Chapter 1

ESTIMATION OF TRANSMISSION LOSS AND ITS UNCERTAINTY

Peter Gerstoft, Chen-Fen Huang, and W.S. Hodgkiss
Marine Physical Laboratory
Scripps Institution of Oceanography
University of California San Diego
La Jolla, CA 92093-0701, USA

Abstract A weakness in sonar performance prediction has been the lack of a means for quantifying the impact of uncertainty in estimates of the ocean environment. This paper describes the estimation of transmission loss and its statistical properties based on observations of ocean acoustic data. This problem is solved by first finding an ensemble of relevant environmental model parameters and then using all of these models to map into the transmission loss domain. In this mapping each environmental model is weighted according to its likelihood function.

This algorithm is illustrated on vertical array data acquired during the ASIAEX 2001 East China Sea experiment. The environmental parameters are first estimated. Based on the likelihood that each of these environmental models fits the ocean acoustic data, each model is mapped into transmission loss. This enables us to compute a full probability distribution for the transmission loss in both range and depth.

Introduction

A weakness in sonar performance prediction has been the lack of a means for quantifying the impact of uncertainty in estimates of the ocean environment. Figure 1.1 offers an overview of posterior estimation of transmission loss from ocean acoustic data. Ocean acoustic data $\mathbf{d}$ is observed on a vertical or horizontal array. The inverse problem is first solved using Bayes rule to compute the posterior distribution $p(\mathbf{m}|\mathbf{d})$ of environmental parameters $\mathbf{m}$. We are not interested in the environment itself but rather in better estimates of the transmission field, the usage

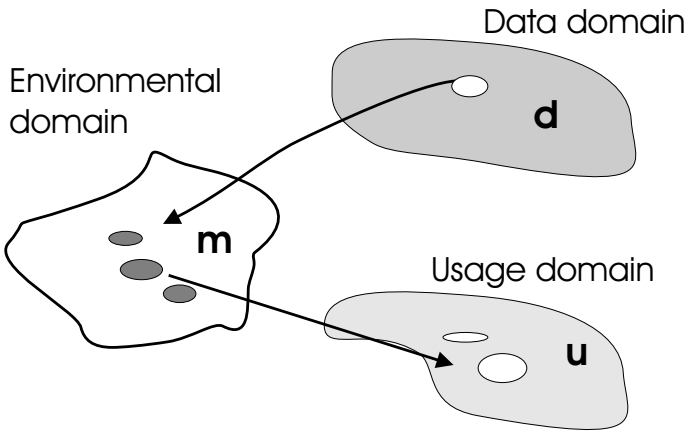

Figure 1.1. An observation **d** is mapped into a distribution of environmental parameters **m** that potentially could have generated it. The environmental parameters are then mapped into the usage domain **u**.

domain $\mathbf{u}(\mathbf{m})$. Based on the posterior distribution $p(\mathbf{m}|\mathbf{d})$, the distribution of $p(\mathbf{u}(\mathbf{m}))$ is obtained via Monte Carlo simulation. From this transmission-loss distribution, all relevant statistics of the transmission loss can be obtained.

The principle of the inversion is indicated in Fig. 1.1. Based on the ocean acoustic data **d** we statistically characterize transmission loss (the usage domain **u**). The vector **d** represents the acoustic data observed at N hydrophones and the vector **u** represents transmission loss at range and depth. As shown in Fig. 1.1, this is mapped via a set of M environmental parameters **m**. The approach involves a number of steps as outlined below:

1 Determine a model for the ocean acoustic environment and select an appropriate propagation model. These two form the mapping $\mathbf{d}(\mathbf{m})$ from the parameter space **m** to data space **d**.

2 Determine the mapping $\mathbf{u}(\mathbf{m})$ from the parameter space to usage space. Except for a change in geometry (source depth), here this is similar to $\mathbf{d}(\mathbf{m})$, but could be any other model (e.g., some other metric of sonar performance).

3 Find acceptable models **m** from the data. As indicated in Fig. 1.1, a region around the data can map into several acceptable solutions in the model domain.

4 Map the acceptable models into the usage domain. Several models can map into the same usage region.

As indicated in Fig. 1.1, the mapping from data to usage domain is non-unique. There are many environmental models that give about the same goodness-of-fit. The maximum likelihood (ML) estimate of the environmental model gives the best fit. Instead of using just one estimated environment, it is proposed to describe the environmental solution probabilistically. This probability is then mapped into the usage domain. Knowing the posterior probability distribution in the usage domain is preferable to having a single point estimate such as the usage domain result corresponding to the ML-solution.

1. INVERSE PROBLEM FRAMEWORK

In the Bayesian paradigm, the solution to determining parameters of interest $\mathbf{m}$ given an observation $\mathbf{d}$ is characterized by the posterior probability $\mathrm{p}(\mathbf{m} \mid \mathbf{d})$. First, the prior information of the model parameter vector is the probability density $\mathrm{p}(\mathbf{m})$. Then, this information is combined with the likelihood function $\mathrm{p}(\mathbf{d}|\mathbf{m})$ provided by the combination of data and the physical model to give the posterior probability density $\mathrm{p}(\mathbf{m}|\mathbf{d})$ of the model parameters. A clear discussion of inverse theory from a probabilistic point of view is given by Tarantola [1]. Additional details of Monte Carlo sampling of posterior probabilities can be found in Ref [2,3,4,5]. The solution to the inverse problem is then

$$\mathrm{p}(\mathbf{m}|\mathbf{d}) = \frac{\mathrm{p}(\mathbf{d}|\mathbf{m})\mathrm{p}(\mathbf{m})}{\mathrm{p}(\mathbf{d})} \ , \tag{1.1}$$

where the probability of the data $\mathrm{p}(\mathbf{d})$ is independent of the environmental model and here can be neglected (set to unity). The posterior distribution $\mathrm{p}(\mathbf{m}|\mathbf{d})$ contains all relevant information and from this distribution all of the relevant features of the environment can be found such as the maximum *a posteriori* (MAP) estimator.

We are not interested in the environment itself but rather better estimates in the information usage domain $\mathbf{u}(\mathbf{m})$. Based on the posterior distribution $\mathrm{p}(\mathbf{m}|\mathbf{d})$, the distribution of $\mathrm{p}(\mathbf{u}(\mathbf{m}))$ is obtained and from this distribution all relevant statistics of the usage space can be obtained. In the present application the usage domain is transmission loss.

The posterior distribution of $\mathrm{p}(\mathbf{m})$ is first obtained and from this the probability distribution of $\mathbf{u}$ is obtained

$$\mathrm{p}(\mathbf{u}) = \int_{\mathcal{M}} \delta[\mathbf{u}(\mathbf{m}) - \mathbf{u}]\mathrm{p}(\mathbf{m}) \, \mathrm{d}\mathbf{m} \tag{1.2}$$

where $\mathcal{M}$ represents the model space. This integral is implemented numerically by using samples from the model space $\mathbf{m}$ based on the posterior distribution $\mathrm{p}(\mathbf{m}|\mathbf{d})$ and then binning $\mathbf{u}(\mathbf{m})$. From the distribution

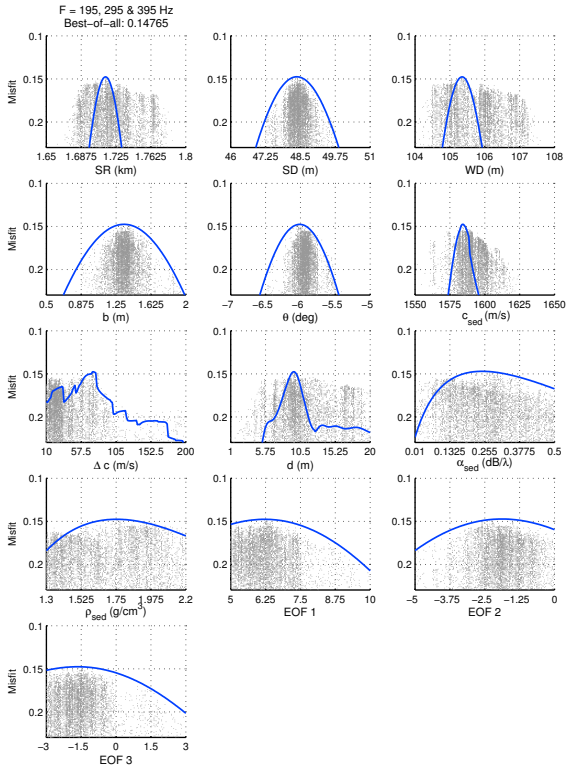

Figure 1.2. Marginal scatter diagrams of the SAGA search for the model parameters. The vertical axis represents the attained misfit on a linear scale. The thick line is the sensitivity curve of the multi-frequency misfit function using the best-fit model as a baseline.

$\mathrm{p}(\mathbf{u})$ all relevant statistical features of the usage domain, e.g., transmission loss, can be obtained, as for example the mean and variance of $\mathbf{u}$.

Assuming a simple linear data model as detailed in Ref [2], we obtain the Bartlett objective function $\phi(\mathbf{m})$ and likelihood function $\mathcal{L}(\mathbf{m})$,

$$\phi(\mathbf{m}) = \mathrm{tr}\mathbf{C} - \frac{\mathbf{d}^\dagger(\mathbf{m})\mathbf{C}\mathbf{d}(\mathbf{m})}{\|\mathbf{d}(\mathbf{m})\|^2} , \tag{1.3}$$

$$\mathcal{L}(\mathbf{m}) = \left[\frac{N}{\pi\phi(\mathbf{m}^{\mathrm{ML}})}\right]^N \exp\left[-N\frac{\phi(\mathbf{m})}{\phi(\mathbf{m}^{\mathrm{ML}})}\right] . \tag{1.4}$$

where $\mathbf{C}$ is the data covariance matrix, N number of sensors, and ML refers to the maximum likelihood estimate.

2. ASIAEX INVERSION

Data from the 2001 East China sea experiment (see Ref [6]) are used to illustrate the method. A 16-element vertical array was deployed in 105-m deep water. For the inversion (Fig. 1.2), a source towed at 48.5-m depth is used. Matched-field geoacoustic inversion using the selected frequencies 195, 295, and 395 Hz was carried out at T = 29 min over a parameter space of 13 parameters including geometrical, geoacoustic, and ocean sound speed EOF coefficients. Based upon the GPS position of R/V Melville, the source was approximately 1.7-km away from the VLA.

Figure 1.2 shows the marginal dot diagrams for the model parameters. The vertical axis is the Bartlett objective function (Eq. (1.3) normalized with tr$\mathbf{C}$) with respect to the parameters sampled during the SAGA inversion [5] with the normal mode code SNAP as a forward model. The thick line superimposed on each scatter plot was obtained by using the best-fit model corresponding to the optimal value of the objective function as a baseline and computing the sensitivity for the optimized parameter. We see that the sampled values for the array bow and tilt parameters (b and θ) are spread mainly inside the sensitivity curve and align mostly with the best-fit values. A similar behavior is observed for the ocean sound speed EOF coefficients but with a wider span. The sensitivity curves is smooth except for Δc and d. The non-smoothness of the curves are due to the number of propagating modes changing when varying the geoacoustic parameters. The consistency between the local (line) and global (dots) searches shows that this set of parameters is weakly correlated with the other parameters. For the geoacoustic parameters, most sampled values wander outside the curve. This reveals the more complicated structure of the multi-dimensional search space. Note that the sampled values for the source range (SR) and the water depth (WD) are spread uniformly throughout the range of the parameter interval. This is due to the strong coupling between these two parameters.

A second inversion is now carried out to determine the uncertainty for two of the most important model parameters. For simplicity, we assume that only water depth and sediment sound speed have any associated uncertainty. All other parameters are fixed at the optimal values found in the inversion detailed above. Varying only the above two parameters gives posterior probability indicated in Fig. 1.3. It is based on the likelihood formulation, Eq. (1.4) and using the same data as in the inversion.

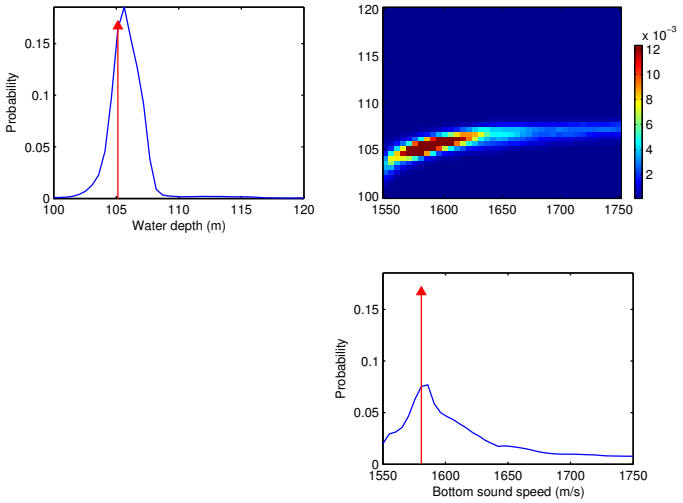

Figure 1.3. Marginal probabilities for water depth and bottom sound speed. The contour plot shows the 2D distribution. The red vertical arrow indicates the ML solution (saga best fit). Note, that the maximum in the marginal distribution might not correspond to the ML maximum.

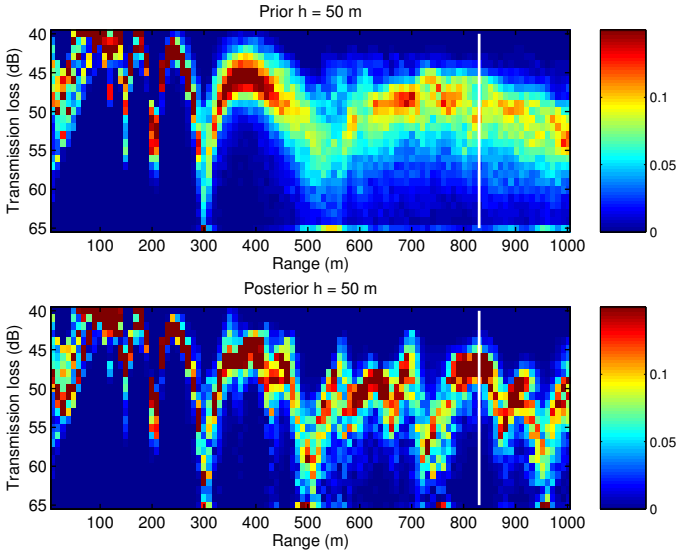

Figure 1.4. Prior (top) and posterior (bottom) probability distributions for transmission loss versus range at 50-m depth.

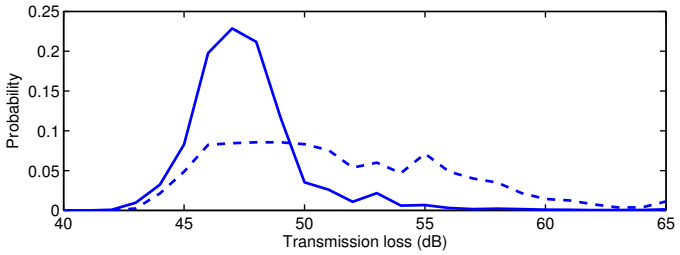

Figure 1.5. Posterior (solid) and prior (dashed) probabilities at 830-m range and 50-m depth. These correspond to a cut (white lines) though the contours in Fig. 1.4.

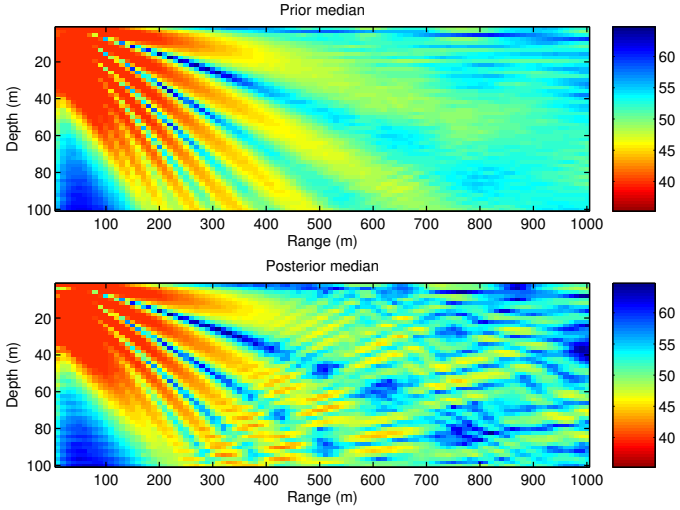

Figure 1.6. Median transmission loss (dB) based on prior (top) and posterior (bottom) distributions of environmental parameters (water depth and sediment sound speed).

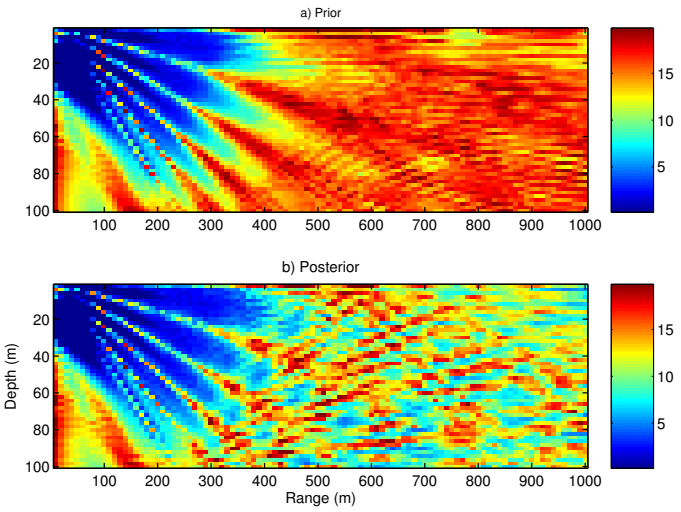

Figure 1.7. Range (dB) between 5th and 95th percentiles of the transmission loss for prior (top) and posterior (bottom) probabilities.

2.1 ASIAEX TL Prediction

The posterior probability (Fig. 1.3) is used to compute the posterior probability of $\mathbf{u}(\mathbf{m})$ using a frequency of 500 Hz and a source depth of 20 m. Except for water depth (bounds 100 to 120 m) and bottom sound speed (bounds 1550 to 1750 m/s), we keep the environment fixed at the values found in the inversion. In the present application, we evaluate $p(\mathbf{u}(\mathbf{m}))$ using grid integration (summing the values at discrete grid points). First the probability for TL at midwater depth (50 m) is evaluated, Fig. 1.4. The prior probability assumes even weighting of all the explored environmental models with the same bounds as above. The prior distribution (top) is spread out over a wide range but the posterior distribution (bottom) is more narrow. For the first 200 m, the propagation is only little influenced by the waveguide parameters and thus there is little difference between posterior and prior distributions. We then examine the probability at one point (50-m depth and 830-m range). This is done by taking a cut trough the contour plots in Fig. 1.4 at 830-m range (indicated by a white line, corresponding to a peak in TL curve), as shown in Fig. 1.5. The posterior (solid) is much more concentrated than the prior (dashed).

Contours of the median TL then are computed for the prior and posterior fields, Fig. 1.6. A good way to understand the uncertainty is to plot the 5th to 95th percentile ranges (gray area) of the prior and pos-

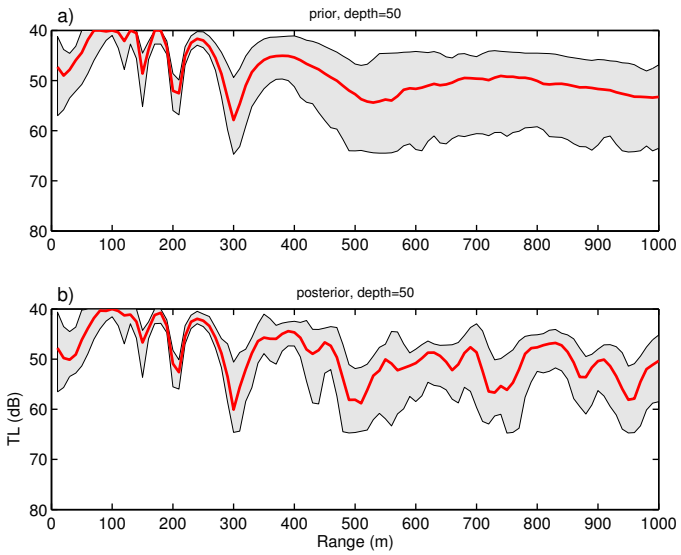

Figure 1.8. The median TL (red) at 50-m depth (dashed) based on all (a) prior samples and (b) posterior samples. The gray area indicates the range between the 5th and 95th percentiles.

terior fields, Fig. 1.7. Close to the source, there is little uncertainty for both prior and posterior fields as the sound field is not influenced by the waveguide parameters. Further away from the source, the prior uncertainty increases earlier in range than the posterior does, as the waveguide parameters are less well determined. It also is seen that around the nulls of the median fields, Fig. 1.6, the variations in the fields are the largest, Fig. 1.7.

The uncertainty is easily conveyed by plotting the median TL (red) combined with the 5th and 95th percentiles (represented by the gray area), see Fig. 1.8. Clearly, the posterior spread has decreased significantly.

3. CONCLUSIONS

An algorithm for estimating the statistical properties of transmission loss based on the output from a geoacoustic inversion has been described using a likelihood formulation. The likelihood function is developed assuming the error in the observed data is Gaussian. The examples presented here used data on vertical arrays, but any data could be used. The approach taken here then maps the environmental parameters via their

probability distributions into a probability distribution of transmission loss.

In the transmission loss domain, we can compute the full posterior distribution at all frequencies, ranges and depths. In the examples, we demonstrated how to use the full transmission loss probability distribution and extracted characteristic features such as median and lower/upper percentiles from this distribution.

Acknowledgments: This work was supported by the Office of Naval Research.

References

[1] A. Tarantola, *Inverse Problem Theory: Methods for Data Fitting and Parameter Estimation,* Elsevier, New York, 1987.

[2] P. Gerstoft and C.F. Mecklenbräuker, "Ocean acoustic inversion with estimation of *a posteriori* probability distributions," *J. Acoust. Soc. Am.* , **104**, pp. 808–817, 1998.

[3] S.E Dosso, "Quantifying uncertainties in geoacoustic inversion. I: a fast Gibbs sampler approach," *J. Acoust. Soc. Am.* , **111**, pp. 129–142, 2002.

[4] K. Mosegaard and A. Tarantola, "Probabilistic approach to inverse problems," Chapter in *International Handbook of Earthquake and Engineering Seismology*, (http://web.ccr.jussieu.fr/tarantola/), 2002.

[5] P. Gerstoft, "Inversion of seismoacoustic data using genetic algorithms and *a posteriori* probability distributions," *J. Acoust. Soc. Am.*, **95**, pp. 770–782, 1994.

[6] C-F. Huang and W.S. Hodgkiss, "Matched Field Geoacoustic Inversion of Low Frequency Source Tow Data From the ASIAEX East China Sea Experiment," *IEEE Oceanic Eng.*, to appear 2004.