

Application of Hybrid Techniques (Self Organizing Map and Fuzzy Algorithm) Using Backscatter Data for Segmentation and Fine-Scale Roughness Characterization of Seepage Related Seafloor along the Western Continental Margin of India

Bishwajit Chakraborty^{1,*}, *Member, IEEE*, A. Menezes¹, S. Dandapath^{1,2}, W. Fernandes¹,
S. M. Karisiddaiah¹, K. Haris¹, and G. S. Gokul¹

¹CSIR-National Institute of Oceanography, Dona Paula, Goa: 403 004

²Shivaji University, Kolhapur, Maharashtra: 416 004

Abstract — In this work, a technique to unravel the number of data classes using multi-beam echosounding backscatter data for seafloor characterization is demonstrated. Application of self-organizing maps (SOMs) to backscatter profile data has been developed to determine the number of classes. Thereafter, fuzzy C-means (FCM) method is implemented using the number of class information for backscatter profile segmentation. The use of the soft computational technique for seafloor backscatter data facilitate in achieving stationary profile data sets suitable for seafloor roughness model application. Power spectral density (PSD) function of the segmented profiles provide power law parameters (β and a') through curve fitting of the PSD function using the power law expression. The data acquired from western continental margin of India (WCMI), off Goa, reveals five distinct classes of backscatter strength having various segment lengths. The estimated roughness parameters (β and a') of segmented profiles provide quantitative information about the area seafloor roughness. A gridded map of the estimated roughness parameter (β) is generated using 'kriging' method. Along with the application of the SOM and FCM for stationary data segmentation, the presentation of the gridded map of the seafloor roughness and class profile overlies on the backscatter map, which is a first time application relevance to understand the seafloor. The present study puts emphasis on how a combination of the soft computational (SOM and FCM) and numerical techniques (power spectral density at short and fine-scale) are effective in recognizing the seafloor processes and associated sedimentological dynamics in a complex geographical environment (involving pockmarks and faulted structures) subjected to strong bottom currents and seasonal upwelling.

Index Terms — Multi-beam backscatter, Seafloor classification and characterizations, Self-Organizing map (SOM), Fuzzy C- means (FCM), Power spectral density.

I. INTRODUCTION

Echo-sounding systems, single beam (SBES) and multi-beam (MBES), allow coincident acquisition of high-resolution seafloor backscatter and bathymetric data [1], [2], which enormously sustains the marine exploration applications. The remotely acquired bathymetric and backscatter data reveal large-scale as well as fine-scale seafloor roughness (at textural level) [3], [4] respectively. In general, the quantitative or inversion models are not directly applicable to the acquired data and hence bathymetric and backscatter data pre-processing is imperative before modeling is employed [5]. Quantitative model application using non-stationary dataset is a major obstacle since most of the models presume that the input data is stationary and therefore the application of segmentation techniques to divide the dataset into stationary segment is indispensable [6]. Therefore, the success of the remote acoustic model parameter estimations engages data segmentation technique. This leads towards the development of remote online survey methods i.e., reduction of time consuming spot sampling frequency necessary for model parameter validation.

It has been established that the “soft-computing technique”, i.e., the use of inexact solutions to computationally hard tasks e.g., the artificial neural networks (ANNs) based self-organizing maps (SOM) architecture can be used for dimensional reduction applications i.e., representing input vectors to a specific output [7]-[9]. In the past, ANN techniques were proposed for hydro-acoustic data classification [10]. The SOM exercises unsupervised competitive learning on the unknown dataset (input) onto coarser clusters i.e., primary classifications [11]. For real time survey applications, the SOM can be utilized to formulate a decision regarding the number of data classes during the online data acquisition, that are then used as an input to the fuzzy C-means (FCM) algorithms for data segmentation [12]. The FCM will require initial information about the number of available data classes obtained using the SOM architecture. Once the profiles are segmented, the spectral techniques e.g., the power spectral density (PSD) function of the spatial data as a function of the frequency [13] - [15] is required to calculate the seafloor roughness parameters. Accordingly, the seafloor spectral parameters such as power law exponents (β) and intercept (a') of the segmented geomorphic regions are estimated. We have used the curve fitting of the power law function to the log-log plot of the PSD versus the wave number (k) curves of the backscatter traces. The present study endeavors to improve the understanding of the fine-scale roughness associated with the seepage related seafloor along the western continental margins of India (WCMI) [16].

Besides the “introduction”, there are five major sections that cover this work. The sections on “materials” include description of the WCMI study site, MBES echo-sounding data and techniques employed. Moreover, data segmentation methods using soft-computational techniques such as SOM and FCM are elucidated along with the numerical estimation of the roughness parameters in the “methods” section. The other two sections on the “results” and “discussion”, describe the number of class determination details using SOM and FCM, as well as the fine-scale roughness parameters of the segmented datasets of the WCMI study site. The overall findings are summarized in the “conclusion” section.

II. MATERIALS

The seafloor characterization using backscatter profile involves segmentation and subsequent fine-scale roughness parameters estimations of the WCMI (off the Goa coast). The water depth of the survey area varies from 145 m in the northeast, to 330 m in the southwest region [Fig. 1(a)] [16]. The survey area covers nearly 72 km² (9.0 km x 8.0 km) revealing a significant numbers of the pockmarks that are produced by the presence of a gas or fluid seepages escaping from the subsurface along the faults especially towards the western end of the area. Presence of such seepages has economic significance as an indicator for hydrocarbon exploration [17], [18].

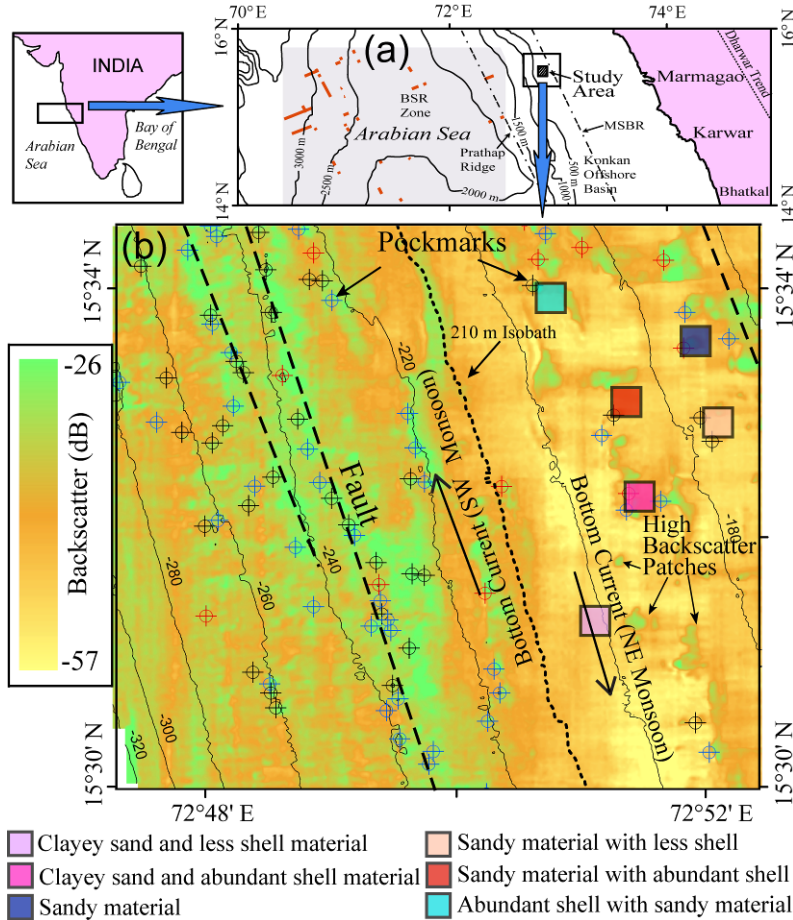


Fig.1. (a) Location of study area including some of the main structural features of the region. On land Dharwarian trend-commonly used for Archaean rocks that have a NNW-SSE trending Precambrian orogenic structure shown as dotted lines. Red highlighted lines and gray shade indicate the identified bottom simulating reflectors (BSR). MSBR refers to Mid-Shelf Basement Ridge, and WCF indicate West Coast Fault. (b) Backscatter map of the study area showing 160-320 m isobaths with an interval of 20 m depth. Pockmarks are indicated by crossed circles. Black, blue and red circles with cross marking represent circular, elliptical and elongated pockmarks respectively. Dashed lines indicate location of identified faults. Black arrows show bottom current directions. Solid colored squares represent the sediment types (see legend). Inset shows outline map of western continental margin of India with the location of the part of the Arabian Sea reproduced from [16].

The calibrated backscatter map presented in Fig. 1b show its variation with the Latitude and Longitude, and the entire map consists of 884930 data points. Fernandes and Chakraborty [19] had developed backscatter data processing technique (PROBASI II) for EM1002 multi-beam system. The data was corrected for propagation and other effects using Neptune software for bathymetry data including tide corrections etc. The backscatter data was processed with PROBASI II algorithm for data normalization and CFLOOR (Cfloor AS) software for improved visualization and gridding (grid resolution: 10 x 10 m). The backscattering strength of this area ranged from -25 to -57 dB. Generally, angular backscatter strength data show higher values at the normal incidence angles compared to the outer-beam angles especially in the case of smooth seafloor. Such backscatter data produce artifacts on-line while using multi-beam system. Even after carrying out sonar related processing of the backscatter data, the artifact around the centre beam direction can be clearly observed (Fig. 2). Therefore, off-line corrections are essential to compensate outer-beam backscatter strength data in such a way that the effect of angular backscatter strength is removed. The improvement in the image data quality can be seen [Fig. 1(b)] after utilizing the four stage image processing technique. The work developed here shows that the employed technique can efficiently classify the survey area using linear data traces (backscatter / bathymetric) varying along the geographical south to the north (Fig. 3). These traces are mentioned as a “profiles”. We have selected seventeen profiles for the present work. Each of the selected seventeen profiles from the backscatter map of the survey area holds 447 data points, and the distance between the two consecutive data points (along the profile) is nearly 20 m. The uniform separation between the seventeen parallel profiles is 0.46 km. The MBES echo-sounding survey data for the present study was acquired using the Simrad Kongsberg EM1002 system operating at 95 kHz. A total of 7599 data points of the processed MBES backscatter data were used from the seventeen selected profiles.

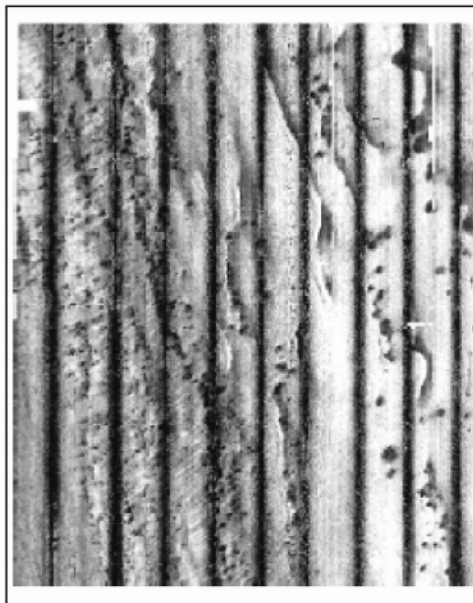


Fig.2. Multi-beam backscatter data show centre beam artifacts (nine dark lines) due to angular backscatter strength

If a statistical technique is applied to the profile data having unpredictable mixtures of the statistical parameters (mean, variance, probability density function) of the smooth and rough seafloor, the estimated model parameter accuracy would be reduced [20]. Therefore, it would be imperative to divide the profile data into stationary segments before any quantitative analyses of the data segment is carried out. Here, we have applied the backscatter profile data segmentation through combined use of SOM and FCM [12]. The SOM uses an unsupervised learning, where the network is unaware of the number of classes in which a particular backscatter dataset would be segregated.

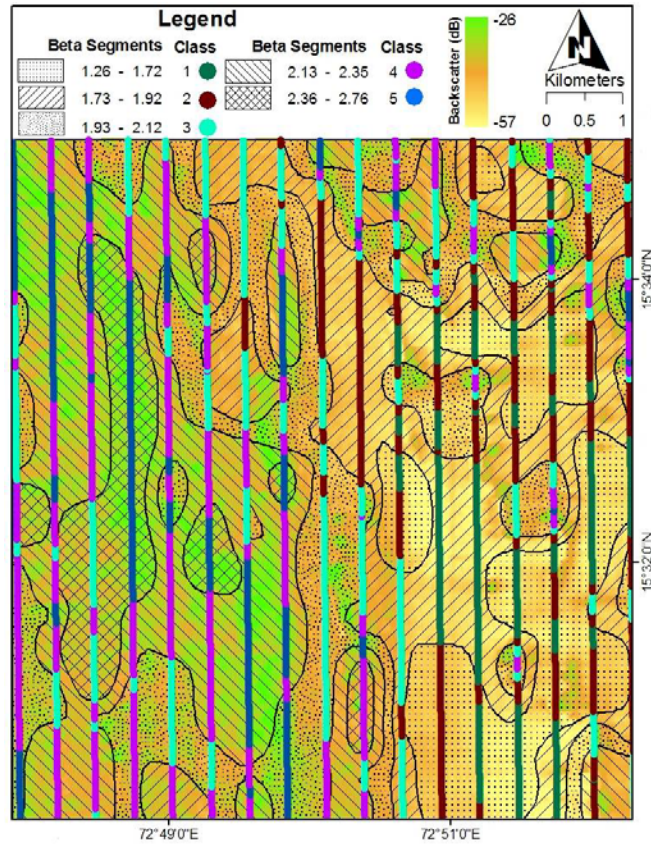


Fig.3. Seventeen backscatter profiles classified into five different classes (depicted in color) overlaid on rasterized map of β values that is also being estimated using the segmented profiles of the five classes.

III. METHODS

The extracted seventeen backscatter profiles from the calibrated image data are subjected to a twenty-points moving averaging. A moving average is a type of filter generally used with temporal or spatial series data to smooth out short-term fluctuations. For a time or space series, the first element of the moving average can be obtained by taking the average of the initial fixed subset of the series. This subset is then modified by “shifting forward”, i.e., by excluding the first number of the series and including the next number following the original subset in the series i.e., average of the data within the sliding window of 400 m along the backscatter profile [Fig. 4(a)]. Thereafter, the normalization of the

data is employed using the given technique [10]. The corresponding depth values are also used to calculate the seafloor local roughness i.e., large scale roughness. In the case of profile segmentation application proposed here, the backscatter and depth data are employed to cover fine and large scale roughness aspects of the seafloor respectively, and these two characteristics are then used as a feature vectors. The ‘roughness’ parameter i.e., the absolute value of the ‘y’, is the deviation of the depth value about the local linear trend of the data. The large scale roughness parameter was estimated for each of the nine data points (180 m) for the depth data profiles. If the deviations with respect to the local linear trends are large, then the seafloor surface is considered as rough, otherwise the surface is smooth. Furthermore, sliding window (400 m) i.e., twenty-point moving average and normalization technique is also applied to the computed roughness values of the depth data as an additional feature [Fig. 4(a)].

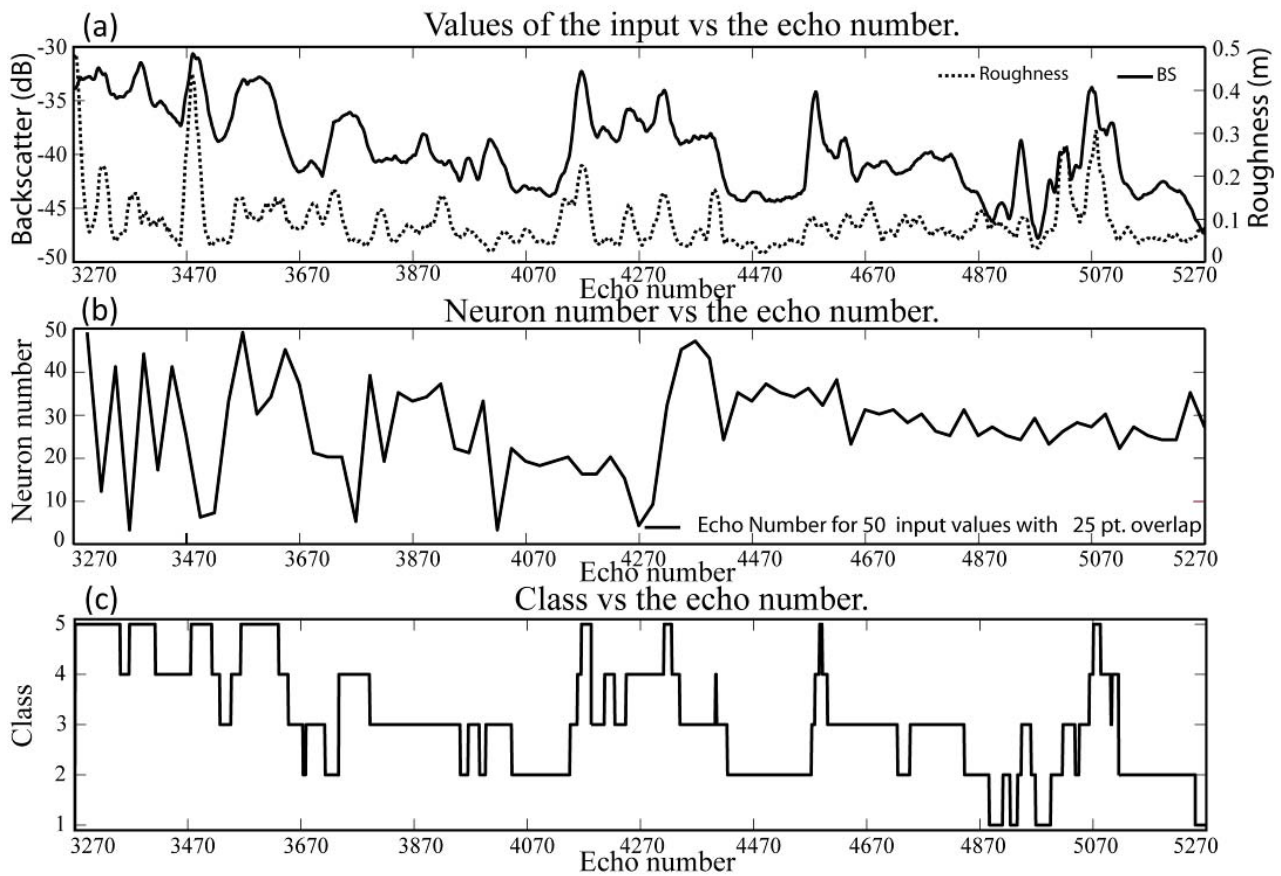


Fig.4. (a) Representative plot of the input values (backscatter and roughness) from a section of the data profile, (b) Firing neurons correspond to the SOM output of the two input feature vectors for above data, (c) Classification of the data points using FCM.

An algorithm provided in Fig. 5 (a flowchart of the entire technique employed) for data segmentation as well as other methodology (PSD) employment for the segmented data. The employed SOM architecture comprised of a flat one-dimensional neuron grid. As already mentioned, the backscatter strength along with the seafloor roughness calculated using the depth data from the profiles are used as

the two input features. The feature vectors are connected to the SOM network, which consists of 50 output neurons in the grid. When the backscatter and roughness data are employed as input vectors to the SOM architecture, the neurons in the grid compete among themselves to get activated. Weights of those neurons in its neighborhood are updated iteratively to form a representative cluster (please refer [10] for more details about the neighborhood reduction technique used in the SOM). This is known as a tuning of the weights in response to a given class of input vectors. The algorithm organizes the nodes in the grid into local neighborhoods that act as an input feature.

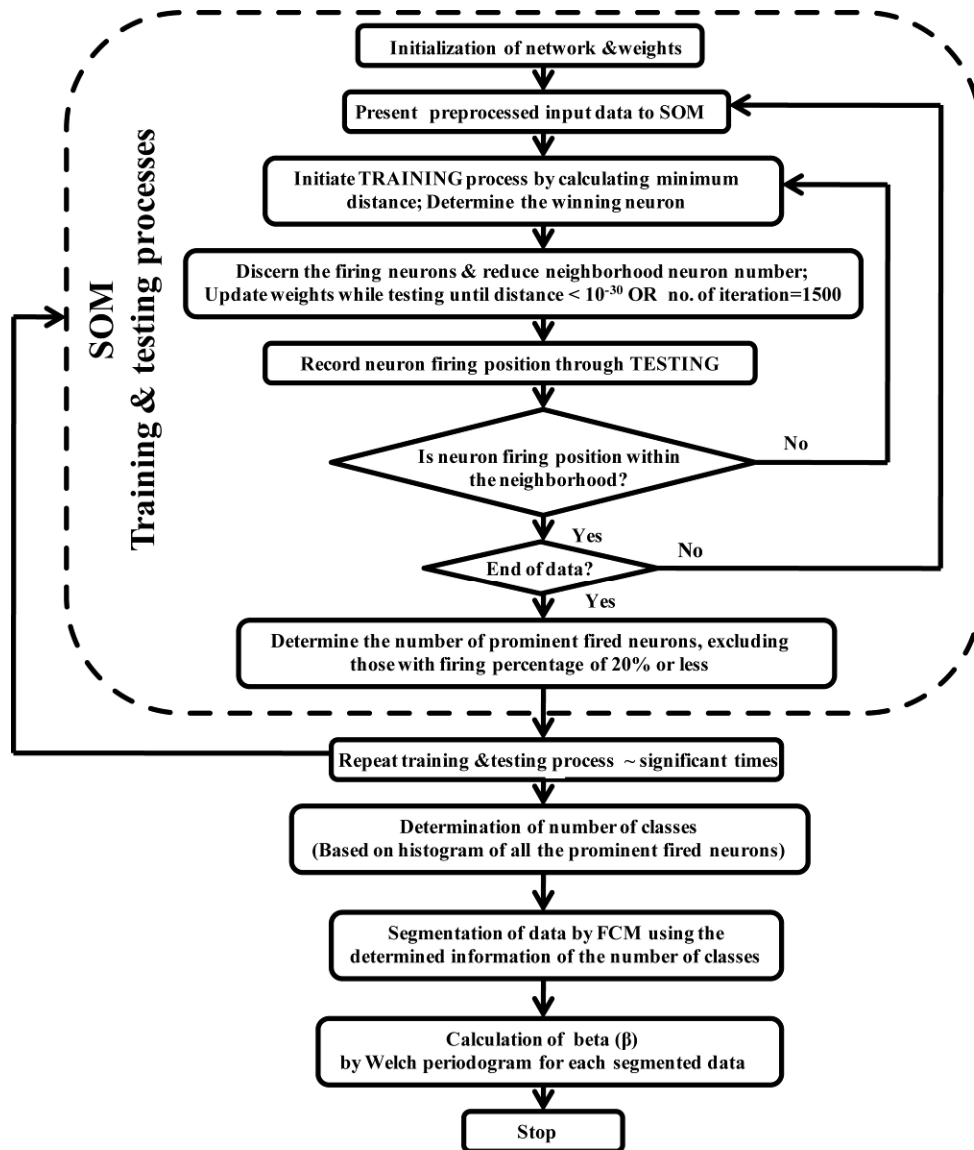


Fig.5. Flowchart of the methodology followed including SOM and FCM for determining the number of seafloor classes, and β parameters using seventeen selected profiles from the backscatter map.

An initial weight matrix having small random value is used. Initially, two input feature vectors are presented as a training sample, and then the euclidean distance (d_j) between the input vector and the weight of the neurons in the one-dimensional grid is computed:

$$d_j = \sum_{i=1}^j [x_i(t) - w_{ij}(t)]^2 \quad (1)$$

where $x_i(t)$ is the input vector to the node i at time ' t ', and $w_{ij}(t)$ ($0 \leq i \leq n-1$) is the random weight vector from input nodes i to j for initialization of the weights from the n inputs to the nodes having small random values. The neuron having the least distance is designated as the 'winner' neuron. The weights of this neuron are updated and the distance matrix is computed iteratively for the same input vectors to minimize the error through use of the expression $w_{ij}(t+1) = w_{ij}(t) + h(t) \times [x_i(t) - w_{ij}(t)]$. This procedure is consistently repeated for a number of times till the value of d_{\min} (minimum distance) is reduced below that of a pre-specified error value. The term $h(t)$ is a learning function ($0 < h(t) < 1$), and gradually reducing the magnitude of the weight update as the error is successively reduced. The neighborhood size also decreases as the number of iterations increase, thus localizing the area of maximum activity in response to the input vectors in the SOM architecture [21]. To determine the number of classes using the SOM architecture, two input feature vectors of each data points were presented to the random weight matrix (1 x 50) and the closest neurons were selected to be the firing neurons for the input data. The closest neurons on either side of the firing neuron is updated using the learning function [$h(t) \sim 0.4/t^{0.2}$] [10]. As t increases, the neighborhood is reduced to one, i.e., at the end, only the firing neurons are updated. Initial value of the ' $h(t)$ ' changes from 0.40. At $t=30$, the value reduces to 0.2, and further reduces below 0.15 for $t \geq 150$. The training stops when the error is less than 10^{-30} or at the 1500th iterations. Each time a new input class is applied to the network, the winning node must first be located. This identifies the region of the map that will have its weights updated. In this study, the number of iterations to reach the pre-specified error value is optimized based on the employed trial of the data. The testing of the input features are initiated once the training is completed and the excited neurons during the testing processes are plotted [Fig. 4(b)]. If testing result show that the winning neuron exists within the cluster (group of five neurons around the central neuron) of the trained neuron, then it is assumed to be belonging to the same class as the data where it was trained earlier. Otherwise, further training is initiated in the search of firing neuron, and the training-testing process is resumed. At the end of the training-testing process, the percentage of the number of times the different neurons have been excited, are plotted as a bar diagram with respect to the neuron numbers. The maximum numbers of classes are equal or above the number of occurrences i.e., 20% of the highest neuron firing (Fig. 6).

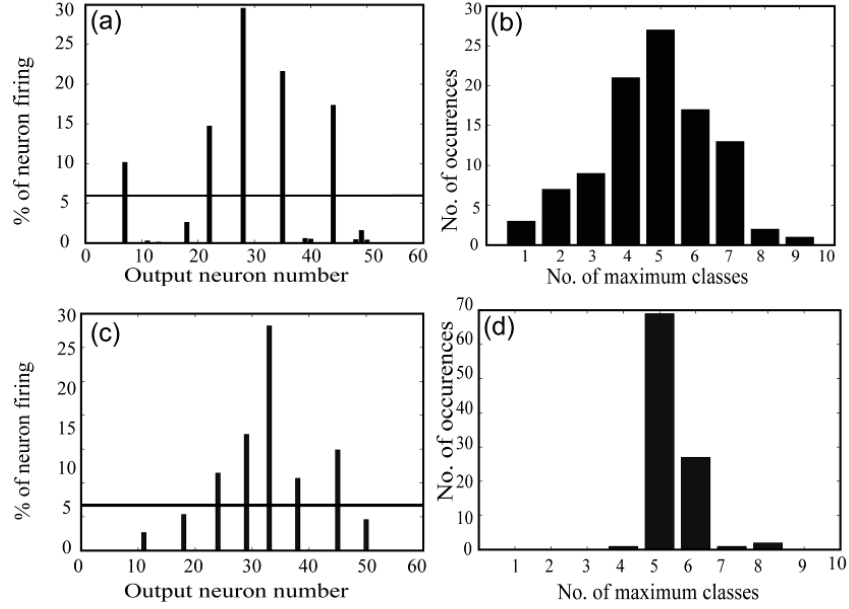


Fig.6. (a and c) Horizontal line represents the line of 20% of the maximum number of neuron firings. Here, there are five bars above the line indicating five classes obtained from one training-testing process for different moving averaging scheme of input data, (b and d) histograms of the number of occurrences of maximum number of classes obtained from the ~ 100 training-testing process employing SOM analysis i.e., indicating no. of classes available in the datasets.

Though, the soft computational technique (SOM) is useful for real time data classification purposes, there is a necessity to validate the extent of success employing other diverse techniques. In order to validate the classes obtained using SOM and FCM, multi-modal statistical techniques were employed here. Histogram of the backscatter data points of the seventeen backscatter profile data are fitted using multi-modal curves considering each curve as Gaussian distribution. The Probability Density Functions (PDFs) of the backscatter strengths are computed. The Gaussian mixture model is expressed as [22]:

$$F(X) = \sum_{n=1}^n a_n \times \exp[-\{(X - b_n)/c_n\}^2] \quad (2)$$

where a_n , b_n and c_n , are the estimated scaling amplitude (to scale the height of the curve), mean, and standard deviations respectively for the PDF components. These parameters are estimated using curve fitting between the experimental and predictive PDFs, which involve comparison between the estimated correlation coefficients and sum square of the residuals (R^2 - SSR criteria) [23]. The mixtures of the normal distributions of the five components are presented (Fig. 7). The highest correlation coefficients (R^2) and the lowest errors (SSR) have been considered to determine the five predictive components and the resultant (mixture) PDFs through use of the experimental data. The study carried out here supports the fact that the backscatter data exhibits the number of available classes in the dataset as a means to support SOM based study.

In the past, extensive comparative work is carried out using unsupervised SOM along with the supervised architecture LVQ (for MBES) [11], and the combination of SOM and FCM as well as SOM and LVQ (for SBES seafloor envelope data) [12]. However, the work presented in [12], comprehensively reported the advantages of hybrid networks SOM and FCM combination over SOM and LVQ. The aim of present work is to develop a system for segmented seafloor classification (based on the SOM and FCM) and subsequent characterization (PSD) of the mapped backscatter data using the selected profiles. This would help in achieving operational advantage for seafloor data acquisition and interpretation. The use of diverse unsupervised algorithms such as ‘‘Adaptive Resonance Theory’’ (ART) and ‘‘Fuzzy ART network’’ where cluster creation is being controlled by use of an explicit parameter known as ‘‘vigilance thresholds’’ [24], are also being used elsewhere for similar applications. In this study, multiple training-testing processes using the SOM, produce neuron numbers. These firing neurons are then plotted in a histogram indicating the five major classes [Fig. 6(a), (c)]. Subsequently the FCM technique is used to segment the datasets [Fig. 4(c)] of the firing neurons in order to obtain the five segmented sets of the original backscatter profile data. The FCM generated segmented profiles are presented in color code using the Geographic Information System software ArcGIS (Fig. 3). In this work, MATLAB based FCM algorithm is employed for clustering of the profile data.

The use of the SOM/FCM techniques along with the spectral estimation to the segmented sections emphasizes the significance of the method employed here. The computations of the PSD of the segmented profiles are employed. The power law equation is given as [15]:

$$\log_{10} P_H = (-\beta)\log_{10} k + \log_{10} a' \quad (3)$$

where P is power [$\text{m}^2/(\text{cycles}/\text{km})$] and k is wave number (cycles/km). A straight line fit of this expression with a PSD provides β (that corresponds to the slope of the straight line) and the intercept a' (m) of the input profile. The estimated β from the above equation is known as the ‘spectral exponent’ of the power law curve. The Welch’s method was implemented using the MATLAB toolbox through the ‘pwelch’ function. A Hamming window was used to compute the modified periodogram of each segment. The PSD is presented on log-log plots of the PSD versus wave number (Fig. 7). Multiple humps in the spectrum in $\log_{10}\text{-}\log_{10}$ [Fig. 7(a)] are generally seen within the ~ -1.5 (cycles/km) to -0.2 (cycles/km). If the profile represents appropriate β (slope) and a' (intercept) values, then its periodogram would be well fitted by a straight line (power law) in $\log_{10}\text{-}\log_{10}$ space. These ‘‘humps’’ are identified as artifacts and are associated with edge effects, which could partially be improved by the edge tapering [25]. However, a clear-cut drop in power occurs somewhere between the wave number (beyond the -0.2 in $\log\text{-}\log$ scale) [Fig. 7(a)]. At higher wave numbers (smaller scales), the periodogram does appear to provide appropriate straight line fitting. The straight-line fitting parameters: β (slope) and a' (intercept) values are estimated within the chosen \log_{10} (wave number) window ranges. Parameters such as correlation coefficient (R^2) and sum of the squared residuals (SSR) of the data points within the windows for the PSD (drawn from the segmented data) and corresponding power law [$W(k)$] function provide β and a' values of the segmented profiles [23] [Fig. 7(b)].

IV. RESULTS

Using SOM architecture, the number of classes were determined by counting the occurrences of number of prominent neurons [bar diagram, Fig. 6(a), (c)]. For example, due to the one time training-testing process of the entire profile data, the highest neuron firings occurred at the neuron position 28, and the rest are observed at 8, 22, 36, and 45 as shown in [Fig. 6(a)]. Similar training-testing process is repeated for a number of times (~ 100) for the entire datasets i.e., the seventeen profiles. For each training-testing process, the numbers of classes are determined from the numbers of occurrences versus the number of classes as plotted in a histogram [Fig. 6(b)]. The number of classes occurring maximum number of times (at the end of repeated cycle for entire data) is considered to be the number of classes existing in the data. This process provides the number of classes available in a given dataset without any prior information. To examine the consistency of the employed technique and the method of pre-processing the variables such as sliding window width etc., the averaging of thirty backscatter data points i.e., 1-30 and so on were also being tested alternatively. It could be observed that there was no change in the number of class estimation using the SOM technique [Fig. 6 (c), (d)].

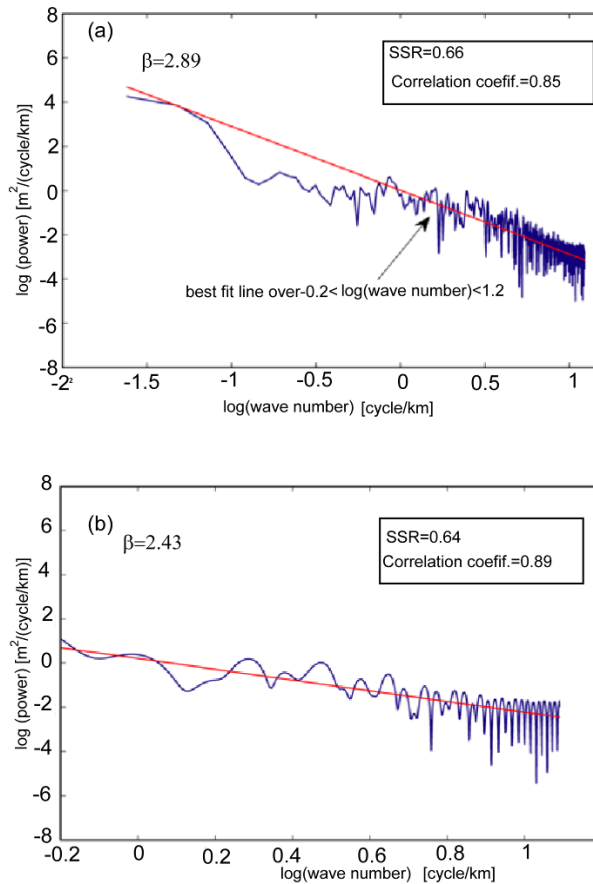


Fig.7. (a) *Log-log* plot of the Welch's averaged modified periodogram (pwelch function in MATLAB) applied to a representative segment, (b) After selecting the appropriate range of wave numbers for curve fittings of the remaining *log-log* plot of the power spectral density versus wave number belonging to the same segment (power law parameter fitting SSR and correlation coefficients are given in the inset).

Though, the soft-computing technique (SOM and FCM) are useful for real time data segmentation purposes, there is a necessity to validate the extent of success employing other diverse techniques such as multi-model PDFs curve fitting mentioned in section II. Five classes (Fig. 8) are observed for the backscatter values of the seventeen profiles. The limits of the backscatter strengths of each class and their overlapping areas are also seen. The five classes obtained using multi-model PDFs matches with the number of classes obtained employing SOM.

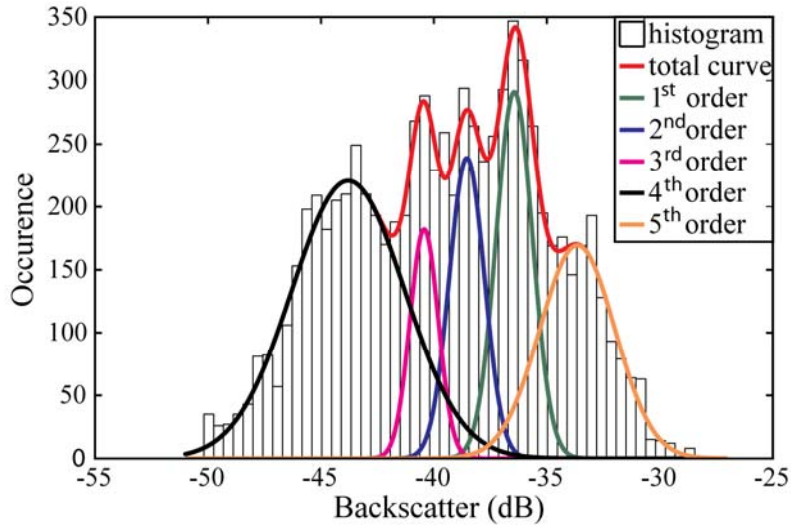


Fig.8. Occurrences of backscatter strength (dB) with respect to the rasterized backscatter pixels (dB) of the study area and fitted multi-modal curves of the total and the five PDF components.

The segmentation of the seventeen profiles provided 243 segments within the five classes. However, the estimation of the seafloor roughness such as β and a' of the 160 segments were determined. Due to the increased inaccuracies of the estimation parameters for the smaller profile lengths, only 160 segmented profiles having varied profile lengths could be used for β and a' estimation. Using these β values a map was generated utilizing the krigging method [26]. The ‘Jenks natural breaks’ (used in the ArcGIS analyst module) algorithm [27] was used to classify the entire map into five gridded blocks. The raster data maximizes the variance between the different β blocks, and minimize the variance within the similar blocks of the β values. A map is presented utilizing the predicted β values (Fig. 3). The results of the segmented classes of the backscatter profiles and the five sets of gridded β values are superimposed over the backscatter raster map using Arc GIS 8.3 (Fig. 3). The predicted β values i.e., map generated data points, are compared with the estimated β values which are situated over the profiles. The histogram of the error values (difference between the estimated and predicted β values) for 7599 data points from the 17 profiles, indicate the accuracy of the prediction [Fig. 9(a)]. This figure shows the extent of fluctuations in the estimated values. Besides, the relationships among the estimated and predicted β values are also indicated by the correlation coefficients (0.862) [Fig. 9(b)]. The present study reveals that, the predicted slope (β) values of the map data are fairly accurate. Overall, the β values are matching with the background map.

In general, the higher β values (2.13-2.76) fall within the fault region of the WCMI, whereas, the lower values (1.26-1.93) indicating small or fine-scale sedimentary region exists towards the shallower eastern end of the shelf. Spatially located intermediate region depicts β within the (1.93-2.12). Moreover, the five classes allocated from the profiles using SOM and FCM techniques are also overlaid on the backscatter map. It can be seen that the number of segments are more towards the comparatively shallow water sedimentary regions. Here the fine-scale roughness is higher (lower β values) and the segments are classified as regions I and II. Similarly, near the fault trace site, the fine-scale roughness is comparatively low (higher β) and the regions are classified as IV and V. Intermediate β values are classified as III i.e., medium roughness, that exists between the fault trace and the shallower region towards the continental shelf. The estimated slope (β) and intercept (a') values of the straight line fitted power law for the 160 profile data reveal the variation from 1.26 to 2.76 and the corresponding intercept a' values from -0.77 to $+0.71$.

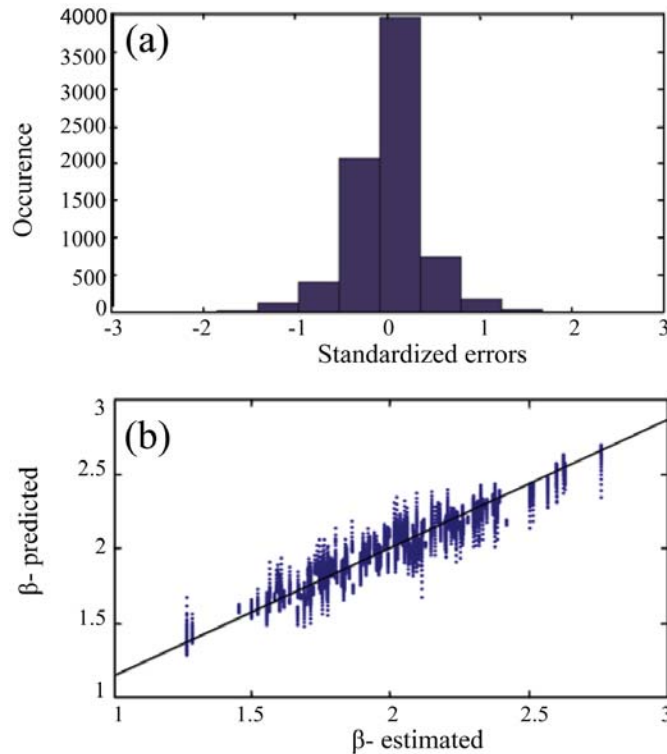


Fig.9. (a) Histogram of standardized errors between the computed and predicted β values (gridded) of the segmented data points (7599 data points) obtained using ArcGIS, (b) scatter plot between the estimated versus predicted β values. Straight line fitted curve indicate 80% correlations.

The three axes presentation of the slope (β), intercept (a'), and classes (SOM and FCM output) of the segmented profiles show a distinct clustering when observed at a view angle of 8.40° (Fig. 10). The mean, maximum and minimum values of the estimated slope (β) and intercept (a') for the classes I, II, III, IV and V are [$\beta=1.67$ (max. 2.07, min. 1.45); $a'=-0.48$ (max. -0.06 , min. -0.75)], [$\beta=1.75$ (max. 2.32, min. 1.26); $a'=-0.43$ (max. 0.42, min. -0.78)], [$\beta=2.05$ (max. 2.76, min. 1.73); $a'=-0.24$ (max.

0.62, min. -0.60)], [$\beta=2.14$ (max. 2.60, min. 1.67); $a'=-0.24$ (max. 0.64, min. -0.80)], and [$\beta=2.30$ (max. 2.63, min. 1.98)]; $a'=-0.14$ (max. 0.71, -0.60)] respectively. The total segments of 160, which is used for PSD based class determination reveal 23, 29, 45, 36 and 20 segments for classes I, II, III, IV, and V respectively. The overlapping between the five classes (β and a') are observed, however, the power law plots (*log-log*) of the mean values within the wave numbers (-0.25 to 1.25 cycles/ km) show a dominant small scale roughness. This affirms that the highest fine-scale roughness (lowest β) to be the class I type [Fig. 11(a)]. The fine or small scale roughness (PSD) is found to be decreasing with the increase in β values within the wave-numbers beginning from class I to V. The segmented profiles having five classes can be observed overlying on the backscatter map (depicted in color) (Fig. 3).

The performances of the segmented backscatter profiles using the SOM and FCM, and the ‘Jenks’ algorithm based gridded blocks of the segmented β values ascertain the success of the employed data segmentation techniques. The percentages of roughness data pixels (β values) reveal 100% matching for the two classes such as I and V. However, 92.2%, 97%, and 97% matching are observed for the classes II, III, and IV respectively. The comparison made here demonstrates that the proposed data-driven approach can be used as a preprocessor to increase the efficiency of the classifier.

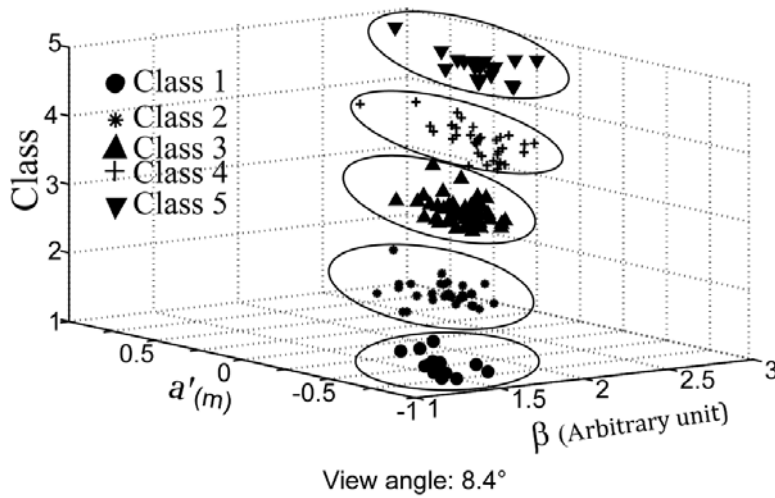


Fig.10. Scatter plot of power law derived parameters such as β and intercept (a') estimated from 160 segments along with the classes determined using SOM and FCM

V. DISCUSSION

The acoustic backscatter strength of the study area ranges from -26 to -57 dB that can be attributed to the seafloor slope, sediment type and relief. In the deeper water (>210 m) region [Fig. 1(b)] several pockmarks in the proximity of the fault zone show high backscatter (-27 to -40 dB). Along the fault zone, the strong seafloor backscatter i.e., increased acoustic impedance is due to the coarse grain seafloor sediment as well as high calcium carbonate content (> 60%) [28]. Earlier study [16] had shown that the areas along the faulted region possess maximum pockmarks [Fig. 1(b)]. Based on the application of box counting technique to the image blocks of the study area, the estimated higher fractal

dimension of the backscatter image blocks show greater roughness (i.e., the rough, fragmented, space-filling seepage distributions) [29]. Structurally, the pockmark shapes are found to be circular, elliptical and composite types with NNE-NNW orientations and the seep direction indicate NNW-SSE orientation. Seafloor pockmarks, bottom simulating reflectors (BSRs), acoustic blanking, and gas-charged sediments were detected during the geological and geophysical surveys. Besides, more than 90% seeps were found to be located in the eastern flank of the pockmarks, which is an important aspect of the seepage dominated fault zone. Interestingly, the present study depict low seafloor roughness i.e., higher β values (2.13 to 2.76) [Fig. 11 (a), (b)] in the area as revealed by the north-south oriented segmented backscatter profile data (SOM and FCM based techniques) that also show dominant IV and V classes. The higher β values are contrary to the previously reported box dimension values where higher seafloor roughness was estimated.

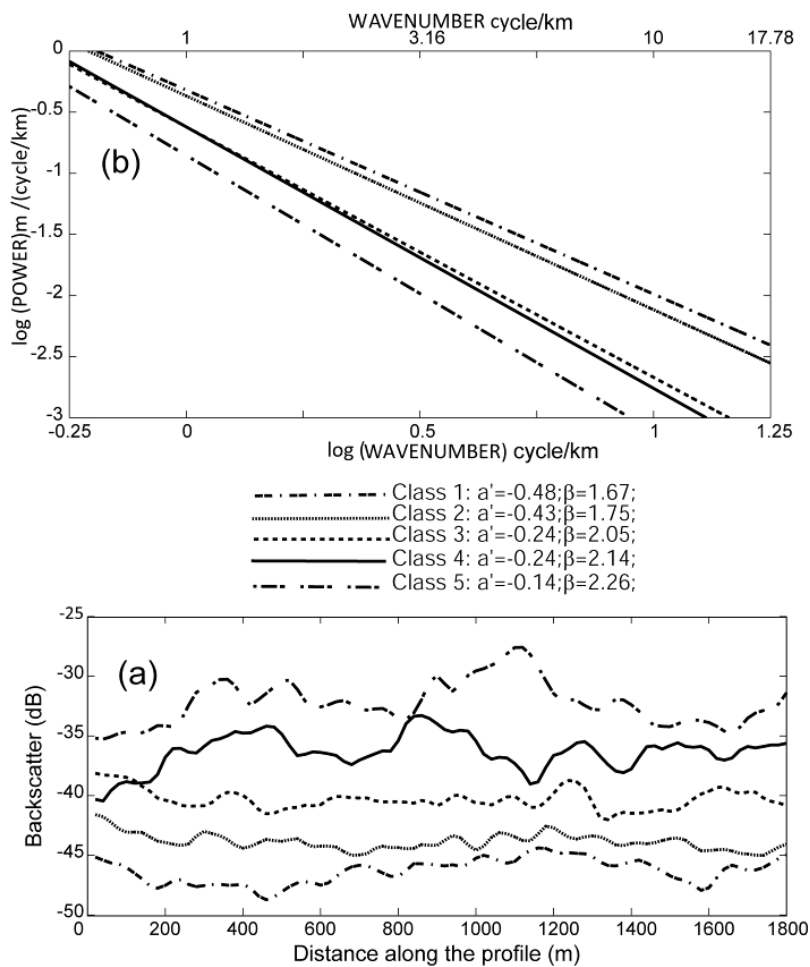


Fig.11. (a) Mean power law parameters (β and a') estimated from the five seafloor classes revealing extent of roughness within the given wave number ranges, (b) Representative profiles of the five classes generated from the rasterized backscatter data indicating the degree of seafloor roughness

Towards the shallower end (< 210 m) where seafloor gradient is gentle, the backscatter strengths are generally low (~ -45 dB). As previously reported by Dandapath et al [16], the area is covered with soft

terrigenous clayey sediment producing average seafloor backscatter strength. However, the sediment data from seven locations in the study area show dominant coarse grained sandy sediments along with abundant shell material [Fig. 1(b)] [29]. Based on the fractal dimensions using the box counting method of the image data of this shallower site, a smooth seafloor roughness was indicated. However, the present study using power law parameter fittings of the power spectral density function to the north-south oriented backscatter profiles show atypical results. The lower β values (1.26 to 1.92) show higher seafloor roughness within the given fine-scale wave-number ranges [Fig. 11 (a)]. Generally the coarse grained sediment along with abundant shell materials reveals high backscatter strength. However due to the prevalent monsoonal bottom currents in this area [30], the existence of seafloor relief is quite obvious. Regional oceanic circulations characterized by seasonal reversal of monsoon driven surface and bottom currents, summer upwelling and winter down welling [31], create an unstable oceanographic condition over this part of the area. Bottom currents are wide (~ 40 km) and run opposite to the direction of surface currents. Measured mean current speed and directions during the southwest monsoons are 12.6 cm/s and 94.5°N respectively. Similarly, during the north-east monsoon, this speed and directions measured here are 12.5 cm/s and 296.6°N respectively [32]. Therefore, we presume that the lower backscatter strengths at the fine-scale level strengthen the existence of seafloor relief due to the bottom currents for coarse grained shelly seafloor sediments.

In this study, the computed β values of backscatter profile data show highest roughness (low β) towards the shallower region (< 210 m). Box dimensions using the box counting computation involves box sizes of 500 x 500 m, which cover backscatter image data due to the different patterns of the seafloor seepage patches. The lower box dimension values reflect lower roughness within the given box sizes [29]. These values are not necessarily comparable to that of fractal dimension computed values using the estimated β for the north-south oriented backscatter profile. The anisotropic seafloor [16] behavior is obvious in this study area due to the north-south oriented dominant bottom currents (as already explained) for the shallower region than the deeper water fault zone. The dominant bottom currents within the shallower area show low backscatter (higher seafloor roughness) even for compacted coarse grained sediments. Moreover, the estimated dimension parameters are box-size sensitive. Therefore the self-affine technique such as spectral method [33] employed in this investigation is more effective than the earlier study using box counting analyses of the seafloor area image [29]. The box counting method appears to be inadequate for analyzing the anisotropic seafloor patterns, but found to be valid for statistically self-similar seafloor type.

VI. CONCLUSION

In this study a novel technique is proposed to characterize the seafloor backscatter data acquired using MBES system that is also applicable to the SBES echo-sounding. A successful analysis has been carried out for the WCMI (off northern part of the Goa coast). This technique provides the real-time seafloor roughness information using remote acoustic method, and is helpful for geological applications having limited sampling data. A soft-computational method was adopted to determine the presence of five classes from the seventeen backscatter and bathymetric profile data. The presence of the five classes has

been further validated employing the multi-modal curve fitting to the histograms of the backscatter profile. This indirectly suggests a logical reasoning for the employment of only two features such as backscatter as well as ‘roughness’ parameter of the depth data. A data-driven approach based on the SOM and FCM based segmentation is used to estimate the fine-scale roughness parameters using PSD of the backscatter profiles. Application of the present technique proffers that, the utilization of calibrated sonar image profiles for real-time classification and characterization is promising, which considerably reduces the survey time.

ACKNOWLEDGMENTS

The data acquisition of this work was carried out under the “Exclusive Economic mapping program” of the Ministry of Earth Sciences (MoES), New Delhi, while the relevant interpretation of the backscatter data was carried out under the “Ocean Acoustics Program” of the National Institute of Ocean Technology (MoES), Chennai. The authors are thankful to Dr. S.W. A. Naqvi, Director, National Institute of Oceanography, Goa for his encouragement. This is NIO contribution xxxx.

REFERENCES

- [1] L. A. Mayer, “Frontiers in Seafloor Mapping and Visualization,” *Mar. Geophys. Res.*, vol. 27, pp.7–17, 2006.
- [2] L. A. Mayer, R. Raymond, G. Glang, M. D. Richardson, P. Traykovski, and A. C. Trembanis, “High Resolution Mapping of Mines and Ripples at the Martha's Vineyard Coastal Observatory,” *IEEE J. Ocean. Eng.*, vol. 32, pp. 133–149, Jan, 2007.
- [3] J. E. Hughes Clarke, “Optimal use of multibeam technology in the study of shelf morphodynamics,” in *International Association of Sedimentology Special Publication # 44, Sediments, Morphology and Sedimentary Processes on Continental Shelves: Advances in technologies, research and applications*, Chichester, UK: Wiley-Blackwell, 2012, pp. 1–28.
- [4] P. Blondel, *Handbook of Sidescan Sonar*, Chichester, Springer/Praxis, 2009.
- [5] K. Haris, B. Chakraborty, C. De, R. G. Prabhudesai, and W. A. Fernandes, “Model-based seafloor characterization employing multi-beam angular backscatter data - A comparative study with dual-frequency single beam,” *J. Acoust. Soc. Am.*, vol. 130, pp. 3623–3632, 2011.
- [6] A. Malinverno, “Segmentation of topographic profiles of the seafloor based on a self-affine model,” *IEEE J. Ocean Eng.*, vol. 14, pp. 348–358, Oct. 1989.
- [7] T. Kohonen, “Self-organising map,” *Proc. IEEE*, vol. 78, pp. 1464–1480, Sep, 1990.
- [8] D. Alexandrou and D. Pantzartzis, “A methodology for acoustic seafloor classification,” *IEEE J. Ocean. Eng.*, vol. 18, pp. 81–86, Apr. 1993.

- [9] Z. H. Michalopoulou, D. Alexandrou, and C. de Moustier, "Application of neural and statistical classifiers to the problem of seafloor characterization," *IEEE J. Ocean. Eng.*, vol. 20, pp. 190–197, Jul. 1995.
- [10] B. Chakraborty, R. Kaustubha, A. Hegde, and A. Pereira, "Acoustic seafloor sediment classification using self organizing feature maps," *IEEE Trans. Geosci. Remote Sens.*, vol. 39, pp. 2722–2725, Dec. 2001.
- [11] B. Chakraborty, V. Kodagali, and J. Baracho, "Sea-floor classification using multibeam echosounding angular backscatter data: A real-time approach employing hybrid neural network architecture," *IEEE J. Ocean. Eng.*, vol. 28, pp. 121–128, Jan. 2003.
- [12] C. De and B. Chakraborty, "Acoustic characterization of seafloor sediment employing a hybrid method of neural network architecture and fuzzy algorithm," *IEEE Geosci. Remote Sens. Lett.*, vol. 6, pp. 743–747, Oct. 2009.
- [13] C. G. Fox and D. E. Hayes, "Quantitative methods for analyzing the roughness of the seafloor," *Rev. Geophys.*, vol. 23, pp. 1–48, 1985.
- [14] J. Berkson and J. Matthews, "Statistical characterization of seafloor roughness," *IEEE J. Ocean. Eng.*, vol. 9, pp. 48–52, Jan. 1984.
- [15] B. Chakraborty, R. Mukhopadhyay, P. Jauhari, V. Mahale, K. Shashikumar, and M. Rajesh, "Fine-scale analysis of shelf – slope physiography across the western continental margin of India," *Geo-Mar. Lett.*, vol. 26, pp. 114–119, 2006.
- [16] S. Dandapath, B. Chakraborty, S. M. Karisiddaiah, A. Menezes, G. Ranade, W. Fernandes, D. K. Naik and K. N. Prudhvi Raju, "Morphology of pockmarks along the western continental margin of India: employing multibeam bathymetry and backscatter data," *Mar. Petrol. Geol.*, vol. 27, pp. 2107–2117, 2010.
- [17] H. King and B. MacLean, "Pockmarks on the Scotian shelf," *Geol. Soc. Am. Bull.*, vol. 81, pp. 3141–3148, 1970.
- [18] M. Hovland and A. G. Judd, *Seabed Pockmarks and Seepages Impact on Geology, Biology and the Marine Environment*, London, Graham & Trotman, 1988.
- [19] W. Fernandes and B. Chakraborty, "Multi-beam backscatter image data processing techniques employed to EM 1002 system," in *IEEE OES Proceedings of the International Symposium on Ocean Electronics (SYMPOL 2009)*. Kochi, India, pp.93–99, 2009.
- [20] A. R. Bansal, and V. P. Dimri, "Gravity evidence for mid-crustal domal structure below Delhi for belt and Bhilwara super group of western India," *Geophys. Res. Lett.*, 26, pp. 2793–2795, 1999.
- [21] B. Chakraborty and K. Haris, "Seafloor roughness estimation employing bathymetric systems: an appraisal of the classification and characterization of the bathymetric data," (Invited paper) in *Advances in Ocean Acoustics (Proceedings of the 3rd International conference on Ocean Acoustics*,

- Beijing), Jixun Zhou, Zhenglin Li. and Jeffrey Simmen, Eds. New York: AIP Conference Proceedings 1495, pp, 283–296, 2012.
- [22] J. D. Pearson, C. H. Morrell and L. J. Brant, “Mixture models for investigating complex distributions,” *J. Quant. Anthropol.*, vol. 3, pp. 325–345, 1992.
- [23] L. Seuront, *Fractals and Multifractals in Ecology and Aquatic Science*, New York, CRC Press, 2010.
- [24] G. A. Carpenter, S. Grossberg, D. Rossen, “Fuzzy ART: Fast stable learning and categorization of analog patterns by an Adaptive Resonance System,” *Neural Networks*, vol. 4, pp. 759–771, 1991
- [25] J. A. Goff, D. L. Orange, L. A. Mayer, J. E. Hughes Clarke, “Detailed investigation of continental shelf morphology using a high-resolution swath sonar survey: the Eel margin, northern California,” *Mar. Geol.*, vol. 154, pp. 255–269, 1999.
- [26] Environmental Systems Research Institute (ESRI) ArcGIS desktop release 8.3. Redlands, CA, 1999.
- [27] M. North, “A method for implementing a statistically significant number of data classes in Jenks Algorithm,” in *Proc. Sixth International Conference on Fuzzy Systems and Knowledge Discovery (IEEE Computer Society)*, 2009, pp. 35–38, (doi 10.1109/FSKD.2009.319).
- [28] M. Thamban, V. P. Rao, and S. V. Raju, “Controls on organic carbon distribution in sediments from the eastern Arabian sea margin,” *Geo-Mar. Lett.*, vol.17, pp. 220–227, 1997.
- [29] S. Dandapath, B. Chakraborty, N. Maslov, S.M. Karisiddaiah, D. Ghosh, W. Fernandes, and A. Menezes, “Characterization of seafloor pockmark seepage of hydrocarbons employing fractal: A case study from the western continental margin of India,” *Mar. Petrol. Geol.*, vol. 29, pp. 115–128, 2012.
- [30] S. R. Shetye, A. D. Gouveia, S.S.C. Shenoi, D. Sundar, G. S. Michael, A. M. Almeida, K. Santanam, “Hydrography and circulation off the west coast of India during the southwest monsoon 1987,” *J. Mar. Res.*, vol. 48, pp. 359–378, 1990.
- [31] S. W. A. Naqvi, H.W. Bange, L. Farias, P. M. S. Monteiro, M .I. Scranton, and J. Zhang, “Marine hypoxia/anoxia as a source of CH₄ and N₂O,” *Biogeosciences*, vol. 7, pp. 2159–2190, 2010.
- [32] P. Amol, D. Shankar, S. G. Aparna, S. C. Shenoi, V. Fernando, S. R. Shetye, A. Mukherjee, Y. Agervadekar, S. Khalap, N. P. Satelkar, “Observational evidence from direct current measurements for propagation of remotely forced waves on the shelf off the west coast of India,” *J. Geophys. Res.*, 117: C05017. doi:10.1029/2011/2011JC007606, 2012.
- [33] A. Malinverno, “A simple method to estimate the fractal dimension of a self-affine series,” *Geophys. Res. Letts.*, vol. 17, pp. 1953–1956, 1990.