Kernel methods for the detection and classification of fish schools in single-beam and multibeam acoustic data

Bart Buelens, Tim Pauly, Raymond Williams, and Arthur Sale

Buelens, B., Pauly, T., Williams, R., and Sale, A. 2009. Kernel methods for the detection and classification of fish schools in single-beam and multibeam acoustic data. – ICES Journal of Marine Science, 66: 1130–1135.

A kernel method for clustering acoustic data from single-beam echosounder and multibeam sonar is presented. The algorithm is used to detect fish schools and to classify acoustic data into clusters of similar acoustic properties. In a preprocessing routine, data from single-beam echosounder and multibeam sonar are transformed into an abstracted representation by multidimensional nodes, which are datapoints with spatial, temporal, and acoustic features as components. Kernel methods combine these components to determine clusters based on joint spatial, temporal, and acoustic similarities. These clusters yield a classification of the data in groups of similar nodes. Including the spatial components results in clusters for each school and effectively detects fish schools. Ignoring the spatial components yields a classification according to acoustic similarities, corresponding to classes of different species or age groups. The method is described and two case studies are presented.

Keywords: classification, detection of fish schools, kernel methods, multibeam sonar.

Received 5 August 2008; accepted 24 October 2008; advance access publication 2 February 2009.

B. Buelens and T. Pauly: Myriax Software Pty Ltd, GPO Box 1387, Hobart, TAS 7001, Australia. R. Williams and A. Sale: School of Computing and Information Systems, University of Tasmania, Private Bag 100, Hobart, TAS 7001, Australia. Correspondence to B. Buelens: tel: +61 3 62315588; fax: +61 3 62341822; e-mail: bart.buelens@myriax.com.

Introduction

The use of multibeam sonar in fishery-acoustic research has led to new insights in fish behaviour. Promising results have been reported on topics including vessel avoidance (Misund and Aglen, 1992; Soria *et al.*, 1996; Gerlotto *et al.*, 2004), schooling behaviour (Gerlotto *et al.*, 1999), predator–prey interaction (Noettestad and Axelsen, 1999; Axelsen *et al.*, 2001; Benoit-Bird and Au, 2003; Brehmer *et al.*, 2006), and fish migration (Hafsteinsson and Misund, 1995). These studies are qualitative, in the sense that no attempt is made to estimate the number of fish or total fish weight from the acoustic backscatter recorded by the multibeam sonar. For such quantitative work to be possible, the data have to be calibrated (Foote *et al.*, 2005). Equally important for quantitative work is the determination of the school volume to allow for the estimation of total fish weight using calibrated sonar data (Tang *et al.*, 2006). School detection is the process of determining the spatial extent of fish schools from sonar data. Besides detecting fish schools, it is of interest to determine which schools are likely to contain fish of the same species. The identification of species is achieved through classification of the sonar data. In this paper, kernel methods are presented as a means of achieving school detection and classification simultaneously, not only in multibeam sonar data, but also in single-beam echosounder data.

Material and methods

Data representation

The data under consideration are acoustic recordings collected by single-beam echosounder and multibeam sonar. A level of abstraction is introduced to represent data from both types of instrument in a unified manner. Such a representation allows for the expression of processing algorithms in general terms. The transformation of data recordings obtained from the instruments into this unified representation is referred to as preprocessing.

A preprocessed acoustic dataset consists of a collection of nodes. A node is a spatially and temporally referenced feature, vector $\mathbf{v}$, $\mathbf{v}(x, t, f)$, with x the spatial coordinates of $\mathbf{v}$, typically longitude, latitude, and depth; t the temporal coordinates of $\mathbf{v}$, typically date and/or time; and f the vector of features of $\mathbf{v}$, typically relating to acoustic energy.

As it is desirable to have datasets as small and compact as possible, redundancy in sets of nodes must be avoided; this will facilitate subsequent analysis. The following sections describe the preprocessing of single- and multibeam acoustic data into a representation by nodes and an approach to establishing a suitable set of features f.

Preprocessing single-beam echosounder data

Digital echosounder data comprise a sequence of pings, with each ping containing a series of data samples obtained by sampling the acoustic-return signal received by the echosounder. Typical sampling rates of the return signal are such that successive samples overlap in the sampling volume corresponding to one pulse length, resulting in high sample-to-sample correlations within pings. Furthermore, ping rates and vessel speeds are typically chosen such that there is significant overlap of sampling volumes from ping to ping, resulting in high ping-to-ping correlations of the sample data. In preprocessing echosounder data

into nodes, within-ping and across-ping correlations should be removed to attain a set of nodes with the least possible redundancy. Two approaches are discussed.

The first approach is a thorough removal of the correlations in the echosounder data. Decorrelating within-ping data can be achieved by deconvolution using the echo of the transmit pulse as observed from a point scatterer as the point-spread function (PSF), in a one-dimensional fashion. Decorrelating between-ping data is achieved by extending the deconvolution process to two dimensions and incorporating between-ping, sampling-volume overlaps in the filter. In practice, this is difficult to realize because of generally non-constant vessel speeds and vessel motion causing the between-ping, sampling-volume overlaps to vary.

The second approach is more pragmatic and easier to perform than the deconvolution approach. Redundancy in the data is removed by downsampling. The general idea is to reduce the number of samples, thereby taking into account the aim of producing nodes that are less correlated than the original echosounder data. This is achieved by averaging samples over a certain depth range and number of pings. To avoid residual correlation in the nodes, the depth range over which the averaging is conducted should be at least the height corresponding to one pulse length. The number of pings over which the average is taken should be chosen such that all pings for which the sampling interval overlaps with a given ping are included in the averaging. The spatial, temporal, and acoustic volume-backscattering coefficient averages are the respective components of the nodes $\mathbf{v}$, as defined above.

Preprocessing multibeam sonar data

Multibeam sonar data differ from single-beam echosounder data in that there is an additional dimension (Gerlotto *et al.*, 1999). The multibeam ping under consideration is two-dimensional, with samples typically arranged on some polar grid, with distances between samples in the angular direction increasing with range. This is a complication with respect to a downsampling approach, but it simultaneously facilitates the application of a deconvolution approach. Applying a deconvolution across pings suffers from the same difficulties as discussed for single-beam echosounder data.

Therefore, it is proposed to decorrelate data at the within-ping level only, by deconvolution using the echo of the transmit pulse observed from a single-point scatterer as the PSF. The PSF to be used can be established during typical calibration procedures, in which the response of a calibration sphere is measured with the sonar (Foote *et al.*, 2005). A common expectation-maximization algorithm for deconvolution is the Lucy–Richardson algorithm (Richardson, 1972; Lucy, 1974). Thresholding the deconvolved data at or above the noise level results in nodes of the form described above. This approach to preprocessing multibeam data is discussed in detail in Buelens *et al.* (2005).

Selecting features

Spatial and temporal components of nodes are implicitly determined through the procedures described above. It is also essential to include a set of features as a feature vector f with the nodes. The features should be chosen such that they allow analysis algorithms to find patterns in the set of nodes. Features typically include at least some known or derived acoustic information. In all approaches discussed above, nodes emerge as derived from a particular set of acoustic samples. Let A_i be the set of samples contributing to node i; the aggregated backscatter value is a feature defined as

$$\beta_i = \sum_{j \in A_i} w_j b_j, \tag{1}$$

with b_j the volume-backscattering coefficient of sample j and w_j the weight of sample j. The sample volumes can be chosen as the weights. Higher statistical moments of the backscatter values, such as standard deviation, skewness, and kurtosis, can be included as additional features. When data collected at multiple frequencies are available, features such as aggregated backscatter can be included for each frequency.

There is no limit or restriction to the type and number of features that can be included in the feature vector of the nodes. If information is known and expected to be useful, it can be incorporated. For example, texture information such as grey-level co-occurrence matrices (GLCMs) may prove useful in differentiating between fish species. GLCMs capture variability in echogram texture and have been used to separate plankton from fish (ICES, 2006). Another possibility is to include non-acoustic information available from sources other than the echosounder or sonar. For example, water temperature or salinity measurements can be included as features in the feature vector of the nodes.

Kernel methods

Background

Preprocessed data are the sets of nodes of the form described earlier. For a set of nodes $\Theta = \{\mathbf{v}_i\}$, $i = 1, \ldots, N$, the goal of a pattern-analysis method is to establish clusters of similar nodes. A cluster is a group of nodes, where it is expected that nodes within clusters are more similar than nodes from different clusters. The purpose of such an analysis in the context of acoustic data is to formulate the clustering problem such that resulting clusters correspond to aggregations of fish, or to species of fish.

Many clustering algorithms are available (Duda *et al.*, 2000; Hastie *et al.*, 2001; Bishop, 2006). A promising recent development in classification and clustering is the emergence of kernel methods (Scholkopf *et al.*, 1999; Muller *et al.*, 2001; Shawe-Taylor and Cristianini, 2004). Kernel methods have the advantage of being based on generally well accepted linear methods, which are rendered non-linear by sound principles from statistical learning theory (Vapnik, 1995). An additional advantage is the capability of handling multidimensional data for which the scales and metrics of different dimensions are completely different, which is the case for spatio-temporal nodes.

Kernel methods use a non-linear mapping ϕ to map the original data into a higher-dimensional space F:

$$\phi : V \mapsto F : \mathbf{v} \in V \mapsto \phi(\mathbf{v}) \in F. \tag{2}$$

The purpose of this mapping is to render problems that are non-linear in V, linear in F. In such situations, well established linear methods can be used in the space F. A key element of kernel methods is the application of the kernel trick. This is based on the observation that many linear methods employ distances between datapoints only and not the datapoints as such. The distance metric on F is derived from the norm induced by the inner product on F. Consequently, distances can be expressed in terms of inner products. To that effect, a kernel κ is defined on V

as

$$\kappa : V \mapsto \Re : \kappa(\mathbf{v}_1, \mathbf{v}_2) = \langle \phi(\mathbf{v}_1), \phi(\mathbf{v}_2) \rangle, \tag{3}$$

with $\langle , \rangle$ the inner product on F. From Equation (3), it is clear that the kernel is a mapping from the original space F to the real numbers directly. This bypassing of F is known as the kernel trick. Formulating expressions for κ directly, the mapping ϕ is defined only implicitly. One of the more common functions often chosen as a kernel is the Gaussian kernel,

$$\kappa(\mathbf{v}_1, \mathbf{v}_2) = \mathrm{e}^{-(\mathbf{v}_2 - \mathbf{v}_1)^2 / 2\sigma^2}, \tag{4}$$

with σ the width of the kernel. Other popular kernels include the polynomial, sigmoidal, and identity kernels (Shawe-Taylor and Cristianini, 2004).

Kernels for acoustic data nodes

Nodes in a space V have three components, namely the spatial-, temporal-, and feature-vector components. Initially, these components are considered separately, with kernels defined on the corresponding subspaces of V. Kernels κ [Equation (4)] are defined on each of the subspaces. It has been demonstrated that such kernels can be extended to kernels over the space V (Buelens, 2008). Let these kernels be κ_x, κ_t, and κ_f, respectively. Given these kernels, the following are kernels also (Muller *et al.*, 2001; Shawe-Taylor and Cristianini, 2004):

$$\kappa_{\mathrm{a}}(\mathbf{v}_1, \mathbf{v}_2) = \kappa_x(\mathbf{v}_1, \mathbf{v}_2) + \kappa_t(\mathbf{v}_1, \mathbf{v}_2) + \kappa_f(\mathbf{v}_1, \mathbf{v}_2), \tag{5}$$

$$\kappa_{\mathrm{m}}(\mathbf{v}_1, \mathbf{v}_2) = \kappa_x(\mathbf{v}_1, \mathbf{v}_2)\, \kappa_t(\mathbf{v}_1, \mathbf{v}_2)\, \kappa_f(\mathbf{v}_1, \mathbf{v}_2). \tag{6}$$

Kernel κ_{a} is additive, whereas κ_{m} is multiplicative. Generally, it is recommended to use additive kernels where similarities in one, two, or all three of the components will contribute to the overall similarity measure individually, whereas the multiplicative kernel requires joint levels of similarity in all three components.

Kernel methods for clustering

A widely used clustering algorithm is k-means (Lloyd, 1982), described in many texts on classification and pattern analysis (Duda *et al.*, 2000; Bishop, 2006). The application of k-means in the space F, through the use of kernels, is known as kernel k-means (Scholkopf *et al.*, 1998; Girolami, 2002). Because the kernel-k-means algorithm can be unstable in certain conditions, an alternative was proposed by Ng *et al.* (2002). This has become known eponymously as the NJW-algorithm. NJW is a spectral relaxation of kernel k-means (Dhillon *et al.*, 2005). Spectral methods are based on the spectrum of a matrix; the spectrum is the set of eigenvalues. In this case, the matrix under consideration is the kernel matrix **K,** defined as

$$\mathbf{K}_{ij} = \kappa(\mathbf{v}_i, \mathbf{v}_j), \tag{7}$$

for $i, j = 1, \ldots, N$. The purpose is to determine k clusters $C_1, \ldots, C_k$ such that

$$\begin{aligned} C_i &\subseteq \{\mathbf{v}_1, \ldots, \mathbf{v}_N\} \\ \cup_i C_i &= \{\mathbf{v}_1, \cdots, \mathbf{v}_N\} \\ C_i \cap C_j &= \{\,\}, \forall\, i, j. \end{aligned} \tag{8}$$

From Equation (4), it can be seen that the kernel function takes high values for similar datapoints and low values for dissimilar datapoints. Alternative kernel functions exhibit the same property. Based on this observation, the cut cost for a set of clusters of a dataset is defined as

$$Q_{\mathbf{K}} = \sum_{y_i \neq y_j} \kappa(\mathbf{v}_i, \mathbf{v}_j), \tag{9}$$

with y_i the cluster membership label of node $\mathbf{v}_i$:

$$y_i = m \Leftrightarrow \mathbf{v}_i \in C_m. \tag{10}$$

The cut cost $Q_{\mathbf{K}}$ is the sum of the values of the kernel matrix for nodes in different clusters. This is used as an optimization criterion to find the clustering with minimal cut cost. In the NJW algorithm, it is customary to normalize the kernel matrix using a diagonal matrix **D** with elements

$$\mathbf{D}_{ij} = \sum_j \kappa(\mathbf{v}_i, \mathbf{v}_j), \tag{11}$$

and defining the matrix **L,** sometimes referred to as the Laplacian, as

$$\mathbf{L} = \mathbf{D}^{-1/2} \mathbf{K} \mathbf{D}^{-1/2}. \tag{12}$$

The cut cost for **L** is defined as

$$Q_{\mathbf{L}} = \sum_{y_i \neq y_j} \mathbf{L}_{ij}. \tag{13}$$

The following procedure finds a clustering that minimizes $C_{\mathbf{L}}$ for a given matrix **L** (Ng *et al.*, 2001). Let $\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_N$ be the N eigenvalues of the matrix **L,** with $p_1, \ldots, p_N$ the corresponding eigenvectors. Retain the k eigenvectors $p_1, \ldots, p_k$ corresponding to the k largest eigenvalues and form the $N \times k$ matrix **P** by stacking the k eigenvectors in columns. Form the matrix **Z** by normalizing the rows of **P** to have unit length:

$$\mathbf{Z}_{ij} = \frac{\mathbf{P}_{ij}}{\sum_j \mathbf{P}_{ij}^{1/2}}. \tag{14}$$

Consider each row of the $N \times k$ matrix **Z** as a point $\mathbf{z}_i$ in a k-dimensional space and cluster these points into k clusters using the regular k-means algorithm (Lloyd, 1982). Finally, assign the original node $\mathbf{v}_i$ to cluster m only if $\mathbf{z}_i$ is assigned to cluster m.

The actual clustering is performed in a space of dimension k, whereas the standard kernel k-means operates in an N-dimensional space. This explains the superior stability of NJW over kernel k-means.

Results

The proposed kernel method is demonstrated through two case studies. Datasets from a single-beam echosounder and a multibeam sonar are analysed. Ground-truthing of the results is not straightforward, because it would require an independent observation of the fish *in situ*, in three dimensions for multibeam

sonar. Consequently, outcomes of accepted alternative methods and expert assessments are used to evaluate the results.

Case study 1: Southern Ocean dataset

Single-beam echosounder data were collected in 2004 by the Australian Antarctic Division (AAD) near Heard Island and the McDonald Islands in the Southern Ocean (dataset courtesy of Toby Jarvis, AAD). A Simrad EK60 echosounder was used with three acoustic frequencies: 38, 120, and 200 kHz. The data used in the present case study were collected for 55 min along a transect of ~10 nautical miles. In line with standard AAD operating procedures (Jarvis, 2006), data are downsampled to 2 m high cells containing sample data from 25 pings.

An additive Gaussian kernel is used, with two spatial coordinates: distance along the cruise track and depth in the water. The features relating to the acoustic energy are the mean volume-backscattering coefficients at each of the three frequencies. Temporal information is not used. The three backscatter levels are predominantly used in the analysis, with smaller contributions from the spatial coordinates, as is observed from the values used for the parameter σ (Table 1).

The AAD processing scheme (Jarvis, 2006) aims at identifying four classes of scatterer: resonant scatterers and fluid-like scatterers; at depths shallower than 100 m the fluid-like scatterers are further divided into two classes, small and large. The AAD scheme includes a classification procedure based on the responses at the three frequencies (Korneliussen and Ona, 2003). This procedure is the combined result of expert knowledge applied in an *ad hoc* manner and independent validation of the outcomes (Korneliussen and Ona, 2003; Jarvis, 2006). The results of the application of this procedure to the data under consideration provide a reference frame for comparison with the results obtained using the kernel methods, which have a statistical basis and are automatic. The NJW algorithm is run with $k = 4$. Through expert assessment, each of the resulting four clusters of nodes is labelled. The results are compared with those obtained by the AAD scheme (Figure 1). Classes representing fluid-like scatterers, small fluid-like scatterers, large fluid-like scatterers, and resonant scatterers are coloured green, orange, red, and blue, respectively. Although there is overall similarity, some differences are observed. The layer of resonant scatterers (in blue) is thinner in the AAD classification (Figure 1a) than in the kernel classification (Figure 1b). Furthermore, the kernel methods classify a number of nodes at greater depths as resonant scatterers, whereas the AAD approach does not. Expert assessment of backscatter data at 120 and 200 kHz provides evidence in favour of the kernel methods outcome. Another difference is the identification of large fluid-like scatterers (in red): the kernel method determines a significant layer of such scatterers, whereas the AAD method only identifies some seemingly randomly distributed scatterers. The most likely reason for this effect is that the kernel method takes the spatial components of the data into account, whereas the AAD method does not.

Table 1. Parameters used in the kernel function in case study 1.

Component	σ
Distance cruise track	0.01
Depth	0.001
Backscatter at 38 kHz	0.8
Backscatter at 120 kHz	0.8
Backscatter at 200 kHz	0.8

Case study 2: Lake Opeongo dataset

In 1999, a Kongsberg Mesotech SM2000 multibeam sonar was deployed on Lake Opeongo, Canada, during a joint project of SonarData Pty Ltd, Kongsberg Mesotech, and the Ontario Ministry of Natural Resources and Scientific Assessment

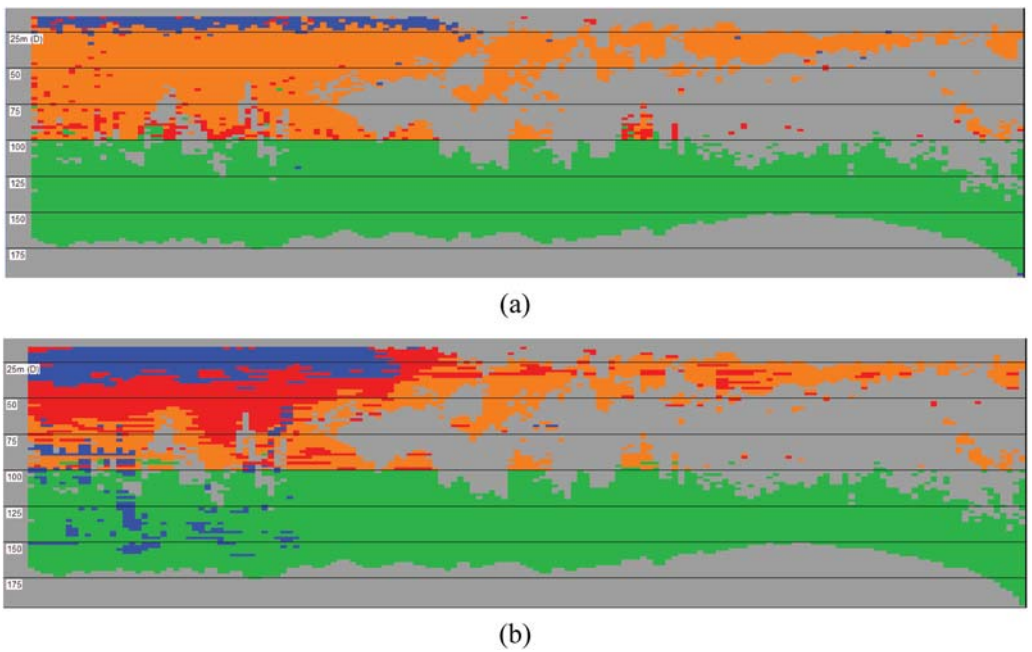

Figure 1. Case study 1: (a) classification according to the AAD scheme; (b) classification using the kernel method. Fluid-like scatterers are displayed in green, small fluid-like scatterers in orange, large fluid-like scatterers in red, and resonant scatterers in blue.

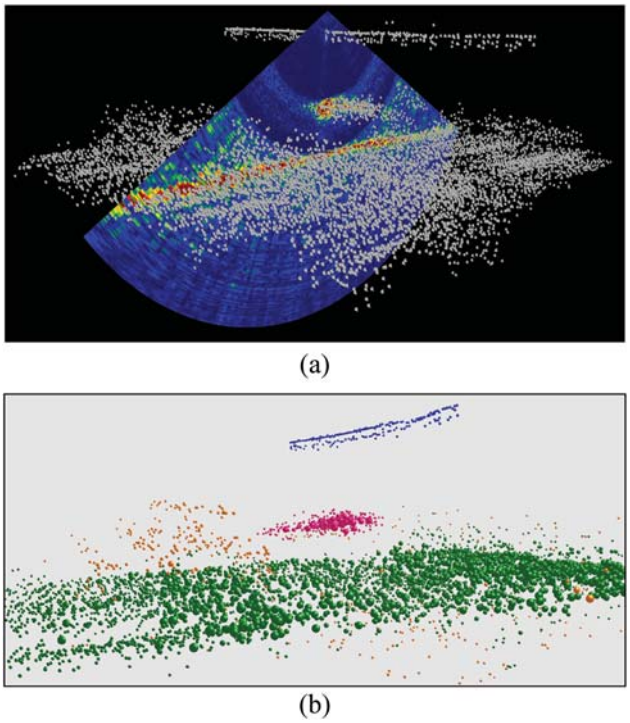

Figure 2. Case study 2: (a) the nodes with some of the original data; (b) four clusters of nodes as detected by the kernel method. The nodes representing the fish school are displayed in pink, the seabed in green, the foreign instrument in blue, and noise in orange.

Table 2. Parameters used in the kernel function in case study 2.

Component	σ
Longitude	10.0
Latitude	10.0
Depth	4.0
Point backscatter	1.0
Mean backscatter	2.0

Technology Laboratory. The sonar operational frequency was 200 kHz, collecting data from 128 beams over a 120° swathe. The data file from this case study contains recordings of a school of *Coregonus artedii* (lake herring or cisco). These data are displayed in plate 3.5 of Simmonds and MacLennan (2005).

Nodes are derived using the deconvolution approach (Figure 2a). The features used are the sample volume-backscattering coefficient at each node and the mean backscatter of all contributing samples as given in Equation (2). The NJW algorithm is applied with parameters as given in Table 2, for a multiplicative Gaussian kernel. The multiplicative kernel is used because the goal of the analysis is to detect schools, represented by clusters of spatially neighbouring nodes that, at the same time, possess similar acoustic features.

The results are illustrated in Figure 2b. Four clusters are distinguished: the fish school (pink), the seabed (green), the foreign instrument present in the acoustic beams (blue), and the noise (orange). The results obtained through the automated-classification method match those obtained through expert visual assessment of the data. Spatial components and first-order statistics of the volume-backscatter coefficients are clearly sufficient to separate echoes from fish, the seabed, and noise in this dataset using kernel methods.

Conclusions

Case study 1 demonstrates that the NJW kernel-clustering method can classify multifrequency, single-beam echosounder data in a fully automated manner, the results of which are comparable with visual classifications done by an acoustician. Case study 2 demonstrates the ability of the method to detect fish schools in multibeam sonar data. The flexibility in the use of the spatial components of the nodes adjusts the algorithm to act either as a pure feature-classification algorithm (no spatial components) or as a purely geometric, schools-detection algorithm (spatial components only). The power of the algorithm is in the combined usage of spatial, temporal, and feature components.

Future work can include research on automated selection of the optimal set of parameters and of the number of clusters to be determined. Comparison with other algorithms for classification, applied to a wide range of datasets, should provide a basis for formulating recommendations for hydroacoustic data-processing guidelines.

Acknowledgements

We thank Toby Jarvis of the AAD for making available the data file used in case study 1.

References

Axelsen, B. E., Anker-Nilssen, T., Fossum, P., Kvamme, C., and Noettestad, L. 2001. Pretty patterns but a simple strategy: predator–prey interactions between juvenile herring and Atlantic puffins observed with multibeam sonar. Canadian Journal of Zoology, 79: 1586–1596.

Benoit-Bird, K., and Au, W. 2003. Hawaiian spinner dolphins aggregate midwater food resources through cooperative foraging. Journal of the Acoustical Society of America, 114: 2300.

Bishop, C. M. 2006. Pattern Recognition and Machine Learning. Springer, New York.

Brehmer, P., Lafont, T., Georgakarakos, S., Josse, E., Gerlotto, F., and Collet, C. 2006. Omnidirectional multibeam sonar monitoring: applications in fisheries science. Fish and Fisheries, 7: 165–179.

Buelens, B. 2008. Scientific data mining for spatio-temporal hydroacoustic data sets. PhD thesis, School of Computing and Information Systems, University of Tasmania, Hobart.

Buelens, B., Williams, R., Sale, A., and Pauly, T. 2005. Model inversion for midwater multibeam backscatter data analysis. *In* IEEE Oceans '05 Europe, Brest, France, pp. 431–435.

Dhillon, I. S., Guan, Y., and Kulis, B. 2005. A unified view of kernel k-means, spectral clustering and graph cuts. Technical Report, Department of Computer Science, University of Texas at Austin, Austin, TX, USA.

Duda, R. O., Hart, P. E., and Stork, D. G. 2000. Pattern Classification, 2nd edn. Wiley-Interscience, New York.

Foote, K. G., Chu, D. Z., Hammar, T. R., Baldwin, K. C., Mayer, L. A., Hufnagle, L. C., and Jech, J. M. 2005. Protocols for calibrating multibeam sonar. Journal of the Acoustical Society of America, 117: 2013–2027.

Gerlotto, F., Castillo, J., Saavedra, A., Barbieri, M. A., Espejo, M., and Cotel, P. 2004. Three-dimensional structure and avoidance behaviour of anchovy and common sardine schools in central southern Chile. ICES Journal of Marine Science, 61: 1120–1126.

Gerlotto, F., Soria, M., and Freon, P. 1999. From two dimensions to three: the use of multibeam sonar for a new approach in fisheries

acoustics. Canadian Journal of Fisheries and Aquatic Sciences, 56: 6–12.

Girolami, M. 2002. Mercer kernel-based clustering in feature space. IEEE Transactions on Neural Networks, 13: 780–784.

Hafsteinsson, M. T., and Misund, O. A. 1995. Recording the migration behaviour of fish schools by multibeam sonar during conventional acoustic surveys. ICES Journal of Marine Science, 52: 915–924.

Hastie, T., Tibshirani, R., and Friedman, J. 2001. The Elements of Statistical Learning. Springer, New York.

ICES. 2006. Report of the Working Group on Fisheries Acoustics Science and Technology (WGFAST), 27–30 March 2006, Hobart, Tasmania. ICES Document CM 2006/FTC: 01. 56 pp.

Jarvis, T. 2006. Hydroacoustic data post-processing using SonarData Echoview v3.50. Australian Antarctic Division, Kingston.

Korneliussen, R. J., and Ona, E. 2003. Synthetic echograms generated from the relative frequency response. ICES Journal of Marine Science, 60: 636–640.

Lloyd, S. P. 1982. Least squares quantization in PCM. IEEE Transactions on Information Theory, 28: 129–137.

Lucy, L. B. 1974. Iterative techniques for rectification of observed distributions. Astronomical Journal, 79: 745–754.

Misund, O. A., and Aglen, A. 1992. Swimming behaviour of fish schools in the North Sea during acoustic surveying and pelagic trawl sampling. ICES Journal of Marine Science, 49: 325–334.

Muller, K. R., Mika, S., Ratsch, G., Tsuda, K., and Scholkopf, B. 2001. An introduction to kernel-based learning algorithms. IEEE Transactions on Neural Networks, 12: 181–201.

Ng, A. Y., Jordan, M. I., and Weiss, Y. 2001. On Spectral Clustering: Analysis and an algorithm. Advances in Neural Information Processing Systems, 14. Proceedings of the 2001 Neural Information Processing Systems (NIPS) Conference. Ed. by T. G. Dietterich, S. Becker, and Z. Ghahramani.

Noettestad, L., and Axelsen, B. E. 1999. Herring schooling manoeuvres in response to killer whale attacks. Canadian Journal of Zoology, 77: 1540–1546.

Richardson, W. H. 1972. Bayesian-based iterative method of image restoration. Journal of the Optical Society of America, 62: 55.

Scholkopf, B., Burges, C. J. C., and Smola, A. J. 1999. Advances in Kernel Method—Support Vector Learning. MIT Press, Cambridge, MA.

Scholkopf, B., Smola, A. J., and Muller, K. R. 1998. Nonlinear component analysis as a kernel eigenvalue problem. Neural Computation, 10: 1299–1319.

Shawe-Taylor, J., and Cristianini, N. 2004. Kernel Methods for Pattern Analysis. Cambridge University Press, Cambridge, UK.

Simmonds, E. J., and MacLennan, D. N. 2005. Fisheries Acoustics: Theory and Practice 2nd edn. Blackwell Science, Oxford, UK.

Soria, M., Freon, P., and Gerlotto, F. 1996. Analysis of vessel influence on spatial behaviour of fish schools using a multi-beam sonar and consequences for biomass estimates by echo-sounders. ICES Journal of Marine Science, 53: 453–458.

Tang, Y., Iida, K., Mukai, T., and Nishimori, Y. 2006. Estimation of fish school volume using omnidirectional multi-beam sonar: scanning modes and algorithms. Japanese Journal of Applied Physics, Part 1: Regular Papers, Brief Communications and Review Papers, 45: 4868–4874.

Vapnik, V. 1995. The Nature of Statistical Learning Theory. Springer, New York.

doi:10.1093/icesjms/fsp004