Received: 28 February 2024 | Revised: 27 May 2024 | Accepted: 8 June 2024

IET Radar, Sonar & Navigation

DOI: 10.1049/rsn2.12608

The Institution of Engineering and Technology WILEY

ORIGINAL RESEARCH

Model selection techniques for seafloor scattering statistics in synthetic aperture sonar images of complex seafloors

Derek R. Olson[1] | **Marc Geilhufe**[2]

[1]Oceanography Department, Naval Postgraduate School (NPS), Monterey, California, USA

[2]Norwegian Defence Research Establishment (FFI), Kjeller, Norway

Correspondence

Derek R. Olson, Oceanography Department, Naval Postgraduate School (NPS), 833 Dyer Rd., Monterey, CA 93933, USA.
Email: derek.olson@nps.edu

Funding information

United States Office of Naval Research, Grant/Award Number: N00014-23-WX01149

Abstract

In quantitative analysis of seafloor scattering measurements, it is common to model the single-point probability density function of the scattered intensity or amplitude. For more complex seafloors, the pixel amplitude distribution has previously been modelled with a mixture model consisting of two K distributions, but the environment may have more identifiable scattering mechanisms. Choosing the number of components of a mixture model is a decision that must be made, using *a priori* information, or using a data driven approach. Several common model selection techniques from the statistics literature are explored (the Akaike, Bayesian, deviance, and Watanabe-Akaike information criteria) and compared to the authors' choice. Examples are given for synthetic aperture sonar data collected by an autonomous underwater vehicle in a rocky environment off the coast of Bergen, Norway, using the HISAS-1032 synthetic aperture sonar system. The Bayesian information criterion aligned most closely with the interpretation of both the acoustic images and the plots of the probability of false alarm.

KEYWORDS

backscatter, reverberation, statistical distributions, synthetic aperture sonar

1 | INTRODUCTION

Acoustic measurements of seafloor backscattering are a source of unwanted sound in seafloor object detection [1–3] but also provide a rich set of information regarding the seafloor properties and structure [4–7]. The intensity in a sonar image (i.e. a spatial map of measured backscattering) can be characterised using a random process [7]. There are a variety of metrics or features, that can be used to describe this random process, including the autocorrelation function, power spectrum [8], wavelet decomposition [9], grey-level co-occurrence matrix [6, 10], the mean intensity (scattering cross section) [11, 12] and in general, the probability density function (pdf) of the amplitude (or intensity) [13–15] and phase [16, 17].

In previous research on acoustic scattering from complex regions such as rocky seafloors, it was found that the scattering cross section is typically higher than sedimentary seafloors due to the presence of large root mean square (RMS) roughness, but in certain locations the cross section was quite low because of locally smooth areas of rocky surfaces (low RMS roughness) [12, 18]. To model the statistics of the scattered field, a mixture[1] pdf with two components was found to be the most appropriate for rocky [14] and other complex areas [21, 22], which was justified by the non-stationary character of the acoustic data. Each sample of the data was modelled as being drawn from one of two distributions, for example, either from the seafloor or man-made structure in the case of Abraham et al. [21], or from horizontal or vertical facets [14]. The choice of the number of components was an a priori assumption but may not be accurate for all environments. Statistical modelling of acoustic measurements from rocky seafloors using mixture models with an unknown number of components (i.e. two or more) and estimating the most appropriate number of components is the focus of this work.

In general, the number of components that make up a non-stationary sonar image is unknown, and the pdf model for each

[1]We also note that mixtures are a popular choice to model population data taking on discrete values, such as the common cold and influenza viruses [19], and more recently the SARS-CoV-2 virus [20].

component is also unknown. Both must be selected prior to choosing a model and estimating the parameters. The more model parameters are used (i.e. more components, or a more complex statistical model for each component), the better the data will be fit, but the parameters may lose meaning, in that the parameters have higher uncertainty and may lack a connection to the physical environment. In this work, we restrict ourselves to the K distribution, with a justification given in Section 3 and focus on choosing the number of components. We explore the use of several model selection techniques, the Akaike information criterion (AIC), Bayesian information criterion (BIC), deviance information criterion (DIC), and the Watanbe-Akaike information criterion ($WAIC$). These techniques penalise more complex models in different ways. We also use the log-likelihood ($\mathscr{L}$), the Kolmogorov–Smirnov (KS) test and the modified upper-tail Anderson–Darling (AD) test to characterise the model-data fit. In this paper, we compare these model selection criteria to intuition gained from examining the sonar images as well as visual inspection of the probability of false alarm plots.

This paper is organised as follows. A description of the sonar data used in this work and example images are given in Section 2. The background statistical modelling and model selection techniques are given in Section 3. Results are presented and discussed in Section 4, and conclusions are given in Section 5.

2 | DATA

The measurements used in this work are synthetic aperture sonar (SAS) images collected off the coast of Bergen, Norway, by the Norwegian Defence Research establishment (FFI). The platform used for these measurements is the HUGIN-HUS autonomous underwater vehicle (AUV), using a HISAS-1032 interferometric SAS. This sonar system has a centre frequency of 100 kHz and a bandwidth of 30 kHz. The beamformed data is oversampled on a grid with 2 × 2 cm resolution. Data that is used for statistical analysis is decimated by a factor of three in each dimension to reduce the correlation between samples due to the point spread function of the sonar system. Independent samples are necessary for the form of the joint-likelihood function described in Section 3. Seafloor textures may have mean intensity correlations due to patchiness, and correlation due to speckle. If neighbouring samples are more closely-spaced than the mainlobe width of the point spread function, then samples may have correlated speckle. The number three was chosen to balance our desire to reduce the correlation between samples and to not throw out any environmental information. We also note that if the two point correlation function is modelled, then it would be appropriate to use correlated non-Gaussian scattering models, such as those described by Oliver and others [23, 24].

An example image is shown in Figure 1. The image consists of an exposed rock outcrop, with sedimented areas in between. To show the detailed environmental structure, two tiles are plotted in Figures 2 and 3, both of which are 600 × 600 pixels, or 12 m per side. The blue and red boxes in this figure show where these tiles were extracted from. These tiles comprise two data ensembles that are used to fit statistical models and apply model selection techniques. Although the SAS images have both magnitude and phase for each pixel, only the complex magnitude (termed "amplitude" in this work) is studied here. The phase may also provide information regarding the nature of scatterers and should be studied in future work [16].

Both tiles show that the rock structure consists of a low intensity uniform scatterer that varies continuously due to undulations in the rock structure. The continuous variations in intensity are punctuated by bright lines due to fractures, and steps created by glacial erosion, and dark regions composed of either shadows or regions tilted away from the sonar array. These features are distinguishable due to their different intensities, and the SAS system has a high enough resolution that discrete scatterers in the environment are resolved. Therefore, a mixture model is appropriate for modelling the pdf of the ensemble consisting of the pixels from each tile.

The empirical distribution of the data is shown as the probability of false alarm (pfa), because it is a common way of presenting sonar reverberation statistics [13, 25, 26]. The pfa, $P_F(A)$ is defined as $1 - P(A)$, where $P(A)$ is the cumulative distribution function (cdf), and A is the amplitude. This quantity is shown versus normalised amplitude for each tile in Figure 4, tile 1 in (a) and tile 2 in (b). The normalised amplitude is calculated by taking the ensemble of data samples for each tile and dividing each point by the RMS amplitude of the ensemble. The data from tile 1 shows a slight "knee" (change in slope) near normalised amplitudes of 2 and 4. These changes in slope of the $\log(P_F(A))$ indicate different components that make up the model. The $\log(P_F(A))$ of tile 2 has more pronounced knees in the curve, near the normalised amplitudes of 2, 6, and possibly 14, although the last one is more uncertain due to the finite sample size being more evident at high amplitudes (i.e. the pfa curve becomes more stair-case like, rather than a smooth curve). In $\log(P_F(A))$ versus A space, a Rayleigh distribution curves downward, a K distribution consists of a straight line, and anything heavier than a K curves upward (such as a generalised Pareto) [14, 21].

3 | BACKGROUND

In this section, the basic definitions, statistical models, and model selection definitions are given. The data for each ensemble are given in the form of an $N_1 \times N_2$ array of amplitude values, where the samples are statistically uncorrelated with each other. Uncorrelated samples are obtained by decimating the beamformed acoustic data so that its spatial bandwidth in the along-track and across-track dimensions are equal to the sampling spatial frequency [27]. In this case, decimation by a factor of three in each dimension achieves this goal. The total number of samples

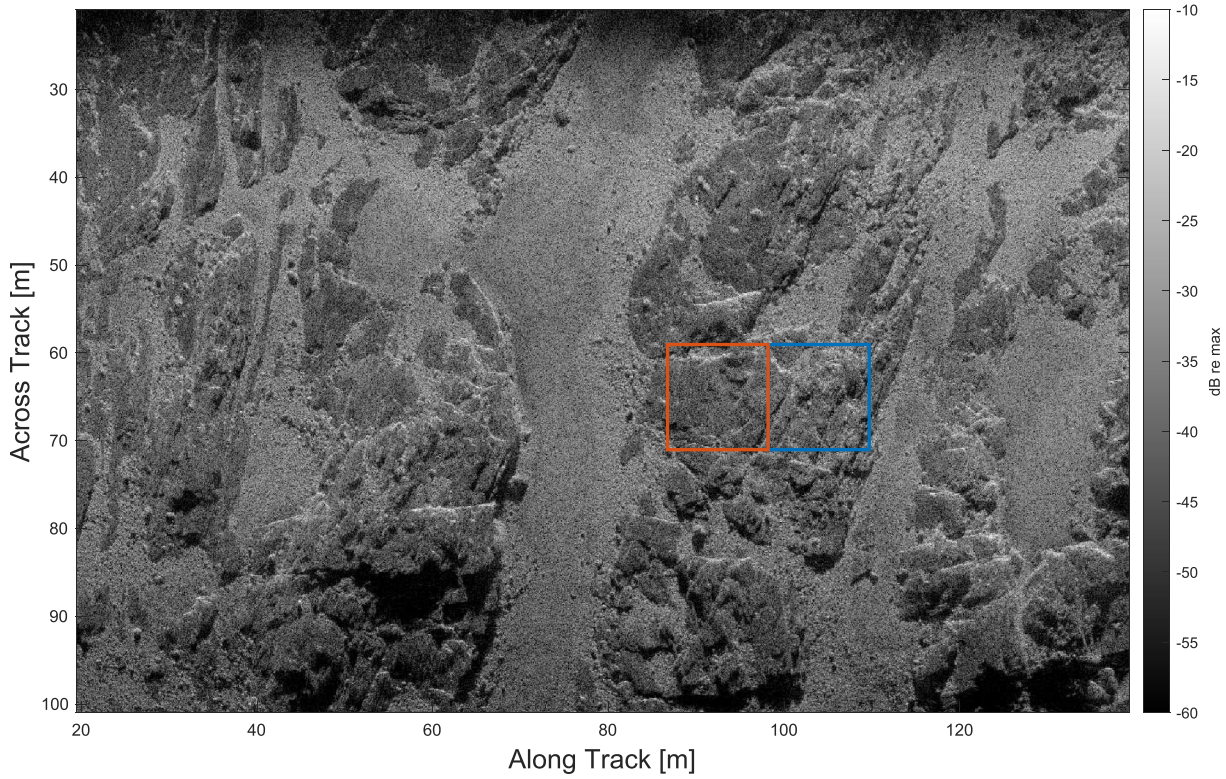

FIGURE 1 An example SAS image, plotted as a function of along track distance on the horizontal axis and across track distance on the vertical axis. The colour scale in decibels is referenced to the maximum intensity in the image, since the system is uncalibrated. The blue and red boxes mark the adjacent regions from which tile 1 and tile 2 were extracted (shown in Figures 2 and 3).

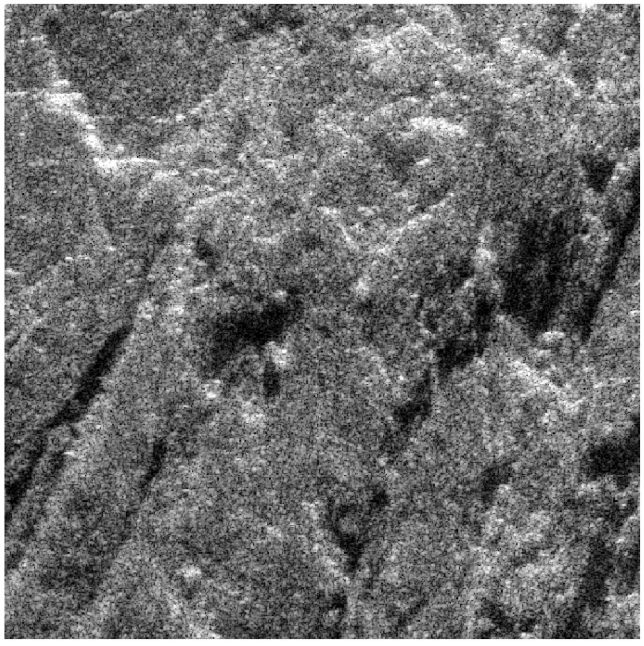

FIGURE 2 Tile 1, taken from the 12 × 12 m blue region of Figure 1. This tile consists of flat portions of smooth rock, dropstones, and a fractured face (in the uppermost region).

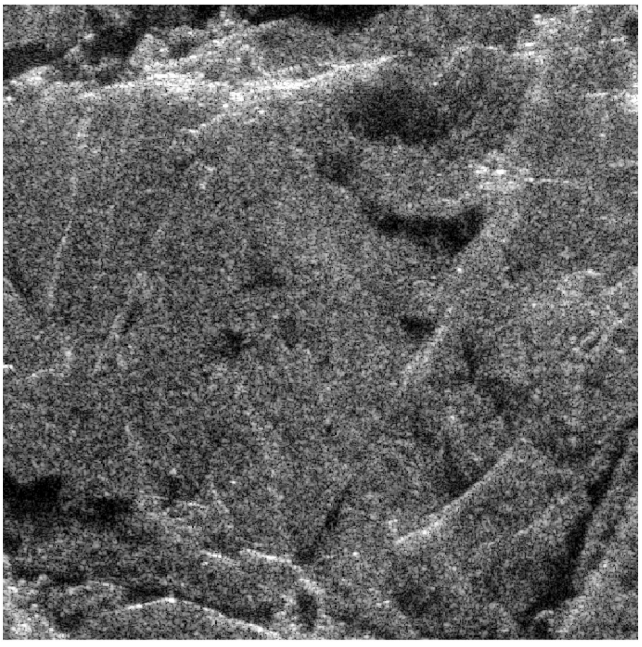

FIGURE 3 Tile 2, taken from the 12 × 12 m red region of Figure 1. This tile consists of flat portions of smooth rock, several small fractures, and a vertical face to the left.

is $N = N_1 N_2$. The random variable for normalised intensity is denoted $\mathscr{I}$, and the normalised amplitude random variable is $\mathscr{A} = \sqrt{\mathscr{I}}$. We denote the pdf of the normalised amplitude (also termed complex envelope in some communities [13, 28]) by $p(A)$, with A being a member of the population.

We use K distributions to model the individual components of the SAS images. In previous work analysing the same

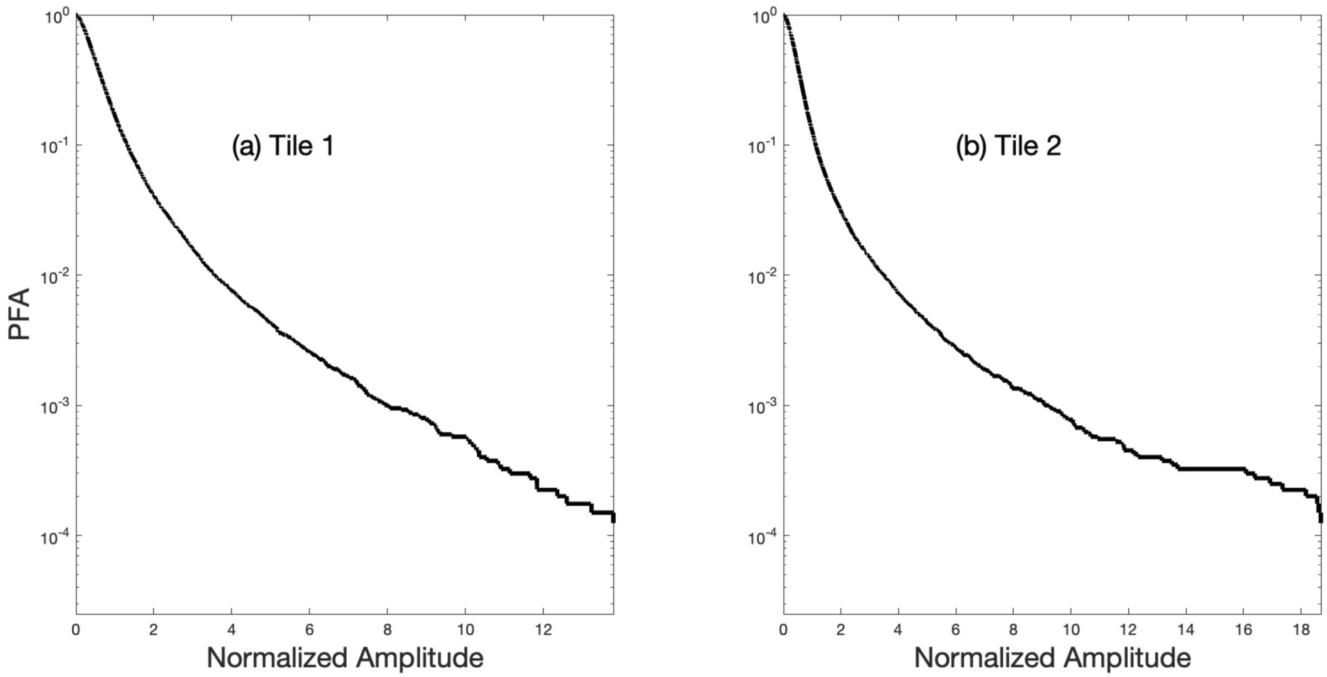

FIGURE 4 The empirical probability of false alarm, $P_F(A) = 1 - P(A)$ for tile 1 (a), and tile 2 (b), as a function of normalised area, A.

dataset [29], a Rayleigh model was used for the low amplitude portions of the image, and K distributions were used for the rest of the amplitude range. A drawback of that work was that the Rayleigh model did not provide a good fit to the lowest amplitudes when the number of components was small, and examination of the data revealed that the darkest portions of the image were not shadows. Note that even in regions of shadow, grating lobes and other imaging artefacts [30] may be present and result in non-Rayleigh statistics.

The K distribution has been shown to be an appropriate model for a wide range of statistically stationary seabed types [13, 27, 28]. It has the additional advantage of having a meaningful physical interpretation resulting from the modulation of the power of a Rayleigh distribution [31–34] or a finite number of scatterers [28, 35]. The K distribution contains the Rayleigh model as a sub-member, so if the lowest amplitude portions of the image are truly composed of additive Gaussian noise, the mixture model will have a high shape parameter for that component and accurately represent a Rayleigh distribution. Compared to other more flexible models (such as the generalised Pareto), it does not have pathological properties such as infinite moments [14, 36, 37].

A K distribution for the scattered field amplitude has the following form [28]:

$$p_K(A|\lambda,\alpha) = \frac{4}{\sqrt{\lambda}\Gamma(\alpha)}\left(\frac{A}{\sqrt{\lambda}}\right)^{\alpha} K_{\alpha-1}\left(\frac{2A}{\sqrt{\lambda}}\right), \qquad (1)$$

where $\Gamma(\cdot)$ is the gamma function, $K_\nu(x)$ is the modified Bessel function of the second kind of order ν with argument x, λ is the scale parameter of the K distribution, and α is the shape parameter. The expected value of the intensity ($I = A^2$) for this model is $E[A^2] = \alpha\lambda$ and is denoted σ. In the results below, the K distribution is parameterised using the pair (σ, α) rather than the scale and shape parameters. When used in a mixture distribution, the parameters have subscripts m to denote which component the parameters correspond to; that is, σ_m and α_m.

Mixture models are formed by a weighted sum of individual pdf components. The physical meaning of this type of model is that every measurement in a population, or sample, can be identified with one of the M components. The weights of the distributions, w_m, are normalised such that $\sum w_m = 1$, and therefore the weights can be interpreted as the fraction of pixels corresponding to each component.

The form of the mixture models used here is as follows [38]:

$$p(A|\theta) = \sum_{m=1}^{M} w_m p_K(A|\sigma_m, \alpha_m), \qquad (2)$$

where M is the number of mixture components, θ is a vector of length $k = 3M - 1$, consisting of the parameters of the model w_m, σ_m and α_m. The weights are subject to the constraint $\sum_m w_m = 1$, and so the weights constitute $M - 1$ independent variables. The variance parameters are sorted from lowest to highest values ($\sigma_1 < \sigma_2 < \ldots < \sigma_M$). The parameters of this mixture model are found using the expectation-maximisation (EM) algorithm [38, 39]. This method maximises a slightly altered version of the log-likelihood for each component, but asymptotically maximises the likelihood function for mixture models [39]. An EM method for mixtures of two components for common clutter models is given in ref. [21], and a generalisation of that method is used in this work. We also differ from Abraham et al. [21] in that a finite-difference based

gradient method is used for the maximisation step for the K distributions, rather than using the moment-based method described in that work (which we have found to be unreliable).

Since the number of components that constitutes the environment is in general unknown, model selection techniques are used to pick M, the number of K distribution components. As M increases, it is better able to match the pdf of the measurement, but at the cost of more uncertainty per parameter [14, 40]. The joint likelihood function, $\ell(\theta|A)$, is a common metric for model-data fit and is a component of many of the model selection techniques described her. It is defined for N independent samples by the following equation:

$$\ell(\theta|A) = \prod_{n=1}^{N} p(A_n|\theta), \tag{3}$$

where A_n is the $n-th$ member of the ensemble, and $p(A_n|\theta) = \ell(\theta|A_n)$ is the likelihood function for an individual data sample. The parameter vector $\hat{\theta}$ that maximises ℓ is called the maximum likelihood estimate. It is common to work with the log-likelihood, $\mathscr{L} = \log(\ell)$, which is given as follows:

$$\mathscr{L}(\theta|A) = \sum_{n=1}^{N} \log(p(A_n|\theta)). \tag{4}$$

Later in this work, these quantities are abbreviated as $\mathscr{L}(\theta) = \mathscr{L}(\theta|A)$ for the joint log-likelihood, $\mathscr{L}_n(\theta) = \log(p(A_n|\theta))$ for the per-sample log-likelihood, and $\ell_n(\theta) = \ell(\theta|A_n)$ for the per-sample likelihood. Other methods for estimating parameters exist, such as the method of moments used by Abraham and others for the K distribution [21, 41]. However, the method of moments is not guaranteed to be unbiased or reach minimum variance, and we prefer use of the maximum likelihood method.[2]

More complex models typically result in a higher likelihood function, but it may be broader in parameter space (implying a large parameter variance and larger uncertainty), so $\mathscr{L}$ alone cannot be used as a basis on which a decision about M can be made. There exist various model selection techniques, some of which are based on Bayesian concepts [40] and are called "information criteria", with smaller values preferred. Two simple metrics are based on the log-likelihood function, but with an additional penalty that depends on the number of parameters, k.

The Akaike information criterion [40, 42, 43], is a simple penalisation of the log-likelihood function evaluated at the maximum likelihood estimate. This form is based on an asymptotic approximation of the posterior probability density function (ppd) of parameters θ given the data, $p(\theta|A)$ and is given by

$$AIC = -2\mathscr{L}(\hat{\theta}) + 2k. \tag{5}$$

[2]It may be possible that some function of the moments could be used to construct an unbiased minimum variance estimator, according to the Lehmann-Scheffe theorem [58].

This criterion is the simplest explored here.

The Bayesian information criterion is defined as follows [44],

$$BIC = -2\mathscr{L}\hat{\theta} + \log(N)k, \tag{6}$$

where $\hat{\theta}$ is the maximum likelihood estimate of θ. This criterion results from an asymptotic (i.e. large N) Gaussian approximation of the marginal likelihood (also called the evidence). The penalty to the log-likelihood is therefore a function of the number of parameters, which is due to the ppd becoming narrower as N becomes asymptotically large. It has been criticised for inappropriately penalising complex models and for improperly treating the prior distribution in the approximation of the ppd [45].

Both the AIC and BIC are based on the log-likelihood function evaluated at $\hat{\theta}$ and therefore may be sensitive to the numerical estimate. Additionally, point estimates do not contain any information about the ppd as a whole (since they are based on asymptotic estimates) and are thus insensitive to whether the likelihood function is narrow or broad near $\hat{\theta}$. The final two model selection criteria are based on Monte-Carlo Markov chain (MCMC) samples drawn from the posterior distribution of the parameters given the data. The details behind MCMC sampling are beyond the scope of this paper, and the reader is directed to Gelman et al., [40] for a more in-depth general treatment, and Dettmer et al., [46] for practical applicaions of MCMC techniques in ocean acoustics. In basic terms, the ppd $p(\theta|A)$ is the pdf of the parameters of a model θ given the data A. Bayes' theorem states that

$$p(\theta|A) = \frac{p(A|\theta)p(\theta)}{p(A)} \tag{7}$$

where $p(A|\theta)$ is the joint likelihood function, $p(\theta)$ is the prior pdf of the models (incorporating anything known about the model parameters that is not inferred from the measured data), and the probability of the data, $p(A)$, in the denominator ensures that integrating the ppd yields unity. In this work, uniform priors are used. The mixture weights are constrained to lie between 0 and 1, and the property $\sum w_m = 1$ is enforced by performing the search on the first $M-1$ values of w_m, and computing $w_M = 1 - \sum_{m=1}^{M-1} w_m$. The K distribution parameters for the Mth component are unaffected by this equality constraint, and so are used in the Markov chain in the typical way. Additionally, the constraint $\sum_{m=1}^{M-1} w_m < 1$ is used as part of the prior to ensure that a model with larger M is not strictly equivalent to models with smaller M, which can happen when one of the weights is zero and can lead to a singular model. The variance and shape parameters were each constrained to the intervals $\sigma_m \in [0, 400]$ and $\alpha_m \in [0, 40]$, respectively. These bounds on the K distribution parameters are set to save computational resources. A K distribution with a shape parameter that tends to infinity approaches a Rayleigh distribution, and very small changes are observed for $\nu > 40$. The

constraint on σ is used to prevent this parameter from being much greater than the largest intensity sample in either data set.

To compute draws of θ from its ppd, the popular Goodman–Weare (GW) affine MCMC sampler [47] is used here, using an implementation in MATLAB [48]. The GW algorithm samples directly from $p(\theta|A)$ given a functional form for the likelihood function and the priors, by using an ensemble of walkers of size N_W that exchange information at each step. MCMC samples drawn from the posterior distribution are denoted by θ_l, where l indexes the sample number and ranges from 1 to L with L being the number of MCMC samples. The initial portion of the MCMC draws from each walker contains a burn-in period of increasing variance of each parameter before it plateaued. The variance was computed by estimating the sample variance over the ensemble of walkers for each MCMC step. The burn-in portion was about 2000 samples for each walker and was discarded. Since $20 \times k$ walkers were used, this means that between 200,000 and 560,000 samples were discarded. Before discarding the burn-in samples, each MCMC run had 1.5 million samples.

The deviance information criterion was proposed by Spiegelhalter and others [49, 50] as a more Bayesian version of the AIC [40]. The DIC is calculated using MCMC samples as follows:

$$DIC = -2\mathscr{L}\left(\hat{\theta}_{Bayes}\right) + 2p_{DIC}, \tag{8}$$

where

$$\hat{\theta}_{Bayes} = E(\theta|A) = \frac{1}{L}\sum_{l=1}^{L}\theta_l \tag{9}$$

is the posterior mean of the MCMC parameter estimates (calculated as the mean over l for each element of the parameter vector), and p_{DIC} is an estimate of the effective number of parameters. This quantity is computed by the following equation:

$$p_{DIC} = 2\left(\mathscr{L}\left(\hat{\theta}_{Bayes}\right) - \frac{1}{L}\sum_{l=1}^{L}\mathscr{L}(\theta_l)\right). \tag{10}$$

DIC appears to be a popular improvement over the point estimates used in AIC and BIC. It is more Bayesian–incorporating the posterior distribution in the penalty term and the posterior mean $\hat{\theta}_{Bayes}$ as an estimate of the parameters in the log-likelihood term. However, it has been criticised for not being fully Bayesian, and that the effective number of parameters p_{DIC} could become negative in certain cases [40, 50]. There exists an alternative version of p_{DIC} based on the variance, $p_{DIC_{alt}} = 2\mathrm{var}[\mathscr{L}(\theta_l)]$, computed as the posterior variance taken over the ensemble of MCMC draws. This results in an alternative formulation of the DIC [40],

$$DIC_{alt} = -2\mathscr{L}\left(\hat{\theta}_{Bayes}\right) + 2p_{DIC_{alt}}. \tag{11}$$

Both DIC and DIC_{alt} are used here as potential model selection criteria.

The last information criterion considered here is an update to the DIC, termed the Watanabe-Akaike information criterion [51, 52] (although it was initially referred to as the "widely-applicable information criterion" by Watanabe). It is motivated by an asymptotic approximation to leave-one-out cross validation and also relies on MCMC samples of θ drawn from $p(\theta|A)$. The WAIC uses the likelihood function calculated for each data sample; that is, instead of a point estimate of the parameters it utilises their posterior distribution. The criterion is calculated as follows:

$$WAIC = -2lppd + 2p_{WAIC}, \tag{12}$$

where the log pointwise predictive density, $lppd$, is computed as follows [53]:

$$lppd = \sum_{n=1}^{N}\log\left(\frac{1}{L}\sum_{l=1}^{L}\ell_n(\theta_l)\right). \tag{13}$$

The effective number of parameters p_{WAIC} is the sum over data samples of the posterior variance of the log-likelihood of an individual sample,

$$p_{WAIC} = \sum_{n=1}^{N} V_n, \tag{14}$$

where

$$V_n = \frac{1}{L-1}\sum_{l=1}^{L}\left(\mathscr{L}_n(\theta_l) - \overline{\mathscr{L}_n(\theta_l)}\right)^2 \tag{15}$$

is the posterior variance for sample n for a given model, and $\overline{\mathscr{L}_n(\theta_l)} = (1/L)\sum_{l=1}^{L}\mathscr{L}_n(\theta_l)$ [53]. WAIC can be considered fully Bayesian. Unlike the other information criteria presented here it can evaluate predictions used for new data in a Bayesian context and tends to be more reliable also for complex models [40].

In the radar and sonar communities, there are other commonly used goodness of fit criteria, that express how well a model fits a data set. The KS test [54] is a comparison between the empirical cdf and the cdf produced by a model. One of metrics of this test is the p-value, which is related to hypothesis rejection. If the p-value is larger than some threshold (typically 0.05), then the hypothesis that some proposed distribution is the correct choice is not rejected. In this framework, higher p-values give an indication of better model fit and the model is not rejected, and smaller ones result in a given model being rejected. Other statistical tests are designed for cases in which the lower or upper part of the data are desired to be weighted more strongly. In sonar and radar measurements, false detections tend to arise from the large amplitude portion of the pdf, and therefore it is

common to use the AD test statistic modified to emphasise the upper tails [55]. p-values from both the KS test and modified upper-tail AD test are presented in this work and denoted p_{KS} and p_{AD}, respectively. Note that the KS and AD tests are not model selection techniques, like the information criteria described earlier but rather are designed to reject models based on some significance level [54].

4 | RESULTS AND DISCUSSION

4.1 | Model-data comparison

K mixture models were fit to the amplitude data shown in Figures 2 and 3, using M varying between 2 and 5 components. Also included are a single K distribution, and the Rayleigh distribution for reference. The model-data fit is shown graphically in Figure 5 for both tiles in terms of the $\log(P_F(A))$.

Model-data fits for both tiles are poor for both the Rayleigh and single K distributions. This behaviour is expected because the images in Figures 2 and 3 contain clusters of very high amplitude pixels, indicating that several different scattering mechanisms are responsible for the scattered field. Next, the K mixture with $M = 2$ (KM2 in Figure 5a,b) fit extremely poorly. The low amplitude region is modelled well, but above a normalised amplitude of 3 or 4 in tile 1, the KM2 and data curves depart quite severely. As the number of components is increased, the model-data fit becomes remarkably good, for both tiles. A mixture of three K distributions fits the tails quite well, and increasing the number of components does not seem to noticeably improve the fit. For tile 2, the $M = 4$ and 5 case (KM4 and KM5 in the legend) provides a slight improvement near the very high amplitude tails, compared with $M = 3$. Even with this slight improvement for models with $M > 3$, the authors own intuition suggests that $M = 3$ is the preferred model. This assessment is compared to the automated model selection results in 4.3.

4.2 | MCMC marginal PDFs

Given MCMC draws from the posterior, the two-dimensional marginal distributions for each parameter combination and the one-dimensional marginal distributions for each parameter (termed a "corner plot") can be used to visually inspect the parameter correlations. For models with low uncertainty in their parameter estimates, the widths of the parameters will be small and uncorrelated with other parameters. In the asymptotic case with a non-singular statistical model, and a large data sample, these tend towards bi-variate Gaussian distributions [40], but in the finite data case they can be non-Gaussian. The marginal distributions are constructed using a two-dimensional kernel smoothed density pdf estimator using the MCMC samples and a one-dimensional histogram, as implemented in ref. [48].

In Figure 6, the marginal pdfs for tile 1 are plotted, with the same values of M as in the pfa plot in Figure 5. The marginals for tile 2 are plotted in Figure 7. The simplest mixture model, $M = 2$ is plotted in the upper left of each plot. The 1D marginals (on the diagonal) appear to be approximately Gaussian, as do the 2D marginals in the off-diagonal subfigures. This behaviour indicates that the $M = 2$ model is statistically very well behaved, although it does not fit the data well. The marginal ppd of w_1 and w_2 is a single line with slope of negative unity, since the constraint $w_2 = 1 - w_1$ is enforced. Next, the $M = 3$ models are plotted in the upper right of each figure. The 1D marginals deviate more from a Gaussian appearance and have significant skew. In particular, σ_3, and

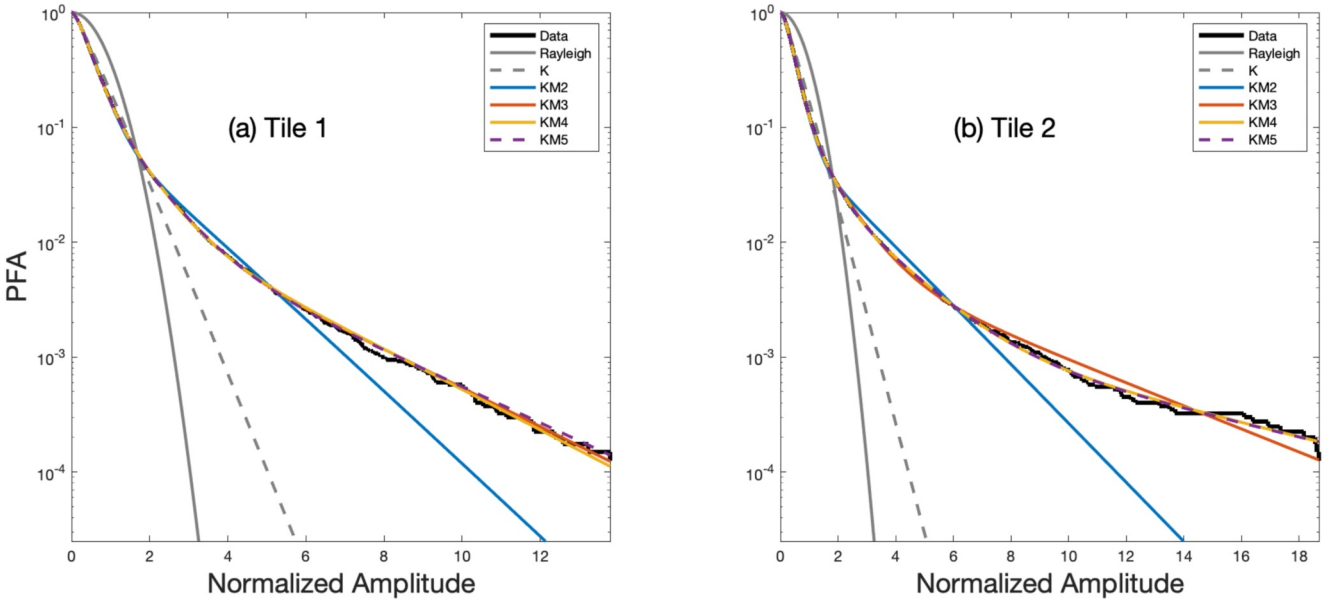

FIGURE 5 The probability of false alarm for the data from tile 1 (a) and 2 (b), compared to the various mixture models explored here.

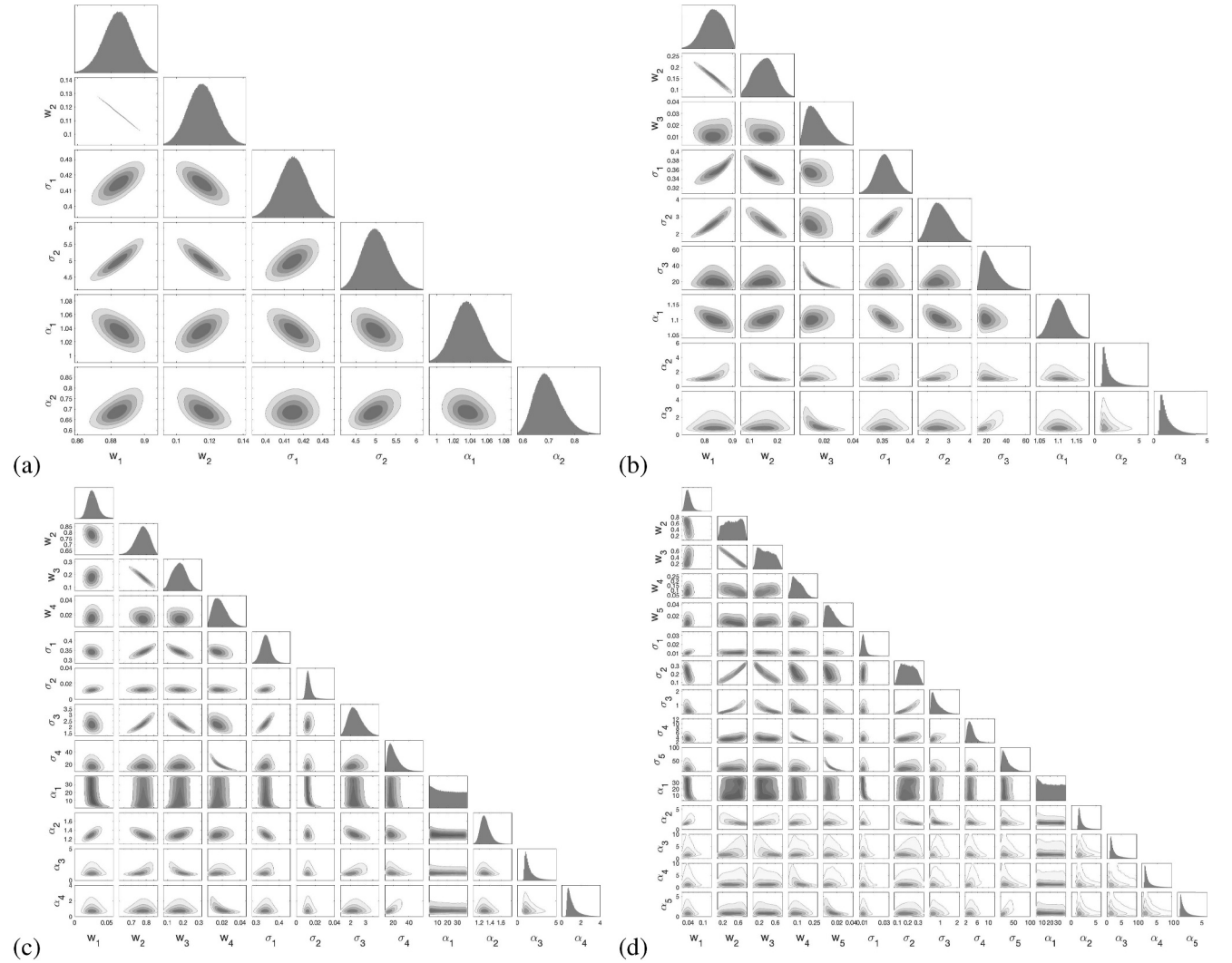

FIGURE 6 One- and two-dimensional marginal posterior probability distributions of the mixture model parameters from tile 1 resulting from MCMC simulations of (a) two K components, (b) three K components, (c) four K components, and (d) five K components. The posterior pdfs of the two component model appear to be Gaussian, and the shapes of the marginal pdfs tend to become less Gaussian and more skewed as the number of components is increased.

especially α_2 and α_3 appear skewed, meaning that there is more uncertainty in the estimates of these parameters.

In the $M = 4$ case for tile 1, the behaviour is broadly similar to the $M = 3$ case, except that the marginal distribution for α_1 (the shape parameter of the mixture component with the lowest amplitude) is almost uniformly distributed. For tile 2, the $M = 4$ marginal distribution has a similar near uniformity for α_1, but also a very large variance of σ_4. The parameters of σ_2 and σ_3 for tile 2 have a bi-modal behaviour in the 1D and 2D marginals, in addition to having a large variance. These parameters indicate that at least part of the information about the parameters of the mixture is highly uncertain. For $M = 5$ in the lower right corner of Figures 6 and 7, the same uniformity is seen in α_1. For tile 2, σ_5 is nearly uniform, while for tile 1 it is much more peaked, although it still has a large variance. In tile 1 and $M = 5$, marginal pdfs for the weights (w_2 and w_3), multiple maxima are evident, indicating that the model uncertainty is not narrow enough to distinguish each component from each other. For tile 2, the pdfs appear unimodal, but are quite wide. This behaviour is possible with singular statistical models [52]. A consequence of this uncertainty is that mixtures with large numbers of components seem to be statistically poorly behaved. Of the models with well-behaved marginal ppds, only the $M = 3$ model performs well in terms of model-data fit, as seen in Figure 5.

4.3 | Model selection results

As model section criteria, the log-likelihood ($\mathscr{L}$), *BIC*, *AIC*, *DIC*, DIC_{alt}, *WAIC*, and p values from the KS and AD tests are shown for both tiles in Table 1. For each tile, the best performing model in each column is shown in boldface. The maximum $\mathscr{L}$ value for both tiles is the five component model. This result is expected, since more complex models in general provide a better fit. The simplest information criterion, *AIC*,

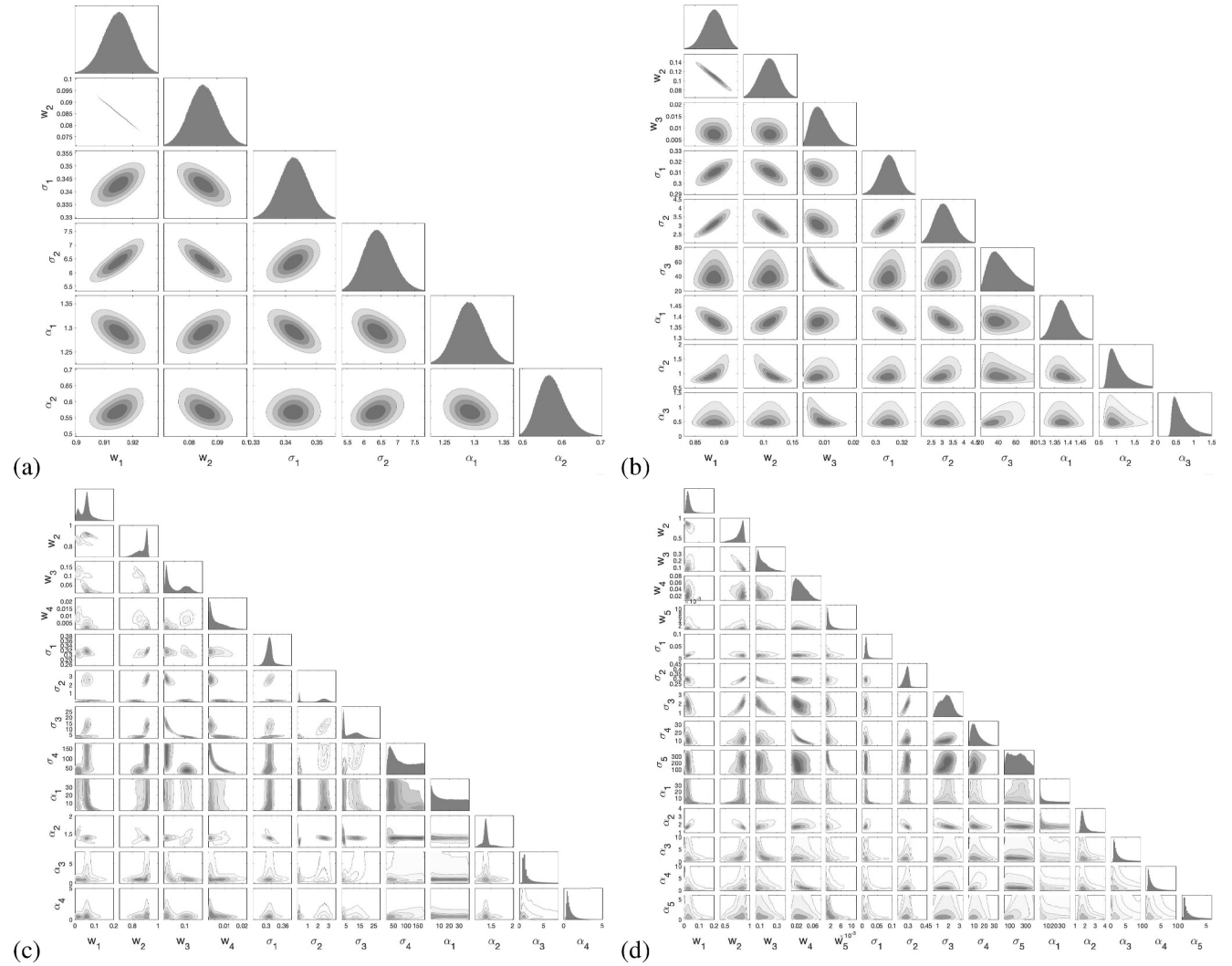

FIGURE 7 One- and two-dimensional marginal posterior probability distributions of the mixture model parameters from tile 2 resulting from MCMC simulations of (a) two K components, (b) three K components, (c) four K components, and (d) five K components. Many of the same conclusions as the corner plot from tile 1 (Figure 6) can be drawn here as well.

TABLE 1 Model selection results for both image tiles, including the parameters of $\mathscr{L}$, *AIC*, *BIC*, *DIC*, DIC_{alt}, *WAIC*, p_{KS}, and p_{AD}.

Data	Model	$\mathscr{L}$	*AIC*	*BIC*	*DIC*	DIC_{alt}	*WAIC*	p_{KS}	p_{AD}
Tile 1	KM2	−12,271	24,552	24,595	24,539	29,323	24,553	0.00	0.00
	KM3	−12,211	24,438	**24,506**	24,400	24,483	24,438	0.65	0.97
	KM4	−12,200	24,423	24,517	24,381	**24,478**	24,423	0.69	0.93
	KM5	**−12,195**	**24,419**	24,539	**23,762**	25,075	**24,417**	**0.99**	**0.99**
Tile 2	KM2	−4752	9513	9556	9516	15,709	9515	0.00	0.00
	KM3	−4655	9327	**9395**	9321	**9337**	9327	0.53	0.74
	KM4	−4648	9317	9412	**8934**	9743	9321	0.89	0.94
	KM5	**−4641**	**9311**	9431	9125	9506	**9311**	**0.99**	**0.99**

Note: In each column, the best value is shown in bold.

has the lowest value for the five component model, in agreement with the maximum likelihood. For both tiles, the smallest *BIC* is for the three component model. The penalty for model complexity (quantified in the *BIC* as the number of parameters) is more severe than for *AIC*, since there are 40,000 data samples in each tile. If $\log(N) > 2$ (it is approximately 10.6 in this case), then the *BIC* will be smaller than the *AIC* for the same model. The *BIC* seems to give a very simple trade off

between model complexity and accuracy that agrees with the intuition described in the interpretation of the pfa plots in Figure 5, and the interpretation of the images (Figures 2 and 3) in Section 2, its theoretical drawbacks non-withstanding [45]. The *DIC* selected best model were $M = 5$ for tile 1 and $M = 4$ for tile 2. Using the variance-based DIC_{alt} gave different results, $M = 4$ for tile 1, and $M = 3$ for tile 2. In both of these cases, DIC_{alt} selected a simpler model than *DIC* and seems to be slightly more biased in favour of simple models, at least for this dataset. Both *DIC* and DIC_{alt} favour either the same or slightly more complex models as *BIC*. *WAIC*, in the last column of Table 1, favours the most complex model $M = 5$, for both tiles.

Although *WAIC* is a commonly used model selection metric, it seems to favour model-data fit more than penalising complexity, for this dataset. The WAIC is not a simple penalisation of the log-likelihood but is motivated by leave-one-out cross validation [52, 53]. Vehtari et al. [53] suggest that if any of the variance terms V_n from Equation (15) are greater than 0.4, then the estimate of p_{waic} is unreliable. For all values of M explored here, two of the V_n terms resulted in values greater than 0.4, with the majority being close to 0.1 or 0.05. The data samples that crossed above the 0.4 threshold were the two highest amplitudes in the ensemble for both tiles. It makes sense that high amplitude outlier data samples were poorly fit by the model (and had large variance) since these samples are few in number and are not represented well by the joint likelihood function that is used to perform the MCMC sampling [47].

In terms of the p values, for both statistical tests, the values increase as the number of terms is increased, as shown in the two rightmost columns of Table 1. The only model that is rejected is the $M = 2$ case, which has a p value close to machine precision for both the KS and AD tests. Interestingly, the $M = 3$ case for tile 1 has a higher p_{AD} than $M = 4$, but $M = 5$ has the largest p_{AD} value. This can be explained by the fact that most of the error between the data and model occurs for normalised amplitudes between 0.1 and 1, which is far from the tails. Since the AD test emphasises the tails, the largest errors are discounted for $M = 3$. The trends in the p-values show that as the complexity of the model grows, the better the fit to the data is obtained. This trend is expected because the KS and AD statistical tests are designed to reject hypotheses, rather than choose between many alternative models and thus do not have any penalty for complexity.

Assuming the intuition of the authors that $M = 3$ is the best balance between complexity and accuracy is correct, then *BIC* is the best model selection metric (in this scenario). All the other information criteria favour more complex models. The $M = 4$ and $M = 5$ models produced a large amount of parameter uncertainty given the results of the *MCMC* simulations in Figures 6 and 7. The $M = 3$ choice is supported by examining the different physical effects that produce the texture shown in Figures 2 and 3. They are composed of medium intensity scattering from (on average) horizontal surfaces, low intensity scattering from portions of the surface that are tilted away from the sonar or in shadow, and high-amplitude portions due to scattering from vertical facets resulting from glacial plucking [12, 56], dropstones, or micro cracks. It is possible that the more complex models can differentiate between different high intensity scatterers. These are composed of, fractured surfaces, vertical facets, and small cracks, and are represented in the tails in Figure 5b at amplitudes greater than 10. However, it is difficult to see the benefit of differentiating these high amplitude scatterers from examination of the pfa curves. Therefore, we suggest that the number of components shall be three.

The only information criterion that reliably produces $M = 3$ as the best model is the *BIC*. Since the *BIC* is based on a single maximum likelihood parameter estimate, it may be numerically unstable. If a single point estimate is deemed unsuitable for a particular statistical modelling application (such as remote sensing or target detection performance prediction), then the ensemble-based DIC_{alt} could be used as a more robust estimate (since it tends to favour simpler models).

We also note that in previous work by the authors [29], use of a Rayleigh distribution for the lowest amplitude component required four components to achieve the same degree of fit as a mixture of three K distributions achieved here. Notably the smallest *BIC* for the $R + KN$ model explored in ref. [29] (recalculated here using the decimation factor from the present manuscript) was for four components (one Rayleigh and three K distributions) and had a value of 24,517 for tile 1 (which is the same as a four component mixture of only K distributions), and a value of 9421 for tile 2 (which is larger than the three component mixture of K distributions). This difference in model selection criteria highlights the importance of selecting an appropriate statistical model for the mixture components. The use of a Rayleigh model for the lowest amplitude scatterers in the image did not have enough complexity for the *BIC* to be able to match an intuitive partition of the image into different scatterers. A Rayleigh distribution would be appropriate for shadows, but the darkest areas in the image may actually be very low amplitude scattering from a rock surface tilted away. Even if shadows are present, then sidelobes and grating lobes from the SAS point spread function [30, 57] could cause a departure from a Rayleigh distribution in these areas.

4.4 | Generalisation to the complete SAS image

The analysis presented in the previous subsections was restricted to two 12 x 12 m tiles without information on how the model selection results generalise to other types. To analyse this, we have performed fits of the mixture models on 300 × 300 pixel tiles (6 x 6m) on the entire SAS image in Figure 1 and used the two simplest model selection criteria, AIC and BIC, to determine the suitable number of components for each patch. Since simple textures were included, the choice of a single mixture distribution was included here as

well. DIC and WAIC were not analysed due to the enormous computational complexity of performing MCMC sampling for 260 patches, each with approximately 10^4 independent samples.

The image contains a variety of seabed textures. By visual inspection some of the textures appear to result from a single component, and some have clustered regions of high or low amplitude. The model selection results for AIC are shown in Figure 8, and for BIC in Figure 9, where the sonar image intensity and model selection results have been encoded in the hue-saturation-value (HSV) colour space. The colours for the preferred model per patch have been converted from the red-green-blue colorspace to HSV and are used for the hue and saturation channels. For the value channel, we use the log intensity scaled between 0 and 1.

In the AIC results in Figure 8, only four tiles were chosen to have a single K-distribution component, even though an appreciable portion of the image is taken up by uniform sediment. Many of the sedimented tiles have $M = 2$ or $M = 3$ as the best model, even though they do not appear to contain much texture or structure. The rocky textures are chosen to have $M = 4$ and sometimes $M = 5$ as the best component.

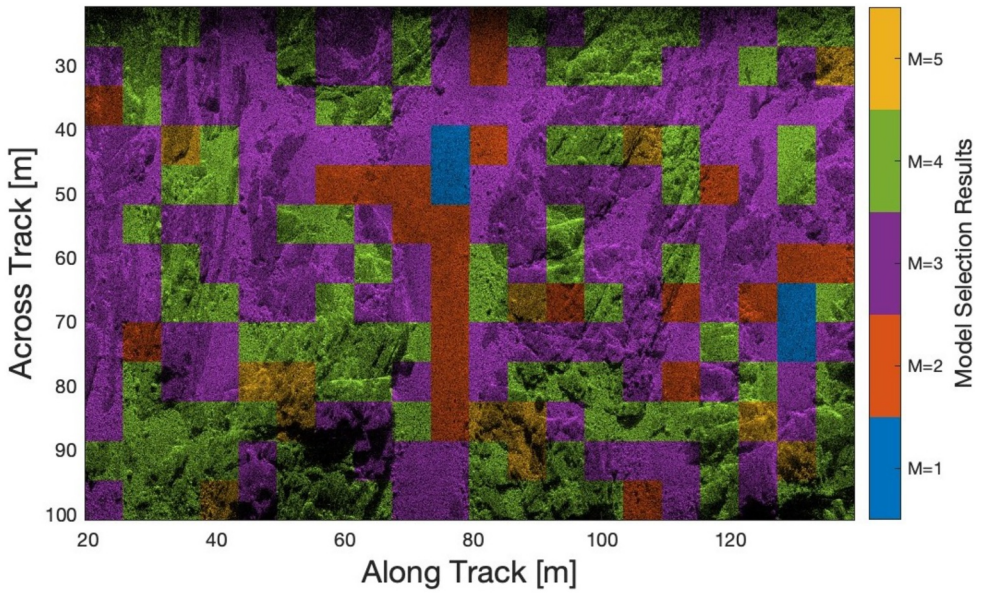

FIGURE 8 Model selection results using AIC for 300 × 300 pixel (6 × 6 m) patches from the entire SAS image in Figure 1. The colorbar indicates the chosen mixture components between a single component, $M = 1$, and five components.

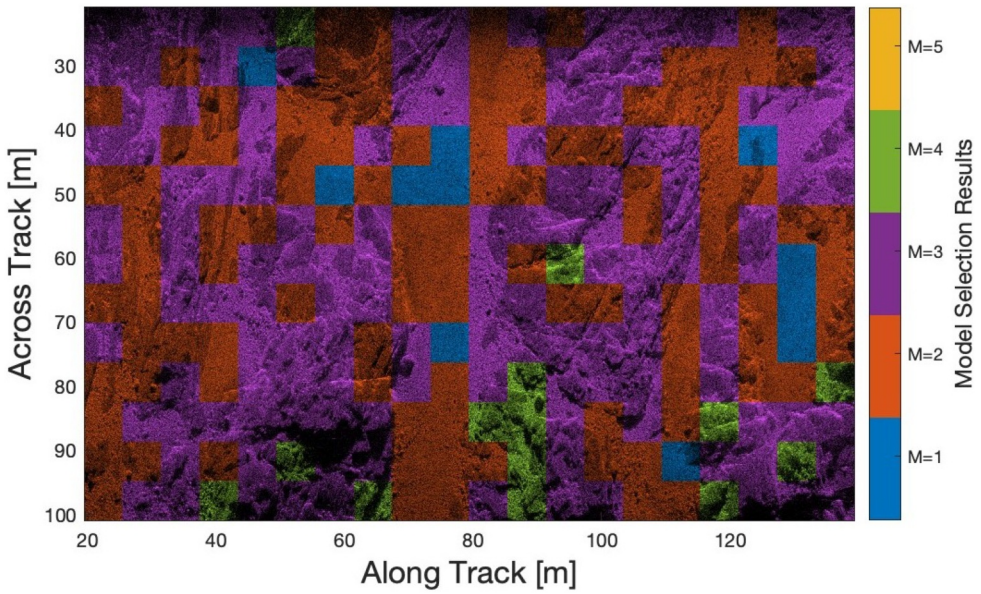

FIGURE 9 Model selection results using BIC for 300 × 300 pixel (6 × 6 m) patches from the entire SAS image in Figure 1. The colorbar indicates the chosen mixture components between a single component, $M = 1$, and five components.

These trends agree with the analysis of tiles 1 and 2 earlier, where AIC tended to prefer more complex models than intuition suggested.

Regarding BIC results presented in Figure 9, 11 tiles were chosen to be well modelled by a single component for the sedimented areas. Compared with four tiles for the AIC, this indicates that the BIC is better at selecting appropriately simple models for uniform textured images. The rest of the sedimented areas has the best BIC for $M = 2$. Many of these tiles contain both rocks and sediment, so the selection of $M = 2$ is appropriate. The rocky areas are mostly modelled with $M = 3$ and occasionally $M = 4$. The $M = 4$ areas tend to have both very bright facets and extremely deep shadows, so the decision or $M = 4$ may be warranted, even though our analysis of the two tiles in the previous two subsections indicated that $M = 3$ was an appropriate model. The two example tiles did not have extremely deep shadows. Overall, BIC tends to favour simpler models when applied to all 260 tiles in the image, which shows that the trends examined for the two tiles do in fact generalise well to the entire image.

5 | CONCLUSION

We presented a statistical model for SAS images of complex, non-stationary, rocky seafloors. This model consisted of a sum of an unknown number of K distributions. The number of K distributions was selected using four different model selection criteria. The results of these criteria were compared against the authors' interpretation of the SAS images, and the comparison between the empirical and model probability of false alarm. We concluded that the *BIC* provided the best balance between model complexity and goodness of fit, since it has a very severe penalty for complex models (at least when the number of samples is large) and is computationally simple to compute. Interesting areas for future work would be applying this methodology to other examples of rock outcrops, other seafloor textures and use with other sensors, or the same sensor in an environment with different noise levels. This method could also be used in the future to select models and partition images into different scatterer classes for biological or geological remote sensing.

AUTHOR CONTRIBUTIONS

Derek R. Olson: Conceptualisation; formal analysis; funding acquisition; investigation; methodology; software; visualisation; writing—original draft; writing—review and editing. **Marc Geilhufe**: Conceptualization, data curation; methodology; writing—original draft; writing—review and editing.

ACKNOWLEDGEMENTS

The authors thank FFI's HUGIN-HUS operator group for collecting the data used in this study. This research was supported by grant number N00014-23-WX01149 from the United States Office of Naval Research under the Young Investigator Program. The views expressed in this document are those of the authors and do not reflect the official policy or position of the U.S. Department of Defense or U.S. Government.

CONFLICT OF INTEREST STATEMENT

The authors have no conflicts of interest to declare.

DATA AVAILABILITY STATEMENT

Research data are not shared.

ORCID

Derek R. Olson https://orcid.org/0000-0002-7928-0468
Marc Geilhufe https://orcid.org/0000-0002-2775-0544

REFERENCES

1. Williams, D.P.: Fast unsupervised seafloor characterization in sonar imagery using lacunarity. IEEE Trans. Geosci. Rem. Sens. 53(11), 6022–6034 (2015). https://doi.org/10.1109/tgrs.2015.2431322
2. Quinn, R. In: Ford, B., Hamilton, D.L., Catsambis, A. (eds.) Acoustic Remote Sensing in Maritime Archaeology. Oxford University Press (2012)
3. Galusha, A.P., et al.: A fast target detection algorithm for underwater synthetic aperture sonar imagery. In: Isaacs, J.C., Bishop, S.S. (eds.) Detection and Sensing of Mines, Explosive Objects, and Obscured Targets XXIII. SPIE (2018)
4. Zare, A., et al.: Possibilistic fuzzy local information c-means for sonar image segmentation. In: 2017 IEEE Symposium Series on Computational Intelligence (SSCI). IEEE (2017)
5. Peeples, J., Xu, W., Zare, A.: Histogram layers for texture analysis. IEEE Transactions on Artificial Intelligence 3(4), 541–552 (2022). https://doi.org/10.1109/tai.2021.3135804
6. Blondel, P., Sichi, O.G.: Textural analyses of multibeam sonar imagery from Stanton Banks, Northern Ireland continental shelf. Appl. Acoust. 70(10), 1288–1297 (2009). https://doi.org/10.1016/j.apacoust.2008.07.015
7. Stewart, W.K., et al.: Quantitative seafloor characterization using a bathymetric sidescan sonar. IEEE J. Ocean. Eng. 19(4), 599–610 (1994). https://doi.org/10.1109/48.338396
8. Weszka, J.S., Dyer, C.R., Rosenfeld, A.: A comparative study of texture measures for terrain classification. IEEE Trans. Syst. Man. Cybern. SMC-6(4), 269–285 (1976). https://doi.org/10.1109/tsmc.1976.5408777
9. Williams, D.P.: Unsupervised seabed segmentation of synthetic aperture sonar imagery via wavelet features and spectral clustering. In: 2009 16th IEEE International Conference on Image Processing (ICIP). IEEE (2009)
10. Haralick, R.M., Shanmugam, K., Dinstein, I.: Textural features for image classification. IEEE Trans. Syst. Man. Cybern. SMC-3(6), 610–621 (1973). https://doi.org/10.1109/tsmc.1973.4309314
11. Jackson, D.R., Richardson, M.D.: High-Frequency Seafloor Acoustics. Springer, New York (2007)
12. Olson, D.R., Lyons, A.P., Sæbø, T.O.: Measurements of high-frequency acoustic scattering from glacially eroded rock outcrops. J. Acoust. Soc. Am. 139(4), 1833–1847 (2016). https://doi.org/10.1121/1.4945589
13. Lyons, A.P., Abraham, D.A.: Statistical characterization of high-frequency shallow-water seafloor backscatter. J. Acoust. Soc. Am. 106(3), 1307–1315 (1999). https://doi.org/10.1121/1.428034
14. Olson, D.R., et al.: Scattering statistics of rock outcrops: model-data comparisons and bayesian inference using mixture distributions. J. Acoust. Soc. Am. 145(2), 761–774 (2019). https://doi.org/10.1121/1.5089892
15. Gauss, R.C., et al.: Moment-based method to statistically categorize rock outcrops based on their topographical features. In: OCEANS 2015 - MTS/IEEE Washington, pp. 1–5 (2015)
16. Xenaki, A., Pailhas, Y., Sabatini, R.: Persistent scatterers detection from phase statistics of multi-view synthetic aperture sonar imaging. J. Acoust. Soc. Am. 148(4 Supplement), 2792 (2020). https://doi.org/10.1121/1.5147771

17. Just, D., Bamler, R.: Phase statistics of interferograms with applications to synthetic aperture radar. Appl. Opt. 33(20), 4361 (1994). https://doi.org/10.1364/ao.33.004361
18. Gruber, J.A., Olson, D.R.: Scattering measurements of rocky seafloors using a split-beam echosounder. JASA Express Letters 4(2) (2024). https://doi.org/10.1121/10.0024755
19. Griffiths, D.A.: Maximum likelihood estimation for the beta-binomial distribution and an application to the household distribution of the total number of cases of a disease. Biometrics 29(4), 637 (1973). https://doi.org/10.2307/2529131
20. Ueki, M.: Beta-negative binomial nonlinear spatio-temporal random effects modeling of covid-19 case counts in Japan. J. Appl. Stat. 50(7), 1650–1663 (2022). https://doi.org/10.1080/02664763.2022.2064439
21. Abraham, D., Gelb, J., Oldag, A.: Background and clutter mixture distributions for active sonar statistics. IEEE J. Ocean. Eng. 36(2), 231–247 (2011). https://doi.org/10.1109/joe.2010.2102150
22. Gelb, J.M., Heath, R.E., Tipple, G.L.: Statistics of distinct clutter classes in midfrequency active sonar. IEEE J. Ocean. Eng. 35(2), 220–229 (2010). https://doi.org/10.1109/joe.2009.2031547
23. Oliver, C.: Correlated K-distributed clutter models. Opt. Acta: International Journal of Optics 32(12), 1515–1547 (1985). https://doi.org/10.1080/713821683
24. Oliver, C., Quegan, S.: Understanding Synthetic Aperture Radar Images, Ser. SciTech Radar and Defense Series. SciTech Publ (2004). [Online]. https://books.google.com/books?id=IeGKe40S77AC
25. Chotiros, N.P., et al.: Acoustic backscattering at low grazing angles from the ocean bottom. Part II. Statistical characteristics of bottom backscatter at a shallow water site. J. Acoust. Soc. Am. 77(3), 975–982 (1985). https://doi.org/10.1121/1.392065
26. Gensane, M.: A statistical study of acoustic signals backscattered from the sea bottom. IEEE J. Ocean. Eng. 14(1), 84–93 (1989). https://doi.org/10.1109/48.16818
27. Abraham, D.A., Lyons, A.P.: Reverberation envelope statistics and their dependence on sonar bandwidth and scattering patch size. IEEE J. Ocean. Eng. 29(1), 126–137 (2004). https://doi.org/10.1109/joe.2004.824039
28. Abraham, D., Lyons, A.: Novel physical interpretations of K-distributed reverberation. IEEE J. Ocean. Eng. 27(4), 800–813 (2002). https://doi.org/10.1109/joe.2002.804324
29. Olson, D.R., Geilhufe, M.: Comparison of model selection techniques for seafloor scattering statistics. In: Proceedings of Synthetic Aperture Sonar and Synthetic Aperture Radar 2023. Lerici, Italy (2023)
30. Cook, D., Brown, D.: Analysis of phase error effects on stripmap SAS. IEEE J. Ocean. Eng. 34(3), 250–261 (2009). https://doi.org/10.1109/joe.2007.907935
31. Ward, K.D.: Compound representation of high resolution sea clutter. Electron. Lett. 17(16), 561–563 (1981). https://doi.org/10.1049/el:19810394
32. Lyons, A.P., Olson, D.R., Hansen, R.E.: Modeling the effect of random roughness on synthetic aperture sonar image statistics. J. Acoust. Soc. Am. 152(3), 1363–1374 (2022). https://doi.org/10.1121/10.0013837
33. Lyons, A.P., Olson, D.R., Hansen, R.E.: Quantifying the effect of random seafloor roughness on high-frequency synthetic aperture sonar image statistics. In: Proc. Inst. Acoust. Conf.: Acoustic and Environmental Variability, Fluctuations and Coherence, Cambridge, UK (2016)
34. Olson, D.R., Lyons, A.P.: Resolution dependence of rough surface scattering using a power law roughness spectrum. J. Acoust. Soc. Am. 149(1), 28–48 (2021). https://doi.org/10.1121/10.0002974
35. Jakeman, E., Tough, R.: Non-Gaussian models for the statistics of scattered waves. Adv. Phys. 37(5), 471–529 (1988). https://doi.org/10.1080/00018738800101419
36. La Cour, B.R.: Statistical characterization of active sonar reverberation using extreme value theory. IEEE J. Ocean. Eng. 29(2), 310–316 (2004). https://doi.org/10.1109/joe.2004.826897
37. Barnard, T., Khan, F.: Statistical normalization of spherically invariant non-Gaussian clutter. IEEE J. Ocean. Eng. 29(2), 303–309 (2004). https://doi.org/10.1109/joe.2004.828204
38. Titterington, D.M., Smith, A.M., Makov, U.E.: Statistical Analysis of Finite Mixture Distributions. John Wily and Sons, Chichester (1985)
39. Dempster, A.P., Laird, N.M., Rubin, D.B.: Maximum likelihood from incomplete data via the EM algorithm. J. Roy. Stat. Soc. B 39(1), 1–38 (1977). https://doi.org/10.1111/j.2517-6161.1977.tb01600.x
40. Gelman, A., et al.: Bayesian Data Analysis. Chapman and Hall/CRC (2013)
41. Abraham, D.A., Lyons, A.P.: Reliable methods for estimating the K-distribution shape parameter. IEEE J. Ocean. Eng. 35(2), 288–302 (2010). https://doi.org/10.1109/joe.2009.2025645
42. Akaike, H.: Information theory and an extension of the maximum likelihood principle. In: Petrov, B.N., Csaki, F. (eds.) Porceedings of the Second International Symposium on Information Theory, pp. 267–281. Akademiai Kiadod, Budapest (1973)
43. Akaike, H.: A new look at the statistical model identification. IEEE Trans. Automat. Control 19(6), 716–723 (1974)
44. Schwarz, G.: Estimating the dimension of a model. Ann. Stat. 6(2) (1978)
45. Weakleim, D.L.: A critique of the Bayesian information criterion for model selection. Socio. Methods Res. 27(3), 359–397 (1999)
46. Dettmer, J., Dosso, S.E., Holland, C.W.: Model selection and bayesian inference for high-resolution seabed reflection inversion. J. Acoust. Soc. Am. 125(2), 706–716 (2009)
47. Goodman, J., Weare, J.: Ensemble samplers with affine invariance. Commun. Appl. Math. Comput. Sci. 5(1), 65–80 (2010)
48. Grinsted, A.: GWMCMC (2024). [Online]. https://github.com/grinsted/gwmcmc
49. Spiegelhalter, D.J., et al.: Bayesian measures of model complexity and fit. J. Roy. Stat. Soc. B Stat. Methodol. 64(4), 583–639 (2002)
50. Spiegelhalter, D.J., et al.: The deviance information criterion: 12 years on. J. Roy. Stat. Soc. B Stat. Methodol. 76(3), 485–493 (2014)
51. Watanabe, S.: A widely applicable Bayesian information criterion. J. Mach. Learn. Res. 14(1), 867–897 (2013)
52. Watanabe, S.: Asymptotic equivalence of bayes cross validation and widely applicable information criterion in singular learning theory. J. Mach. Learn. Res. 11(116), 3571–3594 (2010). [Online]. http://jmlr.org/papers/v11/watanabe10a.html
53. Vehtari, A., Gelman, A., Gabry, J.: Practical Bayesian model evaluation using leave-one-out cross-validation and waic. Stat. Comput. 27(5), 1413–1432 (2016)
54. Papoulis, A., Pillai, S.: Probability, Random Variables, and Stochastic Processes, 4th ed. McGraw-Hill (2002)
55. Sinclair, C., Spurr, B., Ahmad, M.: Modified Anderson darling test. Commun. Stat. Theor. Methods 19(10), 3677–3686 (1990)
56. Zoet, L., et al.: Accelerated subglacial erosion in response to stick-slip motion. Geology 41(2), 159–162 (2013)
57. Hansen, R.E.: Introduction to synthetic aperture sonar. In: Kolev, N. (ed.) Sonar Systems. InTech (2011). ch. 1
58. Lehmann, E.L., Scheffé, H.: Completeness, similar regions, and unbiased estimation. I. Sankhya: the Indian Journal of Statistics 10(4), 305–340 (1950). [Online]. https://www.jstor.org/stable/25048038

How to cite this article: Olson, D.R., Geilhufe, M.: Model selection techniques for seafloor scattering statistics in synthetic aperture sonar images of complex seafloors. IET Radar Sonar Navig. 1–13 (2024). https://doi.org/10.1049/rsn2.12608