

Topographic Visual Analytics of Multibeam Dynamic SONAR Data

Iain Rice
NCRG
Aston University
Birmingham, B4 7ET
Email: ricei@aston.ac.uk

David Lowe
NCRG
Aston University
Birmingham, B4 7ET
Email: d.lowe@aston.ac.uk

Abstract—This paper considers the problem of low-dimensional visualisation of very high dimensional information sources for the purpose of situation awareness in the maritime environment. In response to the requirement for human decision support aids to reduce information overload (and specifically, data amenable to inter-point relative similarity measures) appropriate to the below-water maritime domain, we are investigating a preliminary prototype topographic visualisation model. The focus of the current paper is on the mathematical problem of exploiting a relative dissimilarity representation of signals in a visual informatics mapping model, driven by real-world sonar systems. A realistic noise model is explored and incorporated into non-linear and topographic visualisation algorithms building on the approach of [9]. Concepts are illustrated using a real world dataset of 32 hydrophones monitoring a shallow-water environment in which targets are present and dynamic.

I. INTRODUCTION

SONAR (sound navigation and ranging) is used extensively in underwater acoustics. The fundamental issues commonly encountered in the SONAR domain include low signal-to-noise ratio, multipath reflections from targets, sea surface and floor and the high volume of data for analysis. Large sonar systems can produce hundreds to thousands of beams worth of data, the display of which can easily result in data overload for sonar operators when presented in conventional lofargram displays. The multi-beam time series signals are considered as a set of time-dependent vectors in a high dimensional observation space (where the dimensionality is the number of beams being recorded), and this set of observation vectors is the result of an unknown latent generative model consisting of multiple signals mixed with noise. The desire is to find a low-dimensional latent space representation of the high dimensional data which preserves structure in the original data which would be useful to an operator for anomaly detection. Such ‘topographic’ (structure-preserving) projections need to be generically nonlinear and ideally capable of mapping the projection of new data without being ‘re-trained’ for each block of new data. Additionally, since the observation data is always subject to uncertainty, if there is a framework and information to represent uncertainty in the observation data, the projection method should accommodate the uncertainty in the latent space construct. Finally, since each observation vector (sonar beam scan) only has significance in the context of its relative information content compared to the context of a large number of other scans (e.g., human operators perceptually ‘integrate’ lofargrams to smooth out noise, spot anomalous

behaviour, and occluded tracks) we are interested in methods which use relative measures of dissimilarity between pairs of observations, rather than assuming the observed values of a single isolated observation vector are the fundamental entities. In this paper we introduce one such nonlinear topographic visualisation approach which is flexible enough to use different measures for representing ‘similarity’ between observations vectors. Results are presented on a real-world trial for 32 hydrophones samples in a shallow-water environment with active targets.

A. Methodology

The modelling philosophy adopted in this paper assumes that the ‘information is in the fluctuations’. That is, rather than regarding the noise in observed sensor data as a nuisance to be removed, we prefer to characterise the noise processes and hence any fluctuations from this noisy state must by definition be informative, interesting and possibly a signal of relevance. Therefore we will be characterising the residuals once we have constructed models for possible stationary signal sources. This noise process will typically be in high dimensions and hence we then need a mechanism to compare the relative fluctuations of the noise processes over time, or over beams. This motivates the focus of the paper, to develop the use of high-to-low dimensional projection methods which preserve the relative dissimilarities in the high dimensional space.

II. SIGNAL MODELLING

The dynamic signal vector, s_b , observed in a beam b can be thought of as primarily sensing a noise process, ϵ_b , plus a set of additive target tonals, u_b ; where a target may be present.

$$s_b = u_b + \epsilon_b. \quad (1)$$

The sensor array across all beams can be considered either as a set of one-dimensional time series or groups of vector time series processes. Firstly we consider the case of a beam-by-beam analysis.

A. Beam-by-beam analysis

Our approach involves approximating beam-formed signals with a non-linear auto-regressive model and fitting the residuals with a realistic mixture model, as follows.

When present, an observed dynamic signal can be represented by an auto-regressive process. We have found that

a linear autoregressive process is not capable of sufficiently capturing the dynamics of a signal in this environment. We therefore seek a nonlinear auto-regressive process for the modelled target signal. ie for the time series of beam b , an estimate of the signal is provided by a deterministic function which depends on delayed time series components of the observed signal, $(\hat{u}_b)_t = f_b[s_t; \Lambda]$ where $s_t = (s_{t-1}, s_{t-2}, \dots, s_{t-L-1})$ is a standard delay vector. We choose $L = 14$ here to minimise the mean square error over all beams in a similar fashion to the linear partial autocorrelation function method. The parameterised nonlinear function $f_b[s_t; \Lambda]$ is taken generically to be a linear-output Radial Basis Function (RBF) neural network where Λ denotes the set of parameters in the network, ie the set of weights λ and the set of centres μ used to define the network, ie

$$f_b[s_t; \Lambda] = \sum_{j=0}^N \lambda_j \phi(|s_t - \mu_j|).$$

$N = 100$ centres randomly chosen from a training set of delay vectors were used in this paper to perform adequate interpolation and the nonlinear basis functions were taken to be splines, $\phi(z) = z \ln z$. A separate RBF model is trained on each separate beam.

Once an approximation of the target signal, \hat{u}_b , is obtained we then need to model the estimated noise characteristics of the signal, $\hat{\epsilon}_b = |s_b - \hat{u}_b|$. This is important as the characteristics of certain stealthy undersea contacts such as submarines may only be apparent in the background noise. The residual noise process could be approximated with a Gaussian Mixture Model if no prior knowledge of the domain were known. However we choose to build a more appropriate mixture model incorporating known domain-relevant noise characteristics for the analysis, including:

- Residual signal characteristics \sim Laplace distribution,
- Thermal noise characteristics \sim Rayleigh distribution,
- Surface wave scatter \sim K distribution,
- Rain characteristics \sim Gamma distribution,
- Miscellaneous characteristics \sim Normal distribution.

Thermal, rain and surface scatter have all been shown to have significant impact on target detection and sensor performance [11]. It is popular in the literature to use a Poisson distribution to model rain characteristics, however we choose to use a Gamma distribution in order to ensure all mixture components are continuous distributions. The problematic issue with this mixture comes when trying to optimise the hyper-parameters of the mixture since some of the variables within the log-likelihood derivatives are transcendental or non-invertible, preventing an EM approach. We therefore seek to optimise the hyper-parameters using scaled conjugate gradients. As is discussed in the literature, e.g. [2], optimising parameters from the K distribution using gradient descent from a maximum likelihood approach typically result in unrealistic parameters and convergence issues. We therefore use a bayesian approach for the parameter selection following [1] prior to optimising the remaining hyper-parameters using scaled conjugate gradients. This optimisation is repeated over all beams analysed with the weight for each mixture component ranging between 0 and 1.

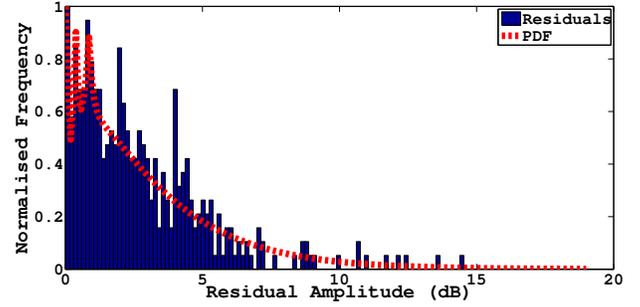


Fig. 1. A representative example of a fit of our compound noise mixture mode to the residual noise in beam 26 in the 16th second. The non-gaussian and multi-peaked nature of the residual noise is evident.

The result of this modelling process is an optimised noise model for a separate noise process in each of the beams. The fit of our compound noise mixture model against the actual residuals over a 1 second interval using beam 26 as an exemplar from the real data we used later is shown in figures 1. The non-Gaussian nature of the residuals is evident. The shape and scale of this distribution varies over each beam. This combination of an NAR signal and noise mixture process takes 1.014 seconds to fit each beam on a 1.7GHz core i5 processor with 8GB of RAM and can be run in parallel.

After discussing the multi-beam case, we will subsequently explain how these high-dimensional noise distributions are then visualised in 2D. It is expected that any anomalous behaviour in the environment would then be apparent as a significant departure from this expected noise model, and hence such a departure should be evident in a low dimensional visualisation of this noise space.

B. Beam grouping

As an alternative to treating each beam signal separately, we can choose to combine groups of beams into neighbouring clusters to create a set of **vector** autoregressive models. For the purposes of the data in this paper, we choose to analyse the beams in overlapping groups of 5 beams via a shifting window, i.e. for the 32 beams analysed there are 28 groups. Using a nonlinear Vector AR process we seek to find the weight matrix, \mathbf{B} , corresponding to the matrix of signal delays, \mathbf{U} , and the nonlinear function, f which matches the observed vector signal, \mathbf{S} , as closely as possible in a mean-squared error sense. There is assumed additive noise, ϵ in the vector process.

$$\mathbf{S}_g = f(\mathbf{B}_g \mathbf{U}_g) + \epsilon_g = \hat{\mathbf{S}}_g + \epsilon_g \quad (2)$$

where g is the group index. We again use an RBF network with 100 centres randomly chosen from the interpolated signal with an 8th order delay used for the nonlinear vector auto-regression. We choose an ' $r \log(r)$ ' activation function for interpolation and train a new network on each group. Unfortunately the noise model specified for the single beam case does not generalise to vector processes. So as an alternative to seeking dissimilarities between noise models, we use the covariance of the approximations in order to compute the dissimilarity between groupings. ie for each group we use

the covariance as a proxy for the noise model:

$$\mathbf{S}_g = \frac{1}{M-1} (\mathbf{S}_g - \hat{\mathbf{S}}_g)^T (\mathbf{S}_g - \hat{\mathbf{S}}_g) \quad (3)$$

where M is the number of approximated points.

Whether we use the single-beam or multi-beam approach, the output in both cases is a high dimensional representation of the uncertainty between our signal model and the observed data. The next phase is to create a low-dimensional representation of these high-dimensional noise processes. Since we need to retain the concept of ‘similarity’ (because the aim of this study is to use departures from what is regarded as normal for the situation captured by the sensors), we focus on a method for topographic representations of the noise models.

III. NEUROSCALE: TOPOGRAPHIC DIMENSIONALITY REDUCTION

We seek a dimension-reducing, *topographic* transformation of the noise data for the purposes of visualisation and analysis. By ‘topographic’, it is implied that the geometric structure as determined by pairwise relationships of the data is preserved in the transformation. This is a requirement that relative ‘dissimilarities’ are preserved on average, if possible. For this work we consider both deterministic and probabilistic approaches to the visualisation mechanism called NeuroScale. Although some alternatives exist, the NeuroScale approach is most suited to the requirements of this problem domain. A constrained Gaussian Process model has been proposed as an alternative topographic visualisation mapping [5], however these impose a gaussian likelihood over the observations and do not work directly with dissimilarity data. Another nonlinear visualisation method using dissimilarity data as input is the Stochastic Neighbour Embedding model [4], however this does not create a prototype trained feed-forward mapping which we need in our problem domain.

A. NeuroScale

The NeuroScale approach [6], [7] employs a nonlinear transformation $\{\mathbf{f} : \mathbb{R}^n \rightarrow \mathbb{R}^m : \mathbf{f}(\mathbf{x}) = \mathbf{y}\}$ from the original configuration space that maps into the feature space. We choose the class of nonlinear parameterised transformations provided by RBF networks. This has the advantage that a *transformation* is obtained, allowing interpolation which distinguishes this approach to other methods of low dimensional visualisation. The model parameters are adjusted to minimise the global STRESS measure analogous to the classical multidimensional scaling method, (but now with the advantage that a transformation is provided and not just a mapping):

$$E = \sum_{p=1}^P \sum_{q < p} \frac{[d_n(\mathbf{x}_p, \mathbf{x}_q) - d_m(\mathbf{y}_p, \mathbf{y}_q)]^2}{d_n(\mathbf{x}_p, \mathbf{x}_q)} \quad (4)$$

where $d_n(\mathbf{x}_p, \mathbf{x}_q)$ is the ‘similarity’ between data objects in the original space (often, but not necessarily, taken to be a Euclidean distance between points $\|\mathbf{x}_p - \mathbf{x}_q\|$) and $\{d_m(\mathbf{y}_p, \mathbf{y}_q) = \|\mathbf{y}_p - \mathbf{y}_q\|\}$ are the distances in the latent visualisation space (which can also be relaxed if prior knowledge dictates otherwise). The visualisation space dimensionality is often, but not exclusively, taken to be $m = 2$. The latent points

\mathbf{y} are generated by the RBF network, given the data points as input. That is, $\mathbf{y}_q = \mathbf{f}(\mathbf{x}_q; \mathbf{\Lambda})$, where \mathbf{f} is the nonlinear transformation effected by the Radial Basis Function model with parameters, i.e. weights and any kernel smoothing factors, $\mathbf{\Lambda}$. The (squared) ‘distances’ in the feature space (assuming euclidean discrepancies) may thus be given by

$$d_m^2 = \|f(\mathbf{y}_q) - f(\mathbf{y}_p)\|^2 = \sum_{l=1}^m \left(\sum_k \lambda_{lk} [\phi_k(\|\mathbf{x}_q - \boldsymbol{\mu}_k\|) - \phi_k(\|\mathbf{x}_p - \boldsymbol{\mu}_k\|)] \right)^2 \quad (5)$$

The topographic nature of the transformation is imposed by the *STRESS* term which attempts to match the relative inter-point dissimilarities in the latent visualisation space with the dissimilarities in the input space. Note that nowhere is an isolated point vector needed; only measures of pairwise dissimilarity. A by-product of this STRESS measure is that the projected points sit on arbitrarily scaled axes which only represent the relative dissimilarity between points and have no bearing on the observed data.

Also, note that central to this transformation is the assumption of provided ‘distance’ functions d_n, d_m in input and latent spaces, and also the choice of interpolating basis spline functions, $\phi(\dots)$. Also, note that this RBF is distinct to the use of the RBF previously. In both sections the RBF is used as a convenient nonlinear interpolation tool, but here its use is explicit for the dimension-reduction aspect, whereas previously it was used as a mechanism to effect a nonlinear autoregressive time series prediction method,

B. Probabilistic NeuroScale

The standard pointwise NeuroScale architecture was extended in [10] to accommodate a restricted level of uncertainty in a probabilistic framework. The uncertainties in the observed high-dimensional space, namely the noise models in the single-beam case, or the covariance matrices in the multi-beam model, are preserved in the mapping where only the means of the low dimensional distributions are changed in the mapping phase. The pointwise dissimilarity measure in the visualisation space, d_m , needs to be altered to incorporate appropriate dissimilarities depending on the objects in the high dimensional space. The measures of dissimilarity in the data space are similarly chosen to reflect the noise objects in the high dimensional space as follows.

For example, for the single-beam case we have 32 explicit mixture noise models in a given time period. For the multi-beam case we have 28 full covariance matrices. In each case we need a measure of (dis)similarity. For probability distributions the Kullback-Leibler divergence (KL) is often considered between distributions P and Q :

$$KL(P\|Q) = \int P(x) \left(\log\left(\frac{P(x)}{Q(x)}\right) \right) dx \quad (6)$$

For mixture distributions there are approximations to the KL divergence, see [3] for a description of these methods, but these still require the computation of the KL divergence between each individual distribution from the mixture. This is problematic as the KL divergence cannot be computed between the K distribution and any other and there are problems

with the integrals required when comparing the Laplace and Rayleigh distributions. There are alternative dissimilarity measures between probability distributions to the KL divergence such as the Bhattacharyya or Hellinger distances [8], however the integrals required are still analytically intractable. For this reason the dissimilarity measure we chose to represent divergences between noise distributions in the single-beam case is the Kolmogorov-Smirnov distance:

$$d_{KS}(\epsilon_{b1}, \epsilon_{b2}) = \sup_x (F_{b1}(x) - F_{b2}(x)) \quad (7)$$

where $F_b(x)$ is the cumulative distribution function over beam b . This measure is tractable since the cumulative distribution function of a mixture is equivalent to the weighted sum of the individual cumulative distributions, which are well defined functions.

For the multi-beam case, the proxy for a noise model is a covariance matrix for each beam group. Taking this as representing the uncertainty in a high dimensional Gaussian noise model, the Kullback-Leibler divergence for such a model leads to a distance function between two covariance matrices which can be expressed as

$$d_n(\Sigma_1, \Sigma_2) = 0.5 \text{Tr}[(\Sigma_1^{-1})\Sigma_2] - 0.5 \log(\det[\Sigma_1]/\det[\Sigma_2])$$

Note that the mean of the distributions is taken to be zero as that aspect has already been modelled by the NLVAR RBF network.

Since the previously mentioned Probabilistic NeuroScale requires a scalar measure for uncertainty representation in the visualisation space we chose to use the deterministic approach to NeuroScale when visualising the dissimilarity between these multi-beam models.

A simple choice of distribution for the low dimensional visualisation space for mapping the single-beam noise models is a set of Normal distributions with identical spherical variances, allowing for the central limit theorem to govern the high dimensional space and mapping, as well as being characterised simply by the first 2 moments. In this case the variances in the visualisation space are derived from the variances of the noise models in the data space, whereas the distances between the means of the Normal distributions reflect the dissimilarities between the respective noise mixture models.

In effect this approach implies a uniformly weighted gaussian mixture model in the visualisation space. We will use this interpretation to provide a continuous probability distribution in the visualisation space later.

IV. RESULTS

A. Data

The real world data used for these experiments was provided by DSTL. We would like to thank DSTL for making this data available for this prototype study. The array consisted of 32 hydrophones configured in an approximately linear array with a target ship transmitting a signal whilst traversing the length of the array. This is a shallow water scenario with low levels of rain and thermal noise expected. The frequency response of the hydrophones was 124 - 249Hz. During the

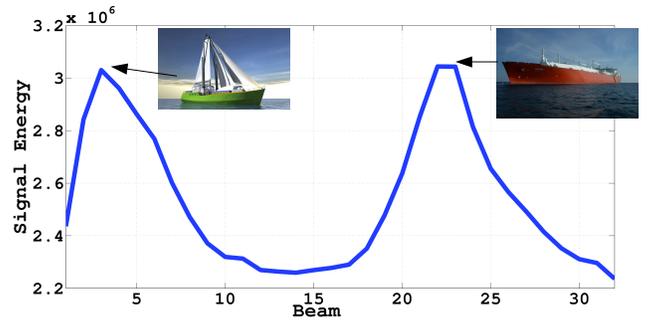


Fig. 2. Signal energy for scenario at 16 seconds into the exercise. We see two targets present in beams 1 to 9 and beams 20 to 29.

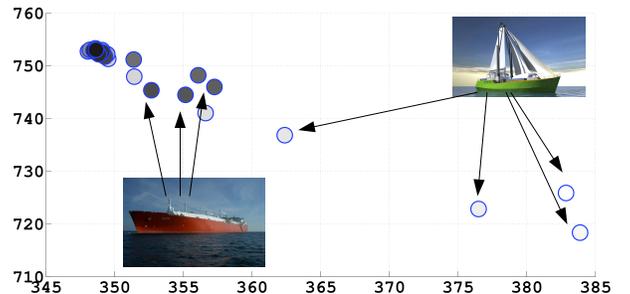


Fig. 3. Projection of 16th second of data using nonlinear vector autoregressive process with KL dissimilarities in high dimensional space and a deterministic NeuroScale mapping into latent space. Shading of beams is white to black ordered from group 1 (beams 1-5) to 28 (beams 28-32) respectively. We see the targets are present in beams 1-6 and 21-25 which appear as anomalies. (The images used are only schematics for the real targets).

scenario another target is present intermittently in the central section of the array. It should also be noted that some low frequency ($\sim 50\text{Hz}$) noise is present from an unknown origin. This 50Hz tonal is fit by the NAR models across all beams. The results shown in this paper relate to a one second segment (16 seconds into the exercise) in which the two targets are present. The signal energy from the array at this time is shown in Figure 2.

B. Beam grouping results

First, we illustrate the results of the NLVAR approach projected using the deterministic approach to Neuroscale in Figure 3. This projection was created by minimising the STRESS measure from equation (4) between the 2 dimensional pointwise Euclidean distances and the KL divergences between groups of the covariance matrices for the 16th second of data, each of which is given by equation (3), in the high dimensional observation space. This visualisation shows the relative similarity of the groups of beams where each point in the plot corresponds to one of the beam groups. The axes in this visualisation space refer only to the relative dissimilarity of the points and therefore have no significance in contrast with the observation space. We see that there is a clear cluster of beams containing only noise and a separate group of ‘outlier’ beams which, as can be verified from the signal energy plot in Figure 2 contain target signal elements.

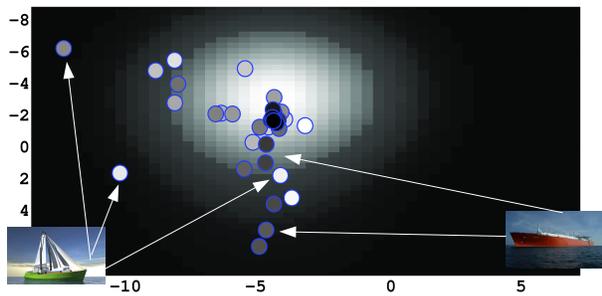


Fig. 4. Projection using nonlinear autoregressive process with probabilistic NeuroScale. We see that areas of high probability (white) density contain beams with primarily noise events, whereas target-containing beams appear as outliers in low probability areas (dark grey). Shading of beams is white to black ordered from beam 1 to beam 32 respectively.

C. Beam-by-beam results

Similarly, the visualisation space for the beam-by-beam approach is constructed using probabilistic NeuroScale.

Once the noise model specified in section IIA is estimated (using scaled conjugate gradients on each beam), we compute the dissimilarity between the noise models across each beam using the KS distance from equation (7). In order to incorporate the signal into the dissimilarity framework as well as in the observation space we also compute the Euclidean distance between the power spectral density of the predicted target signal characteristics from the computed NLAR model. These two dissimilarity measures are scaled and combined equally, providing the high dimensional dissimilarity measure required for the implementation of NeuroScale. We compute the variance of the noise distribution after optimising the hyper-parameters which can then be estimated using the low dimensional Gaussian distribution used as the probabilistic alternative to the deterministic latent points. This low-dimensional latent distribution is then equivalent to an equally weighted 32-component Gaussian Mixture Model. An intensity plot of this distribution is shown in Figure 4 with the means of the distributions plotted as points. Again there is a grouping of noise sources located in regions of higher probability density with the low probability regions being occupied by target signal beams.

V. CONCLUSION

We have shown two approaches in which an arbitrary hydrophone array of signals may be represented in a two-dimensional form which not only preserves relative information between the sensors, but provides a visual analytics in which anomalous behaviour such as potential targets reveal themselves as marked deviations away from expected noise behaviour.

The final stage visual analytics approach was common across the two approaches (the use of a Neuroscale model), but the input data analysis was different (either treating the sensors as a set of one dimensional times series, or as treating them in groups of nonlinear vector autoregressive models). Common to both approaches was the ideology that the modelling priority

was to capture the ‘noise’, or more specifically the uncertainty between the predictor models and the observed data. Visualising these high dimensional noise processes warranted a modified Neuroscale to map distributions to points (the deterministic approach) or distributions to distributions (the probabilistic approach). The probabilistic approach allowed the detection of ‘signals’ as events occurring in the low probability regions of the Neuroscale space. The separation of signals was more apparent in the deterministic approach, although the probabilistic approach offers the potential for a more intuitive and user-friendly mechanism for interpretation and anomaly detection by operators.

Although the primary focus of this paper is on methods, we demonstrated capability on real world sonar data where we demonstrated how realistic, compound mixture models for the noise in the observed data can be used in conjunction with nonlinear dynamical time series predictors to estimate and compare observed signals. The real data example used only 32 beams, but the principle extends to very large sensor arrays in that the low dimensional Neuroscale space does not depend on the input data dimensionality, only on the matrix of relative dissimilarities between the input data channels.

Extensions of this approach currently underway include the incorporation of generic mixture models to allow easy generalisation to permit varying weather conditions and underwater thermal profiles and terrain; the use of Fisher Information to reflect the local information content in the visualisation space; and the interpretation of this visualisation model within the framework of a full generative process.

REFERENCES

- [1] ABRAHAM, D.A. AND LYONS, A.P. 2010 Reliable Methods for Estimating the K-Distribution Shape Parameter *Oceanic Engineering, IEEE Journal of*, **35**(2), 288–302.
- [2] BAREŠ, R. 2011 Environmentally Adaptive Noise Estimation for Active Sonar *In PhD Thesis*
- [3] HERSHEY, J.R. AND OLSEN, P.A. 2007 Approximating the Kullback Leibler Divergence Between Gaussian Mixture Models *Acoustics, Speech and Signal Processing, 2007. ICASSP 2007. IEEE International Conference on*, **4**.
- [4] HINTON, G. AND ROWEIS, S. 2002 Stochastic Neighbor Embedding *Advances in Neural Information Processing Systems* **15**, 833–840.
- [5] LAWRENCE, N. D. AND QUIÑONERO-CANDELA, J. 2006 Local Distance Preservation in the GP-LVM Through Back Constraints. *Proceedings of the 23rd International Conference on Machine Learning*, 513–520.
- [6] LOWE, D. 1993 Novel ‘Topographic’ Nonlinear Feature Extraction using Radial Basis Functions for Concentration Coding in the ‘Artificial Nose’ 3rd IEE International Conference on Artificial Neural Networks 1993, Conference Publication number **372** The Institute of Electrical Engineers, 95–99.
- [7] LOWE, D. AND TIPPING, M. 1996 Feed-forward neural networks and topographic mappings for exploratory data analysis. *Neural Computing and Applications* 1996 **4**, 83–95.
- [8] PEKALSKA, E. AND DUIN, R.P.W. 2005 The Dissimilarity Representation For Pattern Recognition. *World Scientific*.
- [9] RICE, I. & BENTON, R. & HART, L. AND LOWE, D. 2013 Analysis of Multibeam SONAR Data using Dissimilarity Representations. *In Proc. IMA Mathematics in Defence* 2013.
- [10] SIVARAKSA, M. AND LOWE, D. 2009 Probabilistic NeuroScale for Uncertainty Visualisation. *Information Visualisation, 2009 13th International Conference*, 74–79.
- [11] URICK, R. J. 1996 Principles of underwater sound, 3rd edition, McGraw-Hill.