorgioni, A Pignalosa, and S Mazzoli. Distribution and arrest of vertical through-going joints in a seismic-scale carbonate platform exposure (sorrento peninsula, italy): insights from integrating field survey and digital outcrop model. Journal of Structural Geology, 108:121–136, 2018.

[53] Yasutaka Furukawa and Carlos Hernández. Multi-View Stereo: A Tutorial, volume 2. 2015.

[54] Camila Duelis Viana, Arthur Endlein, Ginaldo Ademar da Cruz Campanha, and Carlos Henrique Grohmann. Algorithms for extraction of structural attitudes from 3d outcrop models. Computers & Geosciences, 90:112–122, 2016.

[55] Akbar Esmailzadeh and Kurosh Shahriar. Shape effect of fractures on intensity and density of discreet fracture networks. Periodica Polytechnica Civil Engineering, 63(2):465–479, 2019.

[56] E. Panza, F. Agosta, A. Rustichelli, S. C. Vinciguerra, A. Ogier-Simonin, M. Dobbs, and G. Prosser. Meso-to-microscale fracture porosity in tight limestones, results of an integrated field and laboratory study. Marine and Petroleum Geology, 103(February):581–595, 2019.

[57] Muhammad Usman, Numair A. Siddiqui, Shi-Qi Zhang, Manoj J. Mathew, Ya-Xuan Zhang, Muhammad Jamil, Xue-Liang Liu, and Nisar Ahmed. 3D geo-cellular static virtual outcrop model and its implications for reservoir petro-physical characteristics and heterogeneities. Petroleum Science, 18(5):1357–1369, 2021.

[58] Rushikesh Battulwar, Masoud Zare-Naghadehi, Ebrahim Emami, and Javad Sattarvand. A state-of-the-art review of automated extraction of rock mass discontinuity characteristics using three-dimensional surface models. Journal of Rock Mechanics and Geotechnical Engineering, (xxxx), 2021.

[59] Yaqian Liang, Fazhi He, and Xiantao Zeng. 3d mesh simplification with feature preservation based on whale optimization algorithm and differential evolution. Integrated Computer-Aided Engineering, 27(4):417–435, 2020.

[60] Paul Gillespie, Erik Monsen, Laurent Maerten, David Hunt, John Thurmond, and Dean Tuck. Fractures in Carbonates: From Digital Outcrops to Mechanical Models. Outcrops Revitalized, (July 2015):137–147, 2011.

[61] Gwénaël Caravaca, Stéphane Le Mouëllic, Nicolas Mangold, Jonas L’Haridon, Laetitia Le Deit, and Marion Massé. 3d digital outcrop model reconstruction of the kimberley outcrop (gale crater, mars) and its integration into virtual reality for simulated geological analysis. Planetary and Space Science, 182(October 2019), 2020.

***

This work is licensed under a Creative Commons Attribution 4.0 License. For more information, see https://creativecommons.org/licenses/by/4.0/