Acoustic Multipath Identification with Expectation-Maximization

Brian Bingham, David Mindell,
Dana Yoerger, Brendan Foley, and Warren Seering
Massachusetts Institute of Technology
Email: bbing@mit.edu

***Abstract*— Deep-sea autonomous positioning is susceptible to unmodeled measurement errors. Acoustic range measurements for long baseline navigation are not normally distributed, a situation that erodes the robustness and precision of estimation techniques. A mixed-distribution model captures the combination of direct-path, multipath, and spurious returns common for acoustic localization in limited-range survey. Expectation-Maximization concurrently identifies the parameters of the model and characterizes the observations, converging on a model for homogeneous instrumented environments. The resulting representation aids autonomous navigation for precision applications.**

I. Introduction

Underwater range measurements, using discrete acoustic signals, are not normally distributed about the mean range; they are not Gaussian. When using time-of-flight observations for localization, erroneously assuming Gaussian uncertainty can lead to poor navigation precision or even catastrophic estimation errors; however, a reliable sensor model can eliminate the brittle nature of various estimation-based navigation techniques.

This work concentrates on the classification of acoustic range measurements because such observations are an important component of navigation. Classical long baseline (LBL) localization, the simplest embodiment of navigation using time-of-flight measurements, estimates position through trilateration and range estimates from a set of transponders at fixed, known locations [1]. However an increasing number of navigation approaches incorporate distance measurements and other proprioperceptive sensors for localization, eg., Kalman filter approaches, simultaneous localization and mapping (SLAM) [2], synthetic long baseline (SLBL) [3], etc. The topic of this paper is not a navigation algorithm, but a classification approach for increasing the robustness of acoustic ranging. Through classification and based on a simple empirically identified model, this approach can be incorporated in existing navigation approaches. The approach also enables using more data for localization; by identifying multipath returns these observations can be used to maintain position estimates rather than discarded as spurious. Finally this approach can be used to model the acoustic ranging environment by connecting the probabilistic classification with the survey area.

If every range observation were normally distributed about the true line-of-sight range to a fixed, known location, long baseline positioning would be just simple trilateration. The resulting position uncertainty would be predictable, and autonomous operation would be reliable and robust. A glance at the range data in Fig. 1 is evidence that this simple concept is not realized in application. The EM method, extended to the mixed-distribution model, simultaneously identifies and classifies the range observations. Concurrently identifying the model *and* classifying the observations presents a particular challenge. Given a model, classifying the observations is a straight forward estimation problem with various solution methods, and, given the classification, estimating the parameters of the model is a also straight forward. The EM algorithm, an iterative solution to this *chicken and egg problem*, simultaneously produces a probabilistic classification of the individual data points and an identified set of model parameters, fitting the model to the data and classification..

A. Problem Statement and Approach

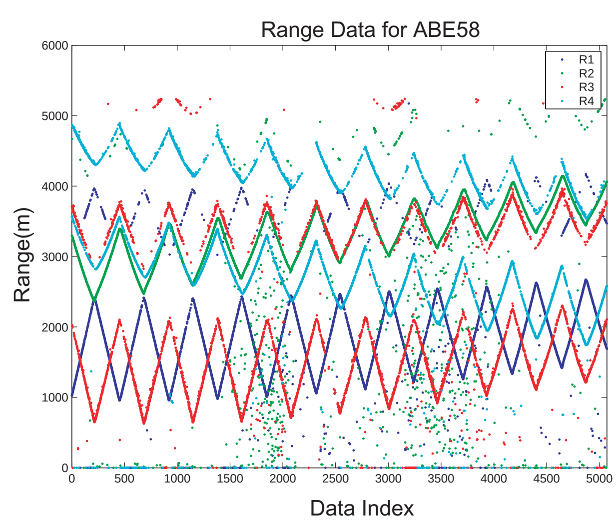

Fig. 1. A selection of 12 kHz LBL data from the ABE58 dive. Time-of-flight is reported as one-way range in meters for successive observations. The four beacons (R1, R2, R3, and R4) are distinguished by color in the legend.

The basic challenge is, given a set of time-of-flight acoustic measurements and uncertain localization estimates, to discern the observations coming from the line-of-sight path between the fixed transponder at a known locations and the mobile host. Long baseline range data in Fig. 1 illustrates this challenge and the typical structure of acoustic range observations.

Concentrating on the range data from the third beacon (R3) illustrates the important archetypes: consistent direct-path returns oscillating between 700 and 2200m in a sawtooth pattern, multipath returns with a similar consistent sawtooth pattern between 3000 and 4000m, and spurious outliers cluttering the data. While the human eye is adept at this classification, autonomously classifying and identifying this type of data is difficult.

This paper proposes an approach to this problem based on the following three steps:

- A mixed-distribution sensor model assuming three possible sources of an observation: a direct-path, a multipath, or an outlier. $\theta = \{DP, MP, OL\}$, where θ is the set of hypothesized sources. By combining Gaussian and uniform distributions the complexity of the range data is captured while retaining the elegance and computational advantages of normal distributions.
- A ray-tracing multipath model with empirically identified geometric parameters. Because the location is estimated with some known confidence, the unknown parameters in the mixed-density model are only those of the multipath environment.
- Classification and identification using Expectation-Maximization (EM). The batch processing approach simultaneously determines a probabilistic classification of each data point, the mutually exclusive probability that a given point is from each of the three sources, and the the identification of an optimal set of model parameters.

We pursue this approach by developing the algorithm in general. As a case-study and to illustrate the concepts we apply the algorithm to the data in Fig. 1 from the ABE58 dive, first examining a subset of the data to illustrate the method and then presenting the results of applying the approach to the entire dataset.

B. Background

Three distinct areas contribute to the development of this method. The next sections present the necessary background on data association, multipath modeling, and the expectation-maximization (EM) algorithm.

1) Classification: Classification is a fundamental challenge for many types of sensor processing. In radar tracking, where measurements are classified based on the multiple sources of a particular return, probabilistic association sorts the returns from using a variety of techniques [4], [5]. Such techniques lead directly to applications for outlier rejection in range observations [6]. Correspondence in robot mapping seeks to match features across disparate sensor scans, i.e., to find a feature known from a previous measurement in the current observation [7]. In machine vision the difficulty is finding the same portions within multiple images - correspondence. A similar challenge in image processing is to classify segments an single image to match the physical properties of the scene [8].

Explicitly classifying acoustic range observations - direct-path, multipath, and outliers - has similarities the general problems of data association and correspondence, but is wholly different than outlier rejection techniques. By capturing the multipath returns, though empirical identification, more data can be incorporated into the navigation algorithms, a better model of the environment can be determined, and there is less opportunity for consistent, but erroneous, data to corrupt autonomous navigation.

2) Multipath Modeling: A general model of the acoustic communication channel is an open research question. At the coarsest level there are two basic multipath models: a probabilistic reflector model prevalent in applications to electromagnetic wave communications [9], [10], and a deterministic ray-tracing model typical in ocean acoustics [11], [12].

Ocean acoustics models are based the large scale propagation of acoustic signals typical of military sonar applications. Deterministic techniques, such as ray-tracing, yield models that are based on the geometry and properties of the medium. When the geometry of the environment is known, Deffenbaugh's approach incorporates multipath returns in the navigation solution [13]. For unknown environments system identification techniques are necessary to learn the parameters of the model based on observation. For small length scales, typical of robotic underwater survey, we assume an acoustically homogenous environment. This assumption simplifies the geometry, neglecting the complexity of large-scale propagation.

3) The EM Algorithm:

> The EM algorithm is a general method for finding the maximum likelihood estimate of the parameters of an underlying distribution from a given data set when the data is incomplete or has missing values. [14]

Background on expectation-maximization (EM) ranges from mathematical fundamentals [14], [15] to applications for robot simultaneous localization and mapping [16] and image segmentation [17], [18]. The key characteristics of the EM for application to multipath identification are the batch approach, iterative solution, and the application to mixed density models. The algorithm processes all the data at once, precluding its use in a real-time scenario. Two steps are repeated to iteratively solve the classification and identification. The *e-step* classifies the data based on the current model, followed by the *m-step* which identifies the model based on the classification. The applications of the algorithm, referred to above, have shown the utility of EM for mixed-distribution models similar to the model proposed for range observations. This work applies the EM algorithm to a new topic, acoustic range observations, and a new model, a mixed model of Gaussian and uniform distributions.

II. Algorithm Development

A. Mixed-Distribution Observation Model

The measurement model combines three possible sources for each observation. For each measurement, indexed in time (k), a hypothesis $(\theta_i(k))$ represents association of the observation with each of three possible sources: a direct-path (DP) range, a multipath (MP) range, or an outlier (OL)[1]. The mixed-density measurement model combines these sources.

$$z(k) = \begin{cases} |\hat{\mathbf{x}}(k) - \mathbf{x_0}(k)|^2 + \nu_{DP} \\ f_{MP}(\hat{\mathbf{x}}(k), \mathbf{x_0}(k), \mathbf{\Phi}) + \nu_{MP} \\ \nu_{OL} \end{cases} \quad (1)$$

$$\begin{aligned} \nu_{DP} &\sim N(0, \sigma_{DP}) \\ \nu_{MP} &\sim N(0, \sigma_{MP}) \\ \nu_{OL} &\sim Uniform \end{aligned}$$

The direct-path measurement is the Cartesian distance metric between the observer's estimated position, $\hat{\mathbf{x}}(k)$, and the known transponder location, $\mathbf{x_0}(k)$. The additive zero-mean Gaussian noise has a standard deviation of σ_{DP}. The multipath observation is a function, f_{MP}, of the observer location, the beacon location, and the model parameter vector $\mathbf{\Phi}$. The outlier observations are modeled as uniformly distributed over a specified range.

B. Multipath Model

An acoustic reflector creates an alternate path for the sonar signal. For specular reflections the length of this path is a function of the relative heights of the source, receiver, and reflecting plane. Fig. 2 illustrates this geometry in a vertical plane containing the host and the beacon. As presented in equation 1, the multipath range is a function of the mobile host position, the fixed beacon location, and the geometry parameters. For a single planar reflector the geometry is parameterized by the location of the reflecting plane. The resulting parameterized multipath model is

$$f_{MP}(x_0, d_0, \hat{x}(k), \hat{d}(k)) = \frac{d_0 + \hat{d}(k)}{\sin\left(\arctan\left(\frac{d_0 + \hat{d}(k)}{|x_0 + \hat{x}(k)|}\right)\right)} \quad (2)$$

where d_0, x_0 and $\hat{d}(k), \hat{x}(k)$ are the depth and horizontal locations for the beacon and host respectively. Depth is used here as the relative vertical distance to the reflecting plane - the *height of the reflecting plane* - and the horizontal location is a scalar distance projected into the plane containing the beacon and the host locations.

C. EM for Multipath Classification

Equations 1 and 2 capture the geometry and uncertainty in the mixed-distribution range measurement model. The EM algorithm alternatively classifies the data probabilistically by estimating the association probabilities ($P(\theta_i(k))$, $\theta_i(k) = \{DP, MP, OL\}$) and identifies the multipath model by estimating the model parameters ($\mathbf{\Phi} = \{d_0, \sigma_{MP}\}$) to arrive at a classification and identification.

[1]Time-of-flight is actually measured, and the range is estimated based on this measurement and an uncertain estimate of the speed of sound.

1) E-Step: Classification: Given a set of model parameters, the estimation step (e-step) classifies the observations. The model parameters, Φ, are 'known' from the previous iteration or from initial conditions, and the algorithm proceeds based on this model. For each data point the association probabilities, $P(\theta_i(k))$ are calculated in the Bayesian sense.

$$\begin{aligned} \gamma_j(k) &= P(\theta(k) = j | z(k), \Phi) & (3) \\ &= \frac{P(z(k)|\theta(k) = j, \Phi) P(\theta(k) = j|\Phi)}{\sum_{i=1}^{M} P(z(k)|\theta(k) = i, \Phi) P(\theta(k) = i|\Phi)} & (4) \\ &= \frac{P(z(k)|\theta(k) = j, \Phi) P(\theta(k) = j|\Phi)}{\rho} & (5) \end{aligned}$$

- Time index: $k = 1, \ldots, N$ for N data points
- Class index: $j = 1, \ldots, M$ for M hypotheses

Bayesian Term	Notation	Description
Posterior	$P(\theta\|z, \Phi)$	Probability of the association given the data
Generative	$P(z\|\theta, \Phi)$	Probability of the measurement given the hypothesized association
Prior	$P(\theta\|\Phi)$	Probability of the association without the measurement evidence
Normalization	ρ	Sum of the probabilities to enforce the mutual exclusive constraint

TABLE I

BAYESIAN TERMINOLOGY AND NOTATION FOR EQUATION 5

The algorithm calculates the generative term for a particular observation based on the probability density function of the particular classification. For DP and MP associations, the distributions are Gaussian, characterized by estimated mean and variance values. Calculating the following values determines the posterior probability:

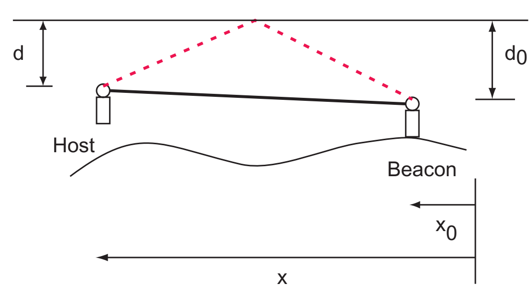

Fig. 2. The geometry of the multipath model. The model is parameterized by the height of the reflecting plan. The dashed (red) line shows the multipath trace while the solid (black) line shows the direct path. Since the range estimated from the round trip travel time the path may consist of both the slant range (line-of-sight) and the specular reflection - the triangle path.

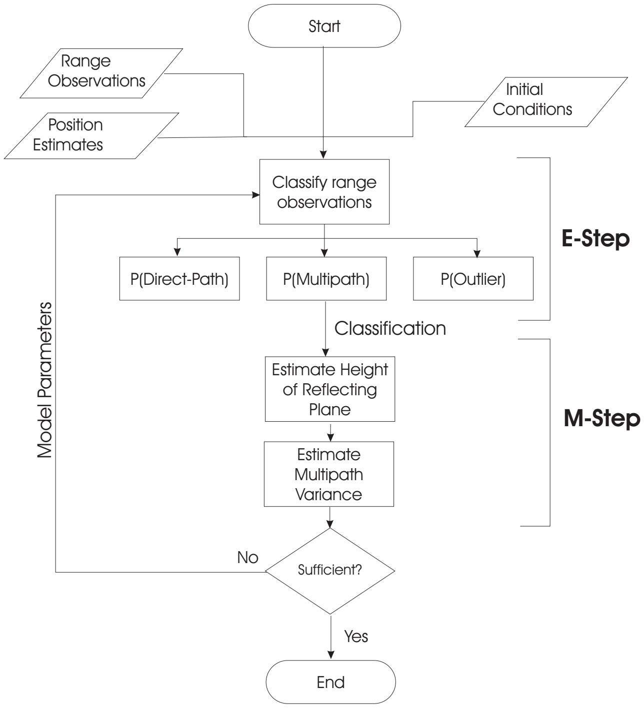

Fig. 3. Flowchart of the EM algorithm applied to range association.

$\hat{z}_{j=DP}(k)$	Estimated range under the direct-path hypothesis from the current position estimate and the known beacon location.
σ^2_{DP}	Variance in the direct-path range estimate - a characteristic of the sensor.
$\hat{z}_{j=MP}(k)$	Estimated range under the multipath hypothesis from the current position estimate, known beacon location, and current iteration of the multipath model parameters.
σ^2_{MP}	Variance in the multipath range estimate - a parameter of the current iteration of the multipath model.

The conditional probabilities are calculated using the cumulative distribution functions for a chi-squared distribution.

$$P(z(k)|\theta = j, \Phi) = 1 - \int_0^u \chi^2(x)dx \tag{6}$$

$$u = \frac{(z(k) - \hat{z}_j(k))^2}{\sigma_j^2}$$

The conditional probability of a spurious measurement is constant.

$$P(z(k)|\theta = OL, \Phi) = P_{OL}$$

The conditional probabilities on the right hand side of equation 5 infer the posterior probabilities on the left hand side. The prior probabilities, $P(\theta_i = j|\Phi)$, are equivalent and can be left out of the calculation because of the normalization. This important detail is discussed more in the implementation and as a topic of future work.

2) M-Step: Model Parameter Estimation: The maximization step (m-step) estimates the model parameters, $\mathbf{\Phi}$ - the height of the reflecting plane and the standard deviation in multipath observations, based on the probabilistic classification from the preceding e-step. The challenge is to find the maximum likelihood estimate of the parameters, given the range observations and their classifications. Since the data model assumes a mixture of Gaussian and a uniform distributions (which can be viewed as a Gaussian with a large variation), the least-squares merit function is appropriate as the likelihood function. Because the only unknown model parameters are those of the multipath portion of the model, only the multipath associations need to be included in the model fitting.

$$\chi^2 = \frac{1}{\sigma_{MP}} \sum_{k=1}^{N} \left[(z(k) - \hat{z}_{j=MP}(k))^2 \gamma_{j=MP}(k) \right] \tag{7}$$

The chosen parameters minimize the residuals, χ^2, in equation 7. Using the classification probability ($\gamma_j(k)$) as a weighting function reinforces the clustering necessary for the algorithm to converge.

The other parameter of the multipath model, the height of the reflecting plane(d_0) , is estimated by calculating the weighted average of the geometry of each data point. The multipath classification probability weights the contribution of each data point to the estimate,

$$d_0 = \frac{\sum_{k=1}^{N} [d_0(k)\gamma_{j=MP}(k)]}{\sum_{k=1}^{N} [\gamma_{j=MP}(k)]} \tag{8}$$

where k is the time index for each observation. The maximum likelihood estimate of the variance is a similarly weighted average of the residuals.

$$\hat{\sigma}_{MP} = \frac{\sum_{k=1}^{N} [(z(k) - \hat{z}_{j=MP}(k))\gamma_{j=MP}(k)]}{\sum_{k=1}^{N} [\gamma_{j=MP}(k)]} \tag{9}$$

The predicted measurements, $\hat{z}_{j=MP}(k)$m are calculated using the estimated height from equation 8.

The EM algorithm uses the alternating e-step and m-step as illustrated in the flowchart of Fig. 3. This iterative solution relies on the consistency of the multipath data to identify both the height of the reflecting plane and the variance in the multipath measurement while converging on a classification of the range observations - the probability that each measurement is the result of the direct-path, the multipath, or an outlier.

III. Application of EM/MP Algorithm

The following example, using data from the autonomous benthic explorer (ABE) in Fig. 4 [19], develops a concrete implementation multipath identification using EM. Long baseline data (12 kHz) from ABE dive number 58 (ABE58) was taken from November 5 to December 4 of 2001 by a research team studying mid-ocean ridge geology.

Fig. 4. The Autonomous Benthic Explorer (ABE) vehicle.

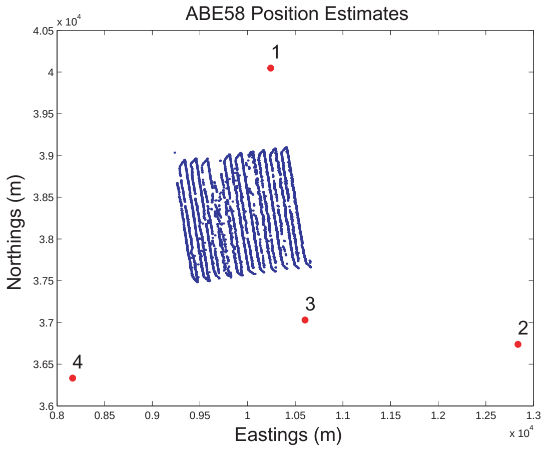

Fig. 5. ABE58 survey in the X-Y plane. Approximately 5,000 data points are plotted from the survey. The 12 kHz LBL beacon locations are shown as labeled red markers

3) Long Baseline Data Overview: The long baseline positioning data from ABE58 is ideal for illustrating and evaluating the hypothesis grid algorithm. The vehicle made range measurements to 4 beacons at known locations using standard 12kHz acoustic transponders. Fig. 5 illustrates the ABE58 survey by plotting the post-processed positions of the vehicle projected into the x-y plane and the long baseline transponder locations.

Measurements corresponding to the direct-path range, the line-of-site between the host and transponder, are modeled with Gaussian uncertainty - characterized by the first and second moments. We estimate the standard deviation of the direct-path range measurement , using least-squares model-fitting, to be **3.97m** for the 12kHz LBL observations.

A. Small Example from the Data

Applying EM to the ABE58 range data illustrates the operation of the algorithm as it converges on a classification and model identification. The process treats each of the four beacons used in the survey separately. The results present the analysis from all four beacons, but observations from beacon #4 are used in this detailed example. Fig. 6 presents a selection of the range data. The direct-path (DP) is the first consistent return ranging from approximately 2500 to 3500 meters. A strong second return, the multipath (MP), is evident from the ranges with values between 4000 to 5000 meters. And spurious returns, outliers (OL), are distributed from 0 to 5000 meters.

1) Initial Conditions: The final values of the algorithm are insensitive to these initial values. The initial parameter vector $\mathbf{\Phi}(0)$ is the starting point for the method. The uncertainty in the multipath observations is initialized at a very high value: $\sigma_{MP}^2(0) \approx 10^7 m^2$. Using a large value is important because it causes the first classification to associate many of the observations as multipath returns. Since the variance is large, the actual value of the height of the reflecting plane parameter does not effect the first classification and this value can also be chosen very large - 4,000 m for this example.

The first e-step, shown in Fig. 7, classifies the measurements based on these initial parameters. Since the direct-path variance is already 'known', the figure shows accurate classification of the direct-path, shown as blue markers in the figure. Because of the very large initial multipath variance, the initial classification considers many of the outliers as possible multipath measurements (red markers in the figure); the algorithm begins with a conservative guess, erring on the side of caution and not throwing away 'good data'. The classification keeps the probability of any of the ranges being outliers very low, as evident by the lack of green markers.

2) Iteration: Following this first classification, the algorithm calculates the parameters of the multipath model based on the least-squares criteria in equation 7. These two parameters are the height of the reflecting plane and the measurement variance - both determine in the m-step. The parameters are a function of the system's geometry and the probabilistic association of the preceding e-step. A convergence factor (α) controls the rate of convergence of the algorithm. The m-step calculates a new estimate of the reflecting plane height - d_0 of equation 8. Blending this new value with the old value ($d_0(i)$) estimates the height of the reflecting plane for the next iteration ($d_0(i+1)$.

$$d_0(i+1) = (1-\alpha)d_0(i) + (\alpha)d_0 \tag{10}$$

Once the new reflecting plane height is available, the algorithm calculates the uncertainty in the multipath using the new plane height and equation 9.

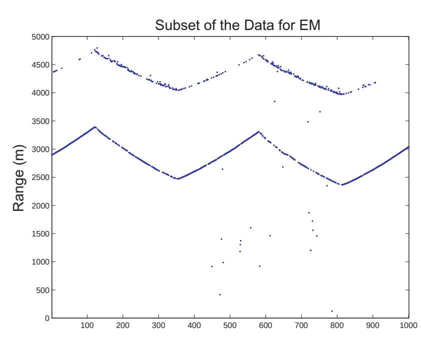

Fig. 6. A selection of 1000 ranges from the ABE58 dataset.

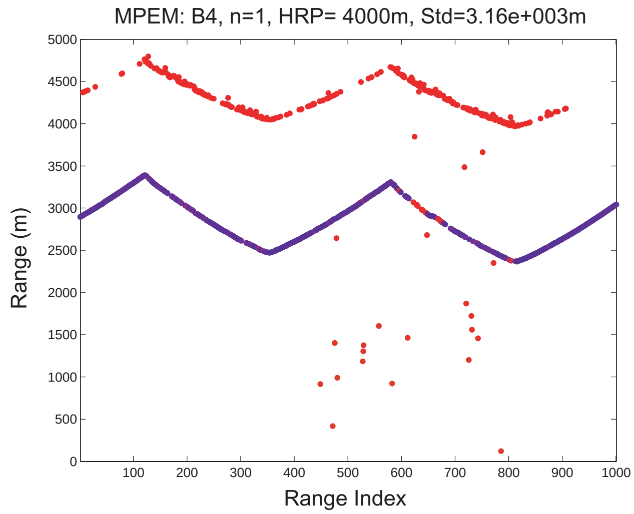

Fig. 7. Classification plot for the first iteration of the EM algorithm. The probabilistic classifications are mapped to blue, red, and green for direct-path, multipath, and outliers (see Fig. 8).

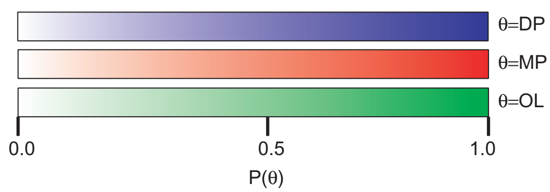

Fig. 8. Three-value colormap for visualizing the probabilistic associations. The probabilistic classifications are mapped to blue, red, and green for direct-path, multipath, and outliers.

3) Final Classification and Identification: The algorithm terminates when the change in multipath parameters is sufficiently small. Fig. 9 shows the final classification. In contrast to the initial classification in Fig. 7, the decreased multipath range variance identifies the consistent multipath data and excludes the outliers. Fig.s 10 and III-A.3 illustrate show the trajectories of the two model parameters as the algorithm converges. As the reflecting plane height decreases, the variance shrinks, creating increasingly conservative bounds on the observations associated with the multipath range. For this example the final values for the multipath model parameters are a reflecting plane height (d_0) of 2483m and a standard deviation (σ_{MP}) of 8.83 m - twice the standard deviation in the direct-path.

4) Normalization with a Uniform Distribution: In contrast to the applications discussed in the background section, applying of the EM algorithm to a mixed-distribution containing both Gaussian and uniform distributions requires special attention to the normalization - the denominator in equation 5. This section explains this challenge, proposes and ad-hoc solution, and illustrates the sensitivity of the final result to a single chosen parameter.

The challenge of normalization stems from the Bayesian requirement for prior probabilities and the challenges of representing ignorance [2]. The difficulty boils down to determining $P(z|OL)$ - the probability of the measurement, given the it is an outlier. The probabilities of the measurement, given it is a direct-path or multipath, $P(z|DP)$ and $P(z|MP$, are calculated from the cumulative distribution functions using equation 6. Upon initialization the model parameters are intentionally far from 'true', and using the mutual exclusive constraint to calculate $P(z|OL) = 1 - P(z|DP) + P(z|MP)$ can cause the solution to diverge. The method must reliably address this ignorance during the initial stages and converge to a solution that satisfies the mutual exclusive constraint of the three conditional probabilities.

A threshold, Pol, is introduced in the normalization of the probabilities. At each iteration and for each observation the following steps are taken:

- Calculate $P(z|DP)$ and $P(z|MP)$ for the current model parameters using equation 6.
- If $1 - P(z|DP) - P(z|MP) < Pol$,
  let $P(z|OL) = 1 - P(z|DP) - P(z|MP)$.
- If $1 - P(z|DP) - P(z|MP) > Pol$,
  let $P(z|OL) = Pol$.

Fig. 12 shows the influence of the threshold Pol on the operation and results of the EM algorithm for this case study.

- Axes a: The height of the reflecting plane is insensitive to the choice of Pol.
- Axes b: Increasing Pol increases the tendency to classify data as outliers, creating a more conservative association of multipath points and reducing the multipath variance.
- Axes c: The average association probabilities change smoothly and consistently as Pol is varied.

[2]This complication is subtle but powerful. There is a strong debate about the difference between the probability of belief and the probability of chance (frequency based) and the inference based on these notions of probability. See Dempster-Shafer [20] and Jeffreys [21].

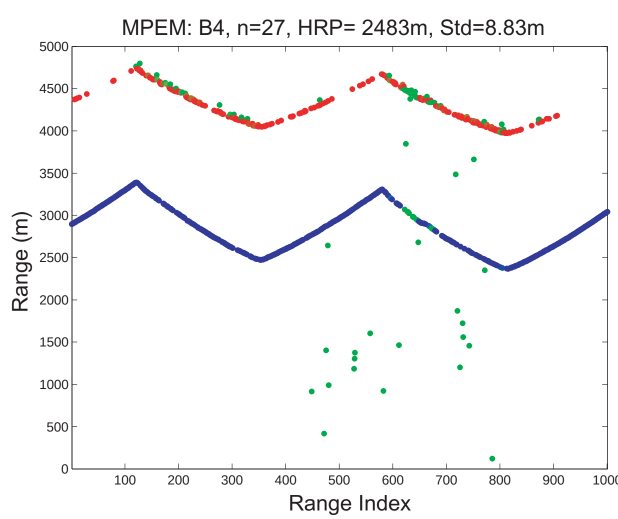

Fig. 9. Classification plot for the last iteration of the EM algorithm. The probabilistic classifications are mapped to blue, red, and green for direct-path, multipath, and outliers

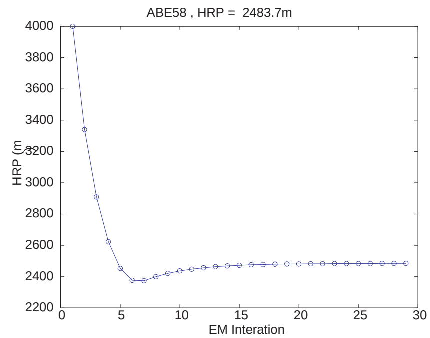

Fig. 10. Convergence of the model parameters.

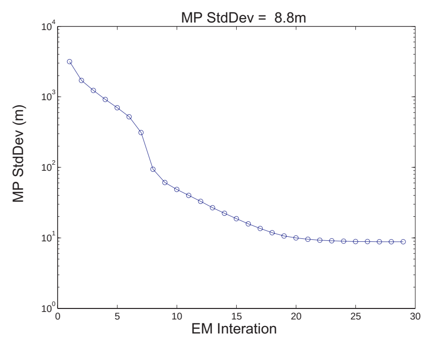

Fig. 11. Convergence of the model parameters.

This normalization is key to the classification of the e-step and hence the identification of the m-step. This threshold is not related to the the empirical probability of an outlier, but is instead a parameter of the algorithm - a single knob that to tune the behavior of the method.

IV. Multipath Identification Results

The previous example illustrates the EM algorithm applied to a limited subset of the measurements from one beacon used in ABE58. The same algorithm is applied to the whole 5,000 point dataset from each of the four transponders. The results, summarized in table II and shown graphically in Fig. 13, identify the four multipath models by determining the height of the reflecting plane (HRP) relative to each beacon and the variance in the acoustic measurements. The final row of the table lists the recorded beacon depths which agree with the location of the reflecting planes determined with the EM method. These results shows that the multipath data is the result of the acoustic signal reflecting off the air-water interface on either the outgoing or incoming path and traveling the line-of-sight path in the other direction - the triangle path in Fig. 2.

V. Conclusions

The previous sections presents a solution to the problem of empirically and automatically identifying a static multipath in acoustic range data. We formulate the EM algorithm for simultaneous model identification and data classification. Applying

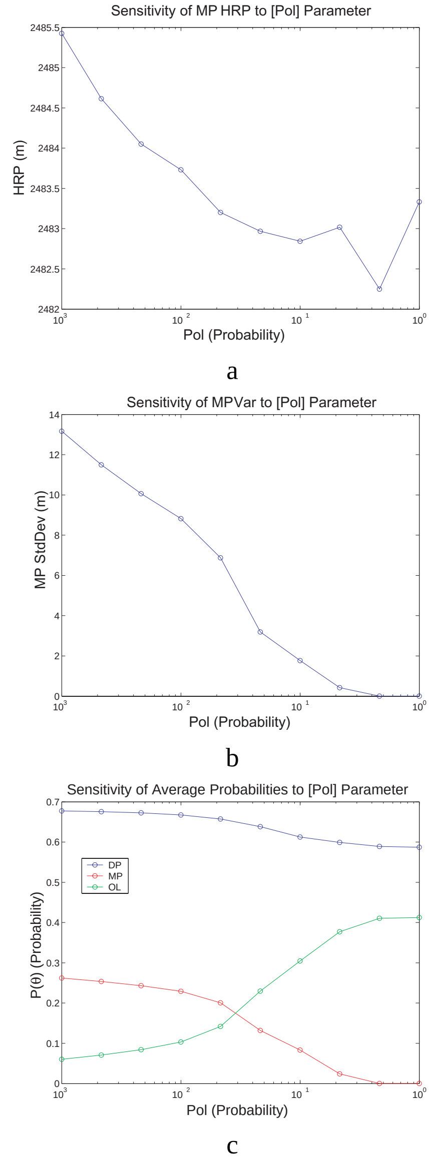

Fig. 12. Sensitivity of algorithm results to the Pol parameter.

this framework to a segment of long baseline data from ABE58 illustrates the operation and demonstrates the feasibility of the algorithm. Because the static multipath is a surface bounce, we can verify the solution for the reflecting plane height with the known location of the air-water interface, i.e. we have a test problem with a known solution to check the method's solution. Two 'knobs' allow for tuning the operation of the algorithm: the convergence factor, α, of equation 10 controls the rate of convergence of the algorithm and threshold parameter, Pol, controls balancing and normalizing association probabilities.

This application of the EM algorithm has some important restrictions. Since the implementation is batch, the entire dataset is treated at once rather than in real-time. The measurement model is a combination of two Gaussian and one uniform distribution - a fundamental assumption. The model also assumes a stable multipath environment amenable to the

Beacon #	1	2	3	4
HRP (m)	2392	2500	2404	2486
$\hat{\sigma}_{MP}$ (m)	4.2	10.5	14.0	6.9
Beacon Depth(m)	2392	2491	2395	2472

TABLE II

MULTIPATH IDENTIFICATION RESULTS: ABE DIVE 58.

proposed ray-tracing model. A more complex model, capturing the physics of larger scale acoustic ranges, could be integrated into the EM framework. Finally, the algorithm is dependent on an independent estimated location associated with each range observation.

This application extends the EM algorithm to a new domain and to a new type of problem by incorporating a uniform distribution. Understanding and quantifying the properties of the environment is critical to developing navigation techniques that leverage environmental awareness. An accurate model of the multipath environment increases the navigation performance and robustness. With the multipath model both direct-path and multipath returns can be used in localization - allowing position estimation from more observations. With an accurate classification of the data navigation algorithms are less susceptible to classification errors which can cause brittle methods to diverge. Applying the EM framework to acoustic ranging data supports navigation using acoustic range observations.

ACKNOWLEDGMENT

The authors would like to acknowledge the support of Ocean Exploration office of the National Oceanic and Atmospheric Administration (NOAA-OE) and thank the Woods Hole Oceanographic Institution (WHOI) ABE group for the data used in this work.

REFERENCES

[1] P. H. Milne, *Underwater Acoustic Positioning Systems*. Houston: Gulf Publishing Company, 1983.

[2] P. M. Newman and J. J. Leonard, "Pure range-only subsea slam," MIT Marine Robotics Laboratory, Tech. Rep., 2002, submitted to the 2003 IEEE International Conference on Robotics and Automation.

[3] M. B. Larsen, "Synthetic long baseline navigation of underwater vehicles," in *OCEANS 2000*, vol. 3, 2000, pp. 2043–2050.

[4] Y. Bar-Shalom and T. E. Fortmann, *Tracking and Data Association*. Academic Press, 1988.

[5] Y. Bar-Shalom and X. R. Li, *Estimation and Tracking: Principles, Techniques, and Software*. Yaakov Bar-Shalom (YBS), 1998.

[6] J. Vaganay, J. J. Leonard, and J. G. Bellingham, "Outlier rejection for autonomous acoustic navigation," in *Proceedings of the 1996 IEEE International Conference on Robotics and Automation*. Minneapolis, Minnesota: IEEE, April 1996, pp. 2174–2181.

[7] S. Thrun, *Exploring Artificial Intelligence in the New Millenium*. Morgan Kaufmann, 2002, ch. Robotic mapping: a survey.

[8] M. R. Anderberg, *Cluster Analysis for Applications*. Academic Press Inc., 1973.

[9] J. G. Proakis, *Digital Communications*. McGraw-Hill, 1989.

[10] J. Tourrilhes, "Linux wireless LAN HowTo," June 2000. [Online]. Available: http://www.hpl.hp.com/personal/Jean_Tourrilhes/Linux/Linux.Wireless.htm%l

[11] F. B. Jensen, W. A. Kuperman, M. B. Porter, and H. Schmidt, *Computational Ocean Acoustics*. New York, NY: Springer-Verlag, 2000.

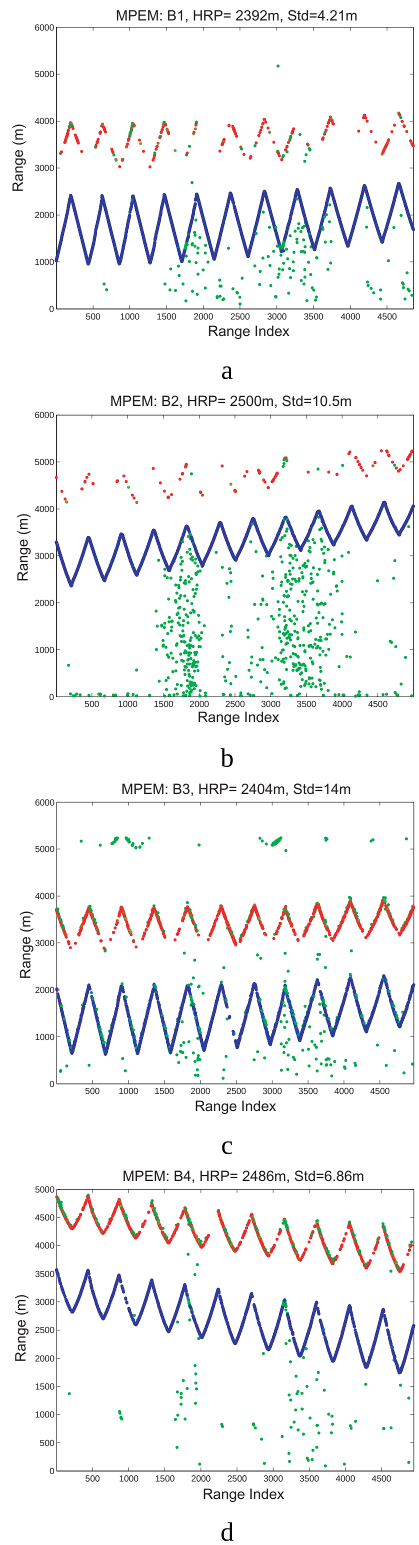

Fig. 13. Results of multipath identification using expectation-maximization for the four beacons in ABE58. The parameter values of the classification are summarized in table II and the colormap is illustrated in Fig. 8.

[12] L. M. Brekhovskikh and Y. P. Lysanov, *Fundamentals of Ocean Acoustics*, 2nd ed. Springer-Verlag, 1991.

[13] M. Deffenbaugh, "Optimal ocean acoustic tomography and navigation

with moving sources," Ph.D. dissertation, MIT WHOI, 1997.
[14] J. Bilmes, "A gentle tutorial on the EM algorithm and its application to parameter estimation for gaussian mixture and hidden markov models," 1997. [Online]. Available: citeseer.nj.nec.com/bilmes98gentle.html
[15] A. Dempster, A. Laird, and D. Rubin, "Maximum likelihood from incomplete data via the EM algorithm." *Journal of the Royal Statistical Society*, 1977.
[16] S. Thrun, D. Fox, and W. Burgard, "A probabilistic approach to concurrent mapping and localization for mobile robots," *Machine Learning*, vol. 31, pp. 29–53, 1998.
[17] M. Saeed, "Maximum likelihood parameter estimation of mixture models and its applicaiton to image segmentation and restoration," Master's thesis, Massachusetts Institute of Technology, 1997.
[18] S. Tadjudin and D. A. Landgrebe, "Robust parameter estimation for mixture model," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 38, no. 1, pp. 439–445, January 2000.
[19] D. R. Yoerger, A. M. Bradley, and B. B. Walden, "Autonomous benthic explorer, deep ocean scientific AUV for seafloor exploration: Untethered, on station one year without support ship," *Sea Technology*, vol. 33, no. 1, pp. 50–54, 1992.
[20] G. Shafer, *A mathematical theory of evidence*. Princeton University Press, 1976.
[21] H. Jeffreys, *Theory of Probability*. Oxford University Press, 1961.