fuzzy clustering of Oceanographic Sound Speed Profiles for Acoustic Characterization

John Dubberley and Robert Zingerelli

Naval Research Laboratory, Bldg. 1005 Rm D-23, Stennis Space Center, MS 39529, USA
john.dubberley@nrlssc.navy.mil

Historic oceanographic sound speed profiles have traditionally been grouped by area and time period, usually one degree square area and monthly time. After grading the profiles, mean profiles and standard deviations are calculated from the accepted profiles and in the acoustics community they are then used to predict the expected acoustic response of the region. Here the historic profiles in NOAA's World Ocean Database 2005 will be divided into the same area and time periods, but in subsets with a sufficient number of profiles, fuzzy clustering will be employed on acoustically relevant oceanographic parameters (mixed layer depth, surface temperature, sound speed gradient, etc) to divide the population into multiple clusters. A parabolic equation acoustic transmission model is then applied on the WOD2005 statistical profiles and on the fuzzy cluster populations. Conclusions will be drawn about the suitability of this clustering to capture the variability of acoustic response at a given time and place.

1 Introduction

Acoustic performance predictions for an area depend on the oceanographic climatology available. Present climatologies such as the National Oceanic and Atmospheric Administration's (NOAA) World Ocean Database 2005 (WOD 2005)[1] break the oceans into regions and time periods and assume that the probability distribution of the profiles is gaussian. A normal probability distribution is usually a safe first guess at an unknown probability distribution, however especially in tropical and subtropical regions it is known that the temperature and salinity profiles during a given month can fall into two distinct groups: those without a significant mixed layer depth and those with significant mixed layer depths caused by some storm event. Here we will introduce fuzzy clustering to break a set of profiles into two populations and then use an average of parabolic equation method (PE) model runs and the variance of this average to show that the two populations and the combined populations lead to similar but statistically different acoustic predictions. These improved climatologies could also be used to improve oceanographic temperature and salinity predictions.

2 Method

2.1 Fuzzy Clustering

Rather than forcing the analysis function to make each point belong to a particular cluster, fuzzy cluster analysis lets the points have partial memberships. It is tempting to see these partial memberships as probabilities of membership, but this view leads one into thinking that the datum is only in one cluster whose location is obscured from us as opposed to the correct idea that the datum is partially in each cluster. Fuzzy cluster analysis finds the degree of membership of a data point in each cluster. With the caveat that the sum of the memberships for a point must equal unity and that the point has to have a measurable membership in each cluster. L.A.Zadeh[2] introduced the fuzzy sets to model imprecise propositions. Later algorithms were developed for using these sets in pattern recognition[3,4]. Here we will be using clustering algorithms developed by C. Borgelt [5] that downloadable from: http://www.borgelt.net/cluster.html.

By way of introducing Fuzzy Clustering for oceanographic profiles we will first examine clustering of a littoral oceanographic areas of interest. The temperature profiles can vary over time and place widely through out the large number of historical measurements available. It may be difficult to get an understanding of the underlying environmental forcing mechanisms by just looking at the raw data sets. Oceanographers and their customers have found that dividing the data into areas and seasons of similar profiles helps explain the environmental variability and forcing functions.

These provinces have been found to be useful, however determining an area's provinces and seasons has traditionally been more of an art than a science. Historically characteristics of the data would be taken and then grouped in some fashion. For example the Naval Underwater Systems Center (NUSC) grouped deep water data by the temperature profiles similarity at given depths [Podeszwa]. There has been cluster analysis used examine sound speed profiles in the Gulf of Alaska [Moustafa]. Where data points are forced into clusters by minimizing the sum of:

$$d_{rs} = \mathrm{sqrt}(\textstyle\sum_{I=1,n}(c_{ri} - c_{si})^2) \qquad (1)$$

over each element in a set of predetermined size clusters where d is the distance between sound-speed profiles and c_{ri} , c_{si} are the respective sound-speeds at the I^{th} depth. These clusters are then grouped repeatedly by the same method until all points in one set. Then one determines which level of clustering the data seems to naturally cluster. This method has been commonly used in deep water where generally after a certain depth all profiles are similar and therefore can be universally trimmed to that depth causing the profiles to be comparable at all points. In shallow water however, point by point comparison is not generally possible given the common occurrence of significant differences below the depth of the shallowest profile in the data.

Here Fuzzy Clustering will be applied to overcome this difficult for an area off the North Carolina coast. The profiles were clustered on the temperature parameters. Temperature was chosen so that Expendable Bathythermographs (XBT) data could be included without having to imply salinity profiles for each XBT. The simplification was determined to be valid except for near shore area effected by fresh water discharges, however all such areas were very shallow in this case and can be seen as one province. The 2860 profiles used were taken throughout the year from 1990 onward.

Each parameter is treated as if it is independent of the other parameters and weighted equally, making it best to limit the number of parameters used. In this case surface temperature, surface iso temperature depth, mean temperature, and summed temperature slope were used as the temperature parameters.

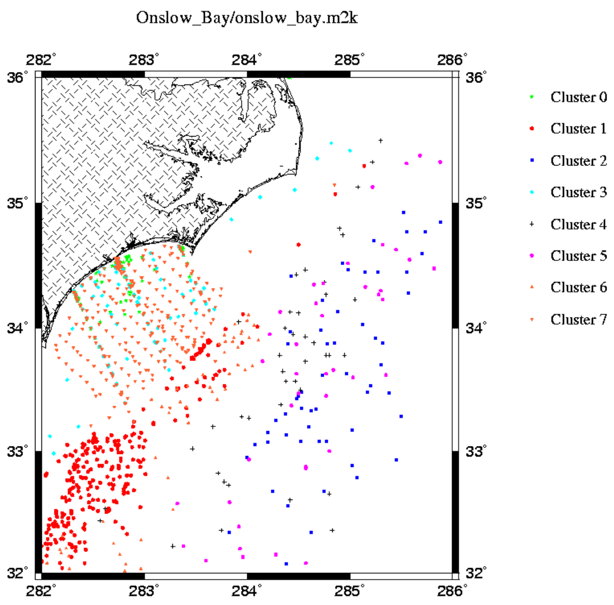

Figure 1 Four cluster case off North Carolina.

Figure 1 shows the area provinced into 4 clusters and figure 2 shows the area provinced into 8 clusters. The four cluster map has three visible regions with the near shore region consisting of two clusters intertwined. On examination of the profiles contained in the clusters, they were seen as summer and winter profiles. The eight cluster case has approximately the same regions with each region subdivided into two or more seasons. Note that there are fewer stations visible on the eight cluster map. As the number of clusters increases the membership percentages are spread between more clusters making fewer of the stations pass the minimum membership threshold. We believe the four cluster view gives the best description of the whole year environment and the eight cluster view shows that the data should be split among the seasons and then provinced into a low number of clusters for the best description of the oceanographic area related to acoustic performance.

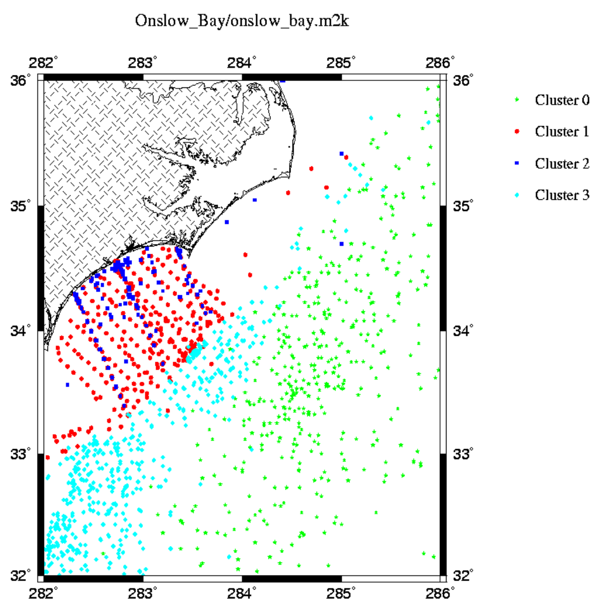

Figure 2. Eight cluster case same data set.

2.2 Acoustic Effects Modeling

In order to show the differences in acoustic propagation for sets of profiles differentiated into clusters as described in Section 2.1, two ensembles of profiles were used as input for the propagation model RAM[9]. Other environmental input parameters were chosen to be fairly generic: flat bathymetry at 1000 meters, and source and receiver depths of 50 meters. The bottom parameters for sound speed, density, and attenuation were isotropic and the standard benchmarking values[10] of 1500 m/s, 1.5 g/cc, and 0.5 dB/λ. Two source frequencies, 50 Hz. and 3 kHz., were chosen to show the effects of the water sound speed profile clusters at low and mid-frequencies.

The resulting transmission loss (TL) curve from each ensemble member was range averaged to simulate third-octave frequency averaging[11], as would be used in a typical ocean survey operation[12]. These TL ensembles were subsequently pressure averaged, and standard deviations about pressure average were generated at each range.

3 Preliminary Results

An area in the South China Sea was chosen for an initial examination of profile fuzzy clustering. The month of January was chosen selection the Conductivity, Temperature, Depth(CTD) profiles available in the WOD database for the chosen one degree square area. Profiles were clustered into two clusters using: surface temperature, surface sound speed duct, mean temperature, and summed sound speed slope. Where needed soundspeed was calculated by converting depth back into pressure as in Leroy[13] and then calculating sound speed at each depth point using the Chen Milleno formulation[14].

Figure 3 displays the profiles from the resultant clustering. As can be seen by the average profiles Cluster 0 has a deeper surface duct than Cluster 1. The upper standard deviation profile surface temperature in Cluster 0 is approximately equal to the average profile surface temperature in Cluster 1. As expected the average profile for all profiles lies between the average profiles for Clusters 0 and 1 and that the standard deviation of all profiles is larger than that of both Cluster 0 and Cluster 1 yet not containing their full ranges.

Oceanographic profiles are sometimes quality controlled by checking if the new measured profile fits within a set number (usually three) standard deviations of the relevant historical profiles. In this case using the appropriate cluster would result in differing quality control determinations than would be found using only the all cluster profile standard deviations.

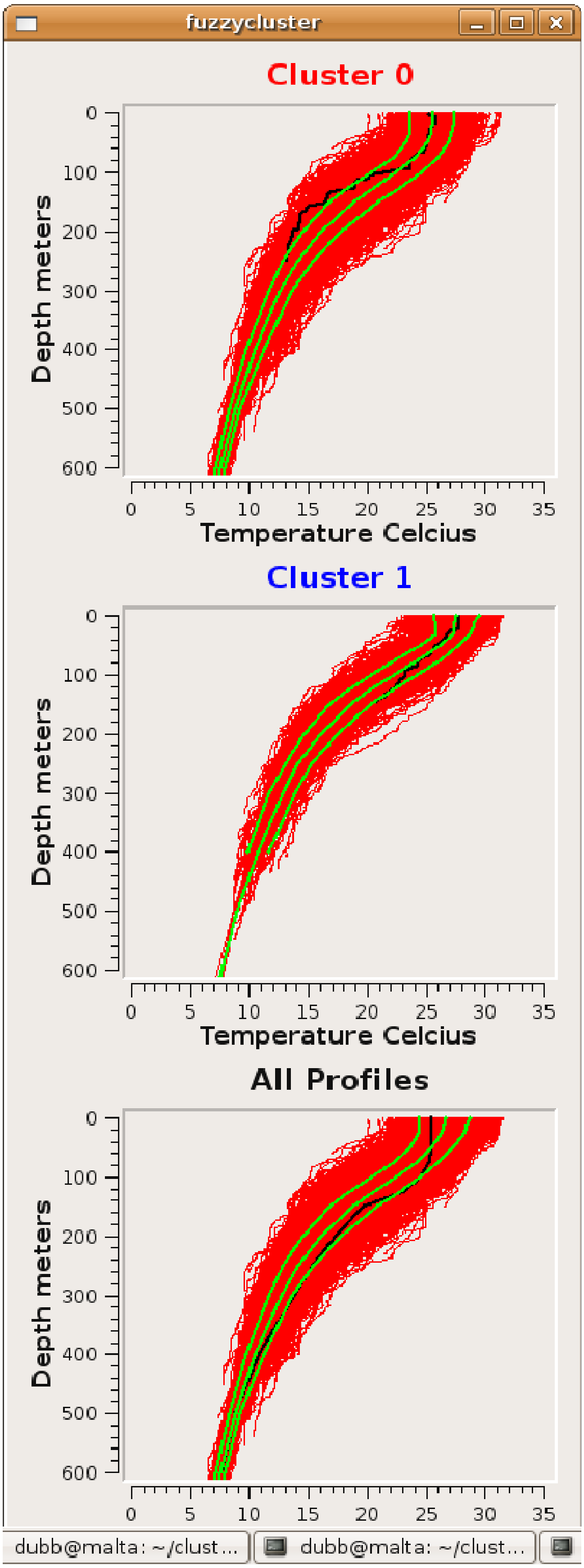

Figure. 3. Profiles in the two clusters displayed above all Profiles.

Red profiles are the individual profiles. Middle green line is the average profile with the surrounding green lines plus and minus one standard deviation. Black line is a sample profile of each set.

Next 100 profiles were extracted from each Cluster. For each selected profile the acoustic transmission loss was calculated as in Section 2.2 with the resulting average transmission loss graphs displayed in Figures 4 and 5. It is not surprising that for the 50 Hz case there is noticeable but small differences detected. At such low frequencies the surface sound speed duct differences between Cluster 0 and Cluster 1 do not greatly effect transmission loss. However as can be seen in Figure 5 at higher frequencies the clusters display meaningful differences in transmission loss.

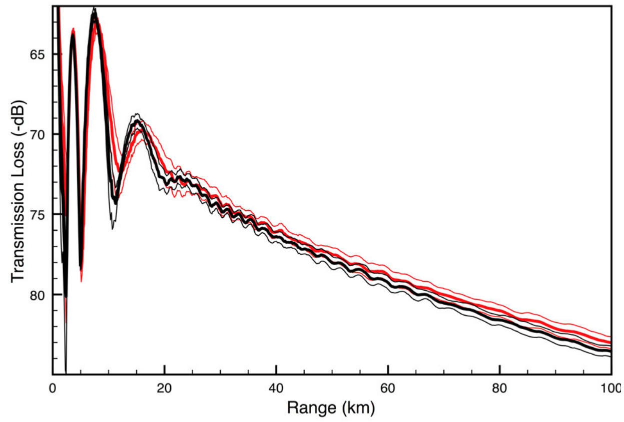

Figure 4. Results from 50 Hz. calculation for ensembles 0 (red) and 1 (black).

Averaged transmission loss is shown as a heavy line, and standard deviations are shown as fine lines for each ensemble. Beyond 15 km, the differences are statistically significant, however they may be too small to discern over noise and experimental uncertainties.

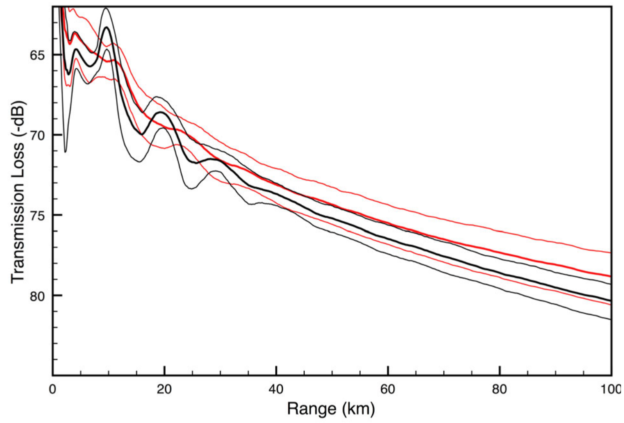

Figure. 5. Results from 3 kHz. calculation for ensembles 0 (red) and 1 (black).

Averaged transmission loss is shown as a heavy line, and standard deviations are shown as fine lines for each ensemble. The results from the two ensembles are both statistically distinct, and sufficiently well separated as to be experimentally resolved.

4 Conclusion

We have shown the fuzzy clustering on a few oceanographic profile features as well as sorting by area and month can produce better climatology where there is sufficient data. The improved climatology would chose between the two available profiles depending of the surface temperature and sea surface height remotely sensed. The chosen profile would then be used as a starter profile for the ocean prediction model. The clustering / PE ensemble run tool can also be used to province near shore regions. Further work is needed to incorporate other data bases such as the ARGO (http://www.argo.net) profiling floats, gliders, and include XBT using estimated salinity.

Acknowledgments

Work done in this paper was sponsored by the Office of Naval Research. The authors would also like to acknowledge Mike Carnes at the Naval Research Laboratory.

References

[1] T.P. Boyer, J.I Antonov, H.E. Garcia, D.R. Johnson, R.A. Locarnini, A.V. Mishonov, M.T. Pitcher, O.K. Baranonva, and I.V. Smolyar, 2006 *World Ocean Database 2005,* S. Levitus, Ed. NODC Atlas NESDIS 60, U.S. Government Printing Office, Wash., D.C., 190 pp., DVDs.

[2] L. A. Zedeh, "Fuzzy sets", *Information and Control*, 8, 338-353, (1965).

[3] J. C. Bezdek, "Pattern Recongnition with Fuzzy Objective Function Algorithm" (1981).

[4] F. Hoeppner, F. Klawonn, R. Kruse, and T. Runkler, "Fuzzy Cluster Analysis: methods for classification, data analysis, and image recognition" (1999).

[5] C. Borgelt R. Kruse , "Speeding up fuzzy clustering with neural network Techniques." *Proc. 12th IEEE Conf. On Fuzzy Systems, IEEE Press*, (2003).

[6] E. Podeszwa, "Sound Speed Profiles for the North Atlantic Ocean" *NUSC Technical Document 5447*, (1976).

[7] M. S. Moustafa and M. D. Mandelberg, "Oceanic provincing through cluster analysis", *Proc. IEEE Oceans*, 478-481 (1991).

[8] S. B. Hooker, N.W. Rees, and J Aiken, "An objective methodology for identifying oceanic provinces", *Progress in Oceanography* 45, p313-338 (2000).

[9] Michael D. Collins, "A split-step Padé solution for the parabolic equation method," *J. Acoust Soc. Am*. 93, 1736-1742 (1993).

[10] F. Jensen and C.M. Ferla, "Numerical solution of range-dependent benchmark problems in ocean acoustics," *J. Acoust. Soc. Am*. 87, 1499-1510 (1990).

[11] C. H. Harrison and J. A. Harrison, "A simple relationship between frequency and range averages for broadband sonar" *J. Acoust Soc. Am.* 97, 1314-1317 (1995).

[12] D.W. Harvey, A. Lowrie, R.D. Filipczyk, "The Generation of N-layered Geoacoustic Descriptions of the Subbottom Using NAVCEANO-Collected Experimental Data," in Proc. *MTS/IEEE Oceans '02 Conference*, pp. 358-362, October 2002.

[13] C. Leroy, F. Parthiot, "Depth-pressure relationships in the oceans and seas", *J. Acoust Soc. Am*. 103, 1346-1352 (1999).

[14] C. T. Chen and F. J. Millero, Speed of sound in seawater at high pressures. *J. Acoust. Soc. Am.*, 62, 1129-1135 (1977).