A Decision Support System to Ease Operator Overload in Multibeam Passive Sonar

Iain Rice and David Lowe

Abstract—Creating human-informative signal processing systems for the underwater acoustic environment that do not generate operator cognitive saturation and overload is a major challenge. To alleviate cognitive operator overload, we present a visual analytics methodology in which multiple beam-formed sonar returns are mapped to an optimized 2-D visual representation, which preserves the relevant data structure. This representation alerts the operator as to which beams are likely to contain anomalous information by modeling a latent distribution of information for each beam. Sonar operators therefore focus their attention only on the surprising events. In addition to the principled visualization of high-dimensional uncertain data, the system quantifies anomalous information using a Fisher Information measure. Central to this process is the novel use of both signal and noise observation modeling to characterize the sensor information. A demonstration of detecting exceptionally low signal-to-noise ratio targets embedded in real-world 33-beam passive sonar data is presented.

Index Terms—Anomaly, fisher Information, neuroScale, SONAR, visualization.

I. INTRODUCTION

THE diversity of signal and noise sources, which make up the underwater acoustic environment, present a significant challenge for data analysis. Traditionally, beamforming has been one of the first steps in attempting to reduce the noise levels in observations, allowing improved target detection. However, some of these beamforming techniques suffer from reliability issues [1]. More recently, adaptive beamforming techniques have been shown to improve the output signal-to-noise ratio (SNR) of passive sonar arrays (see, for example, [2] and the references therein). Extensions of adaptive beamforming beyond the standard linear array structure have been implemented in [3]. Issues of the complexity cost of implementing robust adaptive beamformers can be reduced from cubic to quadratic in the number of parameters, as discussed in [2], and have been made manageable in [4] through implementation using graphical processing units.

Also, recent developments in tracking systems such as the Gaussian mixture model probability hypothesis density (GMM-PHD) filter [5] have been developed to improve the extended Kalman-filter approaches. These developments enhance the digital signal processing elements of the sonar data analysis. However, none of these advances reduce the operator overload problem since the basic requirement to check each one of the many beams for possible targets remains—It just improves the accuracy of target detection. Given that sonar arrays used to analyze this environment can consist of hundreds of individual sensors, data overload for operators is inevitable. The visual analysis and the presentation of multibeam sonar is largely performed in either narrowband or broadband forms. For instance, the LOFARgram (LOw Frequency Analysis Representation) comprises a column of waterfall spectrograms, each of which corresponds to the signal received in a beam. The spectrogram represents the joint time–frequency distribution of a signal, where a frequency distribution in a given time window is evaluated and displayed in such a way that tonals may be detected, and in the case of the LOFARgram, time is displayed vertically and frequency bins are displayed horizontally. The operator needs to visually explore each of these LOFARgrams for every one of hundreds of sonar channels, or beams, plus listening to all the separate acoustic signals to extract interesting tonal signatures. This is an obvious information overload bottleneck for the human sonar operator.

In this paper, we regard beamforming and adaptive filters as just another representation of the passive sonar data, and instead of using LOFARgram representations of this data, we will explore a complete alternative way to visualize and quantify the anomalous or surprising information, vastly reducing the number of important beams a sonar operator needs to check.

One prominent research direction for improving the presentation of information to sonar operators has been creating 3-D representations of active sonar returns (see for instance [6]–[8]). Methods for improving sonar imagery have also been proposed in [9]. However, this does not reduce the information overload issues since no channel prioritization is automatically embedded. A data fusion approach was developed in [10] to present signals from different sources, but this did not highlight anomalies or offer a reduced-signal representation to highlight sonar beams of interest. Barngrover *et al.* [11] presented a novel interface between sonar operator and sensor information, utilizing the experience of the human in the analysis process. Despite generating a summary of the information in a more visually appealing method than standard sonar imagery, this approach failed to generate an aesthetically intuitive visualization of the observed information, which would reduce the information burden on the human user.

In response to the requirement for operator decision support aids to reduce cognitive information overload and so to better represent, condense, and highlight surprise in high-dimensional

structured data [12], this paper presents a novel visualization model, an alternative way to analyze and view multibeam sonar data with the eventual aim of providing simpler representations for skilled operators.

To affect the reduction of high data dimensionality of sonar arrays into a visualization space as well as deal with the noisy and uncertain sonar signals, we will utilize a Probabilistic NeuroScale model [13].

The basic NeuroScale model stems from the Sammon mapping [14] where a configuration of points in a two-dimensional visualization space is generated, which approximately preserves the relative similarities between the original high-dimensional data points. Specifically, a dissimilarity d_{ij} between pairs i, j of original data patterns is defined, and a dissimilarity δ_{ij} between corresponding pairs i, j of configuration points on the visualization space is defined. In the Sammon map, the configuration of visualization points is found by iteratively adjusting the position of each point to minimize the Sammon STRESS (Standardized Residual Sum of Squares)

$$\mathcal{S} = \frac{1}{\sum_i \sum_{j>i} d_{ij}} \sum_i \sum_j \frac{(d_{ij} - \delta_{ij})^2}{d_{ij}}. \qquad (1)$$

This is an effective approach to data dimensionality reduction, especially for visualization purposes, since it attempts to preserve the average interpoint distance relationships between data, and hence tries to preserve topological structure in the visualized projection space. This is valuable in the sonar space since human operators intuitively use relative differences within images and sound to perceive significant events. However, the Sammon map does not provide a transformation between the two spaces. The NeuroScale model [15] was introduced to provide an explicit transformation for each data point using a nonlinear universal function approximator (a radial basis function network [16]) and a simplified STRESS function without the denominator terms in 1. Within NeuroScale, each input data pattern $\boldsymbol{X}_i \in \mathbb{R}^n$ is projected to a visualization point $\boldsymbol{Y}_j \in \mathbb{R}^2$ using a radial basis function $\boldsymbol{Y}_i = h(\boldsymbol{X}_i)$. The precise form used in this paper is defined later in the Method section.

In its basic form, this is a deterministic transformation between vectors. If the input data is actually generated as samples from an underlying noisy probability density function, then a probabilistic version can be implemented where the dissimilarity is taken to be any measure between distribution functions, such as a Kullback–Leibler divergence for example. So the method is not restricted to simply data vectors. Similarly, the projection into the visualization space can be generalized with probabilistic connotations, using the projected point as the mean of a distribution whose variance and higher moments can be derived from the variance and higher moments of the input data. A specific example of this is provided later.

Note that the archetypical approaches to probabilistic visualization, the generative topographic mapping [17] and the Gaussian process latent variable model [18], are unsuitable for this sonar task. These approaches attempt to construct a visualization that, under a set of Gaussian noise model assumptions, generates the observations. These two popular algorithms are not suitable for this application domain due to the constantly changing sonar environment where the observed noise distributions are known to be non-Gaussian [19], as well as the fundamental difference that they do not operate in dissimilarity space and hence are not metric-structure preserving.

II. METHOD

Our alternative approach for the representation of sonar information for the operator, and the presentation of potential anomalous activity, transforms all the multiple time series from all the sonar beams into a small information-preserving low-dimensional feature space in which anomalous behavior becomes apparent and is numerically quantified using information theoretic measures. The framework is composed of four stages: the creation of a signal model, the creation of a noise model, a structure-preserving data projection, and finally an anomaly detection process based on assigning a measure of "surprise" to each projected observation.

A. Signal Model for Each Beam

Consider a set of sonar array data (note that the framework in this paper can work with other sonar representations than the beamformed data, but the example data we use to illustrate the approach happens to be the beamformed data). Denote the time series observed in a given beam at time t by $\{x_t\}$. Typically the time series for a single beam is separated into 1-s segment vectors whose length is equal to the sampling frequency. Assume that the scalar value of the observed time series at time t, x_t is generated according to the following (unknown) noisy dynamical system:

$$x_t = f_t(\boldsymbol{\phi}_{t-1}, \boldsymbol{\Lambda}_t) + \eta_t \qquad (2)$$

where $f_t(\boldsymbol{\phi}_{t-1}, \boldsymbol{\Lambda}_t)$ is the actual possibly time-dependent and stochastic generating function, which drives the system dynamics based on latent variables $\boldsymbol{\phi}_{t-1}$, and where $\boldsymbol{\Lambda}_t, \eta_t$ are sample realizations of a dynamical noise process and an observational noise process, respectively.

These dynamical processes are intrinsically unknown and hence need to be approximated. We make no further assumptions on the true generating process.

The approximating system model to the true process used in this paper is such that

$$x_t = g_t(\boldsymbol{X}_{t-1}, \{W\}) + \epsilon_t \qquad (3)$$

where ϵ_t is the set of residual noise samples from a distribution to be determined, and $\{W\} = \{\boldsymbol{w}, \boldsymbol{\mu}_X\}$ is the set of model parameters of the approximator function used to estimate the dynamics. This paper uses

$$g(\boldsymbol{X}) = \sum_{j=0}^{N} w_j \psi(||\boldsymbol{X} - \boldsymbol{\mu}_X||) \qquad (4)$$

a radial basis function [16], capable of universal function approximation [20], with nonlinear basis functions $\psi(z) = z \log(z)$ centered at positions $\boldsymbol{\mu}_X$ and weighted by parameters $\{w_j; j = 0, \ldots, N\}$. For the experiments used later in this paper, $N = 30$ basis functions were randomly sampled from the observational beamformed vectors to span the signal space

This article has been accepted for inclusion in a future issue of this journal. Content is final as presented, with the exception of pagination.

RICE AND LOWE: DECISION SUPPORT SYSTEM TO EASE OPERATOR OVERLOAD IN MULTIBEAM PASSIVE SONAR 3

based on meeting accuracy, regularization, and efficiency criteria. This choice of N is not generic and as such should be specified for each data set based on these criteria or through Cattell's scree test [21] for example. The input to the approximator function is an embedding vector of delay samples $\boldsymbol{X}_t = \{x_t, x_{t-1}, \ldots x_{t-L}\}$ following Takens' theorem [22] for manifold reconstruction via the Whitney embedding theorem [23]. Methods for selecting the embedding size L are discussed in [24] (also see [25] for an overview). In this paper, in the later experiments, $L = 14$ was selected based on a minimum square error criterion.

Once an approximate signal manifold has been estimated via $g()$, the residuals ϵ_t are regarded as samples from an underlying structured noise model that is central to the uncertainty modeling in our framework.

B. Noise Model

From our dynamical signal model assumptions, the amplitude of the residual sample errors are as follows:

$$|\epsilon_t| \equiv |x_t - g(\boldsymbol{X}_{t-1}, \{W\})|$$
$$= |[f_t(\boldsymbol{\phi}_{t-1}, \boldsymbol{\Lambda}_t) - g(\boldsymbol{X}_{t-1}, \{W\})] + \eta_t|. \quad (5)$$

Estimating the residuals distribution, $q()$ of $|\epsilon_t|$, is an estimation of the fluctuations between the approximate and the true state model. Therefore, any structural changes in the data generating function (such as the appearance of unusual or unexpected targets) as well as changes in the background noise processes (perhaps due to weather variations) will therefore be reflected in this compound residual distribution. The framework relies on exploring the information dissimilarity space between the noise process at one time and subsequent times, as is evident later. Therefore, a good noise model is required.

The data likelihood of the residuals is approximated by a fifth-order compound mixture model [26], i.e.,

$$q(|\epsilon| \| \boldsymbol{\Theta}) = \prod_{d=1}^{D} \sum_{i=1}^{5} \pi_i P_i(|\epsilon| \, |\boldsymbol{\theta}_i) \quad (6)$$

consisting of a weighted linear combination of the following distributions $P_i(|\epsilon| \, |\boldsymbol{\theta}_i)$.

1) $P_1(|\epsilon|; |(\mu, b))$: A Laplace distribution is used to model extraneous signals. Any prominent signals not well described by the signal approximator function should appear in a small area of residual space and as such the sharply peaked Laplace is an appropriate choice.
2) $P_2(|\epsilon|; |\sigma)$ and $P_3(|\epsilon|; |(\alpha_k, \lambda_k))$: Rayleigh and K distributions are used to model a clutter. It is well established in the SONAR literature that K and Rayleigh distributions are suitable for describing background clutter in active SONAR such as biological and environmental effects in the underwater environment [27]. Our approach of introducing a signal model with residuals has parallels with the matched filter envelope described by the K-Rayleigh mixture in [27]. As such, we retain these components in our compound mixture model.

3) $P_4(|\epsilon|; |(\alpha, \beta))$: A Gamma distribution is used to approximate rain. Typically, in the literature, the Poisson distribution is used to describe rain [28]. However, the other distributions in this mixture are continuous and therefore the Gamma distribution, which behaves similarly to the Poisson, is chosen.
4) $P_5(|\epsilon|; |(m, s))$: A normal distribution is used to model the remnants. Any leftover residual elements are represented with a normal distribution, under the assumption that $\{P_1, \ldots, P_4\}$ will model the structured noise processes and all the remaining noise sources are taken as independent distributions modeled collectively by a Gaussian distribution assuming that the Central Limit Theorem applies.

This compound mixture model is optimized to the residual sample data by modifying the parameters to maximize the posterior distribution, given the data. Further details are provided in the appendix. This noise model is motivated by the known biological and physical characteristics of sea clutter sources, which should therefore generalize to a wide range of sea conditions. If an additional clutter source is expected in a region, the additional noise model can be included as an addition to the compound model.

C. Data Projection

The next part of the analysis combines both the previous signal and noise estimation models for each beam in a 1-s interval and creates a low-dimensional mapping specifically optimized to preserve the structural information in the data, based on minimally distorting the average relative dissimilarity (7) between data points. It is well known that the human effort to perform detailed analysis of multiple beams in very large arrays is a significant challenge, which can easily lead to anomalous contacts being missed. Our approach provides an alternative visualization whereby the information from all beams can be represented at once. Operators can then select the most significant beams highlighted as containing anomalies by the system, which may be spread far across the array, and investigate these signals with traditional techniques if desired. The approach generalizes to higher dimensions trivially. We restrict the implementations in this paper to the creation of optimal two-dimensional data representations. Optimality is defined by minimizing a specific distortion measure, the STRESS between the original data space and the projected two-dimensional data representation space, following the approach in [13] and developed in [26]. By construction, this two-dimensional representation will be *structure preserving* such that neighborhoods and the pairwise relative dissimilarities between observations will be preserved.

Specifically, in the original data space, a dissimilarity measure $d_X(i, j)$ between each pair of observations i, j translates to a dissimilarity $d_Y(i, j)$ under the mapping $\boldsymbol{h} : \boldsymbol{X} \to \boldsymbol{Y}$. The STRESS optimality criterion constrains the mapping $\boldsymbol{h}(\boldsymbol{X}) = \boldsymbol{Y}$ to try and preserve these pairwise relationships between the two spaces containing $\{\boldsymbol{X}\}$ and $\{\boldsymbol{Y}\}$ by minimizing

$$E = \sum_{ij} \frac{(d_X(i, j) - d_Y(i, j))^2}{d_X(i, j)}. \quad (7)$$

This article has been accepted for inclusion in a future issue of this journal. Content is final as presented, with the exception of pagination.

4   IEEE JOURNAL OF OCEANIC ENGINEERING

For input signal vectors, $\boldsymbol{X}_i$ and $\boldsymbol{X}_j$ with power spectra $\boldsymbol{S}_i$ and $\boldsymbol{S}_j$, and corresponding estimated noise models q_i, q_j, we choose a dissimilarity measure in data space using a scaled combination of the signal and noise dissimilarities

$$d_X(i,j) = \alpha d_{\text{sig}}(i,j) + (1 - \alpha)d_{\text{noise}}(i,j). \quad (8)$$

Here

$$d_{\text{sig}}(i,j) = ||\boldsymbol{S}_i - \boldsymbol{S}_j||_2 \quad (9)$$

is the Euclidean distance between the corresponding power spectra following [29] and

$$d_{\text{noise}}(i,j) = -\log\left(\sqrt{\boldsymbol{q}_i^T \boldsymbol{q}_j}\right) \quad (10)$$

is the Bhattacharyya distance denoting the overlap between noise distributions $\boldsymbol{q}_i$ and $\boldsymbol{q}_j$ sampled from the estimated distributions q_i and q_j of beams i and j, respectively. The scaling between signal and noise dissimilarities is used to simply give equal weight to each contribution to the overall dissimilarity. As such, for the experiments in Section III, we fix $\alpha = 0.5$. This parameter is application specific and driven by the significance of the signal and noise components in the anomaly detection task. For applications dominated by a single or low number of contacts, the anomaly detection is better performed in the noise subspace, favoring a lower α value. On the other hand, in applications with multiple signals with a low background noise level anomaly detection will be better performed in the signal space, favoring a higher α value. In the underwater environment, there is a significant noise background coupled with the potential for observing multiple contacts, motivating the weighting value chosen for this paper. It should also be noted that in the supervised cases of classification or regression, this parameter could be optimized or treated as a hyperprior in a Bayesian scheme.

Before the dissimilarity in the projected visualization space is defined, we first introduce the specific mapping function $\boldsymbol{h}(\boldsymbol{X})$. An interpolating radial basis function is used between the input data vector $\boldsymbol{X}_i$ and a set of L prototype data samples, $\{\boldsymbol{X}_l; l = 1, \ldots, L\}$ used to evaluate the dissimilarities $\{d_X(i,l); l = 1, \ldots, L\}$, allowing the evaluation of the projected data vectors

$$(\boldsymbol{Y}_i)_k = \boldsymbol{h}(\boldsymbol{X}_i)_k = \sum_l \boldsymbol{A}_{k,l}\psi(d_X(i,l)). \quad (11)$$

The spline function is $\psi(z) = z\log(z)$, and the set of weights $\{(\boldsymbol{A})_{k,l} = (\boldsymbol{a}_l)_k; k = 1, 2; l = 1, 2, \ldots L\}$ determine the mapping function and are numerically determined using a gradient method to minimize the previous STRESS measure in (7). The basis functions $\psi(d_X(i,l))$ are commonly expressed as vectors, $\boldsymbol{\psi}_i$, allowing the outputs to be expressed as $\boldsymbol{Y}_i = \boldsymbol{A}\boldsymbol{\psi}_i$.

This mapping induces a latent probability distribution in the projected visualization space. Specifically, each input data vector is mapped to a point in visualization space, which is taken to define a Gaussian distribution $\mathcal{N}(\boldsymbol{Y}_i, \sigma_i^2 \boldsymbol{I})$ in visualization space, centered around $\boldsymbol{Y}_i = \boldsymbol{h}(\boldsymbol{X}_i) = \boldsymbol{A}\boldsymbol{\psi}_i$, and $\sigma^2 = \text{var}(q_i)$. This is a first-order approximation that will be explored in the future work.

Thus, each input data vector is now mapped to a Gaussian distribution in data visualization space, and hence a conve-

nient dissimilarity measure in the projected data space is the Kullback–Leibler divergence between Gaussian distributions. Explicitly

$$d_Y(i,j) = KL[\mathcal{N}(\boldsymbol{Y}_i, \sigma_i^2 \boldsymbol{I})\|\mathcal{N}(\boldsymbol{Y}_j, \sigma_j^2 \boldsymbol{I})]. \quad (12)$$

Since the positions of the projected latent means $\boldsymbol{Y}_i$ in the visualization space have to be selected to minimize (7), the final part of the process requires an optimization algorithm. Minimization of (7) is an unsupervised task since the eventual positions are unknown. However, an efficient gradient-based optimization algorithm was developed specifically for the NeuroScale model, known as shadow targets optimization. In brief, to minimize (7), the local gradient with respect to the parameters of the radial basis functions are needed. In turn, these depend on the gradients of the STRESS measure with respect to the latent means times the gradient of the latent means with respect to the parameters. Now, if this was a conventional supervised problem minimized by a standard linear least squares process, the cost function would be $E = \sum_i ||\boldsymbol{Y}_i - \boldsymbol{T}_i||^2$, the gradient of which would yield

$$\frac{\partial E}{\partial \boldsymbol{Y}_i} = \boldsymbol{Y}_i - \boldsymbol{T}_i. \quad (13)$$

Therefore, in the unsupervised optimization of (7) if the true target were known, it would satisfy the same equation as (13) but with the cost function being the STRESS cost function. Therefore, by comparing these two situations, we can identify a "ghost" or "shadow" target $\bar{\boldsymbol{T}}_i$ such that if only the STRESS had been optimized, then the shadow target would satisfy

$$\bar{\boldsymbol{T}}_i \equiv \boldsymbol{Y}_i - \frac{\partial E}{\partial \boldsymbol{Y}_i}. \quad (14)$$

This is the basic idea of a recursive algorithm where the full unsupervised gradient descent problem is split into a sequence of stages where each stage can be efficiently solved by efficient linear least squares algorithms, continually refining the best estimates of the shadow targets. Further details of the full shadow targets algorithm can be found in [30] and [13] for a probabilistic version.

This completes the description of all of the components of the STRESS measure (7), and its optimization to create a principled transformation process, which preserves structural relationships in the original data.

D. Anomaly Detection

The key to the anomaly detection, or "surprise" of the information contained in each beam, and the final stage of the framework, is to quantify and represent the unexpected uncertainty of the projected data in the context of the overall data uncertainty in the projection space.

For context, the N total observed training data samples across all sonar beams, when projected into the visualization data space, maps to a set of N induced Gaussian distributions by the process described in the previous section. Thus, the training data defines an uncertainty surface approximated by a uniformly-weighted Gaussian mixture model, thus giving an estimate of the unconditional probability density at an arbitrary point $\boldsymbol{Y}$ in

the projected space as

$$U(\boldsymbol{Y}) = \frac{1}{N} \sum_{i=1}^{N} \mathcal{N}(\boldsymbol{Y}|\boldsymbol{Y}_i, \sigma_i^2 \boldsymbol{I}). \qquad (15)$$

Therefore, if a new data point is projected into a region of low probability density relative to this distribution, it indicates that it is a candidate anomaly with respect to the training data set. Note that this uncertainty surface is primarily to provide context to aid the operator in observing likely anomalous events. However, this observational uncertainty surface is only part of the uncertainty, since it ignores the distortions induced by the nonlinear projection mapping.

A more precise measure of anomalous or unusual data projections is obtained from the Cramer–Rao bound [31]. The covariance matrix of an unbiased estimator is given by the inverse of the Fisher information matrix (FIM), $\mathcal{I}(\gamma)^{-1}$. Generically, for an unbiased estimator $T_\gamma(X)$ of a parameter γ based on data X, we have that

$$\mathrm{cov}_\gamma(T_\gamma(X)) \geq \mathcal{I}(\gamma)^{-1}. \qquad (16)$$

Thus, the FIM provides a lower bound on the uncertainty provided by the data mapping. The FIM for the multivariate Gaussian distribution $\mathcal{N}(\boldsymbol{Y}, \sigma^2)$ with fixed covariance σ^2 is known to be

$$(\mathcal{I}(\gamma))_{mn} = \frac{\partial \boldsymbol{Y}^T}{\partial \gamma_m} \sigma^{-2} \frac{\partial \boldsymbol{Y}}{\partial \gamma_n}. \qquad (17)$$

Since in our framework the projected data mean is parameterically determined by $\boldsymbol{Y}_i = \boldsymbol{A}\boldsymbol{\psi_i}$, for beam i, the FIM corresponding to beam i with respect to the mapping parameters $\boldsymbol{A}$ is the block diagonal matrix

$$\mathcal{I}(\gamma)_i = \begin{bmatrix} \sigma_i^{-2}\boldsymbol{\psi}_i^T\boldsymbol{\psi}_i & \boldsymbol{0} \\ \boldsymbol{0} & \sigma_i^{-2}\boldsymbol{\psi}_i^T\boldsymbol{\psi}_i \end{bmatrix} \qquad (18)$$

where $\boldsymbol{0}$ is an $L \times L$ matrix with all entries 0.

Therefore, the trace of this inverse FIM is a summary measure that provides an A-optimality lower bound of (16) [32]

$$\mathrm{FI}_i = \mathrm{Tr}\left(I(\gamma)_i^{-1}\right). \qquad (19)$$

Large values of FI_i indicate potential new contacts or unusual activity in the beam signatures relative to the training data used to construct the topographic mapping.

In the following section, we illustrate the use of the framework and demonstrate FI_i in tandem with the uncertainty surface, to convey anomalous events in multibeam sonar data.

III. RESULTS

The sonar data set used in this paper was supplied by the Defence Science and Technology Laboratory (DSTL). The recording was collected from a 32 hydrophone line array in the Portland Bay, U.K., beamformed into 33 beams. The sensors operated at a 394-Hz sampling rate with fine weather conditions and only wave backscatter and some surface ships present. During the exercise, a speedboat traveled parallel to the array repeatedly with another ship intermittently present. In this paper, we perform three experiments based on this data set, one with the original
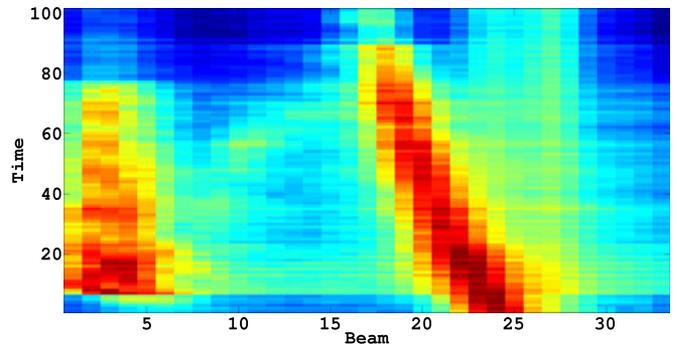
Fig. 1. Broadband energy for SONAR data set. Red indicates high levels of energy and dark blue indicates low energy. The two contacts in beams 1–5 and 21–26 are clearly identifiable in the 100-s recording.

data and the subsequent two involving the introduction of synthetic targets at low SNR. In each case, the training data used to generate the probabilistic NeuroScale mapping consisted of the first 5 s of original data. Note that axes are omitted in each of the visualizations since the NeuroScale transformation is invariant to arbitrary rotations and constant scaling. Therefore, it is only the relative position between points that is important, and therefore the axes are arbitrary and hence not displayed.

A. Experiment I: Normal Data

The signal (broadband) energy in each beam over time, given by

$$\mathrm{Energy}(x^i) = \sum_{t=1}^{T} (x_t^i)^2 \qquad (20)$$

where x_t^i represents the observed time series in beam i at time t, is shown in Fig. 1. From this plot, it is clear that there are two contacts centered in beams 3 (unknown ship) and 21 (DSTL exercise ship). These two contacts have similar observed energies whereas their signal characteristics differ slightly.

A signal model for each beam using the embedding size $L = 14$ is constructed, following which the parameters of the compound mixture model describing the residuals ϵ_t are learned by minimization of the negative log likelihood in (38), as outlined in the appendix. This compound noise model is then used as an approximation of the residual distribution q for each beam.

The two-dimensional representation for this experiment, constructed through the process of Section II-C, is presented in Fig. 2. The signal present in beam 25 is clearly highlighted by the dark gray point, $\boldsymbol{Y}_{25}$ given by (11), through its location in the low-probability cyan region. The linear structure implies a linear latent relationship in the data, which is expected as the array used for this recording was a uniformly spaced line array with standard beamforming, resulting in a single latent parameter here, namely the proximity to the speedboat contact. The uncertainty surface is notably unimodal in this example. This is because for this data set, the training data is dominated by similar uniform noise signals, which when combined with a central limit theorem leads to a skewed unimodal distribution. It is skewed toward the speedboat contact as expected. To show the

This article has been accepted for inclusion in a future issue of this journal. Content is final as presented, with the exception of pagination.

6   IEEE JOURNAL OF OCEANIC ENGINEERING
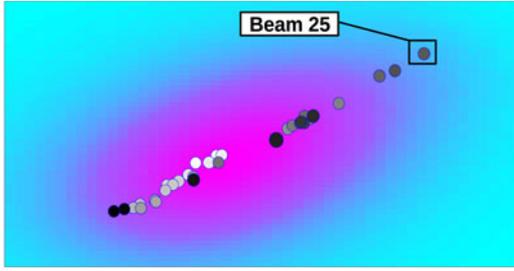
Fig. 2.   Visualization of the data set at 1 s into the recording. The mean of each latent distribution $\boldsymbol{y}_i$ corresponding to the signal and noise models for beam i. The points are colored from white corresponding to beam 1 through gray to black at beam 33. The uncertainty surface upon which these points sit is colored from magenta for high-probability regions to cyan for low-probability observations. Beam 25 and neighboring beams that observe the only present contact are clearly separated from the noise cluster in the high-probability region (magenta) and as an anomaly the beams observing the contact are placed in the low-probability region (cyan).
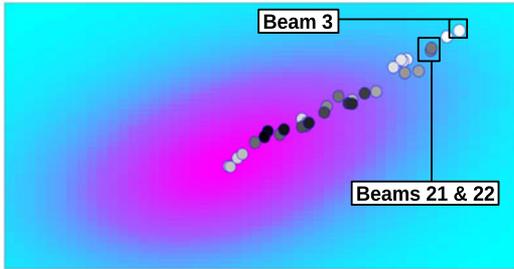
Fig. 3.   Visualization of the data set at 20 s into the recording. The contact centered in beam 25 of Fig. 2 has migrated to beams 21 and 22 (sat atop one-another), which are now located in the low-probability (cyan) region. The second contact centered in beam 3 is also highlighted anomalous with similar signal and noise characteristics to the speedboat contact.
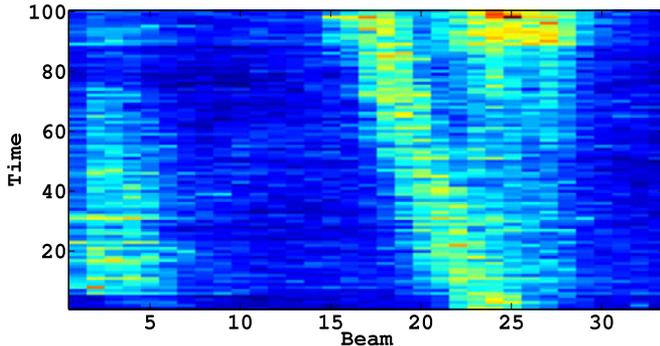
Fig. 4.   Mapping Fisher information for the normal data set. Red and blue regions indicate higher and lower levels of FI_i, respectively, indicating abnormal behavior. This resembles the broadband plot for the data set in Fig. 1, but with narrower concentration of contacts within beams. This highlights the two present contacts as anomalous compared to the prototypes, as would be expected.

evolution of this data representation over time, Fig. 3 presents the visualization 20 s into the scenario. The contact from beam 25 at the beginning of the exercise has migrated to beam 21. This is the same contact with similar noise profile as it is located in the same region of the uncertainty surface. Furthermore, the second contact now appears as the anomalous projection of beam 3.

Anomalies over time are highlighted in Fig. 4 by plotting the mapping uncertainty in each beam FI_i at each instant,

analogous to the broadband representation in Fig. 1. It is clear that the contact centered in beam 25 of Fig. 2 is anomalous from its location on the uncertainty surface, however the mapping uncertainty, and therefore the size of the points, appears relatively flat across the array. This is to be expected because the beams containing the signal and noise were used in the training; so, it is not surprising. They are, however, dissimilar from the noise-only beams; so, they are correctly placed in different regions of the uncertainty surface.

Our framework allows for anomalies to be quickly identified as in Fig. 4 as opposed to merely identifying beams with high signal energy in the broadband plot. Further to this, the relative similarities between beams can be interpreted in Fig. 2. Since $U(\boldsymbol{Y})$ displays all beams simultaneously and it is only those points representing beams in extreme locations or extreme sizes that the operator needs to consider, its workload is significantly reduced by focusing only on those beams automatically highlighted as possibly anomalous. For the visualization of Fig. 2, only the five beams contained within the low-probability region (cyan) need investigating as opposed to the entire 33 actual beams observed. This highlights an important benefit of our approach: the avoidance of data overload that is endemic to sonar data analysis.

The classification of whether an observation should generate an alarm is at the discretion of the operator, avoiding high false alarm rates inherent to automatic detection systems. High values of FI_i indicate that signal and noise sources have been observed, which were unseen in the training data and therefore should be investigated based on the level at which they exceed the baseline level. However, when the location of an observation is in a known region of the uncertainty surface the investigation of a contact with high FI_i is not a priority. An example of this could be the observation of a speedboat in different sea conditions than that of the training data that could make the contact appear more surprising, having a higher value of FI_i than that in Fig. 2, but it would still be located in the same region in $U(\boldsymbol{Y})$. The operator could then assume that this is still a speedboat and investigate only if there are no observations in other beams requiring their attention.

B. Experiment II: Single Injected Signal

The standard data set is modified by inserting a contact $\tilde{x}_t$ with two noisy sine waves at 151 and 163 Hz with relative SNR of -13 dB. This signal is injected into the array, moving from beam 1 to beam 33 uniformly over the 100-s experiment. The signal energy is renormalized in each beam following the signal injection:

$$x_t^i \leftarrow (x_t^i + \tilde{x}_t)\sqrt{\frac{\mathrm{Energy}(x^i)}{\mathrm{Energy}(x^i + \tilde{x})}} \qquad (21)$$

ensuring that the contact cannot be identified by simple fluctuations in the broadband energy. The broadband plot for this experiment is therefore identical to Fig. 1.

Fig. 5 shows the 2-D visualization of the first second of the data set with the injected signal located in beam 1. This visualization was trained with a set of prototypes randomly

This article has been accepted for inclusion in a future issue of this journal. Content is final as presented, with the exception of pagination.

RICE AND LOWE: DECISION SUPPORT SYSTEM TO EASE OPERATOR OVERLOAD IN MULTIBEAM PASSIVE SONAR 7
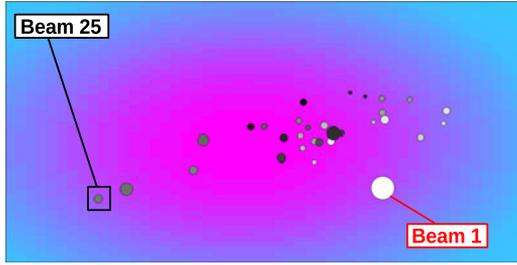
Fig. 5. Experiment II: Visualization of the data set at 1 s into the recording with the injected signal located in beam 1. This beam appears as a large circle and therefore is clearly anomalous despite the low SNR of the injected signal. Beam 25 sits on the edge of the normal (magenta) region as this is more similar to a prototype than the signals in the previous case and therefore less anomalous.
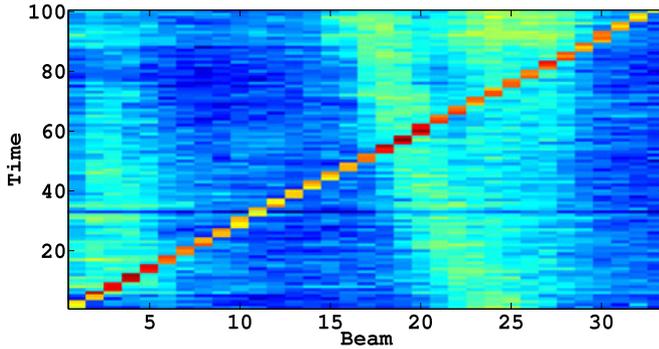
Fig. 6. Experiment II: Plot of FI_i over time for the data set with a single quiet injected signal moving uniformly across the array. This highlights the anomalous injected signal moving across the array while showing the relatively higher levels of mapping uncertainty for the beams containing the two contacts, as shown in Fig. 4 for the normal data set.

Fig. 7. Conventional LOFARgram showing 10 s worth of data for beams 31–33. The injected signal, with tonals at 151 and 163 Hz, is in beam 31 for 3 s following which it is in beam 32 then 33 at the end of the LOFARgram. Due to the extremely quiet −13-dB SNR level, the injected signal is not visually identifiable.

Fig. 8. Experiment III: Visualization of data set with four injected signals in beams 1, 15, 24, and 30. These beams are clearly identified as having a higher FI_i than the other beams in the array. beams 4–6 contain the speedboat contact and are therefore located in the low-probability region.

drawn from the exercise, which may have contained the injected signal and as such the distribution of noise-only beams is different from that of Fig. 2 for the normal data set. The size of the plotted points, whose locations are given by the latent means Y_i, varies dependent upon the level of mapping uncertainty for that beam FI_i plotted over time in Fig. 6. As in the previous experiment, the signal characteristics of the speedboat contact was already observed in the training stage and as such it is not as surprising as the injected signal leading to a lower value of FI_i.

Extended points in the 2-D visual representation indicate more uncertain observations, which should therefore be investigated by an operator.

This example serves to illustrate the significant benefit of our approach over typical broadband display approaches as it clearly indicates normal contacts, separating from noise-only beams more naturally than in Fig. 4, while simultaneously identifying a very quiet contact moving across beams, which can present a challenge for tracking systems. This contact is interesting as it can be thought of as attempting to mask its own signal characteristics with another signal in a narrower beamwidth. This visualization allows anomalies to be identified while not highlighting already known contacts included in the training data to the same level, reducing the operator overload burden.

To further show the ability of this method for highlighting quiet anomalous signals, the conventional LOFARgram visualization using global normalization is shown in Fig. 7 for the end

of the experiment in beams 31–33. These beams are chosen as they contain the lowest level of signal characteristics from the two contacts present in the normal data set. The injected signal is not visible in any of the three beams despite being clear in Fig. 6.

C. Experiment III: Multiple Injected Signals

Typically, in the underwater environment, we are interested not in a single contact but in the observation of multiple signals. When these signals have a low SNR and are narrowband by nature, the archetypical tool for the interrogation of these targets would be the waterfall plot. For large arrays, this places a burden on the operator in terms of both time and data overload. We inject four quiet contacts in beams 1, 15, 24, and 30 at −13, −16, −7, and −13 dB, respectively. The contact in beam 15 is present throughout the modified data set, however the other three contacts are only intermittently present every 10 s to test the ability of the algorithm to highlight sudden signal and residual changes rapidly. The injected signals in beams 1 and 24 contain noisy tonals at 151 and 163 Hz, whereas the injected tonals in beam 15 are sine waves at 151 and 163 Hz. The tonal introduced into beam 30 is fixed at 131 Hz. The beams containing these synthetic signals are normalized as in experiment II to avoid identification in the broadband plot.

Fig. 8 contains the generated visualization at 10 s into the data set, when a random set of mapping prototypes are

This article has been accepted for inclusion in a future issue of this journal. Content is final as presented, with the exception of pagination.

8   IEEE JOURNAL OF OCEANIC ENGINEERING
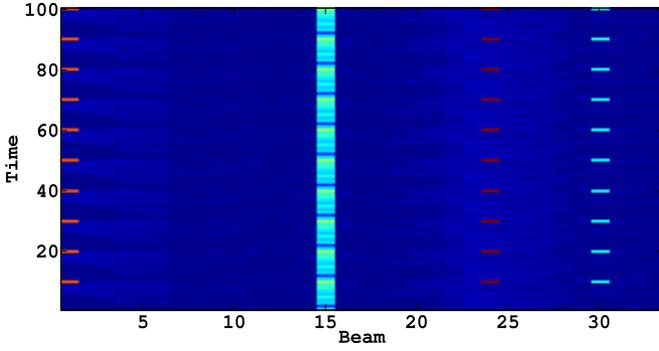
Fig. 9.   Experiment III: Plot of relative surprise over time for each of the beams in the array. The beams with injected signals, beams 1, 15, 24, and 30, all have a raised level of surprise at the correct times in the experiment. The four beams with injected signals contain a raised level of Fisher Information compared to the beams with normal observations only, despite the low SNR of the injected signals.

chosen from the normal data set. The four inserted targets are clearly identified through higher levels of mapping uncertainty, in addition to the fact that beam 15 is located on the cusp of the low-probability region. As with the previous two experiments, Fig. 9 shows the change in mapping uncertainty over time.

This final example demonstrates the ability of the visualization and anomaly detection framework to signify multiple anomalous beams over a background of noise and uninteresting signals. This will allow an operator to select a subset of beams to be investigated based on both the location on the uncertainty surface of Fig. 8, for instance beam 5, and the beams with high levels of surprise, such as the four beams containing injected signals.

IV. Conclusion

This paper has presented a novel visual informatics for anomaly detection in passive multibeam sonar. The approach incorporates recent developments in probabilistic visualization with the signal processing approach called residual modeling. We focused not on merely the signal components, but also on characterizing the distributions over the residuals in each beam. This allows for the identification of anomalies in observations at very low SNR, as shown in the experimental results. The visualization of beam information allows for sonar operators to identify signals with interesting contacts such that they can focus on these beams and neglect the unsurprising beams, preventing data overload.

Further research will test this method on more diverse real-world data sets. In addition to this, we will investigate the set of prototype signal and noise models that allow for identification of anomalous contacts, such as submarines in the visual representation and mapping uncertainty plot without highlighting contacts known to be insignificant, such as merchant ships. We will also look to implement this framework on data sets containing marine animals to test whether the results with the man-made vessels translate to this biological task.

Appendix
Compound Noise Model Optimization

The compound mixture model characterizing the residuals ϵ_t in each beam requires optimization of the K-distribution parameters before fitting the remaining parameters and hyperparameters. This is essential since optimizing the parameters of the K-distribution using gradient descent in a maximum-likelihood (ML) fashion often leads to unrealistic parameters [33]. To remedy this, the parameters of the K-distribution are fit to the residuals in a Bayesian method-of-moments scheme introduced in [19], following which the parameters and hyperparameters of the rest of the compound mixture model are fit using gradient descent over the negative log likelihood.

The priors specified for each distribution were initially set as the conjugate priors, however some of these were found to be a poor fit for the optimized parameters on real data sets. As such, the distributions were chosen to be as representative of the real-world data described in Section III. It should be noted that there are no priors specified over the mixture weights as this has only been found to slow convergence of the gradient learning procedure.

First, the individual distributions and their relevant priors are defined, following which the gradient learning procedure will be described.

A. Model

1) Laplace Distribution: The PDF of the Laplace distribution is

$$P(\epsilon|\mu,b) = \frac{1}{2b}\exp\left[-\frac{|\epsilon-\mu|}{b}\right]. \tag{22}$$

The priors over μ and b are Laplace and Gamma distributions, respectively, i.e.,

$$P(\epsilon|\mu,b) = \text{Laplace}(\epsilon|\mu,b) \tag{23}$$
$$P(\mu|\mu_0,b) = \text{Laplace}(\mu|\mu_0,b) \tag{24}$$
$$P(b|\alpha_b,\beta_b) = \text{Gamma}(b|\alpha_b,\beta_b). \tag{25}$$

2) Rayliegh Distribution: The PDF of the Rayleigh distribution is

$$P(\epsilon|\sigma) = \frac{\epsilon}{\sigma^2}\exp\left[-\frac{\epsilon^2}{2\sigma^2}\right]. \tag{26}$$

The prior over σ is a Gamma distribution

$$P(\epsilon|\sigma) = \text{Rayleigh}(\epsilon|\sigma) \tag{27}$$
$$P(\sigma|a_\sigma,b_\sigma) = \text{Gamma}(\sigma|a_\sigma,b_\sigma). \tag{28}$$

3) K-Distribution: The K-distribution is optimized using the procedure from [19].

4) Gamma Distribution: The PDF of the Gamma distribution used in the SONAR noise model is

$$P(\epsilon|\alpha,\beta) = \frac{1}{\Gamma(\alpha)}\beta^{-\alpha}\epsilon^{(\alpha-1)}\exp\left(-\frac{\epsilon}{\beta}\right). \tag{29}$$

The prior over β is again a Gamma distribution

This article has been accepted for inclusion in a future issue of this journal. Content is final as presented, with the exception of pagination.

RICE AND LOWE: DECISION SUPPORT SYSTEM TO EASE OPERATOR OVERLOAD IN MULTIBEAM PASSIVE SONAR

9

$$P(\epsilon|\alpha, \beta) = \text{Gamma}(\epsilon|\alpha, \beta) \tag{30}$$

$$P(\beta|a_\beta, b_\beta) = \text{Gamma}(\beta|a_\beta, b_\beta). \tag{31}$$

There is no closed-form conjugate prior for α and the distribution over α appears to change on each fit. As such, there is no prior over α.

5) Normal Distribution: The probability density function of the normal distribution is given by

$$P(\epsilon|m, s) = \frac{1}{\sqrt{2\pi s^2}} \exp\left[-\frac{(\epsilon - m)^2}{2s^2}\right]. \tag{32}$$

The priors over m and s are Normal and Gamma distributions, respectively

$$P(\epsilon|m, s) = \mathcal{N}(\epsilon|m, s^2) \tag{33}$$

$$P(m|m_0, s) = \mathcal{N}(m|m_0, s^2) \tag{34}$$

$$P(s^2|a_s, b_s) = \text{Gamma}(s^2|a_s, b_s). \tag{35}$$

6) Likelihood: The probabilistic model can now be written as the maximization of

$$P(\epsilon|\theta_p)P(\theta_p|\theta_h) \tag{36}$$

where $P(\epsilon|\theta_p)$ is the data likelihood of ϵ given the parameters θ_p, and $P(\theta_p|\theta_h)$ is the prior distribution of the parameters θ_p, given the hyperparameters, θ_h. The data log likelihood is given as

$$P(\epsilon|\theta_p) = \sum_i \left[\log\left(\pi_1 \text{Laplace}(\epsilon_i|\mu, b) + \pi_2 \text{Rayleigh}(\epsilon_i|\sigma) \right. \right.$$
$$\left. + \pi_3 K + \pi_4 \text{Gamma}(\epsilon_i|\alpha, \beta) \right.$$
$$\left. \left. + \pi_5 \mathcal{N}(\epsilon_i|m, s^2) + N\left(\sum_j \pi_j - 1 \right) \right) \right] \tag{37}$$

where π_j are the mixture weights and $N(\sum_j \pi_j - 1)$ relates to the Lagrange multiplier used to ensure that the $\pi'_j s$ sum to unity.

Optimization in the maximum *a posteriori* (MAP) context involves maximizing the posterior distribution, typically ignoring the effects of the unknown distribution $P(\epsilon)$, known as the model evidence. In reality, it is often the case that the negative log likelihood and negative log-prior distributions are analyzed. Denoting the set of parameters and hyperparameters $\theta = \{\theta_p, \theta_h\}$, the optimization involves evaluating the gradients for each θ^i

$$\frac{-\partial \log P(\epsilon|\theta_p)}{\partial \theta^i} + \frac{-\partial \log P(\theta_p|\theta_h)}{\partial \theta^i} \tag{38}$$

to find a minimum. This is often performed using a nonlinear gradient optimizer, such as scaled conjugate gradients (SCGs). It should be noted that many of the above distributions require the parameters to be greater than zero and as such the computational implementation of the derivative was taken with respect to the square root of the parameter, following which it was then squared. This notation is omitted here to keep the expressions as simple as possible.

B. Gradients

1) Laplace Parameters and Hyperparameters: The parameters and hyperparameters relevant to the Laplace distribution are $\{\mu, b, \mu_0, \alpha_b, \beta_b\}$ with the following derivatives:

$$\frac{\partial}{\partial \mu} = \begin{cases} -\frac{1}{P(\epsilon|\theta_p)} \left(\frac{1}{2b^2} \exp\left[\frac{|\epsilon - \mu|}{b} \right] \right) \pi_1 & \text{if } \epsilon \geq \mu \\ -(-1)^h \frac{1}{b}, \\ \frac{1}{P(\epsilon|\theta_p)} \left(\frac{1}{2b^2} \exp\left[\frac{|\epsilon - \mu|}{b} \right] \right) \pi_1 & \text{if } \epsilon < \mu \\ -(-1)^h \frac{1}{b}, \end{cases} \tag{39}$$

$$\frac{\partial}{\partial b} = \frac{1}{P(\epsilon|\theta_p)} \left(-\frac{1}{2b^2} + \frac{|\epsilon - \mu|}{2b^3} \right) \exp\left(\frac{-|\epsilon - \mu|}{b} \right) \pi_1$$
$$- \left[\frac{1}{b} - \frac{|\mu - \mu_0|}{b^2} + \frac{1}{2\beta_b^2}(2b - 2\alpha_b) \right] \tag{40}$$

$$\frac{\partial}{\partial \mu_0} = -\frac{1}{b} \tag{41}$$

$$\frac{\partial}{\partial \alpha_b} = \frac{1}{2\beta_b^2}(-2b + 2\alpha_b). \tag{42}$$

where h is the Heaviside function, $h = 1$ if $\mu < \mu_0$ and $h = 0$ if $\mu \geq \mu_0$.

2) Rayleigh Parameters and Hyperparameters: The parameters and hyperparameters relevant to the Rayleigh distribution are $\{\sigma, a_\sigma, b_\sigma\}$ with the following derivatives:

$$\frac{\partial}{\partial \sigma} = \frac{1}{P(\epsilon|\theta_p)} \left(\frac{\epsilon^3}{\sigma^5} - \frac{2\epsilon}{\sigma^3} \right) \exp\left[\frac{-\epsilon^2}{2\sigma^2} \right] \pi_2 - \left[\frac{b_\sigma}{\sigma^2} + \frac{a_\sigma}{\sigma} + \frac{1}{\sigma} \right] \tag{43}$$

$$\frac{\partial}{\partial a_\sigma} = -\log\left(\frac{b_\sigma}{\sigma} \right) + \Psi(a_\sigma) \tag{44}$$

$$\frac{\partial}{\partial b_\sigma} = \frac{1}{\sigma} - \frac{a_\sigma}{b_\sigma}. \tag{45}$$

3) Gamma Parameters and Hyperparameters: The parameters and hyperparameters relevant to the Gamma distribution are $\{\alpha, \beta, a_\beta, b_\beta\}$ with the following derivatives:

$$\frac{\partial}{\partial \alpha} = \frac{1}{P(\epsilon|\theta_p)} \left(\log(\epsilon) - \log(\beta) - \Psi(\alpha) \right)$$
$$\times \left(\frac{1}{\Gamma(\alpha)} \epsilon^{(\alpha-1)} \beta^{-\alpha} \exp\left[\frac{-\epsilon}{\beta} \right] \right) \pi_4 \tag{46}$$

$$\frac{\partial}{\partial \beta} = \frac{1}{P(\epsilon|\theta_p)} \left(\frac{1}{\Gamma(\alpha)} \exp\left[\frac{-\epsilon}{\beta} \right] \right.$$
$$\left. \times \left(\epsilon^\alpha \beta^{(-2-\alpha)} - \epsilon^{-1+\alpha} \alpha \beta^{-1-\alpha} \right) \right) \pi_4$$
$$+ \left[-\frac{(a_\beta - 1)}{\beta} + b_\beta \right] \tag{47}$$

$$\frac{\partial}{\partial a_\beta} = -\log(\beta) - \log(b_\beta) + \Psi(a_\beta) \tag{48}$$

$$\frac{\partial}{\partial b_\beta} = \beta - \frac{a_\beta}{b_\beta}. \tag{49}$$

4) Normal Parameters and Hyperparameters: The parameters and hyperparameters relevant to the normal distribution are $\{m, s, m_0, a_s, b_s\}$ with the following derivatives:

$$\frac{\partial}{\partial m} = \frac{1}{P(\epsilon|\theta_p)}\left(\frac{(\epsilon - m)}{\sqrt{2\pi}s^3}\exp\left[\frac{-(\epsilon - m)^2}{2s^2}\right]\right)\pi_5$$

$$- \frac{(m - m_0)}{s^2} \tag{50}$$

$$\frac{\partial}{\partial s^2} = \frac{1}{P(\epsilon|\theta_p)}\left(\frac{-1}{\sqrt{2\pi}s^2} + \frac{(\epsilon - m)^2}{s^4}\right)\exp\left[\frac{-(\epsilon - m)^2}{2s^2}\right]\pi_5$$

$$+ \left[\frac{1}{2s^2} - \frac{(m - m_0)^2}{2s^4} - \frac{b_s}{s^4} + \frac{a_s}{s^2} + \frac{1}{s^2}\right] \tag{51}$$

$$\frac{\partial}{\partial m_0} = \frac{(m_0 - m)}{s^2} \tag{52}$$

$$\frac{\partial}{\partial a_s} = -\log\left(\frac{b_s}{s^2}\right) + \Psi(a_s) \tag{53}$$

$$\frac{\partial}{\partial b_s} = \frac{1}{s^2} - \frac{a_s}{b_s}. \tag{54}$$

5) Mixture Coefficients: The derivatives with respect to the mixture coefficients $\{\pi_j, j = 1 : 5\}$ are given by

$$\frac{\partial}{\partial \pi_1} = \frac{\text{Laplace}}{P(\epsilon|\theta_p)} - N \tag{55}$$

$$\frac{\partial}{\partial \pi_2} = \frac{\text{Rayleigh}}{P(\epsilon|\theta_p)} - N \tag{56}$$

$$\frac{\partial}{\partial \pi_3} = \frac{\text{K}}{P(\epsilon|\theta_p)} - N \tag{57}$$

$$\frac{\partial}{\partial \pi_4} = \frac{\text{Gamma}}{P(\epsilon|\theta_p)} - N \tag{58}$$

$$\frac{\partial}{\partial \pi_5} = \frac{\mathcal{N}}{P(\epsilon|\theta_p)} - N \tag{59}$$

where the left term in each derivative is the relative responsibility of the mixture weight, and the N (number of datapoints) term is given by the Lagrange multiplier constraint.

These terms for the parameters and hyperparameters can be used by any standard nonlinear optimization algorithm to fit the noise distribution. For the experiments of Section III, SCG was used to optimize the parameters of the MAP compound noise mixture. An example of the residuals' histogram for beam 21 of the normal experiment detailed in Section III-A is shown in Fig. 10.

The mixture model is a good fit to the residuals, smoothing out the effects of the outliers. Fig. 11 shows a comparison between the negative log likelihood of a Gaussian mixture model approximation in addition to ML and MAP fits of the compound mixture model detailed in Section II-B.

The ML and MAP mixtures fit the residuals better than the GMM. The MAP fit is smoother and more reliable than the ML approach. All parameters here are fit with SCG and as such can achieve a local minima. The local minima in the optimization can force the distributions to swap mixture weights without
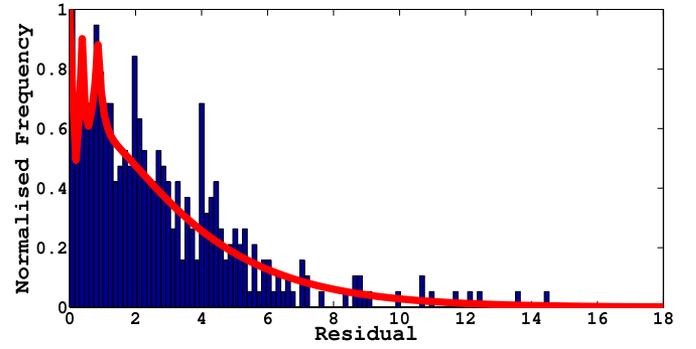
Fig. 10. Histogram of the residuals ϵ_t of beam 21 for 1 s of data. The fit of the noise distribution is shown in red.
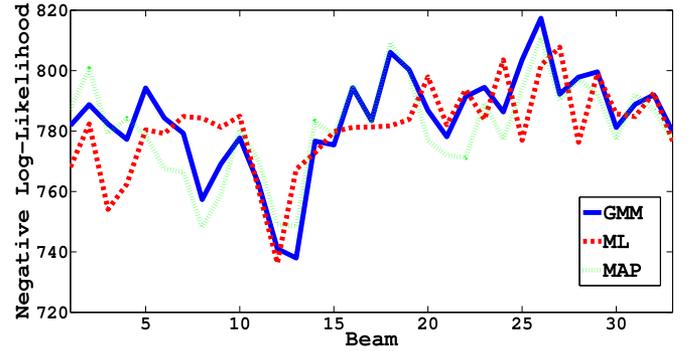
Fig. 11. Negative log likelihoods for the mixture model fit to 1 s of SONAR data. The MAP fit is on average better than that of the ML and GMM.

necessarily changing the overall shape of the distribution, typically for the Gamma and K distributions.

REFERENCES

[1] S. D. Somasundaram and A. Jakobsson, "Degradation of covariance reconstruction-based robust adaptive beamformers," in *Proc. Sensor Signal Process. Defence*, Sep. 2014, pp. 1–5.

[2] S. D. Somasundaram, N. R. Butt, A. Jakobsson, and L. Hart, "Low-complexity uncertainty-set-based robust adaptive beamforming for passive sonar," *IEEE J. Ocean. Eng.*, vol. 99, pp. 1–17, Nov. 2015.

[3] A. E. A. Blomberg, A. Austeng, and R. E. Hansen, "Adaptive beamforming applied to a cylindrical sonar array using an interpolated array transformation," *IEEE J. Ocean. Eng.*, vol. 37, no. 1, pp. 25–34, Jan. 2012.

[4] J. I. Buskenes, J. P. Asen, C. I. C. Nilsen, and A. Austeng, "An optimized GPU implementation of the MVDR beamformer for active sonar imaging," *IEEE J. Ocean. Eng.*, vol. 40, no. 2, pp. 441–451, Apr. 2015.

[5] T. Wood, D. Allwright, P. Bond, S. Long, and I. Moroz, "A new method for processing passive sonar data," in *Proc. 13th Conf. Inf. Fusion.*, Jul. 2010, pp. 1–7.

[6] I. C. Rivera-Collazo, "Looking at the 'Continent divided by water': Coastal and human dynamics, and the potential for submerged landscapes in the Caribbean," in *Proc. IEEE/OES Acoust. Underwater Geosci. Symp. (RIO Acoust.)*, Jul. 2015, pp. 1–10.

This article has been accepted for inclusion in a future issue of this journal. Content is final as presented, with the exception of pagination.

RICE AND LOWE: DECISION SUPPORT SYSTEM TO EASE OPERATOR OVERLOAD IN MULTIBEAM PASSIVE SONAR 11

[7] D. Shea, D. Dawe, J. Dillon, and S. Chapman, "Real-time SAS processing for high-arctic AUV surveys," in *Proc. IEEE/OES Auton. Underwater Vehicles*, Oct. 2014, pp. 1–5.

[8] D. Lu, H. Li, Y. Wei, and T. Shen, "An improved merging algorithm for Delaunay meshing on 3D visualization multibeam bathymetric data," in *Proc. IEEE Int. Conf. Inf. Autom.*, Jun. 2010, pp. 1171–1176.

[9] C. G. Capus, A. C. Banks, E. Coiras, I. T. Ruiz, C. J. Smith, and Y. R. Petillot, "Data correction for visualisation and classification of sidescan SONAR imagery," *IET Radar, Sonar Navig.*, vol. 2, no. 3, pp. 155–169, Jun. 2008.

[10] P. A. M. de Theije, "STARE, a sonar data post-processing and visualisation software package," in *Proc. Eur. Oceans*, Jun. 2005, vol. 1, pp. 481–488.

[11] C. Barngrover, A. Althoff, P. DeGuzman, and R. Kastner, "A brain–computer interface (BCI) for the detection of mine-like objects in sides-can sonar imagery," *IEEE J. Ocean. Eng.*, vol. 41, no. 1, pp. 123–138, Jan. 2016.

[12] DSTL, "Call: Data and information assimilation," 2015. [Online]. Available: http://www.science.mod.uk/events/event_detail.aspx?eventid=105

[13] M. Sivaraksa and D. Lowe, "Probabilistic NeuroScale for uncertainty visualisation," in *Proc. 13th Int. Conf. Inf. Vis.*, Barcelona, Spain, Jul. 15–17, 2009, pp. 74–79.

[14] J. W. Sammon, "A nonlinear mapping for data structure analysis," *IEEE Trans. Comput.*, vol. 18, no. 5, pp. 401–409, May 1969.

[15] D. Lowe and M. Tipping, "NeuroScale: Novel topographic feature extraction using RBF networks," in *Advances in Neural Information Processing Systems 9*, M. Mozer, M. Jordan, and T. Petsche, Eds. Cambridge, MA, USA: MIT Press, 1997, pp. 543–549.

[16] D. Broomhead and D. Lowe, "Multivariable functional interpolation and adaptive networks," *Complex Syst.*, vol. 2, pp. 321–355, 1988.

[17] C. M. Bishop, M. Svensen, and C. K. I. Williams, "GTM: The generative topographic mapping," *Neural Comput.*, vol. 10, no. 1, pp. 215–234, 1998.

[18] N. Lawrence, "Gaussian process latent variable models for visualisation of high dimensional data," in *Advances in Neural Information Processing Systems 16*. Cambridge, MA, USA: MIT Press, Dec. 2003, pp. 329–336.

[19] D. Abraham and A. Lyons, "Reliable methods for estimating the K-distribution shape parameter," *IEEE J. Ocean. Eng.*, vol. 35, no. 2, pp. 288–302, Apr. 2010.

[20] M. D. Buhmann and M. D. Buhmann, *Radial Basis Functions*. New York, NY, USA: Cambridge Univ. Press, 2003.

[21] R. B. Cattell, "The scree test for the number of factors." *Multivariate Behav. Res.*, vol. 1, pp. 245–276, 1966.

[22] F. Takens, *Detecting Strange Attractors in Turbulence* (Series Dynamical Systems and Turbulence, Lecture Notes in Mathematics), vol. 898. Berlin, Germany: Springer-Verlag, 1981, pp. 366–381.

[23] H. Whitney, "The self-intersections of a smooth n-manifold in $2n$-space," *Ann. Math.*, vol. 45, no. 2, pp. 220–246, Apr. 1944.

[24] F. Camastra and M. Filippone, "A comparative evaluation of nonlinear dynamics methods for time series prediction," *Neural Comput. Appl.*, vol. 18, no. 8, pp. 1021–1029, 2009.

[25] H. Kantz and T. Schreiber, *Nonlinear Time Series Analysis*, 2nd ed. Cambridge, U.K.: Cambridge Univ. Press, 2004.

[26] I. Rice and D. Lowe, "Topographic visual analytics of multibeam dynamic SONAR data," in *Proc. Sensor Signal Process. Defence*, Sep. 2014, pp. 1–5.

[27] D. Abraham, J. Gelb, and A. Oldag, "K-Rayleigh mixture model for sparse active SONAR clutter," in *Proc. IEEE OCEANS*, Sydney, NSW, Australia, May 2010, pp. 1–6.

[28] S. Haykin and S. Puthusserypady, *Chaotic Dynamics of Sea Clutter* (Series Adaptive and Learning Systems for Signal Processing, Communications and Control). S. Haykin, Ed. Hoboken, NJ, USA: Wiley, 1999.

[29] I. Rice, R. Benton, L. Hart, and D. Lowe, "Analysis of multibeam SONAR data using dissimilarity representations," in *Proc. 3rd IMA Math. Defence*, Oct. 2013, pp. 1–6.

[30] M. Tipping and D. Lowe, "Shadow Targets: A novel algorithm for topographic projections by radial basis functions," *NeuroComputing*, vol. 19, pp. 211–222, 1997.

[31] C. R. Rao, "Information and the accuracy attainable in the estimation of statistical parameters," *Bull. Calcutta Math. Soc.*, vol. 37, pp. 81–89, 1945.

[32] A. C. Atkinson and A. N. Donev, *Optimum Experimental Designs*. London, U.K.: Oxford Univ. Press, 1992.

[33] R. Bareš, "Environmentally adaptive noise estimation for active sonar," Ph.D. dissertation, School Comput. Sci., Cardiff University, Cardif, U.K., Jan. 2011.

Iain Rice received the B.Sc. degree in mathematics and business administration and the Ph.D. degree in applied mathematics, both from Aston University, Birmingham, U.K., in 2012 and 2015, respectively.

During the Ph.D. degree, he developed a novel framework for visualizing sonar data to identify anomalies as part of a joint project with Thales U.K. Following this, he was a Postdoctoral Researcher in machine learning integrated with the health care domain. He is currently a Postdoctoral Teaching Fellow with Arden University.

David Lowe received the Ph.D. degree in theoretical physics from Warwick University , Coventry, U.K., in 1983. He is currently a coinventor of the radial basis function network, and the NeuroScale visualization scheme and shadow targets optimization process. After a background in the defence scientific civil service, he joined Aston University to establish the Neural Computing Research Group. Working in areas from bioinformatics and medical informatics, defence, finance, and transport sectors, he has been an advisor to international review panels linked to applied data fusion, complex systems, and machine learning.