Autonomous Underwater Vehicle (AUV) Path Planning and Adaptive On-board Routing for Adaptive Rapid Environmental Assessment

by

Ding Wang

B.S. in Automotive Engineering (1997) and S.M. in Physics (2000)
Tsinghua University, Beijing, China
S.M. in Electrical Engineering & Computer Science (2005) and S.M. in Ocean Engineering (2005)
Massachusetts Institute of Technology

Submitted to the Department of Mechanical Engineering
in partial fulfillment of the requirements for the degree of

Doctor of Philosophy in Mechanical Engineering

at the

MASSACHUSETTS INSTITUTE OF TECHNOLOGY

September 2007

Author .....
Department of Mechanical Engineering
June 15, 2007

Certified by .....
Henrik Schmidt
Professor of Mechanical and Ocean Engineering
Thesis Supervisor

Certified by .....
Pierre Lermusiaux
Associate Professor of Mechanical Engineering
and Ocean Science and Engineering
Co-advisor

Accepted by .....
Lallit Anand
Professor of Mechanical Engineering
Chairman, Departmental Committee on Graduate Students

Autonomous Underwater Vehicle (AUV) Path Planning and Adaptive On-board Routing for Adaptive Rapid Environmental Assessment

by

Ding Wang

Submitted to the Department of Mechanical Engineering
on June 15, 2007, in partial fulfillment of the
requirements for the degree of
Doctor of Philosophy in Mechanical Engineering

Abstract

In shallow water, a large part of underwater acoustic prediction uncertainties are induced by sub-meso-to-small scale oceanographic variabilities. Conventional oceanographic measurements for capturing such ocean-acoustic environmental variabilities face the classical conflict between *resolution* and *coverage*. The Adaptive Rapid Environmental Assessment (AREA) project was proposed to resolve this conflict by optimizing the location of in-situ measurements in an adaptive manner.

In this thesis, ideas, concepts and performance limits in AREA are clarified. Both an engineering and a mathematical model for AREA are developed. A modularized AREA simulator was developed and implemented in C++. Philosophies in AREA are discussed. Presumptions about the ocean are made to bridge the gap between the viewpoint in the oceanography community, where the ocean environment is considered to be a deterministic but very complicated system, and that of the underwater acoustic community, where the ocean environment is treated as a random system.

At present, how to optimally locate the in-situ measurements made by a single AUV carrying a CTD (conductivity, temperature and depth) sensor is considered in AREA. In this thesis, the AUV path planning is modeled as a Shortest Path problem. However, due to the sound velocity correlation effect, the size of this problem can be very large. A method is developed to simplify the graph for a fast solution. As a significant step, a linear approximation for acoustic Transmission Loss (TL) is investigated numerically and analytically.

In addition to following a predetermined path, an AUV can also adaptively generate its path on-board. This adaptive on-board AUV routing problem is modeled using Dynamic Programming (DP) in this thesis. A method based on an optimized predetermined path is developed to reduce the size of the DP problem and approximately yet efficiently solve it using Pattern Recognition. As a special case, a thermocline-oriented AUV yoyo control and control parameter optimization methods for AREA are also developed.

Finally, some AUV control algorithms for capturing fronts are developed. A framework for real-time TL forecasts is developed. This is the first time that TL forecasts have been linked with ocean forecasts in real-time.

All of the above ideas and methods developed were tested in two experiments, FAF05 in the northern Tyrrhenian Sea in 2005 and MB06 in Monterey Bay, CA in 2006. The latter MB06 sea exercise was a major field experiment sponsored by the Office of Naval Research and the thesis compiles significant findings from this effort.

Thesis Supervisor: Henrik Schmidt
Title: Professor of Mechanical and Ocean Engineering

Co-advisor: Pierre Lermusiaux
Title: Associate Professor of Mechanical Engineering
and Ocean Science and Engineering

Acknowledgments

First and foremost I would like to thank my advisor Prof. Henrik Schmidt and my co-advisor Prof. Pierre Lermusiaux for all of the guidance, ideas and support that they have graciously provided throughout the course of this thesis. I want to also thank Prof. Daniela Rus and Prof. Nicholas M. Patrikalakis for being my committee members.

Much of my time at MIT was spent in my office and interacting with fellow students, and I would like to acknowledge a few in particular: Tianrun Chen and Harish Mukundan, for helping me a lot in life and listening my complains.

None of this would have been possible without the love and support of my parents and my relatives. I haven't been back to my hometown for almost ten years. I am so grateful for their understanding. The love and support from my girl friend, Ling Li, has seen me through all stages of this thesis, from discussions about methodologies of doing research to biology and artificial intelligence development.

I also want to thank all my friends and all the people supported me. I can always get courage and confidence from them when I felt depressed. Particularly, I want to thank my basketball teammates for giving me a place to release my physical energy.

Finally, I would like to thank the Office of Naval Research for providing funding for this project under Grant N00014-01-1-0817 Capturing Uncertainty and the Adaptive Sampling and Prediction MURI under subcontract 00000917 to Princeton University.

Contents

List of Figures

Chapter 1

Overview

Many Oceanic variabilities exist in the ocean, especially in shallow water, where wind driven flows, tidal currents, river outflow, internal waves, solitary waves, fronts, eddies, thermal changes etc are some of the commonly dominant oceanographic processes. These processes and their intercoupling and interactions with the seabed make the shallow water ocean-acoustic environment highly variable in time and space. In the water column, the Sound Velocity Profile (SVP) etc can vary in complex dynamical ways.

Those variabilities span a wide range of spatial and temporal scales [5, 6]. Conventional oceanographic measurements cannot provide the ability to synoptically observe all those dynamically interlocking, patchy and intermittent processes in coastal ocean, especially for sub-meso-scales short in time and space [7]. Consequently the coastal environment will always be under-sampled at these small and fast scales. The ocean-acoustic environment parameters of the water column and the seabed are generally not known in sufficient detail and with enough accuracy for satisfactory prediction of long-range acoustic propagation in shallow water, even though the shallow water acoustics has been thoroughly investigated both theoretically and experimentally [1].

Modern ocean modeling and assimilation frameworks have a capability of representing the smaller, sub-grid-scale variability statistically [8]. From an acoustic viewpoint, the sub-meso scale variabilities of the order of hundred meters to kilometers make the coastal ocean-acoustic environment largely unknown with many

uncertainties in terms of imperfect sound velocity, depth of the thermocline etc. Such uncertainties can be responsible for a large part of the acoustic prediction uncertainty [9, 10].

To determine the environmental variability of the critical sub-meso scales and short temporal scales, a local, high resolution, rapid deployable in-situ measurement capability has long been recognized as a very important tactical need [9]. By assimilating the in-situ measurement data with ocean modeling, resolution of the ocean filed estimation can be dramatically improved and the acoustic prediction uncertainty can be strongly reduced. The project *Adaptive Rapid Environmental Assessment* (AREA) was proposed for this purpose.

However, the coastal environmental assessment is facing the classical conflict between *resolution*, needed to capture the fine scale variability and *coverage*, needed for the large scale environmental phenomena. Thus, the *Rapid Environmental Assessment* (REA) resources available must focus on the environmental uncertainties critical to the specific acoustic system. Thus optimizing the REA resources deployment pattern becomes the major problem.

In this thesis, the REA resource is a single AUV carrying a CTD (conductivity, temperature, depth) sensor and the problem is focused on how to optimally locate the in-situ measurements made by the AUV, i.e. the AUV path planning problem. The path can be planned before the AUV is launched or the AUV can adaptively determine its waypoints one by another one on-line. For the predetermined path, the path planning problem can be solved by shortest path algorithms, but the computation needed can be very intensive due to the sound velocity correlation effect. In this thesis, an approximate shortest path problem is developed, which is much smaller than the original one and can be solved in real-time.

For the adaptive path, the adaptive on-board AUV path planning problem can be modeled using Dynamic Programming (DP). The associated DP problem can be simplified based on an optimized predetermined path and quickly solved by a specific Approximate Dynamic Programming (ADP) method. In shallow water, the thermocline region is often associated with big sound velocity prediction error. The

thermocline-oriented AUV yoyo control is developed in this thesis, which can lead the environmental sampling focus on the thermocline region and thus capture most critical environmental parameters. As a special case of the adaptive on-board AUV path planning, the thermocline-oriented AUV yoyo control has two parameters need to be optimized. In this thesis, this process is carried by exhaustive search in a small parameter space.

It should be noticed that all the above AUV path planning problems are required to be solved in real-time in AREA, i.e. in several hours. This is a big challenge.

In addition to capturing the environment variabilities, AUV control algorithms for capturing fronts are also developed, including a horizontal zig-zag for surface temperature gradient tracking etc.

This thesis is organized as follows. Chapter 2 introduces some basics in ocean-acoustic environment, ocean field estimation and underwater acoustics. Chapter 3 introduces the motivations, and the engineering and mathematical models in AREA, defined as a path planning problem. Chapter 4 is the overview of the optimization algorithms used in this thesis, including Linear Programming (LP), Network Optimization and DP. Chapter 5 discusses how to model the path planning problem for the predetermined AUV path in network optimization. As a very important step, the linear approximation for transmission loss (TL) is discussed. Chapter 6 discusses how to model the adaptive on-board AUV path planning problem in DP. As a special case, the thermocline-oriented AUV yoyo control is introduced. Chapter 7 discusses how to solve those optimization problems, including the network optimization problem, the DP problem and the AUV yoyo control parameter optimization problem. Chapter 8 introduces the AUV control algorithms developed for capturing fronts. Chapter 9 shows the results from the experiment in Monterey Bay, CA in 2006. Finally, conclusions and future work suggestion are in Chapter 10. The philosophical presumptions used in this thesis are presented in Appendix A.

Chapter 2

Background Introduction

2.1 Overview of Ocean-Acoustic Environment

The ocean is a very complicated dynamic system, evolving on multiple temporal and spatial scales. The ocean-acoustic environment is essentially inhomogeneous and time-varying, which determines sound wave propagation pattern in the ocean.

2.1.1 Basics in Ocean-Acoustic Environment

In the ocean, acoustic propagation is mainly dependent on the ocean Sound Velocity Profile (SVP). According to a simplified formula given by Medwin [11], sound velocity (c) in meters per second can be expressed as a function of temperature (T) in degrees centigrade, salinity (S) in parts per thousand, and depth (z) in meters.

$$c = 1449.2 + 4.6T - 0.055T^2 + 0.00029T^3 + (1.34 - 0.01T)(S - 35) + 0.016z \quad (2.1)$$

Figure 2-1 shows typical patterns of ocean SVP. In nonpolar regions, wind and wave activity mix the top layer water and result in a *mixed layer* with almost constant temperature inside [1]. In this isothermal layer sound velocity increases with depth due to the increasing ambient pressure and hence leads to the *surface duct* region. In a warmer season or warmer part of the day, the sea surface temperature is higher and hence the sound velocity increases toward the surface (this phenomenon is also referred to as "afternoon effect"). Below the mixed layer is the main thermocline where

the temperature and hence the sound velocity decreases with depth. Below the main thermocline, the temperature is constant and the sound velocity increase because of increasing ambient pressure. The *deep sound channel axis* exists between those two regions, which corresponds to the minimum sound velocity. In deep water sound velocity often shows good stability; while in the upper ocean much more oceanographic processes exist, thus sound velocity shows greatest variability. In polar regions, the water is coldest near the surface, so the SVP varies in different patterns. In shallow water, the depth is insufficient for the depth-pressure term to be significant. Thus the winter profile tends to isovelocity, whereas the summer profile has a higher sound velocity near the surface (Figure 2-2). Aside from sound velocity effects, the ocean water is absorptive and will cause attenuation that increases with acoustic frequency.

The ocean is a waveguide bounded by air and seabed. The sea surface is a sim-

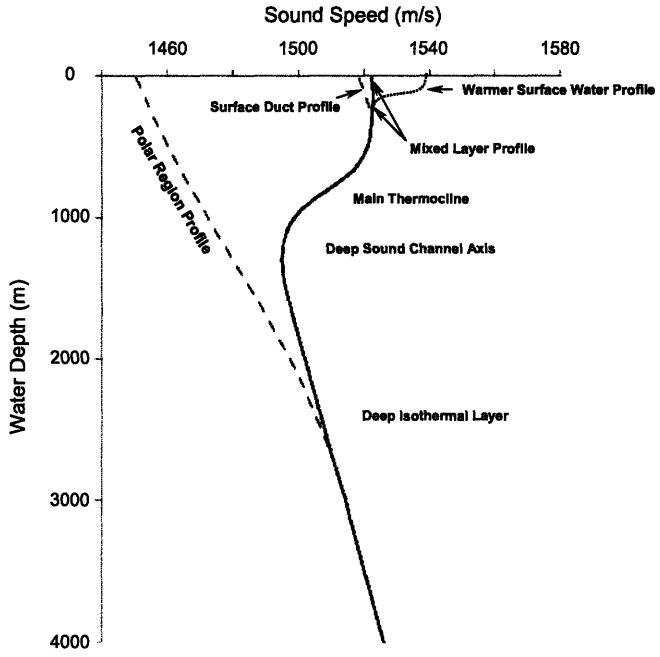


Figure 2-1: Generic sound speed profiles

ple horizontal boundary with nearly perfect reflectivity. The seabed is a quite flat and lossy boundary with layered structure supporting elastic waves. Its geoacoustic properties are summarized by density, compressional and shear wave velocity, and

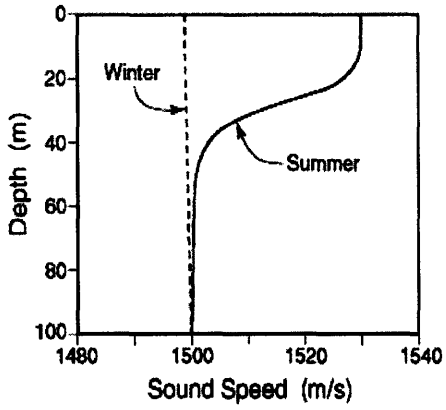


Figure 2-2: Generic sound speed profiles in shallow water

attenuation profiles. The reflectivity characteristics of seabed and topography can strongly vary in different geographical locations [1]. Moreover, both boundaries have small-scale roughness which causes scattering and attenuation of sound.

2.1.2 Ocean-Acoustic Environmental Variabilities

Oceanographic variabilities exist in the water and the seabed, spanning over a wide range in space and time. From an acoustic view point, those oceanographic variabilities inevitably complicate the ocean-acoustic environment and affect acoustic propagation in the ocean more or less.

Water Column

Ocean waters are constantly on the move. As shown in Figure 2-3, in the water column, there are many physical and biological processes with scales covering from 1mm to 1000km in space and from 1sec to 10 year in time. Meteorology-related processes such as the sea surface thermal changes, water mixing caused by wind and wave breaking, surface wave-induced roughness etc may be dominant in the upper layer ocean-acoustic environment. Some other oceanographic processes such as internal waves, internal solitons, internal tides, currents, tides, eddies, fronts, fine structure, microstructure, bubble clouds, plankton migration and distribution etc

may strongly affect acoustic propagation too.

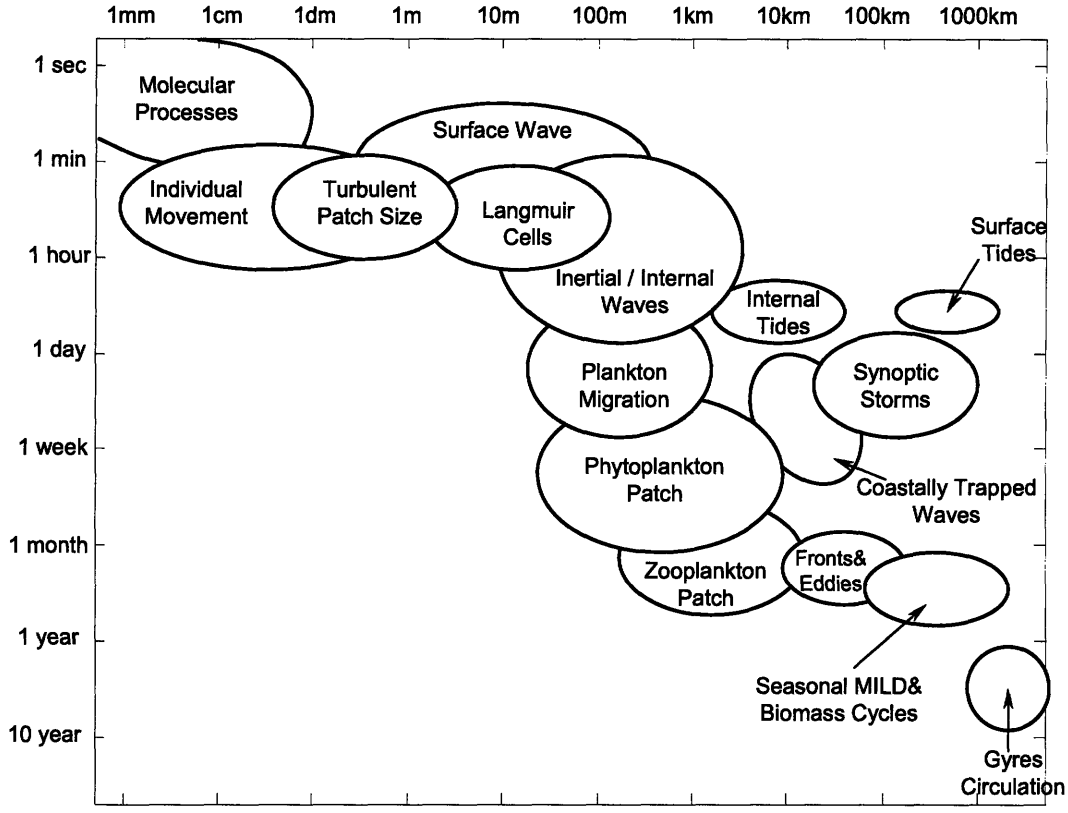


Figure 2-3: Spatial and temporal scales of physical and biological parameters and processes in the sea [3].

Internal Waves Internal waves are gravity waves that oscillate due to the density stratification in the water column, the buoyancy force and the Coriolis force. It has been found that in many locations in the ocean internal wave-induced SVP fluctuations are usually very significant sources or even the dominant causes of sound wave propagation variations [12, 13]. Furthermore, in polar area, internal wave-induced water density variations may also affect the sound wave propagation [14]. The nearly ubiquitous linear internal waves provide a continuous scattering mechanism for redistributing acoustic energy in the oceans. In contrast, nonlinear internal waves, such as internal solitons, provide strong, discrete and possibly azimuth-dependent scattering

events due to their generally higher amplitude and shorter wavelength. Some interesting phenomena such as strong coupling of acoustic normal modes and ducting effect may occur in this case [15]. Generally, internal waves may significantly affect acoustic amplitudes and phases and nearly horizontal ray paths thus affect travel time. Internal solitons may also affect acoustic spatial and temporal coherences [12, 13, 15, 16]. Internal waves can be generated from surface (e.g. by surface wave), bottom (e.g. by a quasisteady current advecting a stratified ocean over bottom topography) and interior of water. Disturbance induced by moving ships may trigger internal waves too [13]. One particularly important type in shallow water is the internal waves produced by tidal currents flowing over a sloping sea floor, which are also called internal tides [15]. Internal waves are characterized by scales from 100m to 10km or more in the horizontal, 1 to 100 m in the vertical, and from about 10 min to 1 day in time [12]. So far, internal waves are usually considered random. The Garrett-Munk statistical model, an empirical model of the internal wave spectrum based on linear internal wave theory has been widely used in deep water. In the upper ocean and in shallow water, the GM model generally turns out to be inappropriate; in the latter situation, the internal wave sometimes appears to have a deterministic nature, characterized by propagation of a soliton [12]. Combination of the GM model and soliton packets was also used in shallow water [16].

Fine- and Microstructure The SVP fluctuations on scales larger than a few meters in the vertical direction are internal wave-related. A smaller scale, the fluctuations are due to the fine- and microstructure [17, 18]. The fine- and microstructure of the temperature/salinity/sound velocity field in the ocean have an irregular stratified nature with layers (within which the water is relatively well mixed), separated by regions of large vertical gradient. Fine- and microstructure variability in SVP involve scales from several meters to hundreds of meters in the horizontal, 1cm to about 10 m in the vertical and milliseconds in time. Such variability would affect sound propagation in the frequency range from approximately 1 to tens of kHz [1] and brings in a large difference in the log-intensity fluctuations, keeping the phase

fluctuations unchanged [18]. So far, the fine- and microstructure variabilities have to be considered as random processes.

Currents Currents in the ocean are caused by wind stress, buoyancy fluxes (due to the water's salinity, heat content and gravity), tides and affected by Coriolis force, bottom topography etc [19, 20, 21, 22, 23]. Ocean currents flow in complex patterns, e.g. wind stress may induce upwellings, wind driven circulations and geostrophic currents in surface ocean; buoyancy fluxes produce thermohaline circulations in deep ocean; with the effect of Corioslis force, ocean currents are in large patterns of rotation called "gyres" in each major ocean. These gyres move in a clockwise direction in the northern hemisphere, and in a counterclockwise direction in the southern hemisphere. Major ocean currents include North & South Equatorial Current, Equatorial Counter Current, North Atlantic Drift, Gulf Stream, Kuroshio Current, Antarctic Circumpolar Current, East Australian Current, Cromwell current etc. These major ocean currents are characterized by horizontal scales of variability limited only by the size of the basin, vertical scales of a few hundred meters, and temporal scales from a few days to seasonal [12]. While in shallow water area, tidal currents caused primarily by the rise and fall of the tide may be more important [23].

The characteristic parameters of major ocean currents are nearly constant in space and time. The horizontal component of the velocity of the currents is much greater than its vertical component. Usually, vertical current profiles show fine structure rather than being smooth, which is caused by vertical layers existing everywhere in the ocean [24].

Velocity of ocean currents may significantly change the phase of a sound wave and hence may cause a noticeable change in the amplitude [24]. If the range from the sound source is sufficiently large, travel time of sound wave may be changed and the principle of acoustic reciprocity may be broken [24, 25, 26]. Furthermore, currents may lead to a qualitative change in sound wave equation if $|\partial v/\partial z| > |\partial c/\partial z|$, where v is the current velocity, c is the sound velocity and z is the depth. Some theoretical results are shown in [24, 27], from which the upstream transmission loss (TL) and downstream TL are

surprisingly different. Ocean currents can be captured by modern oceanographic observation systems and are now considered as deterministic structures.

Tides Tide is the vertical rise and fall of the surface of a body of water on earth, caused primarily by the variation in gravitational forces resulting from the change of position of the sun and the moon relative to points on the earth's surface [23, 19, 28]. There are 3 major types of tides: diurnal tide, semi-diurnal tide or mixed [23]. The horizontal scales of tides are variable and in vertical, tidal wave height can exceed 10m, but subsurface effects may extend to greater distance [12]. In deep water area the elevation change induced by tides has only little effect on the SVP in the water column. However, the associated tidal currents and internal tides may have a profound effect on acoustic propagation: the tidal currents may affect acoustic travel time [28]; acoustic amplitude fluctuations are generally noise-like and are insignificantly dependent of tidal currents; while acoustic phase fluctuations are simply correlated with tidal variations and are proportional to v_R the horizontal velocity of tidal currents [12, 24]. In shallow water area, the tidal effects on acoustics are much greater [12]. Generally, tidal currents become stronger as one approaches the coast, and tidal currents play an increasingly important role in the local circulation [23]. Systematic observations of ocean tides have been made in some locations for several centuries. Accurate prediction of such tides by relating their height and phase to the movements of the sun and moon was introduced by Lord Kelvin in 1870 [19]. However, tidal currents and internal tides can't be accurately predicted. The total effects of tides on the ocean acoustic environment have to be modeled as partially random and partially deterministic.

Eddies An eddy is a rotating parcel of fluid. As such, the eddy concept can be applied to phenomena ranging from momentary vortices in the sea-surface flow to the steady circulation of a basin-wide gyre. For underwater acoustics, however, mesoscale eddies — large coherently rotating bodies of water, which are non-stationary objects — are the most important. A surprising feature of mesoscale eddies is the large

variability in size and current velocity. Diameters vary from 100 to 500 km, with maximum current speeds ranging from less than 20 to more than 150 cm/sec [29, 20]. The effects of mesoscale eddies can reach the whole ocean depth, but primarily stay in the upper region of the ocean. The life of mesoscale eddies varies from one to few months. There are two distinct types of eddies: Cyclonic eddies, consisting of a mass of cold water circulating in the counterclockwise direction (in the northern hemisphere); anticyclonic eddies, consisting of a mass of warm water [29]. Mesoscale eddies may distort the normally horizontal isotherms a lot. Hence substantial perturbations in sound speed, primarily resulting from the large temperature variations, may show up [29]. The amplitude and phase of a sound field can be affected by the variation of sound speed and the water motion in the eddy. The latter factor may cause amplitude variation greater than 10 - 12 dB, and a phase change much greater than π [24]. Moreover, mesoscale ocean features such as fronts and eddies can cause sound to be refracted in the horizontal plane, and hence cause the source's measured bearing to differ from the true one [30, 31]. The mesoscale eddies can be captured and thus can be considered as governed by deterministic processes.

Fronts An oceanic front is the interface between two water masses of different properties. Usually, fronts show strong horizontal gradients of temperature and salinity. These will result in differences in SVP across the front, thus causing changes in acoustic propagation. Some fronts which have weak horizontal gradients at the surface have strong gradients below the surface. In some cases, gradients are weak at all levels, but variability across the front is sufficient to complicate sound transmission. There are several types of fronts: planetary fronts, upwelling fronts, shelf break fronts, shallow sea fronts, plume fronts, estuarine fronts, etc. Fronts' scales can be very variable, e.g. in the deep ocean planetary fronts can span the width of entire ocean basins; fronts in estuaries may be only a few meters wide. However, mesoscale ocean fronts are the most important to underwater acoustics and can induce changes in acoustic propagation path, acoustic intensity, signal travel time and signal shape etc [12]. Mesoscale fronts can be modeled deterministically.

Bubble Clouds Breaking waves and rainfall may produce bubble clouds, which are a major factor in near-surface sound propagation. It has been proved that there can be as many as 10^5 to 5×10^6 bubbles per cubic meter at radii between 15 and 16 microns near the ocean surface even during calm cases. The direct consequence of bubbles at sea have been demonstrated to result in near-surface excess attenuations as great as 60 dB/m and speeds of sound that are tens of meters per second less than 1500m/s [11]. Furthermore, bubbles may exist in sediments.

Biological Processes Marine creatures such as plankton, fish, mammals can backscatter sound wave. Biological backscattering is frequency dependent, and varies with depth, time of day, season, and area. Both resonant and nonresonant scattering effects are involved [32]. Figure 2-4 shows a marine biological pyramid that reveals the immense size range of life at sea.

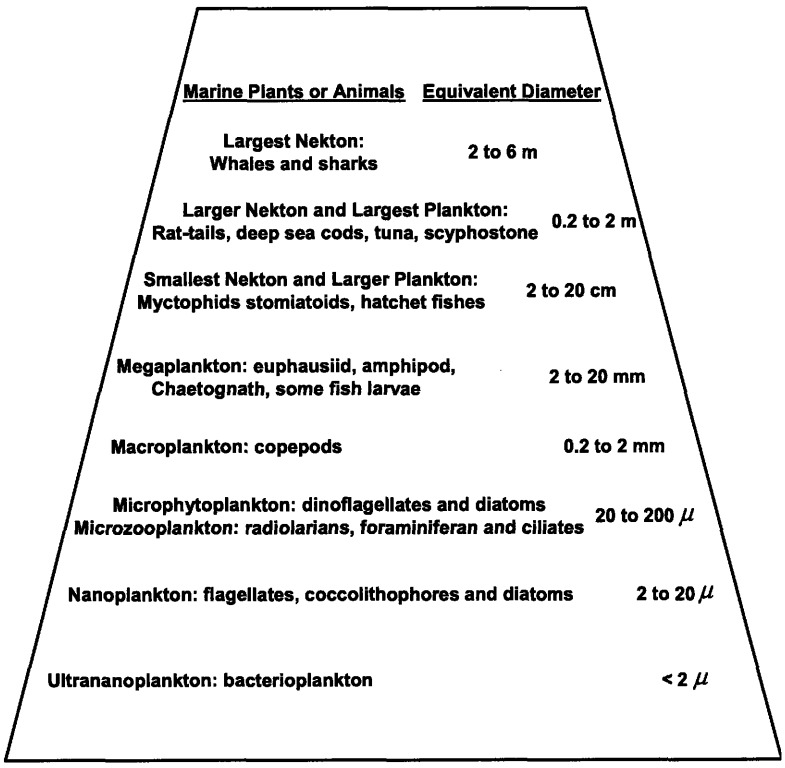


Figure 2-4: Marine biological pyramid with diameter of equivalent spherical volume of the plants or animals [4].

Seabed and Coupling

The seabed is of paramount interest in shallow-water acoustics, particularly at low frequencies and downward refracting environments. Compared with the water column, seabed is less variable. However, the fluctuations in the water column parameters affect not only the sound wave propagation, but also modify the properties of the seabed [12, 33, 34]. In addition, current flow may change the bottom topography, which in turn may affect internal waves and currents etc. Therefore the interaction between currents and seabed make the ocean-acoustic environment more complex.

In fact, coupling widely exists among those physical, biological processes in water and seabed. Eddies are induced by currents, while the boundary of an eddy is often a front. Meteorology-related processes such as wind, surface thermal changes, rainfall are causes of ocean currents. Internal waves may play an important role in mixing processes in the ocean and affect the mean ocean circulation, the ocean temperature and salinity structure [13]. Internal waves and upwelling currents can uplift of phytoplankton into the sunlit layer of the upper ocean, hence affect the biological variability. Tides produce tidal currents, which can produce internal waves (internal tides) by interacting with topography. So, generally speaking, the ocean-acoustic environment is super complex with variabilities over a wide range and coupled together.

2.2 Overview of Ocean-Acoustic Environment Estimation

2.2.1 Ocean Field Estimation

Ocean-Acoustic Environment Estimation can be viewed as a particular case of ocean field estimation. Ocean field estimation requires a knowledge of the distribution and evolution in space and time of the physical, biological and chemical characteristics of the sea [35]. The oceanic physical state variables are usually the velocity components, the pressure, density, temperature and salinity. Examples of biological and chemical state variables are concentration distributions of nutrients, plankton, dissolved and

particulate matter, etc. From observations, basic conservation laws (such as conservation of mass, momentum balance, thermal energy balance, conservation of salt etc) and principles of oceanic physics, biology and chemistry, the ocean dynamical models can be formulated and model parameters can be estimated to approximate those nonlinear, highly variable, wide-scale and interdisciplinary oceanic variabilities. If the state of the ocean at a certain time is given, i.e. boundary and initial conditions are given, the state of the ocean at a later time can be forecasted approximately. However, comprehensive and accurate ocean data acquisition is difficult and costly; ocean field experiments are often uncertain and limited. The complexity of ocean dynamics impedes the ocean models accurate too. Consequently, the ocean forecast/nowcast may be associated with great uncertainties. Sources of the uncertainties include model simplification errors, unresolvable ocean variabilities, boundary condition error, initial condition errors, miscalculation errors etc [8]. To reduce uncertainties, observation data and ocean dynamical models are combined by Data Assimilation (DA).

2.2.2 Data Assimilation

Ocean Data Assimilation refers to the quantitative estimation of marine fields of interest by melding data and dynamics in accord with their specific uncertainties. A data assimilation system consists of three components: the observations, a dynamical model and a data assimilation criterion [35]. By definition, DA can be viewed as an estimation problem. With different use or availability of data, the estimation problem can be divided into 3 types: filtering, forecasting and smoothing. The schemes for solving the assimilation problem have different backgrounds and can generally be related to estimation theory or control theory etc. In this section, principles and methods in data assimilation will not be enumerated, instead the basic ideas in data assimilation will be illustrated through an example — the Objective Analysis (OA), which is often used for ocean nowcast and field initialization. More detailed and comprehensive overview of data assimilation can be found in [35, 36].

Objective Analysis

The objective analysis is a ***Linear Least Square Estimator*** using Gauss-Markov or minimum error variance criterion. Suppose that the approximate ocean dynamical model is linear and discretized:

$$\hat{\psi}_k(-) = A_{k-1}\hat{\psi}_{k-1}(+) \tag{2.2}$$

and the measurement model is linear too:

$$d_k^t = D_k\psi_k^t + v_k^t, \tag{2.3}$$

where, vector $\hat{\psi}_{k-1}(+)$ is the ***posterior*** estimation of the ocean state at stage $k-1$; matrix A_{k-1} is the system evolution coefficient at stage $k-1$; vector $\hat{\psi}_k(-)$ is the ***a priori*** estimation of ocean state at stage k, i.e. if the current stage of time is stage $k-1$, $\hat{\psi}_k(-)$ is the forecast of the ocean state at stage k based on the dynamical model; vector d_k^t is the true measurement data at stage k; vector ψ_k^t is the true ocean state at stage k; matrix D_k is the coefficient linking ocean state and measurement data; vector v_k^t is the true measurement error at stage k, which can be viewed as a realization of random vector v_k.

According to the philosophical presumptions made in Appendix A, ψ_k^t is only partially known, it can be viewed as a realization of random vector ψ_k, though essentially ψ_k^t is thought as deterministic. ψ_k can be modeled as:

$$\psi_k = \hat{\psi}_k(-) + \omega_k(-) \tag{2.4}$$

where $\omega_k(-)$ is the random forecast error associated with the dynamical model uncertainty. Usually, $\omega_k(-)$ and v_k are uncorrelated. Thereafter, d_k^t can be viewed as a realization of the random vector d_k which is modeled as:

$$d_k = D_k\psi_k + v_k. \tag{2.5}$$

If now the ocean forecast $\hat{\psi}_k(-)$, the covariance matrix of $\omega_k(-)$, the true measurement data d_k^t and the covariance matrix of v_k are known, $\hat{\psi}_k(+)$ the nowcast of ocean state at stage k can be obtained via OA as follows:

$$\begin{aligned}
\hat{\psi}_k(+) &= \bar{\psi}_k + Cov(\psi_k, d_k)\, Cov(d_k, d_k)^{-1}\,[d_k^t - \bar{d}_k] \\
&= \bar{\psi}_k + \Lambda_k(-)\, D_k^T [D_k \Lambda_k(-)\, D_k^T + R_k]^{-1} [d_k^t - \bar{d}_k] \qquad (2.6) \\
\Lambda_k(+) &= \Lambda_k(-) - Cov(\psi_k, d_k)\, Cov(d_k, d_k)^{-1}\, Cov(d_k, \psi_k) \qquad (2.7) \\
&= \Lambda_k(-) - \Lambda_k(-)\, D_k^T\, [D_k \Lambda_k(-)\, D_k^T + R_k]^{-1}\, D_k \Lambda_k(-) \qquad (2.8)
\end{aligned}$$

where

$$\begin{aligned}
\Lambda_k(-) &= Cov(\omega_k(-)\,,\ \omega_k(-)) \\
R_k &= Cov(v_k, v_k)
\end{aligned}$$

$\bar{\psi}_k$ is the background (or guess) of ψ_k^t, and most likely $\bar{\psi}_k = \hat{\psi}_k(-)$; $\bar{d}_k$ is the background (or guess) of the measurement data, usually $\bar{d}_k = D_k \hat{\psi}_k(-)$; $\Lambda_k(-)$ indicates the uncertainty associated with $\hat{\psi}_k(-)$; $\Lambda_k(+)$ indicates the uncertainty associated with $\hat{\psi}_k(+)$. Since $Cov(d_k, d_k)$ is usually positive definite, Eq. 2.7 shows that $\Lambda_k(+) < \Lambda_k(-)$, which means that through OA the uncertainty associated with ocean estimation is reduced. R_k is the covariance matrix of the measurement noise.

In this project, OA is used to assimilate the CTD (conductivity, temperature, depth) data with the SVP forecast generated by the **Harvard Ocean Prediction System** via **Error Subspace Statistic Estimation (HOPS/ESSE)** [37, 38]. In this case, the measurement model can be written as:

$$d_k = D_k\, c_k + v_k \qquad (2.9)$$

where, c_k is the random SVP at stage k. D_k is a sparse matrix with only one item

equal to 1 and others equal to 0 in each row, like

$$D_k = \begin{bmatrix} 1 & 0 & \cdots & 0 \\ 0 & 0 & \cdots & 1 \\ \vdots & \vdots & \vdots & \vdots \\ 0 & 1 & \cdots & 0 \end{bmatrix}. \tag{2.10}$$

The associated OA formulas can be written as:

$$\hat{c}_k(+) = \hat{c}_k(-) + \Lambda_k(-) D_k^T [D_k \Lambda_k(-) D_k^T + R]^{-1} [d_k^t - D_k \hat{c}_k(-)], \tag{2.11}$$

$$\Lambda_k(+) = \Lambda_k(-) - \Lambda_k(-) D_k^T [D_k \Lambda_k(-) D_k^T + R]^{-1} D_k \Lambda_k(-). \tag{2.12}$$

$\hat{c}_k(-)$ and $\Lambda_k(-)$ can be generated by ***HOPS/ESSE***.

$$R = \begin{bmatrix} \sigma_n^2 & & & \\ & \sigma_n^2 & & \\ & & \ddots & \\ & & & \sigma_n^2 \end{bmatrix} \tag{2.13}$$

where, σ_n is the standard deviation of CTD noise, which is assumed to be stationary and white.

2.2.3 Variability v.s. Uncertainty

According to the philosophical presumptions made in Appendix A and descriptions in 2.1, in this project the ocean is thought to be a complex but deterministic system with tremendous variabilities over wide-range scales in time and space. Among them, meso-to-large scale variabilities can be adequately formulated by ocean dynamical models [12, 35]; while it is hard to model submeso-to-small scale variabilities in classical ocean dynamical models, due to limited knowledge of oceanic processes within these scales. Therefore, the ocean system should be viewed as a partially known deterministic process. A stochastic mathematical model is suitable for mod-

eling/estimating the ocean system (Appendix A). In this way, the resolvable oceanic variabilities will contribute to the mean values in ocean estimation; the unresolvable variabilities will be transferred into uncertainties associated with the ocean estimation.

Variability and uncertainty are related but different [8]. Oceanic variability is referred to ocean state changes along space or time. Uncertainty in ocean estimation is referred to statistic characterizations of the stochastic ocean modeling. Oceanic variabilities can contribute to uncertainties in ocean estimation; however ocean estimation uncertainties can arise from many other sources [8]:

- To reduce computational expenses, ocean models are simplified and explicit calculations are only performed on a restricted range of spatial and temporal scales (referred to as the"scale window").
- Approximate representations or parameterizations in ocean models due to limited knowledge of oceanic processes within the scale window.
- Initial conditions and model parameters are inexact.
- Interactions between the ocean and earth system are approximate and ocean boundary conditions are inexact.
- Numerical inaccuracy and instability.

2.3 Overview of Acoustic Modeling

Acoustic wave equation is derived from mass conservation equation, Newton's 2nd law and the adiabatic equation of state.

$$\text{Mass conservation:} \quad \frac{\partial \rho}{\partial t} = -\nabla \cdot \rho \vec{v} \tag{2.14}$$

$$\text{Newton's 2nd law:} \quad \frac{\partial \vec{v}}{\partial t} + \vec{v} \cdot \nabla \vec{v} = -\frac{1}{\rho} \nabla p(\rho) \tag{2.15}$$

$$\text{State equation:} \quad p = p_0 + \rho' \left[\frac{\partial p}{\partial \rho} \right]_s + \cdots \tag{2.16}$$

The standard linear wave equation for pressure is

$$\nabla^2 p - \frac{1}{c^2(\vec{x})}\frac{\partial^2 p}{\partial t^2} = f(\vec{x}, t). \tag{2.17}$$

$f(\vec{x}, t)$ represents sound sources. Using the frequency-time Fourier transform, the frequency-domain wave equation, or *Helmholtz equation*, can be obtained:

$$\left[\nabla^2 + k^2(\vec{x})\right] p(\vec{x}, \omega) = f(\vec{x}, \omega) \tag{2.18}$$

with

$$k(\vec{x}) = \frac{\omega}{c(\vec{x})}.$$

Starting from the *Helmholtz equation*, different mathematical methods can be applied to solve the wave equation. There are four types of models often used to describe sound propagation in the sea: ray method, wave number integration method, normal model method and parabolic equation method. In addition, direct, discrete methods such as the *Finite Difference Method* (FDM), the *Finite Element Method* (FEM) can also be used to solve the wave equation. However, their importance in underwater acoustics is rather limited due to excessive computational requirements [1].

2.3.1 Ray Methods

Eq. 2.19 is the Helmholtz equation in cartesian coordinates with a point source at $\vec{x_s}$.

$$\left[\nabla^2 + k^2(\vec{x})\right] p = S_\omega \delta(\vec{x} - \vec{x_s}) \tag{2.19}$$

S_ω is the *source strength* with respect to sound pressure. It is now assumed that the solution of Eq. 2.19 can be expressed as the sum of *ray series* as shown in Eq. 2.20.

$$p(\vec{x}) = e^{i\omega\tau(\vec{x})} \sum_{j=0}^{\infty} \frac{A_j(\vec{x})}{(i\omega)^j} \tag{2.20}$$

where, τ and A_j, $j = 0, 1, \cdots, \infty$ are to be determined.

Substituting Eq. 2.20 into Eq. 2.19, an infinite sequence of equations for the functions τ and A_j can be obtained.

$$O\left(\omega^2\right) \quad : \quad |\nabla\tau|^2 = c^{-2}\left(\vec{x}\right) \tag{2.21}$$

$$O\left(\omega\right) \quad : \quad 2\nabla\tau \cdot \nabla A_0 + \left(\nabla^2\tau\right) A_0 = 0 \tag{2.22}$$

$$O\left(\omega^{1-j}\right) \quad : \quad 2\nabla\tau \cdot \nabla A_j + \left(\nabla^2\tau\right) A_j = -\nabla^2 A_{j-1}, \;\; j = 1, 2, \cdots \tag{2.23}$$

In the standard ray method, only the *eikonal* equation (Eq. 2.21) and the first transport equation (Eq. 2.22) will be considered. A_j, $j = 1, 2, \cdots$ will be ignored by assuming $\frac{A_j}{\omega^j} \approx 0$, which implicates a high-frequency approximation.

$\tau(x)$ can be obtained by solving the *eikonal* equation, which however is a nonlinear Partial Differential Equation (PDE) and difficult to solve directly.

$$\frac{d\mathbf{x}}{ds} = c\nabla\tau \tag{2.24}$$

$$\frac{d\tau}{ds} = \frac{1}{c} \tag{2.25}$$

If we define the *ray trajectory* $\mathbf{x}(s)$ by Eq. 2.24, and transform the cartesian coordinates into the ray trajectory coordinates, the eikonal equation becomes to Eq. 2.25, which is a very simple Ordinary Differential Equation (ODE). where the s is the arc length along the ray. Thus,

$$\tau\left(s\right) = \tau\left(0\right) + \int_0^s \frac{1}{c\left(s'\right)} ds' \tag{2.26}$$

In addition, $\mathbf{x}(s)$ can be easily determined by the initial conditions of a ray and Eq. 2.27, which is obtained from Eq. 2.21 and 2.24.

$$\frac{d}{ds}\left(\frac{1}{c}\frac{d\mathbf{x}}{ds}\right) = -\frac{1}{c}\nabla c \tag{2.27}$$

According to Eq. 2.22 and 2.24, A_0 can be obtained in the ray trajectory coordi-

nates.

$$A_0(s) = A_0(0) \left| \frac{c(s) J(0)}{c(0) J(s)} \right|^{1/2} \tag{2.28}$$

where, J is the *Jacobian* and reflects the spreading of a ray tube [1].

$$J = \left| \frac{\partial \mathbf{x}}{\partial (s, \theta, \varphi)} \right| = \begin{vmatrix} \frac{\partial x}{\partial s} & \frac{\partial x}{\partial \theta} & \frac{\partial x}{\partial \varphi} \\ \frac{\partial y}{\partial s} & \frac{\partial y}{\partial \theta} & \frac{\partial y}{\partial \varphi} \\ \frac{\partial z}{\partial s} & \frac{\partial z}{\partial \theta} & \frac{\partial z}{\partial \varphi} \end{vmatrix} \tag{2.29}$$

θ and φ are respectively the declination and azimuthal take-off angles of the ray. In practice, there exist simple differential equations which provides information about how the ray paths change for infinitesimal perturbations in either the ray take-off angle or the ray source point [1]. These equations form the basis of *dynamic ray tracing* and are used to calculate the J.

After determining the formula of $\tau(s)$ and $A_0(s)$, if the initial conditions of a ray is given, the acoustic pressure field along a ray can be obtained:

$$p(s) = \frac{-S_\omega}{4\pi} \left| \frac{c(s) \cos\theta}{c(0) J(s)} \right|^{1/2} e^{i\omega \int_0^s \frac{1}{c(s')} ds'}. \tag{2.30}$$

However, we are usually interested in the acoustic field at a fixed location in the cartesian coordinates $p(\vec{x})$. To find the $p(\vec{x})$, all eigenrays passing through $\vec{x}$ need to be found, and then $p(\vec{x})$ is equal to the sum of all sound pressures associated with each eigenray. The summing could be coherent, incoherent or semicoherent [1].

The ray theory method is computationally rapid and extends to range-dependent problems. In the operational environment where computation speed is a critical factor, the ray method is used extensively. However, the ray method is an inherent high-frequency approximation method which leads to coarse accuracy in the results, especially for low frequencies. For this reason, in underwater acoustics research community, the ray method is rarely used for low frequencies (below 1kHz); while for high frequencies (a few kHz or above), ray method is the most practical method.

2.3.2 Wavenumber Integration Method

In underwater acoustics, the ocean environment is usually assumed to be azimuthal symmetric about the point sound source [1]. Also the ocean acoustic environment is stratified and varies quickly along depth, slowly along range, so in many cases the environment can be treated as range-independent. If a *cylindrical coordinate* system is constructed with the point source being on the z-axis, the inhomogeneous Helmholtz equation can be written as

$$\left[\frac{1}{r}\frac{\partial}{\partial r}\left(\frac{1}{r}\frac{\partial}{\partial r}\right)+\rho(z)\frac{\partial}{\partial z}\left(\frac{1}{\rho}(z)\frac{\partial}{\partial z}\right)+k^2(z)\right]p=\frac{S_\omega}{\pi r}\delta(r)\,\delta(z-z_s). \tag{2.31}$$

This is a 2-D PDE and z_s is the source depth. The wavenumber k and the density ρ are functions of the depth z. The z-axis can be discretized so that in each layer $k(z)$ and $\rho(z)$ can be treated as constants. Thus, by the Hankel transform pair

$$f(r,z) = \int_0^\infty f(k_r,z)\,J_0(k_r r)\,k_r dk_r \tag{2.32}$$

$$f(k_r,z) = \int_0^\infty f(r,z)\,J_0(k_r r)\,r dr, \tag{2.33}$$

Eq. 2.31 can be transformed into the *depth-separated wave equation*:

$$\left[\frac{d^2}{dz^2}+\left(k^2(z)-k_r^2\right)\right]p(k_r,z)=\frac{S_\omega}{2\pi}\delta(z-z_s), \tag{2.34}$$

where the k_r is the *horizontal wavenumber*. Eq. 2.34 is an ODE, which can be solved for any k_r by taking into account boundary conditions and radiation conditions. Once the spectrum $p(k_r,z)$ is obtained, the $p(r,z)$ can be obtained by the inverse Hankel transform.

$$p(r,z) = \int_0^\infty p(k_r,z)\,J_0(k_r r)\,k_r dk_r \tag{2.35}$$

$$= \frac{1}{2}\int_{-\infty}^\infty p(k_r,z)\,H_0^{(1)}(k_r r)\,k_r dk_r \tag{2.36}$$

Note that for $p(k_r, z)$, poles may exist on the k_r-axis, so the integration can not be exactly operated along the k_r-axis but along a slightly deviated path (see Appendix. B). If $S_\omega = -4\pi$, then the Transmission Loss (TL) is

$$TL(r,z) = -20\, log_{10} |p(r,z)| . \tag{2.37}$$

The wavenumber integration method can generate accurate"near-field" and can includes shear wave effects in elastic media. However, its speed is relatively slow and to extend to range-dependent environments the wavenumber integration method will require much more additional computation efforts.

2.3.3 Normal Mode Method

Let's assume that the solution of Eq. 2.31 can be written in the following format:

$$p(r,z) = \sum_{m=1}^{\infty} \Phi_m(r)\, \Psi_m(z) . \tag{2.38}$$

$\Psi_m(z)$ is the mth eigenfunction of the *Sturm-Liouville* problem consisting of Eq. 2.39, all boundary conditions and radiation conditions.

$$\left[\rho(z)\frac{d}{\partial z}\left(\frac{1}{\rho(z)}\frac{d}{\partial z}\right) + \left(k^2(z) - k_{rm}^2\right)\right]\Psi_m(z) = 0 \tag{2.39}$$

Those eigenfunctions are orthogonal and can be normalized, i.e.

$$\int_0^\infty \frac{\Psi_m(z)\,\Psi_n(z)}{\rho(z)} dz = 0 \text{ for } m \neq n, \tag{2.40}$$

$$\int_0^\infty \frac{\Psi_m^2(z)}{\rho(z)} dz = 1. \tag{2.41}$$

Eq. 2.39 is an ODE. All Ψ_m and k_{rm} can be obtained numerically or analytically. It is assumed that all eigenfunctions of the Sturm-Liouville problem form a complete set. However, this assumption is not valid in many underwater acoustic problems [1] (see more in Appendix C).

Substituting Eq. 2.38 into Eq. 2.39 and applying Eq. 2.40 and 2.41, it will yield

$$\frac{1}{r}\frac{d}{dr}\left(r\frac{d\Phi_n(r)}{dr}\right)+k_{rn}^2\Phi_n(r)=\frac{S_\omega}{\pi r\rho(z_s)}\delta(r)\Psi_n(z_s). \tag{2.42}$$

Combining the radiation condition at $r\to\infty$, solution of the above equation is

$$\Phi_n(r)=-\frac{iS_\omega}{4\rho(z_s)}\Psi_n(z_s)H_0^{(1)}(k_{rn}r), \tag{2.43}$$

(see more in Appendix D). So, finally we find that

$$p(r,z)=-\frac{iS_\omega}{4\rho(z_s)}\sum_{m=1}^{\infty}\Psi_m(z_s)\Psi_m(z)H_0^{(1)}(k_{rm}r). \tag{2.44}$$

In the normal mode method, once all Ψ_m and k_{rm} are available, acoustic fields for all source and receiver configurations are available. Moreover, the normal mode method can be extended to range-dependent environments. In most underwater acoustic problems, the normal mode method can't provide precise "near-field" due to negligence of the continuous spectrum (see Appendix C).

2.3.4 Parabolic Equation Method

The original wave equation Eq. 2.17 is a hyperbolic equation, the Helmholtz equation Eq. 2.18 is an elliptic equation. The Parabolic Equation (PE) method is to replace the elliptic equation with an approximate parabolic equation and solve it by split-step Fourier algorithm etc.

Think about the homogeneous Helmholtz equation in a range-dependent environment:

$$\left[\frac{1}{r}\frac{\partial}{\partial r}\left(\frac{1}{r}\frac{\partial}{\partial r}\right)+\frac{\partial^2}{\partial z^2}+k^2(r,z)\right]p=0. \tag{2.45}$$

Its solution can be assumed to be:

$$p(r,z)=\psi(r,z)H_0^{(1)}(k_0r), \tag{2.46}$$

where the $k_0 = \frac{\omega}{c_0}$ is the reference wavenumber. Utilizing the *paraxial approximation* (small-angle approximation), i.e.

$$\frac{\partial^2 \psi}{\partial r^2} \leq 2ik_0 \frac{\partial \psi}{\partial r}, \tag{2.47}$$

and the Hankel function approximation

$$H_0^{(1)}(k_0 r) \approx \sqrt{\frac{2}{\pi k_0 r}} e^{i(k_0 r - \pi/4)}, \tag{2.48}$$

substituting Eq. 2.46 into Eq. 2.45, an approximate parabolic wave equation can be obtained

$$2ik_0 \frac{\partial \psi}{\partial r} + \frac{\partial^2 \psi}{\partial z^2} + k_0^2 \left(n^2 - 1\right) \psi = 0, \tag{2.49}$$

where $n = \frac{c_0}{c(r,z)}$.

The parabolic equation Eq. 2.49 can be solved by split-step marching algorithm

$$\psi(r + \Delta r, z) = e^{\frac{ik_0}{2}\left[n^2(r_0,z)-1\right]\Delta r}\, \mathcal{F}^{-1}\left\{ e^{-\frac{i\Delta r}{2k_0}k_z^2} \mathcal{F}\left\{\psi(r_0, z)\right\}\right\} \tag{2.50}$$

where the Fourier transform $\mathcal{F}$ is with respect to z.

So if $\psi(0, z)$ is available, $\psi(r, z)$ can be obtained via Eq. 2.50. There are several numerical and analytical methods to construct the starting field $\psi(0, z)$ [1].

The PE method is essentially a small angle approximation. The results asymptotically match the truth in far-field. In practice, the Δr and the Δz used in the FFT are key factors affecting computation accuracy and speed. From experience, the Δz has to be very small to make PE results converge, especially for environments with steep bathymetry. The PE method is presently the most practical and fastest method for range-dependent environments.

Chapter 3

Adaptive Rapid Environmental Assessment (AREA)

As described in the preceding chapters, the ocean-acoustic environment varies in time and space over wide-range scales, and hence the ocean-acoustic environment estimation is often associated with big uncertainties. In this chapter, it will be shown that those estimation uncertainties may induce serious uncertainties in acoustic prediction, especially in shallow water. To minimize the acoustic prediction uncertainties etc., the AREA project was proposed. In this chapter, the motivations, basic ideas, mechanisms and current problems in AREA will be detailed.

3.1 Motivations

3.1.1 Shallow Water Area

The continental masses are surrounded largely by shallow water. Shallow water area (also referred to as "coastal ocean" or "littoral ocean") encompasses about 5% of the world's oceans, roughly the region from the beach to the shelf break, where water depths are about 200 meters [15, 12]. Shallow water is usually a noisy environment because all commercial and military shipping must pass through shallow water when entering or leaving port, or when transiting straits or passages. Shipping lanes ex-

ist along coastlines. Consequently, shallow water, and in particular shallow water acoustics, is an area of major concern to the Navy.

From an acoustic viewpoint, since the SVP in shallow water area is downward refracting or nearly constant over depth (Fig. 2-2), the important ray paths are either *refracted bottom-reflected* or *surface-reflected bottom-reflected* [1]. The properties of the water column and the seabed are all important for acoustic prediction in shallow water. As aforementioned, oceanic variabilities widely exist in the ocean, especially in shallow water, where wind driven flows, tidal currents, river outflow, internal waves, solitary waves, fronts, eddies, thermal changes etc are some of the commonly dominant oceanographic processes. These processes and their intercoupling and interactions with the seabed make the shallow water ocean-acoustic environment highly variable in time and space [10, 39]. In the water column, the temperature profile, salinity profile, plankton distribution profile etc can vary in complex dynamical ways, driven by the variety of coastal oceanographic processes and their coupling. Current flows also interacts strongly with the littoral bottom topography which can be highly variable. In the seabed, bathymetric profiles vary in time and space too, which in turn makes the dynamics of the water column extremely complex. The properties of the seabed are also variable, which impacts acoustic predictions.

Those variabilities span a wide range of spatial and temporal scales [5, 6]. Conventional oceanographic measurements cannot provide the ability to synoptically observe all those dynamically interlocking, patchy and intermittent processes in coastal ocean, especially for sub-meso-scales short in time and space [7]. Consequently the coastal environment will be always under-sampled at these small and fast scales. Oceanographic forecasting by modeling and data assimilation such as the HOPS/ESSE can produce 4-D oceanographic field estimates and the associated uncertainties [40, 41]. However, the spatial and temporal grids used in computation are limited by the available computational resources and the initial conditions can be relatively unknown due to the environmental under-sampling [8]. So, even using nested computational grids, spatial scale smaller than hundred meters in the horizontal, and meters in the vertical cannot be modeled deterministically over large coastal regions (see Fig. 3-1).

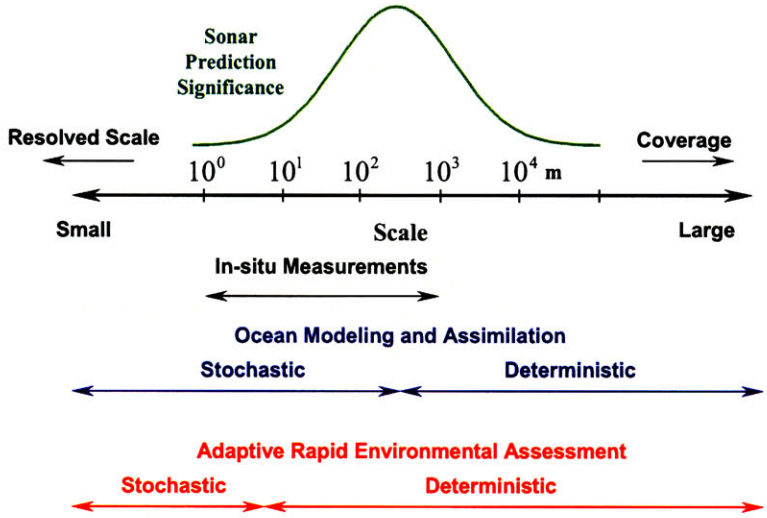


Figure 3-1: Multi-scale environmental assessment. The typical sonar systems performance is dependent on acoustic environment variability over a wide range of scales. Optimal environmental assessment will therefore be a compromise between conflicting requirements of coverage and resolution. By targeting areas of high sensitivity to the sonar system through in situ measurements, the deterministic assessment range will be shifted towards smaller scales.

Modern ocean modeling and assimilation frameworks have a capability of representing the smaller, sub-grid-scale variability statistically [8]. From an acoustic viewpoint, very small scale variabilities is averaged out by the acoustic wave length; while the sub-meso scale variabilities of the order of hundred meters to kilometers make the coastal ocean acoustic environment largely unknown with many uncertainties in terms of imperfect sound velocity, depth of the thermocline etc. Such uncertainties can be responsible for a large part of the acoustic prediction uncertainty [9, 10].

In summary, due to the existence of sub-meso-to-small-scale ocean-acoustic environmental variabilities, great uncertainties may exist in shallow water ocean field estimation, which is based on conventional oceanographic measurement systems and current ocean prediction systems; and the environment parameters of the water column and the seabed are generally not known in sufficient detail and with enough accuracy for satisfactory prediction of long-range acoustic propagation in shallow water, even though the shallow water acoustics has been thoroughly investigated both theoretically and experimentally [1]. An example is shown in next section.

3.1.2 Acoustic Prediction Uncertainty

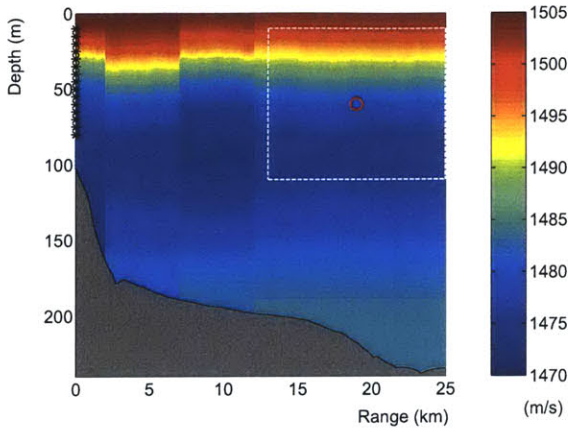
Strictly speaking, the ocean-acoustic environment does not only include the sound velocity, density, attenuation coefficient in the water column and the seabed, but also includes the bottom topography, roughness of the surface and the bottom, current velocity, scattering sources distributions etc. Of those factors, the SVP in the water column is usually the most variable one and plays a very important role in acoustic computation. If big uncertainties exist in the water column SVP prediction, it often leads to serious acoustic prediction uncertainties (see Fig 3-2).

Fig. 3-2 shows a scenario of the Georges bank in the gulf of Maine, where a thermocline and internal waves exist. Fig. 3-2(a) shows the Principal Estimation (P.E.) of the water column SVP. Standard deviations of the associated errors are shown in Fig. 3-2(b), from where it can be found that the P.E. has biggest uncertainties at the thermocline area. Fig. 3-2(c) and 3-2(d) illustrate two possible SVP realizations associated with Fig. 3-2(a) and 3-2(b). In this case, the seabed environment is assumed to be deterministic. If a 100 Hz Continuous Wave (CW) sound source is located at $range = 0km, depth = 10m$ or at $range = 0km, depth = 80m$, the corresponding TL predictions at $depth = 50m$ and $depth = 100m$ will have serious uncertainties as shown in Fig. 3-2(e), due to the water column SVP prediction uncertainties.

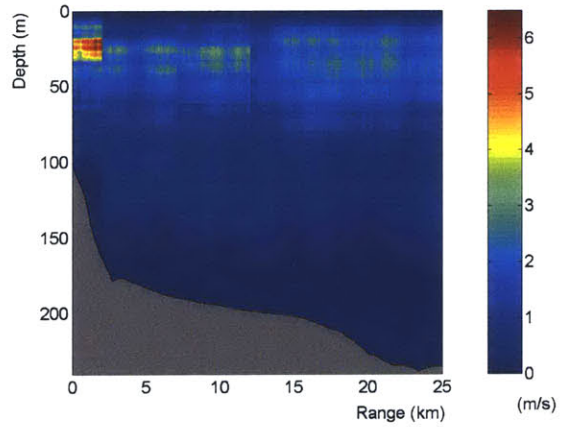
From a viewpoint of sonar system, the acoustic prediction uncertainties will definitely affect sonar performance and sonar performance prediction. The uncertainty of the acoustic predictability is critical to the dB-budget of classical sonar systems by directly affecting the detection and false alarm probabilities [9]. It is also one of the major obstacles to adapting new model-based sonar processing frameworks, such as Matched Field Processing (MFP) [42], to the coastal environment.

For a non model-based sonar system, sonar performance is dependent on sound propagation pattern in the ocean waveguide, the ocean-acoustic environment. Therefore sonar performance can be written as

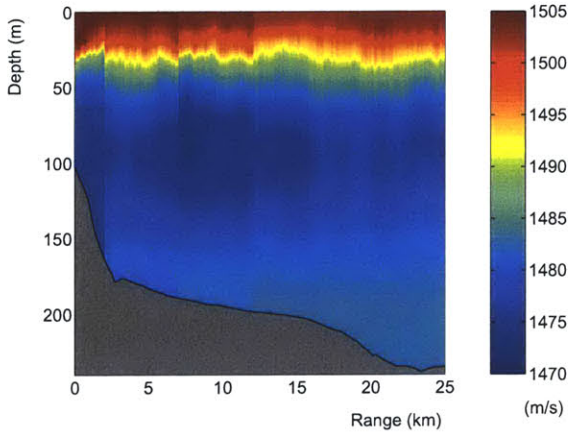
$$SP = f(O), \tag{3.1}$$



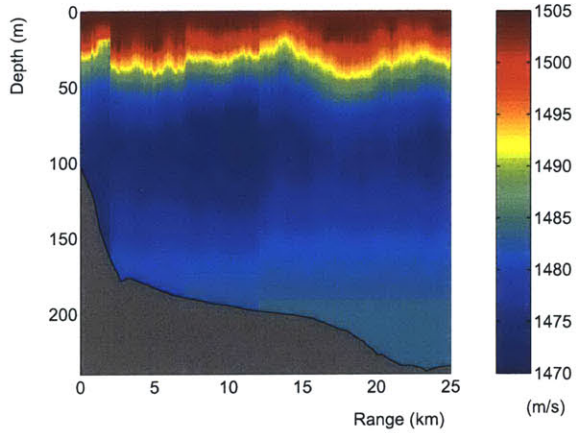
(a) Principal estimation of water column SVP



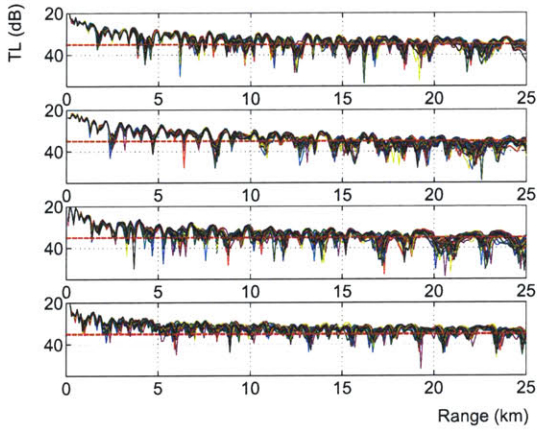
(b) Estimation error standard deviation



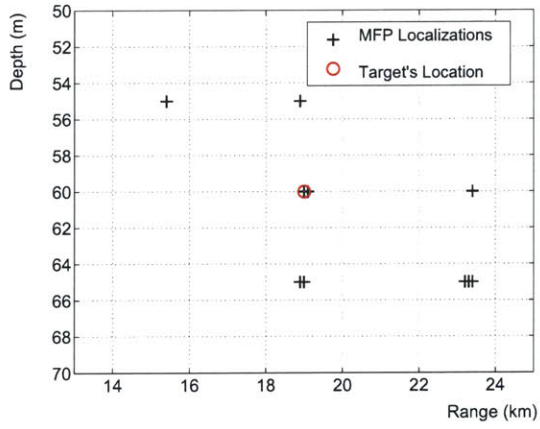
(c) SVP realization 1



(d) SVP realization 2



(e)



(f)

Figure 3-2: In (a), for MFP, the black asteroids indicate the location of the vertical sonar array; the red circle is the location of the sound source; the white rectangular box is the searching domain. In (e), the four TLs plots are respectively associated with $(rs = 0km, zs = 10m, zd = 50m)$, $(rs = 0km, zs = 10m, zd = 100m)$, $(rs = 0km, zs = 80m, zd = 50m)$, $(rs = 0km, zs = 80m, zd = 100m)$, where rs, zs, zd are source range, source depth and receiver depth respectively; those red lines are at $35dB$.

where SP represents a sonar performance metric and O represents the ocean-acoustic environment. The sonar performance prediction uncertainty is connected with the ocean-acoustic environment prediction uncertainties through the function f, which is usually highly nonlinear. *Sonar range* is a typical non model-based sonar performance metric. If, for example, the threshold of a sonar is $35dB$, from Fig. 3-2(e) it can be seen that the *sonar range* could be associated with great uncertainty. For a model-based sonar system such as MFP, sonar performance is dependent on both the true ocean-acoustic environment and the environment prediction,

$$SP = f(O, O'), \tag{3.2}$$

where O' is the environment prediction. If the uncertainty of O' becomes smaller, that means O' is closer to O, the SP will be better. For example, if now the 100 Hz CW sound source is located at $range = 19km, depth = 60m$ and a vertical sonar array with 15 hydrophones is located at $range = 0km$ and uniformly distributed from $depth = 10m$ to $80m$ (see Fig. 3-2(a)), the MFP without any measurement noise will localize the source with mismatched displacements up to several kilometers (see Fig. 3-2(f)). If the standard deviation of SVP prediction error becomes smaller, the expectation of the MFP mismatched displacement will be smaller, that means MFP localization is more accurate and the sonar performance is better.

The acoustic uncertainty associated with the spatially and temporally varying sound speed and the random characteristics of the bottom are also of critical influence to acoustic communication systems, which with the integration of new Autonomous Ocean Sampling Network (AOSN) [43] concept in the operational Navy is becoming of increasing tactical significance.

3.2 Basic Ideas

To determine the environmental variability of the critical sub-meso scales and short temporal scales, a local, high resolution, rapid deployable in-situ measurement capa-

bility has long been recognized as a very important tactical need [9]. By assimilating the in-situ measurement data with ocean modeling, resolution of the ocean filed estimation can be dramatically improved (see Fig. 3-1). As a result, the acoustic prediction uncertainty may be strongly reduced and the model-based sonar performance may be highly improved.

However, its implementation is being constrained by limited resources. The ocean area of interest is usually large, whereas in-situ measurement coverage is very limited due to cost, time and performance constraints (See Fig. 3-1). The consequent limited availability of high-resolution in-situ measurement data for assimilation into the modeling framework may severely limit the usefulness of the forecasts to the acoustic environment prediction. Acknowledging that the size of the ocean area relevant to an acoustic problem is usually as large as tens of kilometers, the acoustic-purposed coastal environmental assessment is facing the classical conflict between *resolution*, needed to capture the fine scale variability and *coverage*, needed for the large scale environmental phenomena. Thus, the *Rapid Environmental Assessment (REA)* resources available must focus on the environmental uncertainties critical to the specific acoustic system. A quantitative and adaptive approach is necessary. Optimizing the REA resources deployment pattern, namely the sampling strategy optimization, becomes thus the major problem. Different sampling strategies may make a significant difference in predicting sonar performance or improving sonar performance.

Adaptive Rapid Environmental Assessment (AREA) — a new adaptive acoustical-environmental sampling approach based on coupled oceanic-acoustic forecasts is currently being developed in connection with the emergence of the new *Autonomous Ocean Sampling Network (AOSN)* technology[44]. In principle AREA is a probabilistic approach to the adaptive sampling problem of littoral REA and envisioned as a real time tactical tool for not only capturing, but also minimizing the acoustic uncertainty of significance to specific sonar systems. In AREA, with ocean forecast providing large-scale coverage, identifying regions and features with strong uncertainty such as coastal fronts, the limited high-resolution tactical resources can be deployed in a manner which is optimal to the acoustic forecast [40, 39]. Consequently, the

limit of deterministic characterization may be shifted significantly towards smaller scales; a much finer resolution can be obtained in the ocean forecast without sacrificing coverage and this will make the acoustic forecast uncertainty minimized (see Fig. 3-1).

The AREA framework can also be used to minimize oceanic uncertainties, biological uncertainties etc [39, 45], or to objectively evaluate the performance of new REA concepts, such as Acoustically Focused Ocean Sampling (AFOS) [44] and Acoustic Data Assimilation (ADA) [46, 6]. To investigate mechanisms and performance limits in AREA, an engineering model and a mathematic model for AREA are developed.

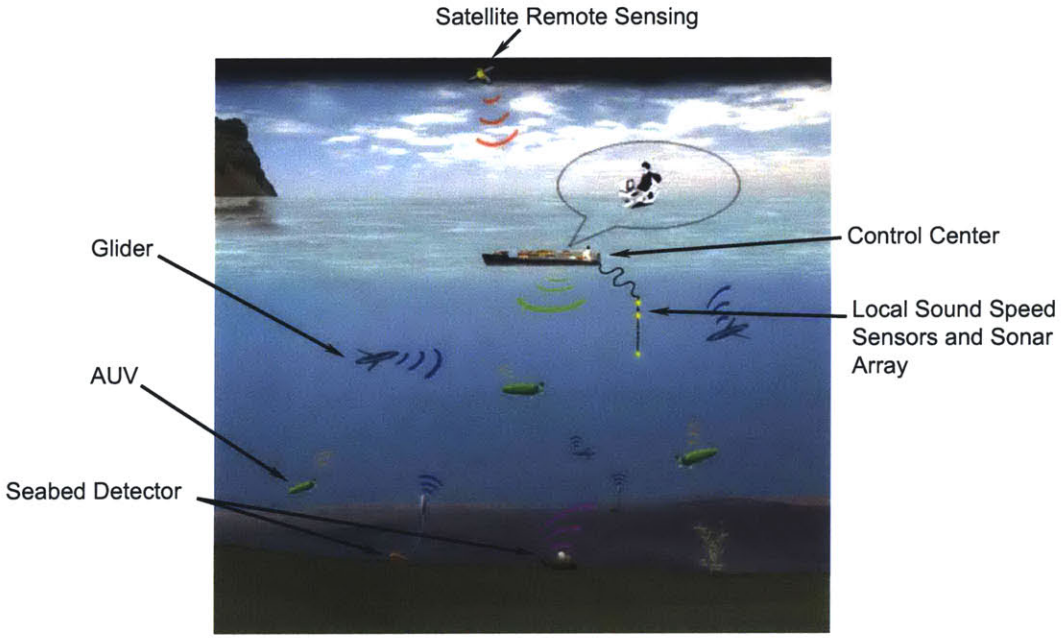


Figure 3-3: Illustration of Adaptive Rapid Environmental Assessment System

3.2.1 Engineering Model

Fig. 3-4 shows the structure of the AREA system and connections with ocean environmental models. In the oceanographic modeling, ocean database, remote sensing and local in-situ measurement data etc. are assimilated with ocean model via HOPS/ESSE; while in the seabed modeling, Geographic & Geological database, local in-situ measurements are assimilated with Geographic & Geological modeling. After

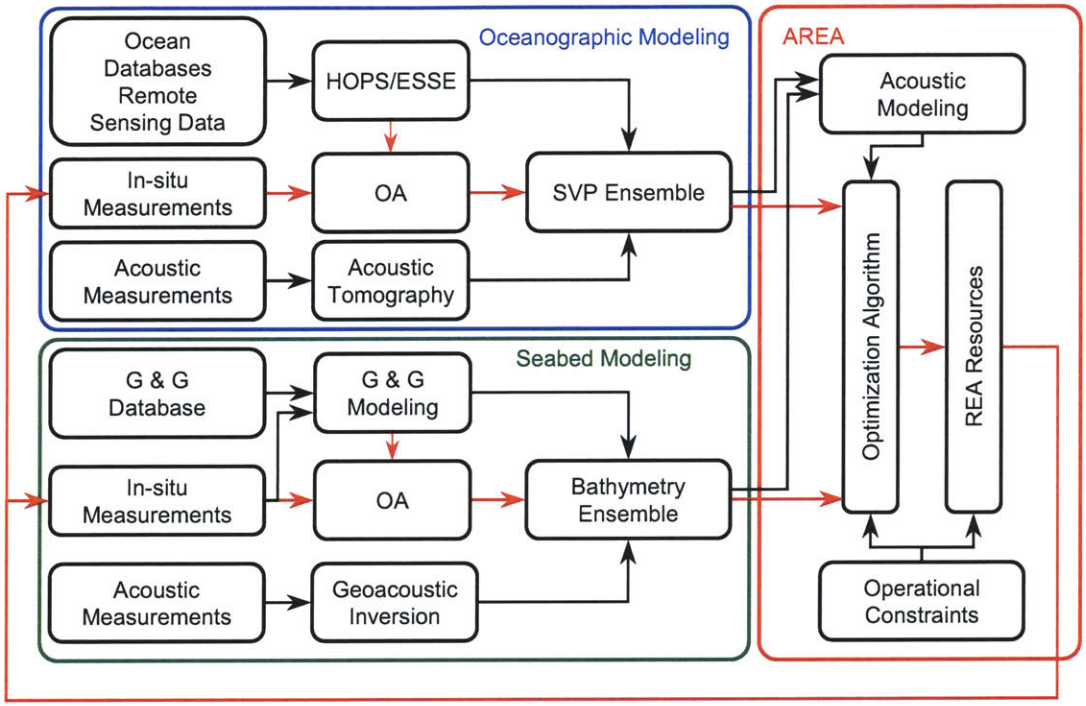


Figure 3-4: AREA wiring diagram. Fore- and now-casts of the local oceanography and geology are producing spatial and temporal environmental statistics in the form of realization ensembles. These ensembles are then used as input to environmental acoustic models to provide associated realizations for the sonar performance, e.g. in the form of probability of detection and false alarms. To minimize the uncertainty of the acoustic prediction and therefore improve the probability of detection to false alarm ratio, the realization ensemble of ocean-acoustic environment and the operational constraints are used to determine an optimal deployment strategy for the REA resources. The REA data are then objective analyzed based on the forecast spatial scales. The resulting reduced uncertainty now-casts are then used for the acoustic prediction.

the data assimilation, the oceanographic modeling and the seabed modeling produce an ensemble of environmental realizations for the water column forecast and the seabed forecast respectively. The acoustic measurement and inversion methods can be utilized to improve both of the water column and seabed environmental predictions. The quantitative uncertainty-maps provide guidance for locating large uncertainties, e.g. determining the bearings associated with maximum uncertainties, and so guide the sampling plans that are computed by AREA.

In the acoustic modeling, the Range-dependent Acoustic Model (RAM PE Code)

— a popular wave-theory technique for solving range-dependent propagation problems in the ocean is used [1]. By coupling the oceanographic, seabed and acoustic models, acoustic prediction uncertainties can be generated via Monte Carlo simulations. The weighted sum of the acoustic prediction uncertainties can thus be used as the objective function in the AREA optimization algorithm which aims to select the sampling plan that reduces these integrated predicted uncertainties. This optimization problem is the most important focus of the thesis.

Under operational REA, the optimization algorithm generates an optimal plan for allocating the REA resources, such as an optimal AUV path, in real-time. Thereafter, REA resources are deployed according to this optimal plan and in-situ measurement data focusing on the most critical uncertainties are collected and passed back to the oceanographic model and seabed model in a short time. Those new local data are rapidly assimilated in the models [40, 6], and ocean environmental and acoustic predictions for the next day are then generated. This process is the *Daily AREA Cycle*, which updates the optimal REA resources allocation pattern everyday.

The *Daily AREA Cycle* constitutes the first level of adaptivity in AREA. In addition to the static optimal REA deployment, the optimization problem can also be treated as a *Sequential Decision Making* Problem (see Fig. 3-6) and modeled in the *Dynamic Programming* (DP) framework, in which the REA resources allocation pattern is not predetermined but generated on-board. An optimal adaptive sampling strategy is then produced, as a function of the data sampled by the autonomous data-collecting platforms, instead of a predetermined optimal sampling pattern. As indicated by the red lines in Fig. 3-4, the dynamic optimization algorithm only outputs the optimal sampling pattern for the next step; after the local data in the next step is collected and rapidly mapped by objective analysis or assimilated in real-time, a new ocean prediction is computed to optimize the subsequent sampling pattern. The whole optimal REA resources allocation pattern is adaptively generated step by step on-board. This is the second level of adaptivity.

This second level of adaptivity in AREA involves dynamic programming. However, it is known that a DP problem is usually $\mathcal{NP}$ hard [47]. Determining an opti-

mization approach for the adaptive sampling strategy that can be computed on-board can thus be extremely difficult. Nevertheless in some particular cases, this difficult problem can be avoided by in-direct methods (see Chapter 5).

From an operational viewpoint, the AREA system can be structured as Fig. 3-5. An operational AREA system usually involves 5 components: the real ocean environ-

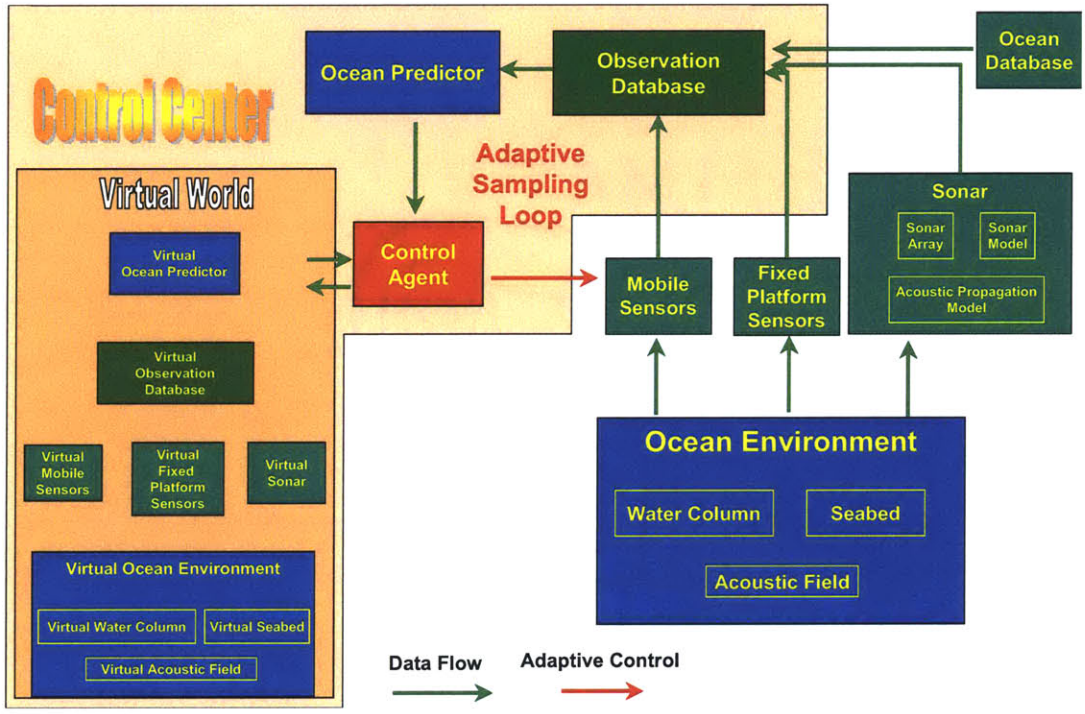


Figure 3-5: AREA wiring diagram from an operational viewpoint.

ment; the sonar system; mobile sensors such as AUVs carrying CTDs; fixed platform sensors such as local XBT, local CDT, satellite or acoustic remote sensing system and seabed detector etc. (see Fig. 3-3); and the control center. The control center is the heart of AREA. People or computers can directly operate and control the whole AREA system through the control center. Basically, it consists of 3 modules: observation database, ocean predictor and control agent. Observation database module includes a data transferring interface/pipeline and data storage, which can communicate with all sensors and the sonar system, receive and save data. The ocean predictor module includes just the oceanographic modeling and the seabed modeling, which utilizes the saved data and provides environmental forecasts to the control agent. The

control agent module works as a decision maker using some sort of *Artificial Intelligence Optimization* methods. This module is very complicated. Depending on the decision making algorithm, the control agent may be structured differently. For most sophisticated algorithms, it usually possesses a virtual world - a mirror of the whole AREA system - and "play" all possible controls in the virtual world, then select the one with optimal virtual consequence as command. This module is the main object of attention in the AREA project and this thesis.

In operations, the AREA system starts with initialization - updating observation database according to the latest ocean database, latest measurements by the fixed platform sensors and sonar configuration information etc. After initialization, the control center will run the ocean predictor module and generate preliminary environmental predictions. All initial information and analysis results will then be collected by the control agent module where a sampling strategy program will be run and work out commands such as predetermined sampling waypoints (in the daily adaptivity) or adaptive sampling strategy (in the on-board adaptivity. See Fig. 3-6) for the mobile sensors. Those commands will be sent to the mobile platforms through communication channels. Following the commands, the mobile sensors will approach sampling points one by one and capture the local uncertainties.

Compared with sound velocity in the water column, variabilities in bathymetry are usually less variable and less rapid. In-situ measurement platforms such as side-scan/sub-bottom profiling AUV, water depth detection etc can capture those variabilities in good resolution and coverage. Presently, AREA thus focuses more on the water column and treats bathymetry deterministically. **In this thesis, we will only consider the water column and focus on the SVP variability.** However, all models and algorithms developed in the thesis can be extended to capture any other variability in the water column.

3.2.2 Mathematic Model

In this section, the basic ideas in AREA will be represented in mathematical formulas. Mechanisms, performance limits in AREA will be mathematically clarified.

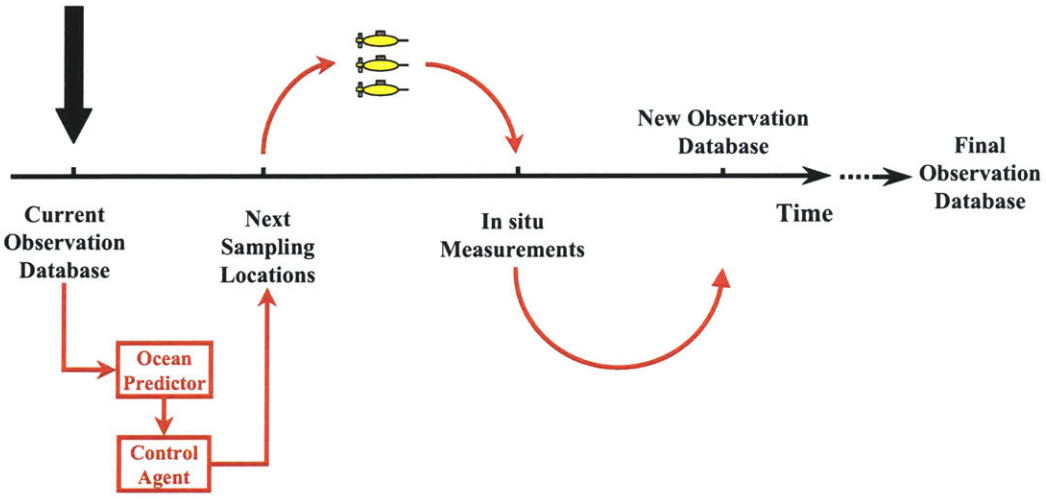


Figure 3-6: Sequential diagram of the on-board adaptivity. At each stage, the Observation Database will be first updated and the Ocean Predictor will do analysis; then the Control Agent will determine next sampling locations; following those commands, mobile sensors will do in situ measurements and new measurement data will be collected, by which the Observation Database will be updated again. This loop is call Adaptive Sampling Loop. Repeating this loop, sampling points locations will be determined sequentially based on all the newest observations.

Deterministic-Stochastic Model for Ocean SVP

Based on the philosophical presumptions made in Appendix A, the true ocean SVP $C^t(x, y, z, t)$ is in nature a deterministic but very complicated dynamic process varying on multiple scales. It can be decomposed into two parts:

$$C^t(x, y, z, t) = \overline{C}(x, y, z, t) + C'(x, y, z, t), \tag{3.3}$$

where, $\overline{C}(x, y, z, t)$ is associated with the part of oceanographic processes in meso-to-large scales, and can be modeled by ocean modeling; $C'(x, y, z, t)$ is associated with the part of oceanographic processes in submeso-to-small scales, and can not be modeled deterministically. Tremendous observations and experiments have been made to investigate the mechanisms and dynamics in $C'(x, y, z, t)$; huge data have been collected. Based on those observations and data, a stochastic model $\widetilde{C}(x, y, z, t)$ for modeling $C'(x, y, z, t)$ can be created and then a deterministic-stochastic model

$C(x,y,z,t)$ for $C^t(x,y,z,t)$ can be obtained,

$$C(x,y,z,t) = \overline{C}(x,y,z,t) + \widetilde{C}(x,y,z,t). \tag{3.4}$$

$C^t(x,y,z,t)$ can thus be viewed as a sample path of $C(x,y,z,t)$. This deterministic-stochastic model doesn't have to be unique. Actually, different people may use different observation data and different methodology, therefore get different $\widetilde{C}(x,y,z,t)$. For example, Regional Ocean Modeling System (ROMS) and Harvard Ocean Prediction System (HOPS) may give us discrepant results [48, 37, 38]. Furthermore, even if using the same observation data, the same ocean model, but with different computational grid size, the deterministic-stochastic SVP model will be different too. It should be noticed that the mean of $\widetilde{C}(x,y,z,t)$ must be always zero. Since if it is not zero, the mean part can be put into $\overline{C}(x,y,z,t)$.

In AREA project, $\widetilde{C}(x,y,z,t)$ in a duration T is considered. T may last for from hours to a day. From a mathematical viewpoint, the objective of this project is to best improve the stochastic model $\widetilde{C}(x,y,z,t)$ for duration T in regard to acoustic purpose, by assimilating in-situ measurement data. During T, we can decompose $\widetilde{C}(x,y,z,t)$ into two parts and rewrite Eq. 3.4 as

$$C(x,y,z,t) = \overline{C}(x,y,z,t) + C_1(x,y,z,t) + C_2(x,y,z,t), \tag{3.5}$$

where, $\widetilde{C}(x,y,z,t) = C_1(x,y,z,t) + C_2(x,y,z,t)$.

$C_1(x,y,z,t)$ is the slowly time-variant part in $\widetilde{C}(x,y,z,t)$, which is highly auto-correlated during T; $C_2(x,y,z,t)$ is the other part, which has lower auto-correlation during T. C_1 and C_2 are both stochastic process with zero mean, but C_1 does not significantly change during T. Therefore, we may approximately treat C_1 as time-invariant. Furthermore, to simplify the problem, it is usually assumed that C_1 and C_2 are not cross-correlated, and C_2 is wide-sense stationary within T. It can be seen that if T is shorter, more components in $\widetilde{C}$ will be counted in C_1 and less will be in C_2, and vice versa. Now C_1 is treated as a time-invariant stochastic process and its Probability Density Function (PDF) can be generated by ocean modeling, if we can

quickly measure $C_1(x,y,z,t)$ at some acoustic-critical locations within T and quickly implement data assimilation, we can then dramatically improve the estimation for C_1, i.e. sharper the PDF of C_1, for duration T. Consequently, the stochastic model $\widetilde{C}(x,y,z,t)$ can be best improved in regard to acoustic prediction uncertainties.

In-situ Measurement and Objective Analysis

In operations, the ocean is discretized in time and space. Let vector $\overline{c}$, c_1 and c_2 denote the spatially discretized $\overline{C}(x,y,z,t)$, $C_1(x,y,z,t)$ and $C_2(x,y,z,t)$ at a certain time within T respectively. Because $\overline{C}$ is the deterministic part, which varies even slower than C_1, C_1 is now assumed to be time-invariant in T, the in-situ observations within T can be modeled as

$$d = D \cdot (\overline{c} + c_1 + c_2') + v \tag{3.6}$$

where, D is the sparse measurement matrix as the one in Eq. 2.10; v is the CTD noise vector, which is assumed to be white. Since $C_2(x,y,z,t)$ may significantly change during the in-situ measurement process, we thus use c_2' to denote the nominal vector corresponding to the time-varying $C_2(x,y,z,t)$ in T such that $D \cdot c_2'$ is just equal to the measurement data. In Eq. 3.6, ocean modeling can deterministically model $\overline{c}$ and generate the PDFs for c_1 and c_2. Thus $\Lambda_{c_1}(-)$ the a priori covariance matrix of c_1 and $\Lambda_{c_2}(-)$ the a priori covariance matrix of c_2 can be known. Since we assume that C_2 is a wide-sense stationary random process in T, $D \cdot \Lambda_{c_2'}(-) \cdot D^T$ the a priori covariance matrix of $D \cdot c_2'$ can get known too by taking into account the temporal correlation. To capture c_1, we need to rewrite Eq. 3.6 as

$$d = D \cdot (\overline{c} + c_1) + (D \cdot c_2' + v), \tag{3.7}$$

i.e. we treat c_2' as a sort of measurement noise. Therefore, by the OA equations (Eq. 2.11, 2.12) we can get

$$\hat{c}_1(+) = \Lambda_{c_1}(-) D^T \left[D \cdot \Lambda_{c_1}(-) \cdot D^T + D \cdot \Lambda_{c_2'}(-) \cdot D^T + R\right]^{-1} [d - D \cdot \bar{c}], \quad (3.8)$$

$$\Lambda_{c_1}(+) = \Lambda_{c_1}(-) - \Lambda_{c_1}(-) D^T [D \cdot \Lambda_{c_1}(-) \cdot D^T + D \cdot \Lambda_{c_2'}(-) \cdot D^T + R]^{-1} D\Lambda_{c_1}(-), \quad (3.9)$$

where, it should be noticed that c_1 has zero mean; c_1, c_2' and v are uncorrelated. Thus for the random vector $c = \bar{c} + c_1 + c_2$, we have

$$\hat{c}(+) = \bar{c} + \Lambda_{c_1}(-) D^T \left[D \cdot \Lambda_{c_1}(-) \cdot D^T + D \cdot \Lambda_{c_2'}(-) \cdot D^T + R\right]^{-1} [d - D \cdot \bar{c}], \quad (3.10)$$

$$\Lambda_c(+) = \Lambda_{c_1}(-) - \Lambda_{c_1}(-) D^T [D \cdot \Lambda_{c_1}(-) \cdot D^T + D \cdot \Lambda_{c_2'}(-) \cdot D^T + R]^{-1} D\Lambda_{c_1}(-) + \Lambda_{c_2}(-). \quad (3.11)$$

while,

$$\hat{c}(-) = \bar{c}, \quad (3.12)$$

$$\Lambda_c(-) = \Lambda_{c_1}(-) + \Lambda_{c_2}(-). \quad (3.13)$$

Comparing Eq. 3.13 and Eq. 3.11, it can be seen that the SVP prediction uncertainty is reduced from $\Lambda_c(-)$ to $\Lambda_c(+)$ by removing some part of prediction uncertainty associated with C_1. However, for those prediction uncertainties associated with C_2, there is no way to mitigate them in AREA. So, if a longer duration T is considered, a worse water column SVP nowcast will be produced; while if T is shorter, the water column SVP nowcast will be more accurate. From another viewpoint, in $\Lambda_c(+)$ the proportion of prediction uncertainties associated with C_1 is lower than that in $\Lambda_c(-)$ and because the variability scale of C_1 is larger than the scale of C_2, the average correlation length in $\Lambda_c(+)$ should be shorter than that in $\Lambda_c(-)$. Thus in Fig. 3-1, the deterministic assessment range after AREA is shifted towards smaller scale.

Acoustic Prediction Uncertainty

The acoustic prediction uncertainty can be used as the objective function in the optimization problem in AREA, which from a mathematic viewpoint is to find the optimal measurement matrix D under some constraints such that the posterior acoustic prediction uncertainty is minimized. However, as aforementioned, the acoustic-SVP relation is often highly nonlinear. So even if in this thesis the SVP prediction error is assumed to be a Gausssian random vector, there is no any closed mathematic form for the acoustic prediction uncertainty, and thus in practice it is very difficult to calculate it quickly. Actually, at present this problem is one of the major obstacles to solving the AREA optimization problem in real-time.

3.3 The AUV Path Planning Problem

So far, the basic ideas, mechanisms, mathematic model and current problems etc. in AREA have been introduced. In this section, the AREA scenarios considered in this thesis and the associated optimization problem will be introduced.

As mentioned before, in this thesis we only consider the water column SVP estimation uncertainties, any other uncertainty in water column and seabed will be ignored. Furthermore, due to the fast progress in AUV techniques, nowadays the AUV has ranges that are comparable to most spatial scales of significant oceanographic processes to acoustics; with also taking into account its excellent mobility, AUV is used in ocean engineering more and more. Therefore in this thesis, AUV carrying a CTD sensor is considered to be the only REA resource. In addition, to simplify the problem, only one single AUV is considered so far.

In summary, **in this thesis the problem to be considered is how to route a single AUV carrying a CTD sensor to measure sound velocities in the water column such that after assimilating those in-situ measurement data the posterior acoustic field nowcast uncertainties will be minimized.**

In this scenario, the whole procedure in AREA is illustrated in Fig. 3-7. At time t_0, a forecast for the SVP at time t_1 ($t_1 > t_0$) is produced by HOPS/ESSE in the

form of SVP realizations ensemble. Around t_1, an AUV with CTD will be launched to do in-situ measurements on a vertical plane along with a certain bearing. The in-situ measurements will be finished quickly, and then those data will be passed back to the control center. After data assimilation, a SVP nowcast will be generated. From the forecast to the nowcast, the associated uncertainty is reduced. From the SVP nowcast, the acoustic prediction uncertainty can be estimated by running Monte Carlo simulations and acoustic model such as RAM. Finding the optimal AUV path under the constraints so as to minimize the acoustic prediction uncertainty is the objective. However, since the true ocean environment around t_1 can not be completely

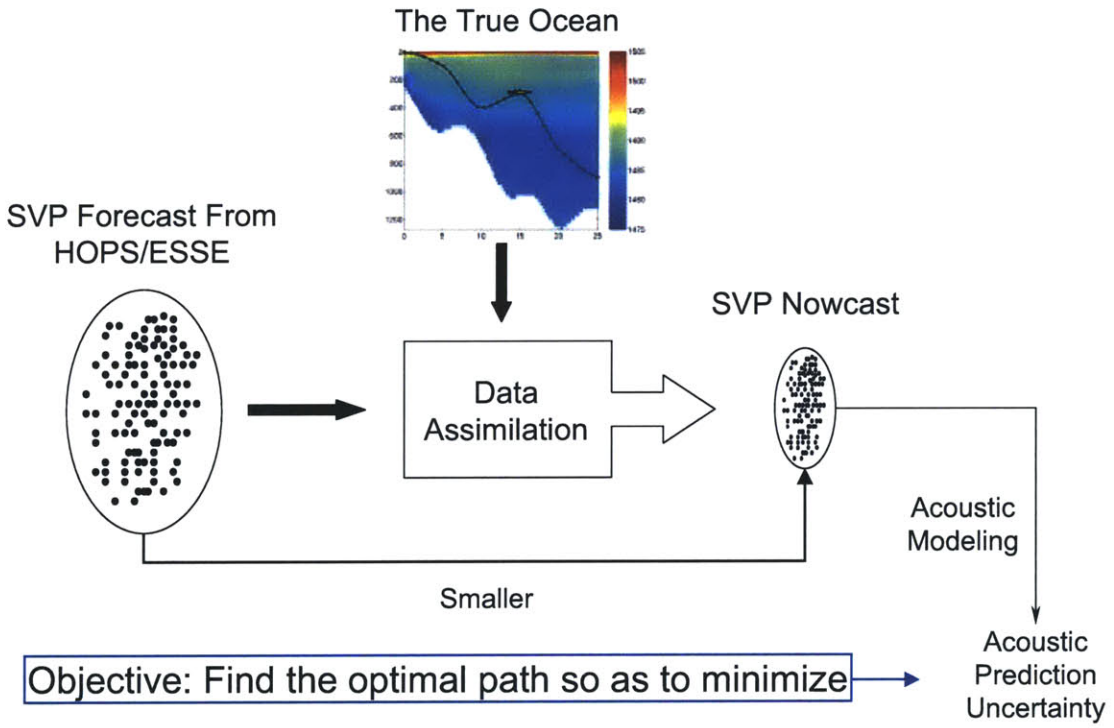


Figure 3-7: Demonstration of the whole process in AREA.

predicted, any SVP realization in the forecast ensemble could be the true one. For different SVP realization, the optimal AUV path may be different. So, which optimal path is the real optimal one for AUV is a question.

One way to define the real optimal AUV path is to implement an AUV path in all SVP realizations in the forecast ensemble and use the sample mean of the acoustic prediction uncertainty as the objective function. The path associated with minimum

objective function value is the real optimal one (see Fig. 3-8). So, in this case, the real optimal AUV path is a predetermined path and only the daily adaptivity exists in AREA.

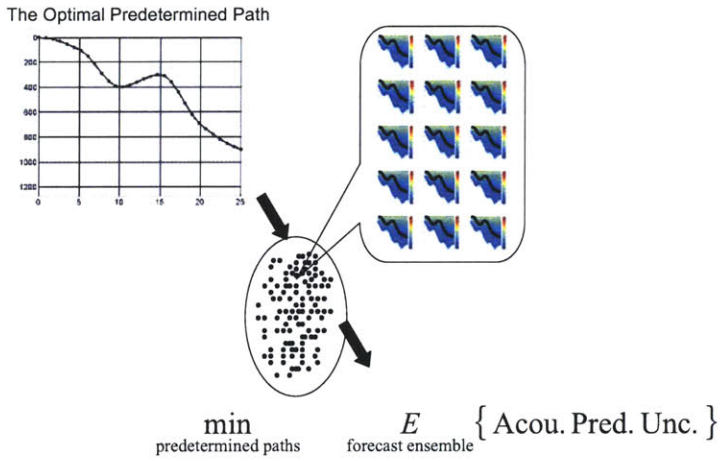


Figure 3-8: The optimal predetermined path.

Another way to resolve this problem is to create an adaptive on-board AUV routing strategy — an algorithm, in which the input is the accrual in-situ measurement data, the output is the next AUV waypoint. So when AUV is moving, the CTD will collect sound velocity data and the AUV can adaptively determine its next waypoint one by one based on all collected data (see Fig. 3-9). If a routing strategy is implemented in all SVP realizations in the forecast ensemble, many different AUV paths may be generated. To find the optimal routing strategy, the sample mean of the posterior acoustic prediction uncertainty can be used as the objective function. The adaptive on-board AUV routing strategy associated with minimum objective function value is then the optimal one (see Fig. 3-10). In this case, AREA can have both daily adaptivity and on-board adaptivity.

In fact, a predetermined path can be viewed as a fixed routing strategy — no matter what happened before, the same next waypoint will be given. Therefore, the predetermined path space is contained by the adaptive AUV routing strategy space and theoretically the optimal routing strategy must lead to a smaller posterior acoustic prediction uncertainty.

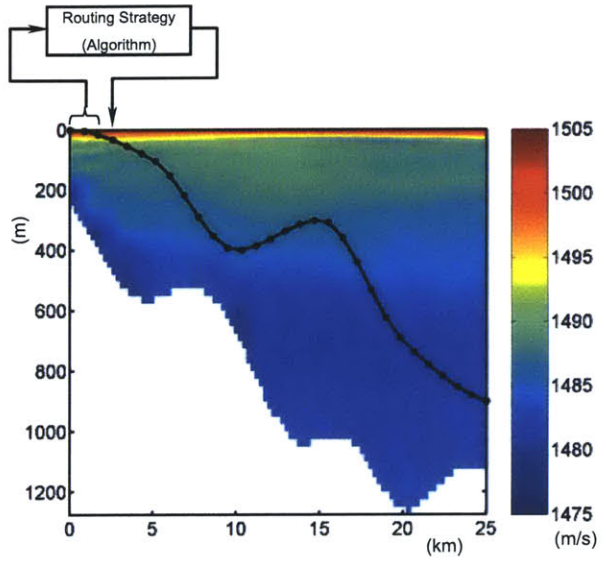


Figure 3-9: Demonstration of an AUV routing strategy.

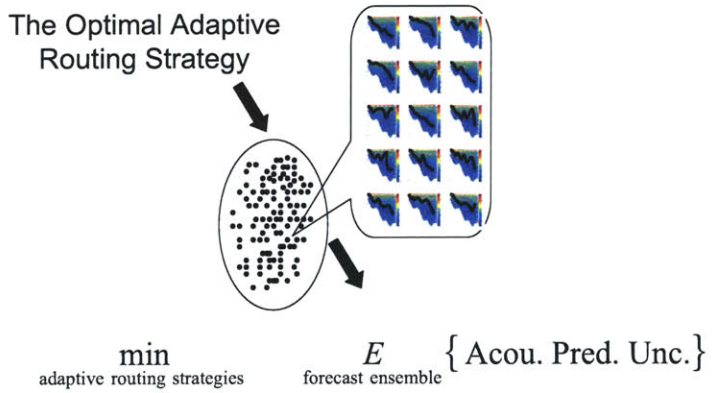


Figure 3-10: The optimal AUV routing strategy.

How to model the above AUV path planning and AUV routing strategy optimization problem in mathematical form and how to solve them in real-time etc. are discussed in chapter 5 and 7.

3.4 Previous Work

In Operational Research (OR) and Management Science (MS), intensive research has been conducted in Vehicle Routing Problem (VRP) [49, 50, 51], which is to find

an optimal route of one or more vehicles through a graph [47]. This is a difficult combinatorial problem. It can be linked with the Traveling Salesman Problem (TSP) or even scheduling.The AUV path planning problem is related to VRP. However, the constraints and the objective function are very different.

The most related previous research is the one conducted by Yilmaz [52, 53]. In this work, path planning for one or more AUVs on a horizontal plane such as the sea surface is considered. The objective function is the integration of a priori SVP prediction uncertainty along the AUV path and no SVP correlation effect is considered. In this work, two optimization methods based on Mixed Integer Programming (MIP) are developed and multi-vehicle multi-day mission is considered. In addition, some similar work can be founded in Unmanned Aerial Vehicle (UAV) path planning research [54].

From the text description, Yilmaz's work is very similar to the AUV path planning discussed in this thesis. In fact however, they are very different in mathematical representation including constraints, decision variables and objective function. In this thesis, the AUV moves on a vertical plane and the AUV path planning is aiming at the minimization of the posterior acoustic prediction uncertainty with taking into account the SVP correlation effect. Therefore the problem addressed in this thesis is a new problem and the practical method for solving this problem will be very different.

Chapter 4

Optimization Algorithm Overview

In mathematics, the term optimization, or mathematical programming, refers to the study of problems in which one seeks to minimize or maximize a real function by systematically choosing the values of real or integer variables from an allowed set [55]. This problem can be represented in the following way,

$$\begin{aligned} \text{minimize} \quad & \mathbf{f}(\mathbf{x}) \\ \text{subject to} \quad & \mathbf{x} \in \mathbf{X}, \end{aligned} \tag{4.1}$$

where, function $\mathbf{f}(\mathbf{x})$ is called an objective function, or cost function; $\mathbf{X}$ is some subset of the Euclidean space $\mathbf{R}^n$, often specified by a set of constraints, equalities or inequalities that the members of $\mathbf{X}$ have to satisfy.

According to properties of $\mathbf{f}(\mathbf{x})$ and $\mathbf{X}$, optimization problem can be categorized into many subfields, such as linear programming, integer programming, quadratic programming, nonlinear programming, stochastic programming, dynamic programming etc. In this chapter, we briefly introduce the optimization algorithms used in AREA project, including linear programming, network optimization, dynamic programming.

4.1 Linear Programming

4.1.1 The Representation of Linear Programming

Linear Programming (LP) is the problem of minimizing a linear cost function subject to linear equality and inequality constraints [56]. For a LP problem, formulas 4.1 can be written as

$$\begin{aligned} \text{minimize} \quad & \mathbf{c}'\mathbf{x} \\ \text{subject to} \quad & \mathbf{A}\mathbf{x} \geq \mathbf{b}. \end{aligned} \tag{4.2}$$

This is a general form and in practice, some special nonlinear programming problem, such as problem with a piecewise linear convex function and linear constraints, can be converted to formulas 4.2. To solve the LP problem, formulas 4.2 is usually transformed into a standard form as in 4.3, by eliminating free variables and inequality constraints.

$$\begin{aligned} \text{minimize} \quad & \mathbf{c}'\mathbf{x} \\ \text{subject to} \quad & \mathbf{A}\mathbf{x} = \mathbf{b} \\ & \mathbf{x} \geq \mathbf{0} \end{aligned} \tag{4.3}$$

4.1.2 The Geometry of Linear Programming

From a geometric viewpoint, LP problems have two characteristics.

1. In the Euclidean space $\mathbf{R}^n$, surfaces with equal objective function values are parallel hyper planes.

2. If it's nonempty, the feasible set of a LP problem constitutes a polyhedron. Each facet of a polyhedron is plane and for the standard form, at least one corner (i.e. extreme point) must exist.

Thus the solutions to LP problems can be summarized in the following [56].

1. If the feasible set is nonempty and bounded, at least one optimal solution exists and there exists one optimal solution which is an extreme point.

2. If the feasible set is unbounded, there are following possibilities:

   (a) There exists an optimal solution which is an extreme point.

   (b) There exists an optimal solution, but no optimal solution is an extreme point. (This is impossible for the standard form LP.)

   (c) The optimal cost is $-\infty$

4.1.3 The Simplex Method

The simplex method applies to the standard form LP problem, in which if the optimal cost is not $-\infty$, there must be one extreme point being an optimal solution. In this case, feasibility and nonnegativity of the reduced costs (formula 4.4) are the optimality conditions for a basic solution $\mathbf{x}$ or a basis matrix $\mathbf{B}$ [56].

$$\begin{aligned} \mathbf{B^{-1}b} &\geq \mathbf{0} \\ \bar{\mathbf{c}}' = \mathbf{c}' - \mathbf{c}_\mathbf{B}'\mathbf{B^{-1}A} &\geq \mathbf{0}' \end{aligned} \tag{4.4}$$

So the simplex method just simply moves from one extreme point of the feasible set to another one along an edge of the polyhedron, each time reducing the cost, until an optimal solution is reached. This process is implemented by performing basis changes (i.e. change one column in matrix $\mathbf{B}$) whenever the optimality conditions are violated. Rules for choosing edge to move are called pivoting rules, which may significantly affect the efficiency of the simplex method. If a basic feasible solution is degenerate, cycling may happen; and furthermore if it's an optimal solution, it may not satisfy the nonnegativity of the reduced costs, i.e. some corresponding basis matrices violate the optimality conditions. But anyway there must exist one basis which satisfies the optimality conditions. Anticycling methods such as lexicography and Bland's rule can help simplex algorithm reach the optimal basis and prevent cycling.

In the details of implementation of the simplex method, different ways to organize the required computations, e.g. different ways to calculate the left sides in formula 4.4, will lead to different efficiency. The revised simplex method and the full tableau implementation are usually used in practice.

Starting the simplex method requires an initial basic feasible solution and an associated tableau. These problems can be solved in two-phase simplex method, in which an auxiliary LP problem is solved first or big-M method, in which the two phases are combined by introducing a positive infinite multiplier in the objective function.

In practice, the simplex method is a rather efficient algorithm. While in the worst case, the number of pivot can be an exponential function of the number of variables and constraints, its average behavior is pretty good [56].

4.1.4 The Dual Simplex Method

For any LP problem, an associated dual LP problem can be established according to some mechanical rules. Formulas 4.5 and 4.6 are a pair of primal and dual LP problem.

$$
\begin{array}{rl}
\text{minimize} & \mathbf{c}'\mathbf{x} \\
\text{subject to} & \mathbf{a_i'x} \geq b_i,\ i \in M_1, \\
& \mathbf{a_i'x} \leq b_i,\ i \in M_2, \\
& \mathbf{a_i'x} = b_i,\ i \in M_3, \\
& x_j \geq 0,\ j \in N_1, \\
& x_j \leq 0,\ j \in N_2, \\
& x_j \text{ free},\ j \in N_3 \qquad (4.5)
\end{array}
\qquad
\begin{array}{rl}
\text{maximize} & \mathbf{p}'\mathbf{b} \\
\text{subject to} & p_i \geq 0,\ j \in M_1, \\
& p_i \leq 0,\ i \in M_2, \\
& p_i \text{ free},\ i \in M_3 \\
& \mathbf{p'A_j} \leq c_j,\ j \in N_1, \\
& \mathbf{p'A_j} \geq c_j,\ j \in N_2, \\
& \mathbf{p'A_j} = c_i,\ j \in N_3 \ (4.6)
\end{array}
$$

From the duality theorem, it is known that if the primal problem has an optimal solution, then so does the dual and the respective optimal costs are equal; if the primal problem is infeasible, then the dual one may be infeasible too or unbounded; if the primal is unbounded, then the dual one is infeasible [56].

The dual simplex method is similar to the primal simplex method applied to the dual problem. For a primal LP problem and the associated dual one, every basis matrix determines not only a primal basic solution but also a dual basic solution. So, in the primal simplex method, as we move from one primal basic feasible solution to another one, we are simultaneously moving from one dual basic nonfeasible solution to another until the optimal one, which is a basic feasible solution in both primal and dual. In the dual simplex method, we move from one dual basic feasible solution to another one, associated with moving from one primal basic nonfeasible solution to another until the optimal one. The dual simplex method is usually used when the optimal basis for a similar LP problem, in which only the right-hand side vector **b** is different, is known.

In fact, the duality theorem can have more general forms and more profound mathematical meaning. It can also be applied to non-linear programming problem under some conditions [56].

4.1.5 The Ellipsoid Method

The simplex method is very effective in solving LP problems arising in applications. In the worst case, however, the simplex method can take an exponential number of iterations [56]. The ellipsoid method is known as a polynomial time algorithm, but it didn't lead to a practical algorithm.

The basic idea in the ellipsoid method is to solve a feasibility problem by constructing a series of ellipsoids, containing the polyhedron and decreasing in volume. In each iteration, the center of the ellipsoid will be checked if it's a feasible solution. If yes, then the ellipsoid method is terminated; if no, a halfspace bounded at the center and containing the polyhedron will be established. A new ellipsoid which contains the interception of the halfspace and the old ellipsoid will be constructed in a simple and quick way. Volume of the new ellipsoid is smaller than the old one and once its volume is below a threshold, a conclusion that the polyhedron is empty can be obtained.

Based on the LP problem and the corresponding dual problem such as Eq. 4.7 &

4.8, a feasibility problem can be constructed (Eq. 4.9).

$$\begin{aligned} \text{minimize} \quad & \mathbf{c}'\mathbf{x} \\ \text{subject to} \quad & \mathbf{A}\mathbf{x} \geq \mathbf{b}, \end{aligned} \tag{4.7}$$

$$\begin{aligned} \text{maximize} \quad & \mathbf{p}'\mathbf{b} \\ \text{subject to} \quad & \mathbf{p}\mathbf{A}' = \mathbf{c}' \\ & \mathbf{p} \geq \mathbf{0} \end{aligned} \tag{4.8}$$

$$\begin{aligned} &\text{In the space of } (\mathbf{x}, \mathbf{p}), \\ &\text{if feasible: } \mathbf{p}'\mathbf{b} = \mathbf{c}'\mathbf{x}, \mathbf{A}\mathbf{x} \geq \mathbf{b}, \mathbf{p}'\mathbf{A} = \mathbf{c}', \mathbf{p} \geq \mathbf{0}. \end{aligned} \tag{4.9}$$

Thus, by implementing the ellipsoid method on Eq. 4.9, the original LP problem (Eq. 4.7) can be solved.

4.1.6 Interior Point Methods

Interior point methods are the hottest research area in LP. They combine the advantages of the simplex method and of the ellipsoid algorithm. From a theoretical point of view, they lead to polynomial time algorithms and in practice, they often outperform the simplex method for large, sparse problems [56].

The affine scaling algorithm

In the ellipsoid method, we approximate the polyhedron by a series of ellipsoids which contains the polyhedron. In the affine scaling algorithm, however, we create a series of ellipsoids contained by the polyhedron. The basic idea of the affine scaling algorithm is as follows.

1. Create an ellipsoid in the interior of the feasible set.

2. The optimization problem over the ellipsoid can be easily solved. Closed forms of the optimal solution $\mathbf{x}^*$ and the duality gap can be obtained.

3. If the duality gap is smaller than the requirement or the problem is found to be unbounded, then algorithm terminates.

4. If the duality gap is bigger than the requirement, then create a new ellipsoid centered at $\mathbf{x}^*$ in the interior and repeat from step 2.

At each iteration, the objective function value associated with $\mathbf{x}^*$ is strictly decreased. At the end, the affine scaling algorithm generates a near-optimal solution. In practice, the affine scaling algorithm, especially the long-step method, has excellent performance and it can be initialized by a method similar to the big-M method [56].

The potential reduction algorithm

The biggest flaw in the affine scaling algorithm is that when the objective function value associated with $\mathbf{x}^*$ decreases, $\mathbf{x}^*$ approaches the boundary of the feasible set quickly and then the algorithm is forced to take very small steps as the approximating ellipsoids become smaller and smaller [56].

In the potential reduction algorithm, however, a nonlinear potential function (Eq. 4.10) is created to balance decreasing the objective function value and staying away from the boundary of the feasible set. It can be proved that if the potential function is decreased at each step by a certain amount, an ϵ-optimal solution can be obtained after a small number of iterations.

$$G(\mathbf{x}, \mathbf{s}) = q \log \mathbf{s}'\mathbf{x} - \sum_{j=1}^{n} \log x_j - \sum_{j=1}^{n} \log s_j \tag{4.10}$$

In practice, at each iteration, the update potential reduction direction can be quickly calculated through a LP problem over an ellipsoid, in which the nonlinear function $G(\mathbf{x}, \mathbf{s})$ is approximated by a linear function. Thus a closed form exists for the potential reduction direction, which is similar to that in the step 2 in the affine

scaling algorithm. The primal step or the dual step will then be operated to make sure the potential function is decreased at each step by a certain amount.

The primal path following algorithm

So far, we can see that in the affine scaling algorithm and the potential reduction algorithm, we don't work on the original LP problem directly, but work on an approximate simpler problem. This approximate problem can be solved easily and when we continuously tune the approximate problem or say reset some parameters in it, the associated optimal solution will converge to the original optimal solution.

The primal path following algorithm utilizes the same methodology, in which a barrier function (Eq. 4.11) is created to force any variable away from the boundary.

$$B_\mu(\mathbf{x}) = \mathbf{c}'\mathbf{x} - \mu \sum_{j=1}^{n} \log x_j \tag{4.11}$$

Solve the barrier problem (Eq. 4.12) and let $\mu \rightarrow 0$, the optimal barrier problem solution $\mathbf{x}(\mu)$ will converge to the original optimal solution $\mathbf{x}^*$.

$$\begin{aligned} \text{minimize} \quad & B_\mu(\mathbf{x}) \\ \text{subject to} \quad & \mathbf{A}\mathbf{x} = \mathbf{b}. \end{aligned} \tag{4.12}$$

However, the barrier problem is a nonlinear programming problem, which is hard to solve. Langrange multiplier can be used to solve the quadratic approximation of the barrier problem and a sub-optimal solution $\mathbf{x}'(\mu)$ (or say the primal Newton direction) can be obtained. When $\mu \rightarrow 0$, $\mathbf{x}'(\mu)$ also converges to $\mathbf{x}^*$.

The primal-dual path following algorithm

The primal-dual path following algorithm is similar to the primal path following algorithm. Both of them use the idea of approximating the central path by taking Newton steps. While in the primal path following algorithm, we only uses the primal Newton direction; in the primal-dual path following algorithm it finds Newton directions

not only in the primal but also the dual space. Thus, this algorithm has excellent performance in large-scale applications and it is the method of choice in commercial implementations of interior point methods. More details can be found in [56].

4.2 Network Optimization

4.2.1 Introduction

Network flow problems are one of the most important and most frequently encountered class of optimization problems. The network optimization problem is usually modeled by a directed graph $\mathcal{G} = (\mathcal{N}, \mathcal{A})$: the supply, demand and transshipment points are modeled by the nodes of the graph $\mathcal{N}$; the routes are modeled by the arcs of the graph $\mathcal{A}$ [57]. Several major classes of network optimization problems arised in practice are introduced in the following.

The Minimum Cost Flow Problem

This problem is to find a set of arc flows that minimize a linear cost function, subject to some constraints; that is,

$$\text{minimize} \quad \sum_{(i,j)\in\mathcal{A}} a_{ij}x_{ij} \tag{4.13}$$

$$\text{subject to} \quad \sum_{\{j|(i,j)\in\mathcal{A}\}} x_{ij} - \sum_{\{j|(j,i)\in\mathcal{A}\}} x_{ji} = s_i, \ \forall i \in \mathcal{N}, \tag{4.14}$$

$$b_{ij} \leq x_{ij} \leq c_{ij}, \ \forall (i,j) \in \mathcal{A}, \tag{4.15}$$

where, x_{ij} is the arc flow of arc (i,j); a_{ij} is the cost coefficient of (i,j); s_i is the supply of node i; b_{ij} and c_{ij} are the flow bounds of (i,j). Eq. 4.14 is the conservation constraints and Eq. 4.15 is the capacity constraints. From the above equations, it can be seen that the minimum cost flow problem is a special case of linear programming problem.

The minimum cost flow problem has many applications, for example, the shortest path problem, the assignment problem, the max-flow problem and the transportation problem [57].

The Shortest Path Problem This is a problem of finding a shortest path from node s to node t in a graph. It can be cast as follows.

$$\text{minimize} \quad \sum_{(i,j)\in\mathcal{A}} a_{ij}x_{ij} \tag{4.16}$$

$$\text{subject to} \quad \sum_{\{j|(i,j)\in\mathcal{A}\}} x_{ij} - \sum_{\{j|(j,i)\in\mathcal{A}\}} x_{ji} = \begin{cases} 1 & \text{if } i = s, \\ -1 & \text{if } i = t, \\ 0 & \text{otherwise,} \end{cases} \tag{4.17}$$

$$0 \leq x_{ij}, \ \forall (i,j) \in \mathcal{A}. \tag{4.18}$$

Moreover, another constraint should be satisfied:

$$x_{ij} = \begin{cases} 1 & \text{if } (i,j) \text{ belongs to } P, \\ 0 & \text{otherwise,} \end{cases} \tag{4.19}$$

where P is a forward path. However, due to the property of the linear programming problem that the optimal solution is usually at an extreme point, it can be shown that Eq. 4.19 is nonnecessary and can be satisfied implicitly.

In AREA project, the original AUV path planning problem can be transformed into a shortest path problem in the first stage. The shortest path problem will be discussed in more details later.

The Assignment Problem Suppose that there are n persons and n objects that we have to match on a one-to-one basis. There is a benefit a_{ij} for matching person i with object j, and we want to assign persons to objects so as to maximize the total benefit. This is a typical assignment problem, in which there are two groups of nodes: node i corresponds to person i; node j corresponds to object j. Such kind of problem

can be formulated as follows.

$$\text{maximize} \quad \sum_{(i,j)\in\mathcal{A}} a_{ij}x_{ij} \tag{4.20}$$

$$\text{subject to} \quad \sum_{\{j|(i,j)\in\mathcal{A}\}} x_{ij} = 1,\ \forall i = 1,\cdots,n, \tag{4.21}$$

$$\sum_{\{i|(i,j)\in\mathcal{A}\}} x_{ij} = 1,\ \forall j = 1,\cdots,n, \tag{4.22}$$

$$0 \le x_{ij} \le 1,\ \forall\,(i,j) \in \mathcal{A}, \tag{4.23}$$

Actually we should further restrict x_{ij} to be either 0 or 1. However, similar to the shortest path problem, this constraint can be satisfied implicitly [57].

The Max-Flow Problem In this problem, we have a graph with two special nodes: the source s and the sink t. The objective is to find a flow vector that makes the divergence of all nodes other than s and t equal to 0 while maximizing the divergence of s. To formulate this problem, an artificial arc (t,s) is introduced and mathematically this problem is casted as:

$$\text{maximize} \quad x_{ts} \tag{4.24}$$

$$\text{subject to} \quad \sum_{\{j|(i,j)\in\mathcal{A}\}} x_{ij} - \sum_{\{j|(j,i)\in\mathcal{A}\}} x_{ji} = 0,\ \forall i \in \mathcal{N} \text{ with } i \neq s \text{ and } i \neq t \tag{4.25}$$

$$\sum_{\{j|(s,j)\in\mathcal{A}\}} x_{sj} = \sum_{\{i|(i,t)\in\mathcal{A}\}} x_{it} = x_{ts}, \tag{4.26}$$

$$b_{ij} \le x_{ij} \le c_{ij},\ \forall\,(i,j) \in \mathcal{A} \text{ with } (i,j) \neq (t,s)\,. \tag{4.27}$$

Network Flow Problems with Convex Cost

As aforementioned, the minimum cost flow problem is a special case of linear programming problem. In practice, however, the cost function may not be linear. An important special case is that the cost function is convex and the feasible set is also convex, i.e.

$$\begin{aligned} \text{minimize} \quad & f(x) \\ \text{subject to} \quad & x \in F, \end{aligned} \tag{4.28}$$

where F is a convex subset of flow vectors in a graph and f is a convex function over F. The cost function $f(x)$ and the constraints F can be separable (Eq. 4.29, 4.30, 4.31)or nonseparable.

$$f(x) = \sum_{(i,j)\in\mathcal{A}} f_{ij}(x_{ij}), \tag{4.29}$$

$$F = \left\{ x \in X \,\middle|\, \sum_{\{j|(i,j)\in\mathcal{A}\}} x_{ij} - \sum_{\{j|(j,i)\in\mathcal{A}\}} x_{ji} = s_i,\ \forall i \in \mathcal{N} \right\}, \tag{4.30}$$

$$X = \{x \,|x_{ij} \in X_{ij}, (i,j) \in \mathcal{A}\} \tag{4.31}$$

It is known that separability is the most important structural characteristic of convex network problems. For nonseparable network optimization problems, the solutions are much more difficult since some algorithms and nice properties do not apply in the absence of a separable structure [57].

Discrete Network Optimization Problems

In many linear or convex network flow problems, there may be integer constraints on the arc flows. The most famous example is the traveling salesman problem. For such kind of problems, the solutions are extremely difficult and strict optimal solution is often not available in practice [57].

4.2.2 Network Flow Algorithms Overview

Linear and convex network optimization problems are special cases of linear and nonlinear optimization problems respectively. General purpose linear and nonlinear algorithms can thus be applied. However, the network structure can be exploited to speed up the solution. In practice, network optimization problems can often be solved hundreds and even thousands of times faster than the general optimization programs of comparable dimension [57].

The algorithms for linear and convex (separable) network problems can be grouped in three main categories:

1. Primal cost improvement. Here a sequence of feasible flows is constructed to improve the primal cost iteratively. The simplex method is an example.

2. Dual cost improvement. Here a dual problem is constructed and a sequence of prices is developed to improve the dual cost iteratively. In this category, the optimality condition is the complementary slackness. The dual simplex method is an example.

3. Auction. The auction algorithm is like an approximate dual cost improvement process. However, there is no primal or dual cost improvement. The auction algorithm is very much similar to the real-life auction process. For example, in the assignment problem, a_{ij} can be viewed as the internal value of object j to person i. The price p_j of object j in the corresponding dual problem can be viewed as the current bidding price of object j. Our target is to make the complementary slackness (Eq. 4.32) satisfied for all i.

$$a_{ij_i} - p_{j_i} = \max_{j \in \mathcal{A}(i)} \{a_{ij} - p_j\} \tag{4.32}$$

The bidding process in the naive auction algorithm is roughly like that:

- let person i choose the most valuable object j_i, which satisfies the complementary slackness.

- increase p_{j_i} such that person i is indifferent between j_i and the second best object.
- repeat the above processes until all persons are assigned and thus the complementary slackness is satisfied.

In practical auction algorithm, the price increase at each time is required to be bigger than a small number ϵ and at the end, the ϵ-complementary slackness (Eq. 4.33) is satisfied.

$$a_{ij_i} - p_{j_i} \geq \max_{j \in \mathcal{A}(i)} \{a_{ij} - p_j\} - \epsilon \tag{4.33}$$

For the network optimization problems with integer constraints, the popular methods include branch-and-bound method, local search methods and rollout algorithms etc. The local search methods include genetic algorithms, tabu search, simulated annealing etc. More details about those methods can be found in [57].

4.2.3 Shortest Path Problem

The shortest path problem appears in a large variety of contexts. In the AREA project, it plays an important role. In this subsection, some often-used shortest path algorithms are briefly introduced.

Complementary Slackness

For the shortest path problem, the complementary slackness (CS) conditions is as follows.

$$d_j \leq d_i + a_{ij}, \ \forall (i,j) \in \mathcal{A}, \tag{4.34}$$

let P be a path starting at a node i_1 and ending at a node i_k,

$$d_j = d_i + a_{ij}, \ \forall (i,j) \text{ in } P. \tag{4.35}$$

d_i is the label for node i, which actually indicates the shortest distance from node i_1 to i.

Generic Algorithm

To find the shortest distance from i_1 to all other nodes, we can start with some vector of labels $(d_1, d_2, \cdots, d_N)$, and then successively select arc (i, j) that violates the CS condition, i.e., $d_j > d_i + a_{ij}$, and set $d_j := d_i + a_{ij}$. This process can be repeated many times until the CS condition is satisfied for all arcs. With $i_1 = 1$, the generic algorithm can be formulated as follows.

Initialization: $V = \{1\}$, $d_i = 0$, $d_i = \infty$, $\forall i \neq 1$.

Remove a node i from V.

$\forall (i, j) \in \mathcal{A}$, if $d_j > d_i + a_{ij}$, set $d_j := d_i + a_{ij}$ and add j to V if $j \notin V$.

In practice, more advanced initialization could be used. The most important thing in the generic algorithm is how to select node i to be removed from V. Different selecting rules may lead to very different computation speed.

Label Setting (Dijkstra) Methods

The Dijkstra method is the special case of the generic algorithm where the node i removed from the candidate list V at each iteration has minimum label, i.e.,

$$d_i = \min_{j \in V} d_j. \tag{4.36}$$

If all arc lengths are nonnegative, any node can be removed from V for at most once, so the number of iterations required by the Dijkstra method is equal or less than N. On the other hand, however, the overhead for finding the minimum d_i may require $O(N^2)$ operations. The binary heap method and Dial's algorithm can be used to minimize the overhead.

Label Correcting Methods

In label correcting methods, the selection of the node to be removed from the candidate list V is simpler and requires less overhead than in label setting methods, at the expense of multiple entrances of nodes in V [57].

The Bellman-Ford Method is the simplest label correcting method, in which a first-in first-out (FIFO) rule is adopted and the candidate list V is maintained in a queue. The Bellman-Ford method is actually very close to the deterministic dynamic programming method.

The D'Esopo-Pape Algorithm In this method, a node is always removed from the top of the queue of V. A node, upon entrance, is placed at the bottom of the queue if it has never been in the queue before; otherwise it is placed at the top [57].

The SLF and LLL Algorithms In the Small Label First (SLF) method, a node is always removed from the top of a double ended queue Q. Whenever a node j enters Q, its label d_j is compared with the label d_i of the top node i of Q. If $d_j \leq d_i$, node j is entered at the top of Q; otherwise j is entered at the bottom of Q. In the Large Label Last (LLL) method, at each iteration, when the node i at the top of Q has a larger label than the average node label in Q, i.e. $d_i > \frac{\sum_{j \in Q} d_j}{|Q|}$, then node i is not removed from Q but moved to the bottom of Q. The SLF and LLL methods can be combined together and thereby obtaining a method referred to as SLF/LLL.

The Threshold Algorithm In this method, the candidate list V is partitioned into 2 queues Q' and Q'' using a threshold s, i.e. $d_j \leq s$, $\forall j \in Q'$ and $d_j > s$, $\forall j \in Q''$. At each iteration, a node is removed from Q', and any node j to be added is inserted at the bottom of Q' or Q'' depending on whether $d_j \leq s$ or $d_j > s$. When the queue Q' is exhausted, Q'' will be repartitioned. So, the threshold algorithm can be viewed as a block version of Dijkstra's method. Furthermore, the threshold method can be combined with SLF/LLL methods. In practice, the combination performs extremely well.

The auction algorithm can also be applied to the shortest path problem. More details can found in [57, 47].

4.3 Dynamic Programming

Before we get into the topic of Dynamic Programming (DP), let's first introduce the stochastic programming problem.

4.3.1 Stochastic Programming

In most optimization problems, the objective function $\mathbf{f}(\mathbf{x})$ only depends on the decision variable $\mathbf{x}$, i.e. once $\mathbf{x}$ is determined then $\mathbf{f}(\mathbf{x})$ is determined completely. However, in some problems, stochastic disturbances exist in $\mathbf{f}(\mathbf{x})$, e.g. if $\mathbf{x_1}$ and $\mathbf{x_2}$ are the moneys that we invested into two different stocks one month ago and $\mathbf{f}(\mathbf{x_1}, \mathbf{x_2})$ is the value of asset that we possess in next month, then $\mathbf{f}(\mathbf{x_1}, \mathbf{x_2})$ does not only depends on the money we invested but also depends on the stock prices in next month, which are not completely known right now but statistical models usually exist. If now we require that $\mathbf{x_1} + \mathbf{x_2} \leq 100$, then how to invest the \$100 into those two stocks so as to maximize $\mathbf{f}(\mathbf{x_1}, \mathbf{x_2})$ is a typical stochastic programming problem. Since $\mathbf{f}(\mathbf{x_1}, \mathbf{x_2})$ is random, thus $E[\mathbf{f}(\mathbf{x_1}, \mathbf{x_2})]$ is usually used as the objective function instead of $\mathbf{f}(\mathbf{x_1}, \mathbf{x_2})$ itself.

$$\begin{aligned} \text{maximize} \quad & \mathbf{E}[\mathbf{f}(\mathbf{x_1}, \mathbf{x_2})] \\ \text{subject to} \quad & \mathbf{x_1} + \mathbf{x_2} \leq \mathbf{100}, \\ & \mathbf{x_1} \geq \mathbf{0}, \\ & \mathbf{x_2} \geq \mathbf{0}, \end{aligned} \tag{4.37}$$

where the expectation is based on the statistical model. More accurate is this model, more benefit return comes in reality.

Methods to solve such kind of optimization problem is essentially the same as methods to solve ordinary optimization problems. The only difference may be that

the expectation will lead to much intensive computation.

4.3.2 Sequential Decision Making Under Stochastic Disturbance

The preceding problem is actually a decision making problem under stochastic disturbance. The decision is made in just one step, i.e. $\mathbf{x_1}$ and $\mathbf{x_2}$ are determined at the same time. If now $\mathbf{x_1}$ is needed to be determined one month ago, while $\mathbf{x_2}$ should be determined by today, then the problem becomes as a decision making problem under stochastic disturbance over a finite number of stages.

$\mathbf{x_2}$ can be determined as early as $\mathbf{x_1}$ is determined, then the result will be the same as that in the previous problem. However, if we delay the determination of $\mathbf{x_2}$ until today, more information about the price of the second stock will be available and better statistical model will be available. Consequently, $\mathbf{E}\left[\mathbf{f}\left(\mathbf{x_1}, \mathbf{x_2}\right)\right]$ will be better optimized.

The decision making problem under stochastic disturbance over a finite (or infinite) number of stages is usually formulated in the frame of dynamic programming[47]. The basic problem is described as follows.

A discrete-time dynamic system is given

$$x_{k+1} = f_k\left(x_k, u_k, \omega_k\right), \quad k = 0, 1, \cdots, N-1, \tag{4.38}$$

where the state x_k is an element of a space S_k, the control u_k is an element of a space $U_k\left(x_k\right)$, and ω_k is the random disturbance, whose PDF is $P_k\left(\omega_k | x_k, u_k\right)$. N is the total number of stages.

The cost function associated with each stage is $g_k\left(x_k, u_k, \omega_k\right)$ and at the stage N, the termination cost is $g_N\left(x_N\right)$. So the objective function is

$$E\left\{g_N\left(x_N\right) + \sum_{k=0}^{N-1} g_k\left(x_k, u_k, \omega_k\right)\right\}, \tag{4.39}$$

where the expectation is over all possible x_k and ω_k.

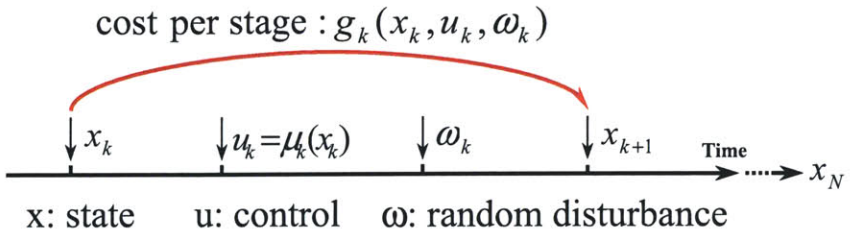


Figure 4-1: Time table of a DP problem.

We consider the class of policies that consist of a sequence of functions

$$\pi = \{\mu_0, \mu_1, \cdots, \mu_{N-1}\}, \tag{4.40}$$

where μ_k maps states x_k into controls $u_k = \mu_k(x_k)$ and is such that $\mu_k(x_k) \in U_k(x_k)$ for all $x_k \in S_k$. For a given initial state x_0 and policy π, the associated objective function is

$$J_\pi(x_0) = \underset{\substack{\forall x_k, \omega_k, \\ k=0,1,2,\cdots,N}}{E} \left\{ g_N(x_N) + \sum_{k=0}^{N-1} g_k(x_k, \mu_k(x_k), \omega_k) \right\}. \tag{4.41}$$

The goal in the DP problem is to find the optimal policy π^* such that

$$J_{\pi^*}(x_0) = \min_{\pi \in \Pi} J_\pi(x_0), \tag{4.42}$$

where Π is the set of all admissible policies.

4.3.3 Dynamic Programming Algorithm

Based on the principle of optimality, the DP problem can be solved by the DP algorithm, which is described in the following [47].

For every initial state x_0, the optimal cost $J^*(x_0)$ of the basic problem is equal to $J_0(x_0)$, given by the last step of the following algorithm, which proceeds backward in

time from period $N-1$ to period 0:

$$J_N(x_N) = g_N(x_N) \tag{4.43}$$

$$J_k(x_k) = \min_{u_k \in U_k(x_k)} E_{\omega_k} \left\{ g_k(x_k, u_k, \omega_k) + J_{k+1}\left(f_k\left(x_k, u_k, \omega_k\right)\right) \right\}, \quad k = 0, 1, \cdots, N-1, \tag{4.44}$$

where the expectation is taken with respect to the probability distribution of ω_k, which depends on x_k and u_k. Furthermore, if $u_k^* = \mu_k^*(x_k)$ minimize the right side of Eq. 4.44 for each x_k and k, the policy $\pi^* = \left\{\mu_0^*, \cdots, \mu_{N-1}^*\right\}$ is optimal.

4.3.4 Deterministic Dynamic Programming Problem

Deterministic DP problems are problems where each disturbance ω_k can take only one value. An important property of deterministic DP problems is that, in contrast with stochastic problems, using feedback results in no advantage in terms of cost reduction. In other words, minimizing the cost over admissible policies $\{\mu_0, \mu_1, \cdots, \mu_{N-1}\}$ results in the same optimal cost as minimizing over sequences of control vectors $\{u_0, u_1, \cdots, u_{N-1}\}$, since once the initial state x_0 is determined, the state sequence and control sequence corresponding to an admissible policy can be known prior [47]. For deterministic DP problems, the expectation operator in the DP algorithm can be removed, and thus the deterministic DP algorithm is much faster than ordinary DP algorithm.

An deterministic finite-state DP problem can be posed as an equivalent shortest path problem and solved by shortest path algorithms, which are often faster than the deterministic DP algorithm [47]. On the other hand, a shortest path problem can be transformed into an equivalent deterministic finite-state DP problem and solved by the deterministic DP algorithm.

4.3.5 Approximate Dynamic Programming Methods

It is well known that for many important DP problems in engineering, the computational requirements of DP algorithm are overwhelming, because the number of states, controls and disturbances is very large. This is called the Bellman's "curse of dimensionality", which motivates the pursuit of sub-optimal solution and the appearance of approximate DP methods [58].

Generally speaking, all approximate DP methods are based on the approximation of the cost-to-go function $J_k(x_k)$ in Eq. 4.44. If the approximate cost-to-go $\tilde{J}_k(x_k),\ \forall k = 0, 1, \cdots, N$ is available, a sub-optimal policy can be obtained by

$$\tilde{\mu}_k(x_k) = \underset{u_k \in U_k(x_k)}{\arg\min}\ \underset{\omega_k}{E}\left\{ g_k(x_k, u_k, \omega_k) + \tilde{J}_{k+1}\left(f_k\left(x_k, u_k, \omega_k\right)\right)\right\},\ \ k = 0, 1, \cdots, N-1. \tag{4.45}$$

$\tilde{J}_k(x_k)$ is often much less computationally intensive, so $\tilde{\mu}_k(x_k)$ can be computed very quickly. $\tilde{J}_k(x_k)$ can be constructed based on artificial neural networks, then it is called neuro-dynamic programming method. Moreover, for a certain problem, some heuristic methods may exist for constructing the approximation $\tilde{J}_k(x_k)$. On this basis, the rollout algorithm can be applied and it often leads to a good sub-optimal policy [47, 58]. The rollout algorithm based on greedy algorithm is pretty popular for on-line policy optimization.

In engineering problems, the Q-factor is often used to replace the $J_k(x_k)$ in the policy optimization. The optimal Q-factor is defined as follows,

$$Q_k^*\left(x_k, u_k\right) = E\left\{ g_k(x_k, u_k, \omega_k) + J_{k+1}^*\left(f_k\left(x_k, u_k, \omega_k\right)\right)\right\}. \tag{4.46}$$

Once the $Q_k^*\left(x_k, u_k\right)$ is available, the optimal policy at stage k is

$$\mu_k^*(x_k) = \underset{u_k \in U_k(x_k)}{\arg\min}\ Q_k^*\left(x_k, u_k\right), \tag{4.47}$$

and the optimal cost-to-go at stage k is

$$J_k^*(x_k) = \min_{u_k \in U_k(x_k)} Q_k^*(x_k, u_k). \tag{4.48}$$

Q-factor can also be approximated through heuristics. In some dynamic systems, there is no explicit model of the system and the cost structure. For such kind of system, the cost-to-go function and the Q-factor can be estimated by some simulation-based methods such as temporal difference method and Q-learning method etc [58].

Combine the Eq. 4.46 and Eq. 4.48, we can obtain the Bellman's equation in terms of Q-factor,

$$Q_k^*(x_k, u_k) = \underset{\omega_k}{E}\left\{ g_k(x_k, u_k, \omega_k) + \min_{u_{k+1}} Q_{k+1}^*\left(f_k\left(x_k, u_k, \omega_k\right), u_{k+1}\right)\right\}, \quad k = 0, 1, \cdots, N-1. \tag{4.49}$$

Using the value iteration method [58] and replacing the expectation with a single sample, the Q-learning equation can be obtained as follows,

$$Q_k(x_k, u_k) := (1-\gamma) Q_k(x_k, u_k) + \gamma\left(g_k(x_k, u_k, \omega_k) + \min_{u_{k+1}} Q_{k+1}\left(x_{k+1}, u_{k+1}\right)\right), \quad k = 0, 1, \cdots, N-1. \tag{4.50}$$

γ is between 0 and 1. It can be proved that under some conditions, after infinite iterations of Eq. 4.50, the $Q_k(x_k, u_k)$ produced by Q-learning method will converge to Q^* for all states and controls at all stages [58].

The approximate dynamic programming is actually a very general concept, which includes many other methods not mentioned in this section. For more details, please refer to [58, 47]. Moreover, the approximate dynamic programming can be viewed as a branch in Machine Learning (ML) field. It is often called Reinforcement Learning (RL) [59]. If we compare the approximation methods and the decision making in the approximate dynamic programming with how the human being thinks and makes decision, some surprising similarity can be found. However, human being's brain is a low-speed but highly parallel computing system, very different from the current

computers. It may be just this reason that makes human being able to make much smarter decisions than current Artificial Intelligence (AI) techniques in very large-scale and very complicated problems. I believe that on someday, by combining the newest neuro-biological achievement and AI techniques, the approximate dynamic programming methods will perform better than human being in most problems.

Chapter 5

Modeling The AUV Path Planning Problem

In the AREA project, the AUV path planning problem can be formulated as an optimization problem as follows.

$$\begin{aligned} \min \quad & \mathbf{f}(\mathbf{x}) \\ \text{s.t.} \quad & \mathbf{x} \in \mathbf{X}, \end{aligned} \tag{5.1}$$

where, three items — $\mathbf{f}$, $\mathbf{x}$, $\mathbf{X}$ — are needed to be modeled or determined.

$\mathbf{x}$ is the decision variable; $\mathbf{f}$ is the objective function; $\mathbf{X}$ is the feasible set of $\mathbf{x}$. In the AUV path planning problem, $\mathbf{x}$ represents the AUV path, $\mathbf{f}$ may represent the posterior acoustic prediction uncertainty, posterior sonar performance prediction uncertainty or posterior SVP prediction uncertainty etc, and $\mathbf{X}$ is the set of all feasible AUV paths, which are constrained by AUV performance limits etc.

In this chapter, we will firstly discuss the constraints and requirements on the AUV path and then talk about how to represent the AUV path in a mathematical form — the decision variable $\mathbf{x}$. After that, how to select the objective function $\mathbf{f}$, how to model it and how to approximate it will be discussed.

5.1 Constraints and Requirements on The AUV Path

So far, in the underwater acoustic community, most people ate still using 2-D acoustic models. This is because in most scenarios, 2-D acoustic models can provide good enough acoustic field estimation with much less intensive computation, while 3-D acoustic models can provide higher precise but the computation is very intensive [60]. Therefore, at present only 2-D acoustic models are considered in the AREA project. This suggests that the in-situ measurements can be constrained on a vertical plane and the corresponding AUV path is a 2-D curve.

For the AUVs considered in AREA, the speed is about 3 knots (about 1.54 m/s) and the battery can last for 8 hours. In the AREA project, the AUV moves on a vertical plane along with a selected bearing and once it reaches the maximum horizontal range (about 10 to 15 km), the AUV will stop moving and float up and stay on the surface or make a "U" turn to come back. In this scenario, there's no any implicit constraint on the total distance of the AUV path but there is a limit on the horizontal distance — the maximum range. This requirement is consistent with the reality. Since if AUV goes too far in the horizontal, the telecommunication and control may get lost. Also, in shallow water, the geometry of ocean is like a paper sheet, thin and wide. The maximum pitch angle of AUV is about 10 to 20 degree; the upper bound and the lower bound of AUV path are around $5m$ and few hundred meters respectively (see Fig. 5-1). Therefore, the AUV's path is always like a horizontal line with some deviations. Moreover, the biggest advantage of the fixed maximum range is that it will make decision variable selection easier and dramatically decrease the search space dimensionality in the optimization problem.

In addition, it is assumed that an AUV always starts to move from a location close to the surface.

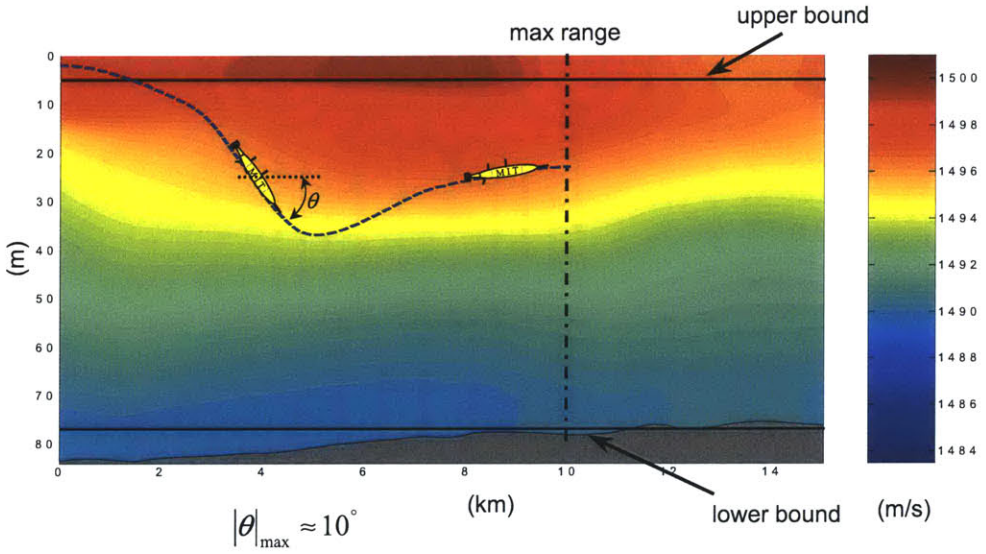
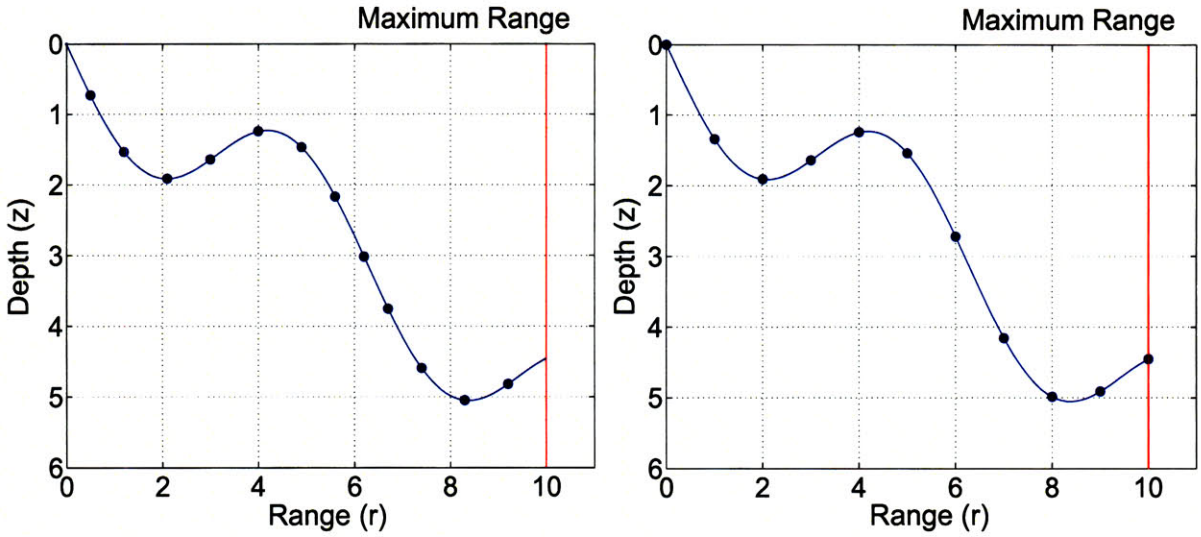


Figure 5-1: Illustration of constraints on the AUV path.

5.2 Select The Decision Variable

For such kind of 2-D curve with maximum horizontal distance fixed, several methods can be used to mathematically model the AUV path. The most intuitive one is to discretize and represent the path using a sequence of equal-distant waypoints $\{(r_1, z_1), (r_2, z_2), \cdots, (r_n, z_n)\}$, where r is the range and z is the depth (see Fig. 5-2(a)). In this way, it's easy to implement waypionts in the AUV lower level control, but unfortunately the number of waypoints n is not fixed and the search space dimensionality is $2n$. Therefore, the corresponding optimization problem is hard to model and hard to solve. Moreover, for such kind of path representation, it will be hard to discretize the ocean so that the waypoints are all located at grid points. Another way to model the path is to discretize and represent it with a sequence of equal-horizontal-distant waypoints $\{(0, z_1), (\Delta r, z_2), \cdots, (n\Delta r, z_n)\}$, where Δr is predetermined and $n\Delta r$ is equal to the maximum horizontal distance (see Fig. 5-2(b)). It can be seen that for this method the number n is fixed and the search space is n-dimensional — the space of $\{z_1, z_2, \cdots, z_n\}$. The corresponding optimization problem is much easier to construct and solve.



(a) Equal-distant waypoints

(b) Equal-horizontal-distant waypoints

Figure 5-2: Illustration of two ways to represent the AUV path.

5.2.1 Ocean Discretization for AUV Path

Once we determined the way to represent the AUV path, another essential problem comes — how to discretize the ocean. Our mission is to find the optimal path, so it may be a good try to put the original optimization problem (Eq. 5.1) in the network optimization framework. Therefore, we need to discretize the ocean vertical plane horizontally and vertically, and construct a graph for all feasible AUV paths [57]. Fig. 5-3 shows the way to discretize the ocean and construct a directed graph. The resolution of the discretization is very essential in the optimization problem, since the horizontal and vertical resolutions determine the size of the graph, which judges the real-time feasibility of the whole AREA system.

Let's use $\mathcal{G} = (\mathcal{N}, \mathcal{A})$ to represent a directed graph, where $\mathcal{N}$ is the set of nodes and $\mathcal{A}$ is the set of arcs. The number of nodes and arcs are denoted by N and A respectively. If it is assumed that for the AUV path the maximum range is $10km$, the upper bound is $5m$, the lower bound is $300m$, the maximum pitch angel is 3.5^o and furthermore if the horizontal resolution is 1.667km, the vertical resolution is about $49.167m$, then the directed graph is just like the one shown in white lines and arrows in Fig. 5-3, and $N = 39$, $A = 124$. If now the horizontal resolution is 0.833km, the

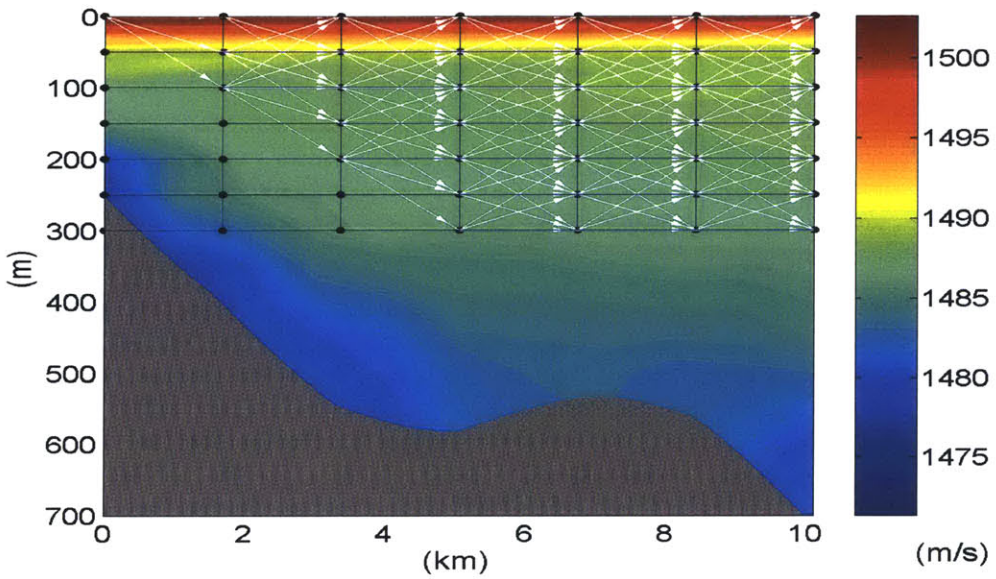


Figure 5-3: Illustration of ocean discretization for AUV path.

vertical resolution is about $24.58m$, N and A will increase to about 300 and 1500 respectively. The complexity of the graph is

$$N = O\left(\frac{LB-UB}{VR}\cdot\frac{MR}{HR}\right) \tag{5.2}$$

$$\begin{aligned} A &= O\left(\frac{2\cdot tan\,(MP)\cdot HR}{VR}\cdot\frac{LB-UB}{VR}\cdot\frac{MR}{HR}\right) \\ &= O\left(\frac{tan\,(MP)\cdot(LB-UB)\cdot MR}{VR^2}\right), \end{aligned} \tag{5.3}$$

where LB and UB are the lower bound and upper bound respectively, VR and HR are the vertical and horizontal resolution respectively, MR is the maximum range and MP is the maximum pitch angle in radius. From Eq. 5.2, it is seen that the number of nodes is related to both HR and VR, while from Eq. 5.3 it can be seen that the number of arcs is related to VR only.

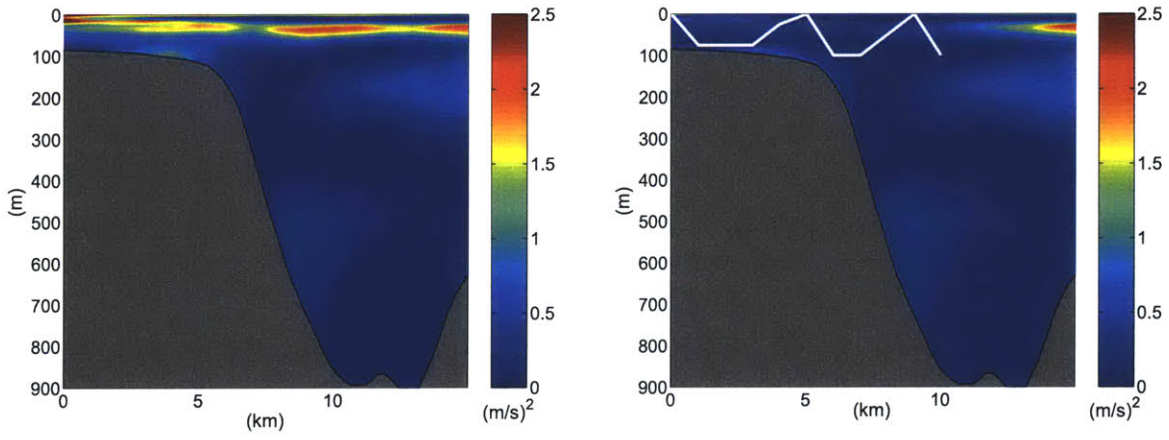
Although Eq. 5.2 and 5.3 indicate that the complexity of the graph is polynomial with respect to $\frac{1}{HR}$ and $\frac{1}{VR}$, it will be shown later that this is not true in practice due to the correlation effect of SVP.For this reason, HR and VR must be selected very carefully. However, it should be noticed that if the resolutions are too coarse, AUV performance such as the maximum pitch angle will have to be sacrificed a lot.

5.3 Objective Function

As discussed in chapter 3, AREA framework can be a multi-purpose system by choosing different objective functions. In this section, we will discuss some possible objective functions and some other very essential issues.

5.3.1 SVP Prediction Uncertainty

The AREA framework can serve for oceanographic purpose, e.g. SVP prediction. In this case, the objective function is the summation of the posterior SVP prediction error in the ocean area concerned. Fig. 5-4 shows an a priori SVP prediction error map and the posterior error map associated with an AUV path. Both of the standard deviation and the variance can indicate the error. The objective function can be either the summation of posterior standard deviation or the summation of posterior variance. In practice, however, only the latter one is adopted due to its additivity. More details is discussed in 7.



(a) A priori SVP prediction error variance

(b) Posterior SVP prediction error variance

Figure 5-4: A priori and posterior SVP prediction error variance map. The white curve corresponds to an AUV path.

The error map in Fig 5-4(a) is associated with the diagonal items of $\Lambda_{c_1}(-)$ — the a priori covariance matrix — in Eq. 3.9. Through this OA Equation, the posterior

covariance matrix — $\Lambda_{c_1}(+)$ can be obtained by detracting the uncertainty reduction part UR:

$$UR = \Lambda_{c_1}(-) D^T [D \cdot \Lambda_{c_1}(-) \cdot D^T + D \cdot \Lambda_{c_2'}(-) \cdot D^T + R]^{-1} D\Lambda_{c_1}(-). \quad (5.4)$$

The uncertainty reduction part is only dependent on the AUV path via the measurement matrix D. Thus the AREA problem can be casted as a minimization problem with respect to the trace of the posterior covariance matrix $tr(\Lambda_{c_1}(+))$ or an equivalent maximization problem with respect to $tr(UR)$. From Eq. 5.4, it can be seen that $tr(UR)$ can be highly nonlinear with respect to D and thus highly nonlinear with respect to $\{z_1, z_2, \cdots, z_n\}$ the representation of the AUV path. Furthermore, $tr(UR)$ should generally be neither convex nor concave. Therefore, in this case **the optimization problem is a nonlinear programming problem with non-convex and non-concave objective function**.

The Operational OA and The Operational Measurement Model

Here, it should be noticed that in real operations the measurement equation (Eq. 3.6), $d = D \cdot (\overline{c} + c_1 + c_2') + v$, is not so practical. This is because that the 2-D SVP has to be discretized and due to the concern about computation intensity the grid of SVP discretization is somewhat sparse in horizontal and vertical [8]. In-situ measurement made by AUV carrying CTD is usually much denser than that grid. The sound velocity measurements d can't thus be represented as in Eq. 3.6. Moreover, d is usually a very big vector so that it can't be directly used in OA and some preprocessing is necessary, otherwise inverse of a very large matrix will be encountered. So far, this problem is solved by using another discretization grid — the OA grid, as shown in Fig. 5-5.

The OA grid is independent on and usually denser than the SVP grid. In Fig. 5-5, each in-situ measurement point is projected to the nearest OA grid point and thus only those red OA grid points are treated as measured points. For each of them, the measured sound velocity is the average value of all associated in-situ measurements

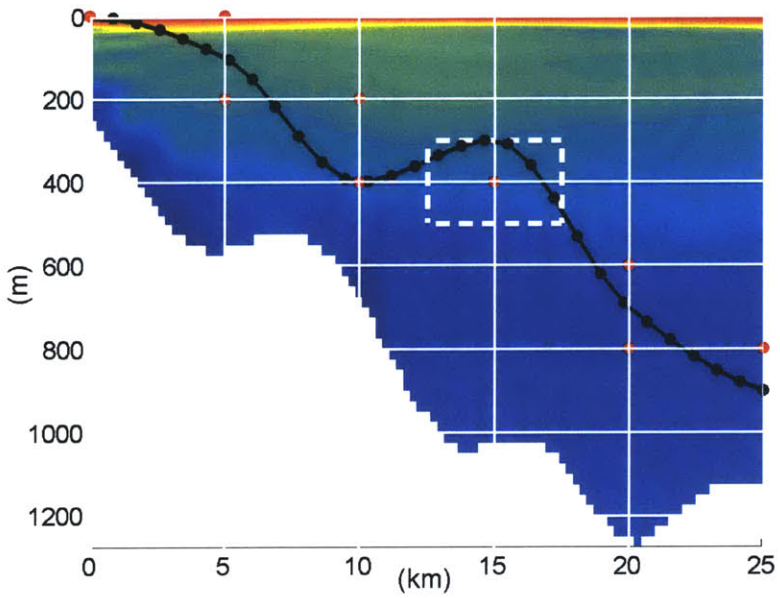


Figure 5-5: Illustration of the OA grid.

and the corresponding measurement error variance is accordingly decreased.

As shown in Eq. 3.7, in AREA we treat c_2' as some sort of measurement noise and combine it with the CTD noise v. We also assume that at one point the value of c_2' is independent of that at another point. Although this assumption is not very accurate, it is reasonable to some extent, since c_2' is weakly auto-correlated in time and space. Thus the real measurement model used in practice is

$$d_i = \overline{C}\left(p_i, t_i\right) + C_1\left(p_i, t_i\right) + \omega_i, \tag{5.5}$$

where, d_i is the ith measurement datum, p_i and t_i are the location and time of the ith measurement, ω_i stands for the gross measurement noise, which includes the CTD noise and C_2. It is assumed that for any two points i and j, ω_i and ω_j are independent and identical random variables. Moreover, ω_i is independent of $\overline{C}\left(p_i, t_i\right)$and $C_1\left(p_i, t_i\right)$.

After the pre-processing, the measurement model for the ith measured OA grid

point is

$$d_i' = \frac{\sum_{j \in S_i} d_j}{N_i} \tag{5.6}$$

$$= \frac{\sum_{j \in S_i} \left(\overline{C}\left(p_j, t_j\right) + C_1\left(p_j, t_j\right)\right)}{N_i} + \frac{\sum_{j \in S_i} \omega_j}{N_i}, \tag{5.7}$$

where the S_i is the set of in-situ measurement points associated with the ith measured OA grid point, N_i is the number of elements in S_i.

So for the example in Fig. 5-5, the grid point $(15, 400)$ is associated with 6 in-situ measurement points, located in the white box. The measurement datum at point $(15, 400)$ is the average of the 6 in-situ measurement data and the measurement error variance is $\frac{var(\omega_i)}{6}$. In this example, the size of d' is dramatically reduced from about 40 to 9.

In AREA, the operational OA equations are

$$\hat{c}(+) = \overline{c} + Cov\left(c_1, d'\right) Cov\left(d', d'\right)^{-1} \left[d' - \overline{c}_{d'}\right], \tag{5.8}$$

$$\Lambda_c(+) = \Lambda_{c_1}(-) - Cov\left(c_1, d'\right) Cov\left(d', d'\right)^{-1} Cov\left(c_1, d'\right)^T + \Lambda_{c_2}(-). \tag{5.9}$$

Here, the $\overline{c}_{d'}$ is the interpolation values of $\overline{c}$ at those measured OA grid points. To implement Eq. 5.8 and 5.9, we need to know $Cov\left(c_1, d'\right)$ and $Cov\left(d', d'\right)$. Let's denote

$$d' = d^1 + d^2, \tag{5.10}$$

and for the vector d^1 and d^2, we have

$$d_i^1 = \frac{\sum_{j \in S_i} \left(\overline{C}\left(p_j, t_j\right) + C_1\left(p_j, t_j\right)\right)}{N_i} \tag{5.11}$$

$$\approx \overline{C}\left(p_i^{OA}, t\right) + C_1\left(p_i^{OA}, t\right) \tag{5.12}$$

$$d_i^2 = \frac{\sum_{j \in S_i} \omega_j}{N_i}. \tag{5.13}$$

In Eq. 5.12, p_i^{OA} is the location of the ith measured OA grid point, t is an arbitrary time, since $\overline{C}$ and C_1 are time-invariant in the duration concerned. From Eq. 5.11 to 5.12 is because that we assume in the neighbor of an OA grid point, $\overline{C}$ and C_1 don't change much since they are highly auto-correlated in space. c_1 is uncorrelated with d^2, then $Cov\,(c_1, d^2) = 0$. Furthermore, since d^1 and d^2 are independent, we have

$$\begin{aligned} Cov\,(c_1, d') &= Cov\,(c_1, d^1) && (5.14) \\ Cov\,(d', d') &= Cov\,(d^1, d^1) + Cov\,(d^2, d^2) \\ &= Cov\,(d^1, d^1) + R_\omega. && (5.15) \end{aligned}$$

In AREA, $R_\omega = Cov\,(d^2, d^2)$ is set semi-empirically. The only problem now is how to construct $Cov\,(c_1, d^1)$ and $Cov\,(d^1, d^1)$.

Normal Distribution Assumption

In the real ocean, the random vector C_1 follows a very complicated stochastic model, which is not feasible in AREA due to the very intensive computation for Λ_{c_1}. The 2-D normal distribution assumption is made to simplify the computation.

We assume that the correlation coefficient function with respect to C_1 at points $p_1 = (r_1, z_1)$ and $p_2 = (r_2, z_2)$ is

$$\rho_{p_1,p_2} = \exp\left(-\frac{(\frac{r_1-r_2}{Lr})^2 + (\frac{z_1-z_2}{Lz})^2}{2}\right), \qquad (5.16)$$

Where r_1, r_2 are ranges of p_1 and p_2 respectively; z_1, z_2 are depths; Lr and Lz are **correlation length** of C_1 in horizontal and vertical respectively. Based on Eq. 5.12, the i,jth item of $Cov\,(c_1, d^1)$ is

$$\sigma(p_i^{c_1})\sigma(p_j^{d^1})\rho_{p_i^{c_1},p_j^{d^1}}$$

where $p_i^{c_1}$ is the location of the ith point in c_1, $p_j^{d^1}$ is the location of the jth point in

d', σ is the standard deviation of C_1. Similarly, the i,jth item of $Cov\left(d^1, d^1\right)$ is

$$\sigma(p_i^{d^1})\sigma(p_j^{d^1})\rho_{p_i^{d^1},p_j^{d^1}}.$$

Let's use the following notations:

$\sigma\left(p^{c_1}\right)$: a column vector of the standard deviation associated with all points in c_1.

$\sigma\left(p^{d'}\right)$: a column vector of the standard deviation associated with all points in d'.

$\rho_{p^{c_1},p^{d'}}$: the correlation coefficient matrix associated with all points in c_1 and d'.

$\rho_{p^{d'},p^{d'}}$: the correlation coefficient matrix associated with all points in d'.

$\odot$: elementwise matrix multiplication (the same as the .* in Matlab).

Then we have

$$Cov\left(c_1, d^1\right) = \left[\sigma\left(p^{c_1}\right) \cdot \sigma\left(p^{d'}\right)^T\right] \odot \rho_{p^{c_1},p^{d'}} \tag{5.17}$$

$$Cov\left(d^1, d^1\right) = \left[\sigma\left(p^{d'}\right) \cdot \sigma\left(p^{d'}\right)^T\right] \odot \rho_{p^{d'},p^{d'}}. \tag{5.18}$$

Optimization Problem Summary

In the scenario of the objective function being the SVP prediction uncertainty, the optimization problem can be expressed as:

$$\max tr\left\{\left[\sigma\left(p^{c_1}\right) \cdot \sigma\left(p^{d'}\right)^T\right] \odot \rho_{p^{c_1},p^{d'}} \left[\left[\sigma\left(p^{d'}\right) \cdot \sigma\left(p^{d'}\right)^T\right] \odot \rho_{p^{d'},p^{d'}} + R_\omega\right]^{-1} \left[\left[\sigma\left(p^{c_1}\right) \cdot \sigma\left(p^{d'}\right)^T\right] \odot \rho_{p^{c_1},p^{d'}}\right]^T\right\} \tag{5.19}$$

$$\text{s.t. } \{z_1, z_2, \cdots, z_n\} \text{ constitutes a feasible path.} \tag{5.20}$$

Please note that in Eq. 5.19, $p^{d'}$ is the only variable, which is implicitly but completely determined by the decision variable $\{z_1, z_2, \cdots, z_n\}$ via the OA grid. Therefore, this problem is a **non-linear deterministic optimization problem**.

5.3.2 Acoustic Prediction Uncertainty

In this section, we discuss the scenario in which the objective function is the posterior acoustic prediction uncertainty. Above all, we need to clarify the representation of the acoustic prediction uncertainty.

Representation of The Acoustic Prediction Uncertainty

Based on the philosophical presumptions made in this thesis, in AREA P — the sound pressure in the water column at a certain time and a certain location is a deterministic but partially known variable. Mathematically, the acoustic intensity $|P|^2$ and the phase ϕ can be modeled using random variables. In underwater acoustics, ϕ is often much more random and unpredictable than $|P|^2$ and hence in many implementations only the acoustic intensity is useful. In AREA only the $|P|^2$ is considered currently. **In this thesis, uncertainty of a random variable is defined as some statistic characterization, such as variance, of the random variable**. For the uncertainty of $|P|^2$, we can use either $-10log\left(var\left(|P|^2\right)\right)$ or $var\left(TL\right)$ to represent it. Note that TL is actually the $|P|^2$ in dB. Facing these two choices, one may ask which representation is more physically meaningful in implementation?

It is well known that human auditory perception to the sound intensity stimulus is logarithmic response [61]. That is to say if the acoustic intensity is doubled, the loudness will not be doubled, and the relation between loudness and acoustic intensity is in a logarithmic format. In the AREA project, the dB-budget of classical sonar systems is a very important concern [9]. From these two perspectives, $var\left(TL\right)$ reflects more about the underwater acoustic prediction uncertainty. Therefore, **in this thesis** $var\left(TL\right)$ **is defined as the acoustic prediction uncertainty**.

$var\left(TL\right) = E\left[TL^2\right] - E\left[TL\right]^2$. Someone may argue that $E\left[TL\right]$ is actually related to the geometric average of $|P|^2$ not the arithmetic average, thus it is not physically meaningful. While in this thesis, we think TL is more like a measure of the acoustic loudness to the sonar system not just the $|P|^2$ in dB. So TL should be treated as an independent physically meaningful variable and $var\left(TL\right)$ can be the indicator of how

accurate is the acoustic prediction with respect to the sonar system.

TL curves for underwater sound are usually calculated for harmonic sources. In contrast, most sonar systems operate over a spread of frequencies rather than one single frequency [62]. To take into account the frequency bandwidth, TL curves are usually smoothed through the method introduced in [62]. If f_0 is the central frequency of the sonar and α is the fractional bandwidth [62], Eq. 5.21 stands for the frequency-average intensity, where $I(f, r_0)$ is the acoustic intensity for a single frequency.

$$I_f = \frac{\int I(f, r_0) \exp\left[-(f - f_0)^2 / (\alpha f_0)^2\right] df}{\int \exp\left[-(f - f_0)^2 / (\alpha f_0)^2\right] df} \tag{5.21}$$

Compared with $I(f_0, r)$, I_f means more to sonar operations. However calculation for I_f will definitely take much longer time, since for each frequency the corresponding $I(f, r)$ must be computed. The idea to solve this problem is that a frequency average can often be approximated closely by a variable width running range average in which the width or window size is proportional to range.

$$I_r = \frac{\int I(f_0, r) \exp\left[-(r - r_0)^2 / (\alpha r_0)^2\right] dr}{\int \exp\left[-(r - r_0)^2 / (\alpha r_0)^2\right] dr} \tag{5.22}$$

In this thesis, we use range-average transmission loss $TL_r = -10log(I_r)$ to replace the original TL and let $var(TL_r)$ be the representation of the acoustic prediction uncertainty.

A priori TL_r Prediction Uncertainty

As mentioned before, the underwater acoustic field is usually highly nonlinearly related to SVP in the water. It implies that even if we assume c_1 is a Gaussian random vector, it is still very hard to calculate $var(TL_r)$ analytically. For such a problem, the most popular and perhaps the most efficient way is to do Monte Carlo simulations.

So once the a priori SVP prediction, which contains not only the P.E. of SVP but also many possible SVP realizations, is generated, the corresponding TL_r realizations can be simply computed for each SVP realization via RAM PE code. Thereafter,

the associated sample variance map can be obtained by statistics. Fig. 5-6 shows an example of the a priori TL_r prediction error variance map. The red point corresponds to the sound source location.

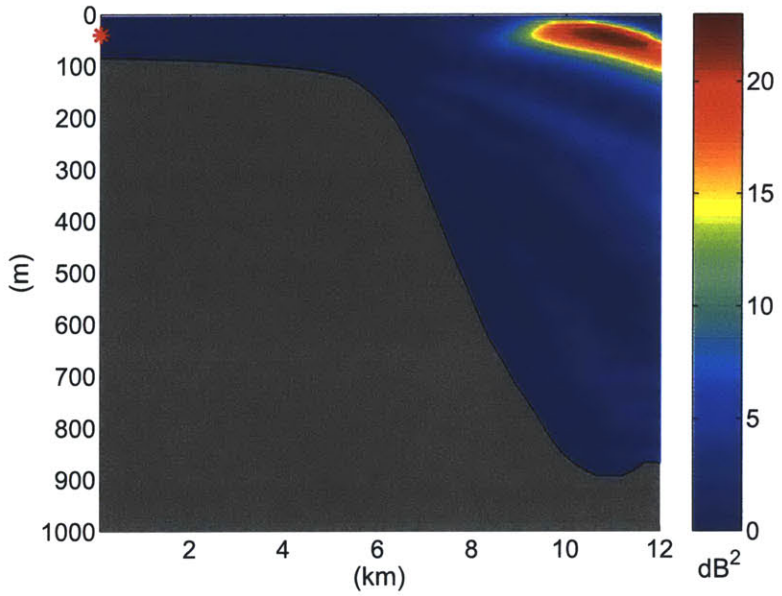


Figure 5-6: Illustration of an a priori TL_r prediction error variance. The red point corresponds to the sound source location. Frequency is 100Hz.

From the viewpoint of a sonar system, one objective of AREA is to minimize the TL_r prediction uncertainty at the location of the hydrophone for all possible sound source locations in the ocean area concerned. In Fig. 5-6, any point in the water from 0km to 12km in range and from 0m to 900m in depth could be the sound source location. So in this case we have to repeat running RAM PE code for each possible source location. However, the acoustic reciprocity theorem states that an acoustic response remains the same when the source and receiver are interchanged [1]. We can thus put the sound source at the location of the hydrophone and calculate TL_r for all depths in the water. In this way, the RAM PE code is needed to be run only once.

Here it should be pointed out that not all the locations in the water are equivalently significant to the sonar and some of them may not even need to be concerned. For example, those locations above 5m in depth or within 2km in range from the sonar

are less interesting to us. While for those depths below 400m, there may be no any submarine can reach.

So far, one may think that for the scenario in which the objective function is the acoustic prediction uncertainty, once we construct a connection between the AUV path and the posterior $var\,(TL_r)$, an objective function similar to Eq. 5.19 can be obtained and hence the optimization problem should be of the same type as Eq. 5.19 and 5.20. However, very unfortunately, this is not true.

For the case of SVP prediction uncertainty being the objective function, we emphasized that the objective function Eq. 5.19 only depends on the AUV path and not depends on the in-situ measurement values at all. This is because in this case only the OA equation Eq. 5.9 is needed, which has nothing to do with the real value of d'. While in the case of acoustic prediction uncertainty being the objective function, the process to calculate the posterior TL_r prediction uncertainty includes:

1. Do in-situ measurements.
2. Pre-process the in-situ measurements data. Generate d'.
3. Input d' and compute the posterior SVP estimation $\hat{c}\,(+)$ via Eq. 5.8.
4. Input $p^{d'}$ and compute the posterior SVP estimation error covariance $\Lambda_c\,(+)$ via Eq. 5.9.
5. Use $\hat{c}\,(+)$, $\Lambda_c\,(+)$, Lr, Lz and the Gaussian distribution assumption to bootstrap many SVP realizations.
6. For each SVP realization, run RAM PE code. Generate many TL_r realizations.
7. Calculate the posterior sample $var\,(TL_r)$.

In step 3, d' is used to generate $\hat{c}\,(+)$ and in step 5, $\hat{c}\,(+)$ is used to generate posterior SVP realizations. So finally the posterior $var\,(TL_r)$ is dependent on d' and therefore a paradox exists: the posterior $var\,(TL_r)$ can not be calculated prior to the in-situ measurements even if the AUV path is determined, but the posterior $var\,(TL_r)$ associated with a certain AUV path is necessary for solving the optimization problem,

which however should be prior to the in-situ measurements. The way to solve this paradox is by introducing the **stochastic optimization** concept, which is discussed in Chapter 4.3.

Now suppose that an AUV path is selected, then based on the a priori SVP prediction, the semi-empirical R_ω, the AUV speed and the CTD sampling frequency, the in-situ measurement data can be predicted. In implementations, the in-situ measurement data prediction is an ensemble including many realizations. For each realization, the associated posterior $var\left(TL_r\right)$ is obtainable. Therefore by doing sample average, $E\left[var\left(TL_r\right)\right]$ can be obtained. The $tr\left\{E\left[var\left(TL_r\right)\right]\right\}$ can thus be the objective function in the stochastic optimization problem. In one word, in the scenario of minimizing the posterior acoustic prediction uncertainty, the objective function is $tr\left\{E\left[var\left(TL_r\right)\right]\right\}$ rather than $tr\left\{var\left(TL_r\right)\right\}$, where the expectation is over all possible a priori SVP realizations and all possible measurement errors. Using the expectation value in the objective function implies that the optimized AUV path is only optimal in regard to the averaged result and it is usually not the real best one for the real situation.

A very similar case is the stock trading strategy. The optimized stock trading strategy can only lead to the highest return ratio on average. That is to say if you implement it in the stock market for infinite many times (and if the stock market is modeled accurately), the gross return ratio will be the highest; while if you implement it for only once, you may actually lose a lot of money.

So the stochastic optimization method can not guarantee the best result for every implementation in the scenario of minimizing the posterior acoustic prediction uncertainty, but anyway it is already the best we can do in the framework of static optimization.

Optimization Problem Summary

In the case of the acoustic prediction uncertainty being the objective function, the optimization problem can be expressed as:

$$\min tr\left\{E\left[var\left(TL_r\right)\right] \odot W\right\} \tag{5.23}$$

$$\text{s.t. } \{z_1, z_2, \cdots, z_n\} \text{ constitutes a feasible path.} \tag{5.24}$$

Where $var\left(TL_r\right)$ is the posterior TL_r prediction error covariance matrix, E is over all possible a priori SVP realizations and all possible measurement errors, W is the diagonal weight matrix

$$\begin{bmatrix} w_1 & & & \\ & w_2 & & \\ & & \ddots & \\ & & & w_n \end{bmatrix}.$$

w_i is the weight for the ith TL_r point.

$var\left(TL_r\right)$ implicitly depends on $\{z_1, z_2, \cdots, z_n\}$ and d'; while $E\left[var\left(TL_r\right)\right]$ only depends on $\{z_1, z_2, \cdots, z_n\}$. The objective function is highly non-linear with respect to $\{z_1, z_2, \cdots, z_n\}$ and strictly speaking, it can only be calculated via Monte Carlo simulations. This optimization problem is a **non-linear stochastic optimization problem**.

5.3.3 Sonar Performance Prediction Uncertainty and Sonar Performance

The long term goal of AREA is to best improve the sonar performance prediction for the non model-based sonar or best improve the sonar performance for the model-based sonar. However, objective function with respect to sonar performance prediction uncertainty (e.g. the uncertainty of sonar range) or sonar performance metric (e.g. the displacement of MFP localization) is generally based on the objective function in Eq. 5.23, and usually much more complicated and non-linear. On current PCs,

even the computation for Eq. 5.23 is rather time-consuming due to the Monte Carlo simulations and thus the real-time feasibility of AREA can only be reached barely. Therefore, in this thesis the objective function with respect to sonar performance prediction uncertainty or sonar performance metric is not considered at present.

However the posterior TL_r prediction uncertainty is a very fundamental and very important measure for the sonar performance prediction uncertainty and sonar performance metric. Loosely speaking, if $tr\{E[var(TL_r)] \odot W\}$ is less, then the sonar performance prediction uncertainty is often less and the sonar performance is often better. To some extent, TL_r can be viewed as a substitute.

5.4 The Approximate Objective Function

For acoustic prediction uncertainty, computing $var(TL_r)$ needs to run Monte Carlo simulation many times, which is fairly slow and is one of the bottlenecks of the real-time feasibility of AREA. Thus, any good approximate and quick way is highly desirable. If such a kind of method can be found, the optimization process can be accelerated a lot and a good sub-optimal solution may be obtained in real time.

5.4.1 Linear Approximation for Transmission Loss

It is well known that the relation between TL and water column sound velocities is highly nonlinear (Eq. 5.25). However, based on the Taylor series expansion, when Δc is small enough, Eq. 5.25 can be linearly approximated by Eq. 5.28.

$$TL = f(c), \tag{5.25}$$

$$c = c_0 + \Delta c, \tag{5.26}$$

$$TL = TL_0 + \Delta TL, \tag{5.27}$$

$$\approx f(c_0) + A \cdot \Delta c, \tag{5.28}$$

where $TL_0 = f(c_0)$ and TL, c are all vectors, A is a matrix. In the real ocean, $\Delta c/c_0$ is usually not higher than $1/100$, so a conjecture can be made that in most

scenarios in AREA, Eq. 5.28 could be a good approximation. If this conjecture is true, then with the assumption that the SVP prediction error is a Gaussian random vector, $var\,(TL)$ can be calculated very quickly by Eq. 5.29,

$$var\,(TL) \approx A \cdot var\,(\Delta c) \cdot A^T. \tag{5.29}$$

If Eq. 5.29 can substitute for Monte Carlo simulations in AREA, the bottleneck is resolved. In next sections, we are going to investigate the conjecture numerically and analytically.

5.4.2 Examples

In this section, two typical examples in AREA are presented and the associated linear TL models are investigated numerically. Both the two examples are located at the shelf break at the Monterey bay, CA.

Example 1: Cross the Shelf Break

Fig. 5-7(a) shows part of the topography of the Monterey bay, CA. The green point is supposed to be the sound source location and in this example the bearing 5 is considered, which crosses the shelf break and is often associated with big uncertainties in SVP prediction. Fig. 5-7(b), 5-7(c), 5-7(d) are respectively the SVP, density profile, attenuation coefficient profile in the seabed. Fig. 5-7(e) is an example of the a priori SVP principal estimation generated by HOPS. The corresponding error standard deviation map is shown in Fig. 5-7(f) and the correlation lengths are $Lr = 2500\ m$ and $Lz = 3\ m$.

TL_0 — the TL field associated with Fig. 5-7(e) is shown in Fig. 5-8(a), where the 100Hz single frequency sound source is located at (0 km, 15m) and only the depths above 300 m are considered. In this thesis, the coefficient matrix A is calculated by finite difference method:

$$A_{i,j} = \frac{f_i\,(c_0 + \Delta c^j) - f_i\,(c_0 - \Delta c^j)}{2\sigma_j}, \tag{5.30}$$

where $A_{i,j}$ is the i,j th item in A; σ_j is the sound velocity prediction error standard deviation at the j th SVP grid point; $\Delta c_k^j = \delta_{j,k} \cdot \sigma_j$, which is equal to 0 except at the j th SVP grid point ($k = j$); f_i is the i th TL points. To compute A by Eq. 5.30, the RAM PE code should be run twice for each SVP grid point. In practice, this process usually takes about 30 minutes. Another popular way to compute A is the *Linear Least Square Fitting* method. But in practice in AREA, the overfitting problem always happens, since we don't have enough time to generate enough training data.

Once the TL_0 and A are known, Eq. 5.28 can be used to approximate the TL. Fig. 5-8(b) shows TLs associated with 200 independent SVP realizations which are generated on the basis of Fig. 5-7(e) and 5-7(f). The receiver depth is 165 m. Those black curves in the upper plot are from RAM PE code, while the blue curves in the lower plot are from linear approximation. The red curve is the corresponding TL_0 at this depth. By comparing these two plots, it can be seen that the linear approximation has pretty good accuracy at most ranges except at those TL nodes, where the linear approximate TLs may be really bad and far away from the true ones or even be negative numbers. The sample variances of the black and blue curves at all ranges are shown in Fig. 5-8(d), from where the same phenomenon can be observed that the sample variance difference is pretty small at most ranges but at those nodes it can be really huge. At all receiver depths, the same thing happens, while the depth of 165 m corresponds to the worst situation.

In this example the linearity conjecture is partially proved and also partially denied. In fact, in this case the linear approximate $tr\,(var\,(TL))$ is equal to 16321.8 (dB^2), while 200 Monte Carlo simulations give $tr\,(var\,(TL))$ equal to 7240.1 (dB^2). The relative error is about 125%. The convergence test of Monte Carlo simulations is shown in Fig. 5-8(c), which shows that in this example 200 Monte Carlo simulations are enough. From the above results, we can see that although the linearity is maintained at most ranges, the high non-linearity at nodes will definitely deteriorate the accuracy of Eq. 5.29. The first way to solve this problem is to remove those nodes from consideration. This is very reasonable, since TLs at nodes don't mean much in practice.

However, it may not be so practical, since locations of those nodes must be determined prior. The second way is to replace TL with the range-averaged transmission loss TL_r, by which the TL is smoothed and nodes are mitigated. The test is shown in Fig. 5-11.

The relation between linearity and the magnitude of the SVP prediction uncertainty is investigated and the result is shown in Fig. 5-9. Instead of using the original SVP prediction uncertainty, we increase or decrease the SVP errors in Fig. 5-7(f) by multiplying it with different multipliers but keep the Lr and Lz unchanged. The results from 200 Monte Carlo simulations are compared with the corresponding linear approximation results in Fig. 5-9(a). The blue line is associated with the linear approximation. The black line is associated with Monte Carlo simulations. The red line indicates the relative error. These two plots clearly shows that as increase the SVP prediction uncertainty, the linearity between ΔTL and Δc gets worse. This is consistent with the characteristics of the Taylor series expansion. The changes in linearity can also be observed from Fig. 5-9(b), 5-9(c), 5-9(d), where it can be seen that the linearity around the TL nodes is deteriorated very quickly as the multiplier increases.

Although in this example, Eq. 5.29 can't provide good accuracy, it is found that the linear approximation result is strongly positively correlated with the Monte Carlo simulation result, i.e. if the $tr\left(var\left(TL\right)\right)$ from Monte Carlo simulations is decreased or increased, the corresponding linear approximation value is very possibly decreased or increased too. Fig. 5-9(e) shows a preliminary test, in which the SVP prediction error was randomly adjusted for 23 times, the corresponding Monte Carlo simulations' results and linear approximation results were generated. A positive correlation can be seen. This implies that if in the objective function, $tr\left(var\left(TL\right)\right)$ is replaced by Eq. 5.29, the optimal solution of the new optimization problem will be a good sub-optimal solution to the original one. Here, one thing should be noticed that **on a 3.80GHz Intel Pentium 4 CPU, the 200 Monte Carlo simulations take almost 10 minutes, while Eq. 5.29 only takes about 2 seconds**. Therefore, it implies that in AREA, it is possible to solve the approximate optimization problem

in real time to get a good sub-optimal solution for the AUV path planning problem.

The effects of correlation lengths on the TL linearity are investigated in Fig 5-10. In Fig. 5-10(a), L means the $tr\,(var\,(TL))$ value from linear approximation; MC means the value from 200 Monte Carlo simulations; err is the error between L and MC, and Re is the relative error. From this table, it can be seen that Lr and Lz do affect the relative error, but in a complicated and coupled way.

Example 1 continue: using TL_r

Now let's replace the original TL with the range-averaged transmission loss TL_r and set $\alpha = 0.1$. The results are shown in Fig. 5-11. Fig. 5-11(a) and Fig. 5-11(b) clearly show that nodes of TL are smoothed a lot and hence the linear approximation around those points gets much better. In this case, the receiver depth of 75 m corresponds to the worst situation. Compared with Fig. 5-8(d), the sample variance difference in Fig. 5-11(d) is dramatically decreased. In this case we can say that the linearity conjecture is proved. In fact, the linear approximate gives $tr(var(TL))$ equal to 780.437 (dB^2), while 300 Monte Carlo simulations give it equal to 784.2822 (dB^2). The relative error is about only 0.49%. The convergence test of Monte Carlo simulations is shown in Fig. 5-11(c), which shows that in this case the convergence is slower and 300 Monte Carlo simulations are needed. The correlation test is shown in Fig. 5-11(e). The Lr, Lz effects on the linearity are shown in Fig. 5-12(a). In this case, when the Lz is very large, the linearity gets much worse and the correlation between the linear approximation result and the Monte Carlo simulation result is less strong.

Example 2: Parallel to the Shelf Break

In the second example, the bearing 4 is considered, which is parallel to the shelf break as shown in Fig. 5-13(a). In this case, let's first suppose the frequency of the sound source is 100Hz and it's located at $(0\ km, 80\ m)$. The correlation lengths are $Lr = 2000\ m$ and $Lz = 3.5\ m$.

From Fig. 5-16(a) and 5-14(b), it can be seen that in this example, due to the

environment and the sound source depth, more normal modes exist. Therefore, the spatial interference is more complicated and more TL nodes exist. Fig. 5-14(b) and Fig. 5-14(d) show that the linear approximation is not good in this case, but from Fig. 5-14(c) the strong positive correlation between the linear approximation result and the Monte Carlos simulation result is still held. Furthermore, in this example, the number of Monte Carlo simulations is 300, which is enough to converge.

Fig. 5-15 shows the range-averaged transmission loss scenario. Again, the linearity at those TL nodes gets better and the strong positive correlation is still held. In this case, $tr\left(var\left(TL\right)\right)$ from the linear approximation is equal to 116.959 (dB^2). 300 Monte Carlo simulations give $tr\left(var\left(TL\right)\right)$ equal to 135.464 (dB^2). The relative error is about 13.66%.

So far, we only considered 100Hz sound source and a preliminary conclusion was made that the linear approximation can be used to accelerate the objective function calculation. Now, let's suppose the sound source frequency is 400Hz and check the TL linearity again. The sound source location is still $(0\ km, 80\ m)$. Fig. 5-16(a) and 5-16(b) show that there are much more normal modes and the spatial interference is messed up much more. As a results, the TL linearity is much worse as the frequency is higher but Fig. 5-16(c) still shows the strong positive correlation.

With $\alpha = 0.1$, the TL_r case is shown in Fig. 5-17. By smoothing TL, the linearity is better. But comparing with the 100Hz cases, it can be seen that increasing sound frequency deteriorates the TL_r linearity. At those TL_r nodes, linear approximation leads to over-estimated result. However once again, the strong positive correlation between the linear approximation result and the Monte Carlo simulation result is still held.

5.4.3 Ideal Waveguide

In this section, the ideal waveguide scenario is considered and the analytical formula of TL will be discussed. Although the ideal waveguide doesn't exist in the real ocean, many properties from the ideal waveguide will carry through to more general ocean waveguide [1].

In the ideal waveguide, the acoustic intensity is given by the following equation:

$$I(r,z) = \frac{8\pi}{rD^2}\left[\sum_m A_m^2 + \sum_m \sum_{n>m} 2A_m A_n cos\left(k_{mn} r\right)\right] \tag{5.31}$$

where,

$$k_{mn} = k_{rm} - k_{rn} \tag{5.32}$$

$$A_m = \frac{sin\left(k_{zm} z_s\right) sin\left(k_{zm} z\right)}{\sqrt{k_{rm}}} \tag{5.33}$$

$$k_{zm} = \frac{(m - 1/2)\,\pi}{D}, \quad m = 1, 2, \cdots \tag{5.34}$$

$$k_{rm} = \sqrt{\left(\frac{2\pi f}{c}\right)^2 - \left[(m - 1/2)\,\pi\right]^2}. \tag{5.35}$$

If there exists small sound velocity variations $\Delta c(r,z)$, based on the adiabatic theorem [63] Eq. 5.32 can be rewritten as

$$I(r,z) = \frac{8\pi}{rD^2}\left[\sum_m A_m^2 + \sum_m \sum_{n>m} 2A_m A_n cos\left(k_{mn} r + \Delta k_{rm}(r) - \Delta k_{rn}(r)\right)\right] \tag{5.36}$$

$$\Delta k_{ri}(r) = -\frac{1}{k_{ri}} \int_0^r \int_0^D \frac{sin^2\left(k_{zi} z\right)\left(\frac{2\pi f}{c}\right)^2}{c} \Delta c(r,z)\, dz\, dr. \tag{5.37}$$

In far field, we only consider those modes with real k_{rm} value. If the water depth D and sound velocity c are given, and if now the sound frequency f is so low that only 1 normal mode exists, then in far field we have

$$I(r,z) = \frac{8\pi}{rD^2} A_1^2. \tag{5.38}$$

This equation implies that $\Delta c(r,z)$ doesn't change the acoustic intensity and TL field.

If now the sound frequency is increased so that 2 normal modes exist, then Eq. 5.32 is like

$$I(r,z) = \frac{8\pi}{rD^2}\left[A_1^2 + A_2^2 + 2A_1 A_2\, cos\left(k_{12} r + \Delta k_{r1}(r) - \Delta k_{r2}(r)\right)\right]. \tag{5.39}$$

Let's write

$$A = \frac{8\pi}{rD^2}\left[A_1^2 + A_2^2 + 2A_1A_2\, cos\,(k_{12}r)\right], \tag{5.40}$$

$$\Delta B = \frac{8\pi}{rD^2}\left[2A_1A_2 cos\,(k_{12}r + \Delta k_{r1}(r) - \Delta k_{r2}(r))\right] - \frac{8\pi}{rD^2}\left[2A_1A_2\, cos\,(k_{12}r)\right], \tag{5.41}$$

$$\text{then } TL = -10log_{10}(A + \Delta B). \tag{5.42}$$

By Taylor series expansion, we have

$$TL \approx -10log_{10}A - 10\frac{\Delta B}{A} + 10\frac{\Delta B^2}{A^2} + o(\Delta B^2). \tag{5.43}$$

If $A >> \Delta B$, then linearity of TL with respect to ΔB will be very good. However, because of the modes interference, the minimum value of A could be very close to 0. Those points just correspond to TL nodes, where TL is very non-linear with respect to ΔB. In the 2-mode case, ΔB is connected to $\Delta k_{ri}(r)$ through a cosine function, which is not linear. While from Eq. 5.37, $\Delta k_{ri}(r)$ is a linear function of $\Delta c(r, z)$. In practice it is found that most non-linearity of TL with respect to $\Delta c(r, z)$ is still from the logarithm function. Therefore, the TL uncertainty estimation from the linear approximation method is always far away from the truth around those TL nodes. In multi-mode scenarios, the same thing happens but the modes interference is more complicated and TL has more nodes. This explains why for higher frequency, the linear approximation method gives worse estimation.

For TL_r at a point, the associated A is averaged in the neighborhood and since A is always nonnegative, thus A_r — the range-averaged A is less close to 0. At TL_r nodes the linearity is better than that at TL nodes. This explains why the TL_r uncertainty estimation from the linear approximation method is always better than that of TL. In addition, Eq. 5.40 shows that if $A_1 >> A_2$, at TL nodes A won't be very close to 0, so linear approximation can give better estimation. While the values of A_1 and A_2 are dependent on modes' shapes and source depth, receiver depth. Eq. 5.37 shows that the correlation lengths Lr and Lz will influence $\Delta k_{ri}(r)$, but the effect is coupled

with mode's shape and receiver depth. If r is now very large, it could be expected that some $\Delta k_{ri}(r) - \Delta k_{rj}(r)$ will be saturated, i.e. bigger than 2π. In such kind of situation, linear approximation method will not be valid, however in this thesis r is not so large and is usually about $10 \sim 15\ km$.

5.4.4 Linear Approximation Summary

From the above analysis and preceding numerical examples, it seems that the linear approximation method can provide a good and quick estimation for the TL_r uncertainty in low frequency scenarios in AREA. While it is also found that TL and TL_r uncertainty estimation from the linear approximation method is strongly correlated with the Monte Carlo simulation result. So generally speaking, in AREA linear approximation method is a very good and practical way to estimate the acoustic prediction uncertainty. It could be hundreds times faster than Monte Carlo simulation method. In fact, when the Error Subspace Statistical Estimation (ESSE) method is used to estimate the acoustic field uncertainty directly from the ocean environment prediction uncertainty, the above linear approximation method is implicitly adopted [6, 64].

Once the $var\,(TL_r)$ can be quickly estimated, the next problem in Eq. 5.23 is how to quickly estimate the expectation, which is taken over all possible a priori SVP realizations and all possible measurement errors. For each SVP realization and each measurement error, the posterior SVP estimation is usually different and hence the associated coefficient matrix A in Eq. 5.28 is different. In practice, it's not feasible to run A for several different SVP realizations and measurement errors and then run the sample mean, since running A once will take about 30 minutes. **The most practical way here is to use the A associated with the SVP principal estimation and zero measurement error to replace the expectation**. Again, this is an approximation but this is the best we can do so far.

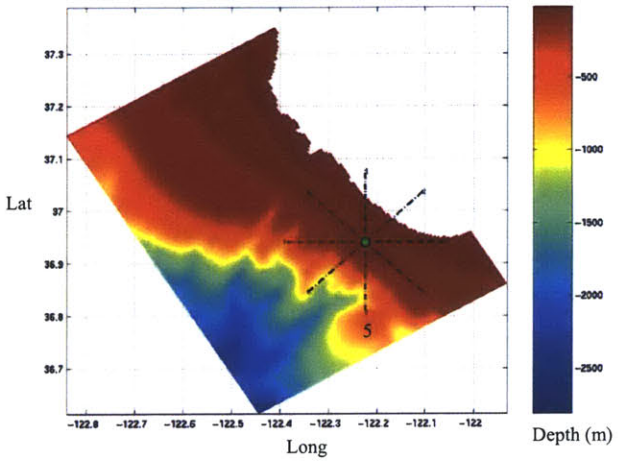
5.4.5 Approximate Optimization Problem Summary

Using the linear approximation of TL_r and replacing the expectation operator with the SVP P.E. scenario, the optimization problem for acoustic purpose can be expressed as:

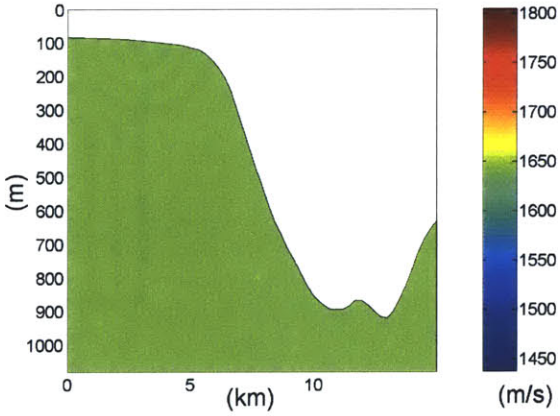
$$\max tr\left\{A\cdot\left[\sigma\left(p^{c_1}\right)\cdot\sigma\left(p^{d'}\right)^T\right]\odot\rho_{p^{c_1},p^{d'}}\left[\left[\sigma\left(p^{d'}\right)\cdot\sigma\left(p^{d'}\right)^T\right]\odot\rho_{p^{d'},p^{d'}}+R_\omega\right]^{-1}\left[\left[\sigma\left(p^{c_1}\right)\cdot\sigma\left(p^{d'}\right)^T\right]\odot\rho_{p^{c_1},p^{d'}}\right]^T\cdot A^T\right\} \tag{5.44}$$

$$\text{s.t. } \{z_1,z_2,\cdots,z_n\} \text{ constitutes a feasible path.} \tag{5.45}$$

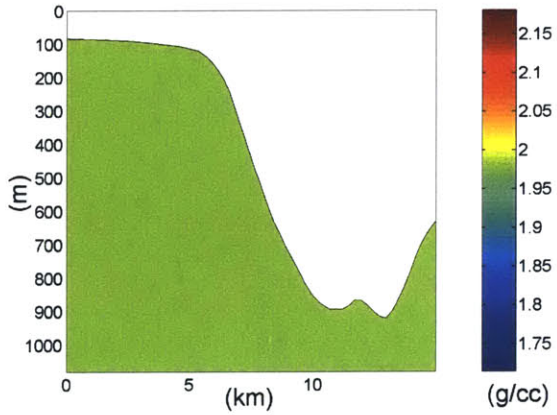
Please note that Eq. 5.44 is very similar to Eq. 5.19 except the multiplication of the matrix A, which is now more like individual significance weights for SVP grid points. Eq. 5.44 means that in linear approximation, we only consider the individual significance of each sound velocity in the water column but ignore the coupling effect among them.



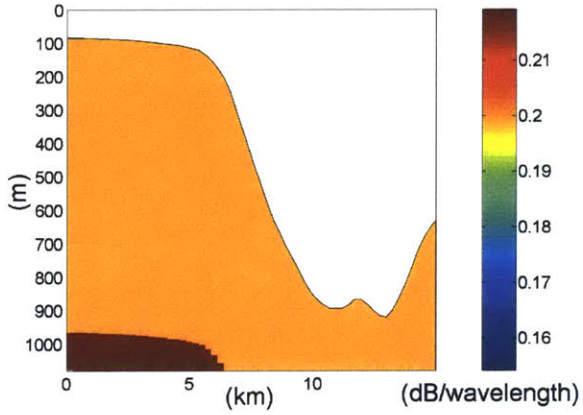
(a) Topography of Monterey bay, CA



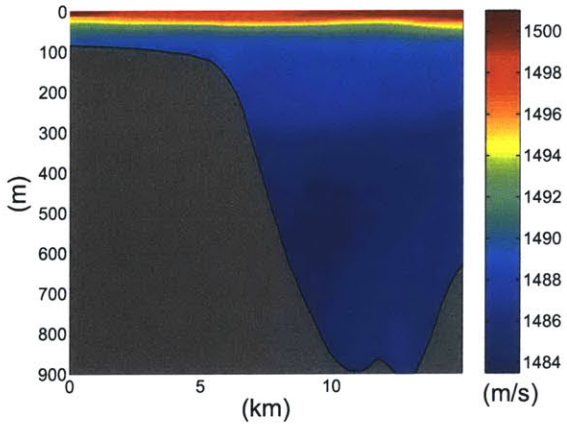
(b) SVP of seabed



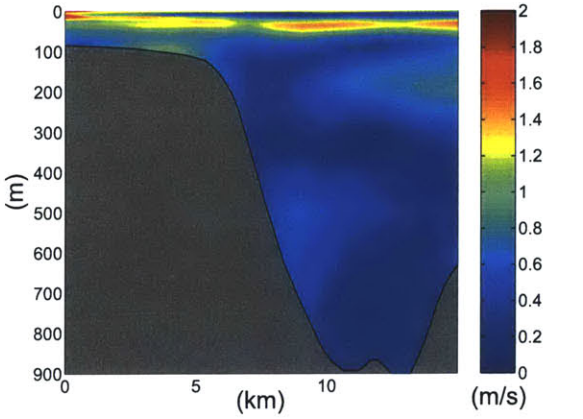
(c) Density profile of seabed



(d) Attenuation coefficient profile of seabed

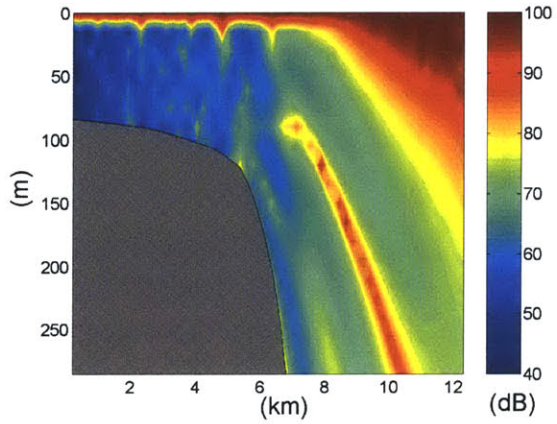


(e) Water column SVP principal estimation

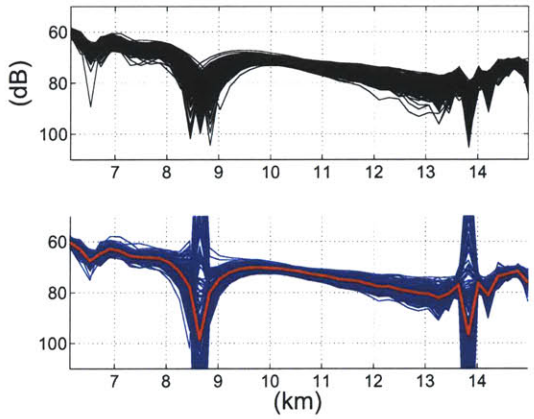


(f) Standard deviation of the SVP estimation error

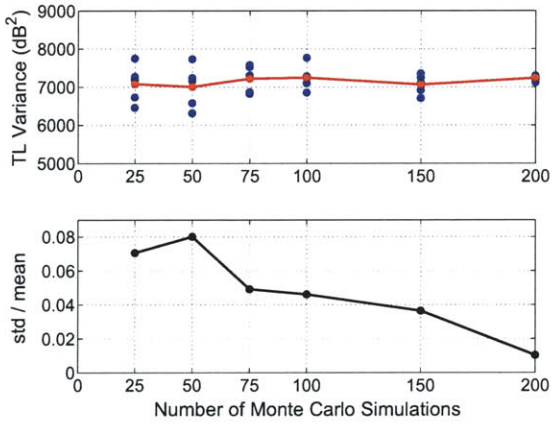
Figure 5-7: Profile of the bearing 5: cross the shelf break. $Lr = 2500\ m$ and $Lz = 3\ m$.



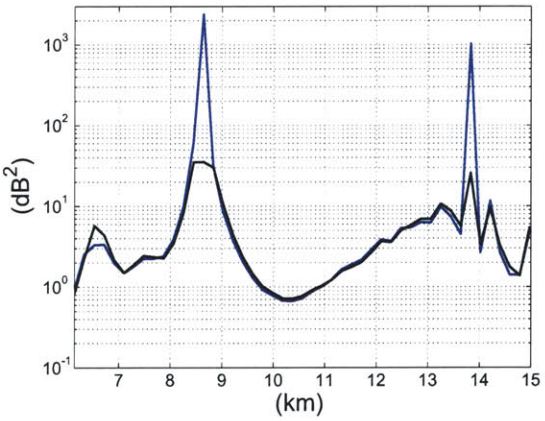
(a) TL_0



(b) TLs comparison at the depth of 165m

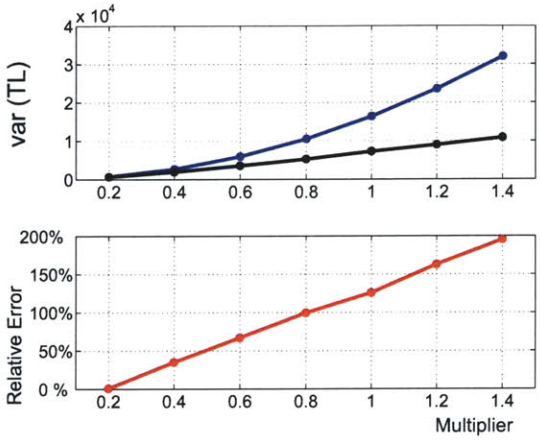


(c) Convergence test of Monte Carlo simulations

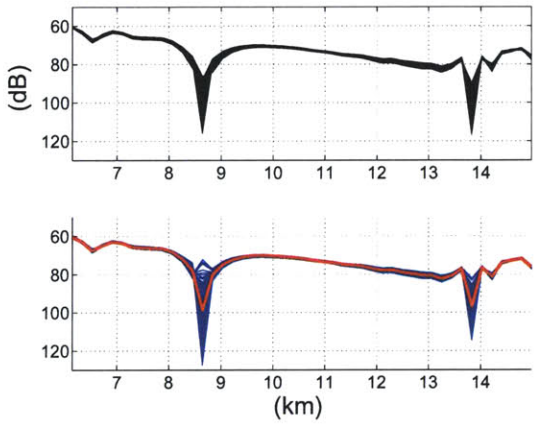


(d) TL sample variance comparison at the depth of 165m

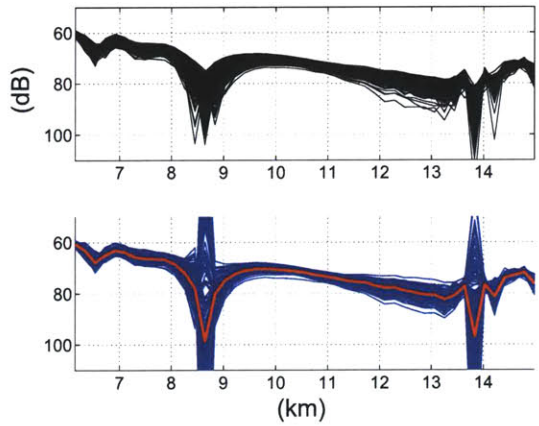
Figure 5-8: The 100Hz sound source is located at $(0\ km, 15m)$. $tr\,(var\,(TL))$ from the linear approximation is equal to 16321.8 (dB^2). 200 Monte Carlo simulations give $tr\,(var\,(TL))$ equal to 7240.1 (dB^2). The relative error is about 125%.



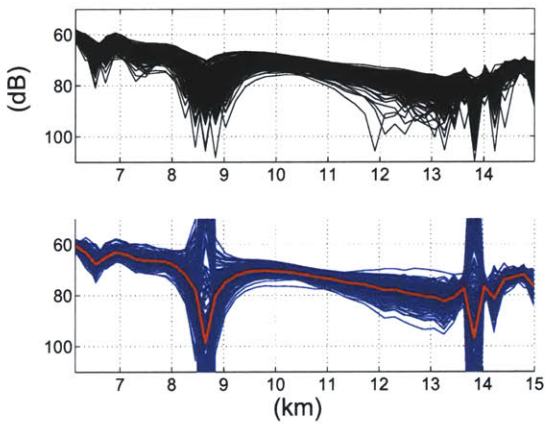
(a) Approximation error v.s. multiplier



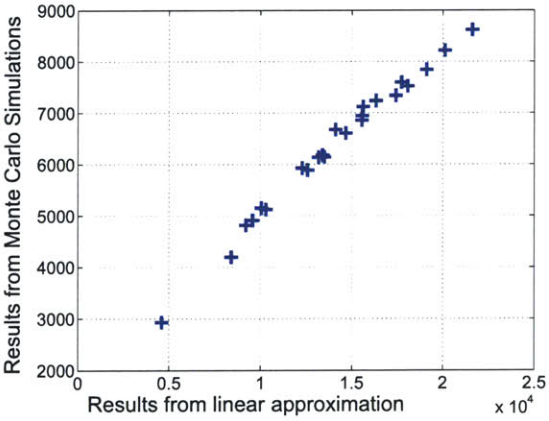
(b) Multiplier=0.2, Receiver Depth=165m



(c) Multiplier=0.8, Receiver Depth=165m



(d) Multiplier=1.4, Receiver Depth=165m

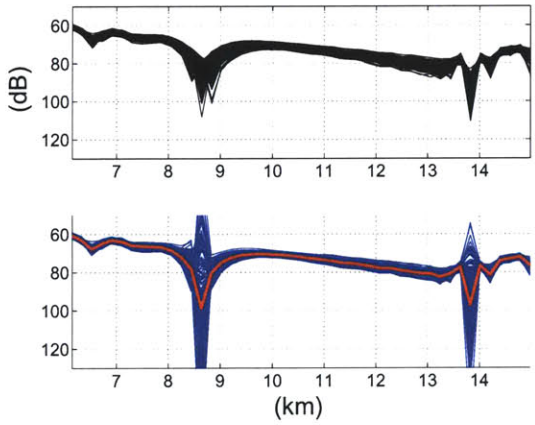


(e) Correlation test

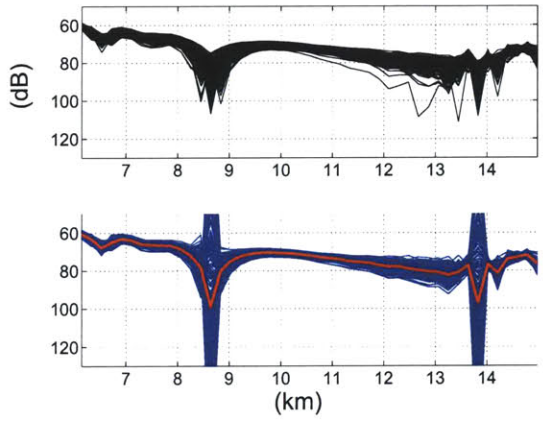
Figure 5-9: (a) is to test the approximation error v.s. the magnitude of SVP prediction uncertainty. (b), (c), (d) are associated with different SVP prediction uncertainty magnitudes. (e) shows the correlation between the linear approximation result and the Monte Carlo simulation result.

	Lr=25m	Lr=2500m	Lr=25000m
Lz=0.1m	L: 9451.09 MC: 5005.06 err=4446.0 Re=88.83%	L: 14687.1 MC: 6550.21 err=8136.9 Re=124.22%	L: 15954 MC: 6224 err=9730 Re=156.33%
Lz=3m	L: 10223.9 MC: 5442.16 err=4781.7 Re=87.9%	L: 16321.8 MC: 7240.1 err=9081.7 Re=125.44%	L: 17927.4 MC: 7091.17 err=1083.6 Re=152.8%
Lz=1000m	L: 5311.69 MC: 1848.91 err=3462.8 Re=187.29%	L: 7448.3 MC: 2969.48 err=4478.8 Re=150.8%	L: 7441.2 MC: 3559.85 err=3881.3 Re=109%

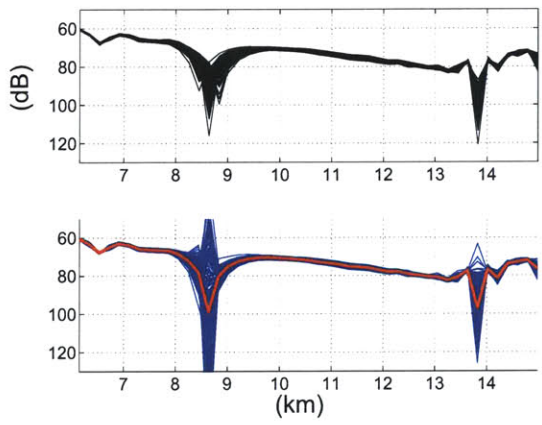
(a) Lr, Lz effects on the TL linearity. L: the linear approximation result; MC: the result from 200 Monte Carlo simulations; err: error; Re: relative error.



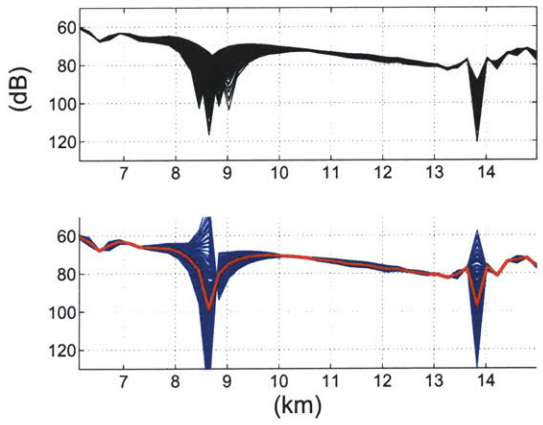
(b) $Lr = 25m$, $Lz = 0.1m$, Receiver Depth=165m



(c) $Lr = 25000m$, $Lz = 0.1m$, Receiver Depth=165m

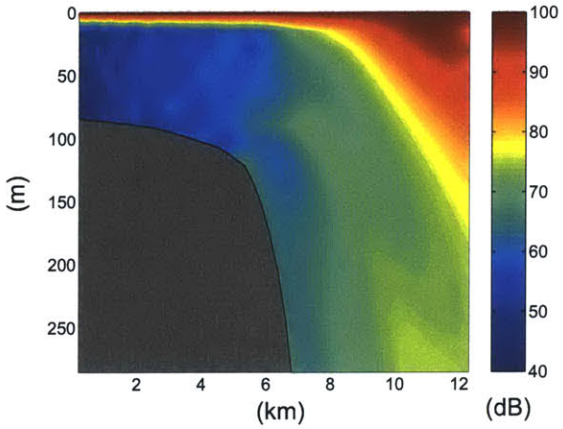


(d) $Lr = 25m$, $Lz = 1000m$, Receiver Depth=165m

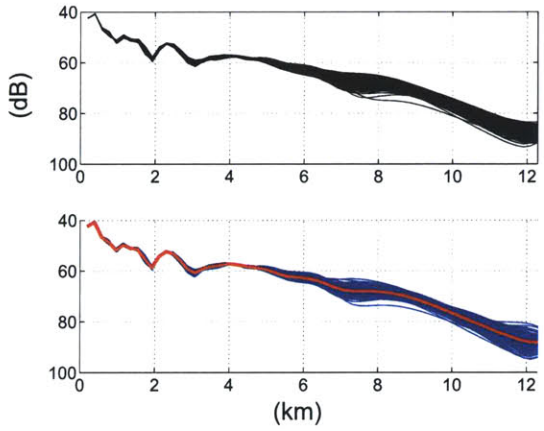


(e) $Lr = 25000m$, $Lz = 1000m$, Receiver Depth=165m

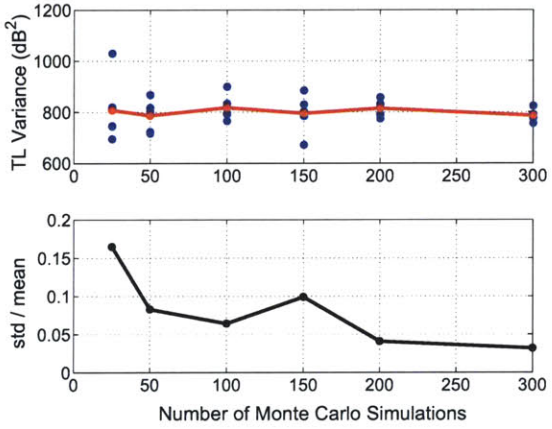
Figure 5-10: The 100Hz sound source is located at $(0\ km, 15m)$.



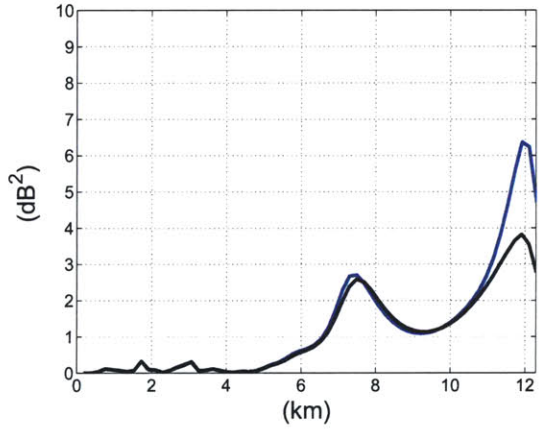
(a) TL_{r0}



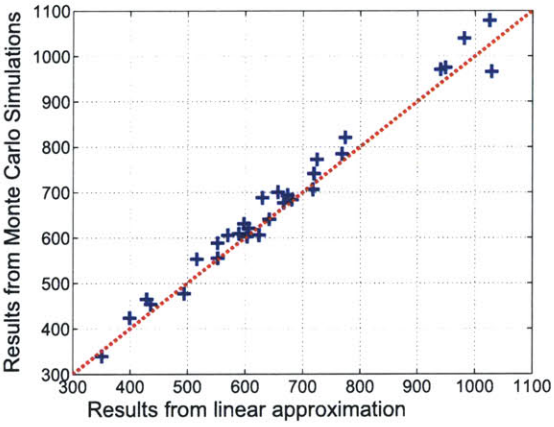
(b) TLs comparison at the depth of 75m



(c) Convergence test of Monte Carlo simulations



(d) TL sample variance comparison at the depth of 75m

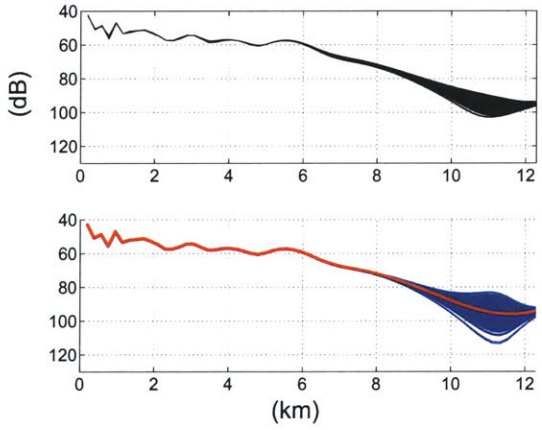


(e) Correlation test

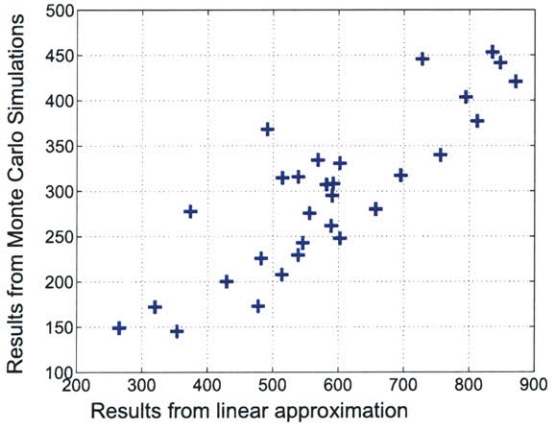
Figure 5-11: Using TL_r, $\alpha = 0.1$. The linear approximate gives tr (var (TL)) equal to 780.437 (dB^2), while 300 Monte Carlo simulations give it equal to 784.2822 (dB^2). The relative error is only about 0.49%.

	Lr=25m	Lr=2500m	Lr=25000m
Lz=0.1m	L: 405.8 MC: 393.8 err=12 Re=3.05%	L: 697.617 MC: 684.989 err=12.6 Re=1.84%	L: 724.819 MC: 622.707 err=102.11 Re=16.4%
Lz=3m	L: 451.88 MC: 452.532 err=0.652 Re=0.14%	L: 780.437 MC: 784.2822 err=3.8452 Re=0.49%	L: 809.883 MC: 797.39 err=12.5 Re=1.57%
Lz=1000m	L: 287.321 MC: 174.679 err=112.64 Re=64.49%	L: 660.372 MC: 381.869 err=278.5 Re=72.93%	L: 677.795 MC: 340.807 err=336.988 Re=98.88%

(a) Lr, Lz effects on the TL linearity. L: the linear approximation result; MC: the result from 300 Monte Carlo simulations; err: error; Re: relative error.

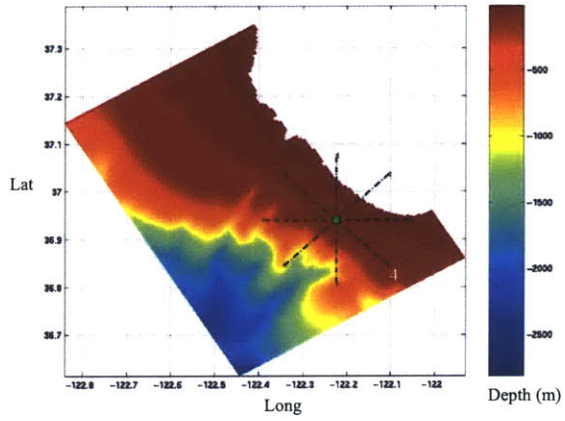


(b) $Lr = 25\ km$, $Lz = 1\ km$. TLs comparison at the depth of 45m

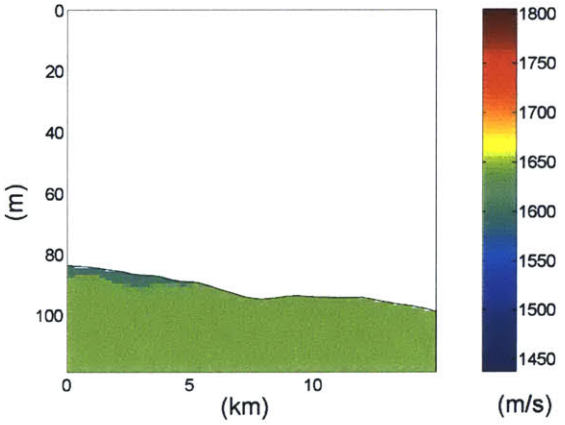


(c) $Lr = 25\ km$, $Lz = 1\ km$. Correlation test

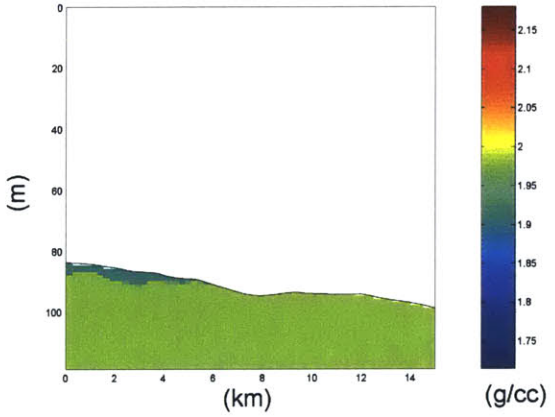
Figure 5-12: The 100Hz sound source is located at $(0\ km, 15m)$. TL_r is used in this case and $\alpha = 0.1$.



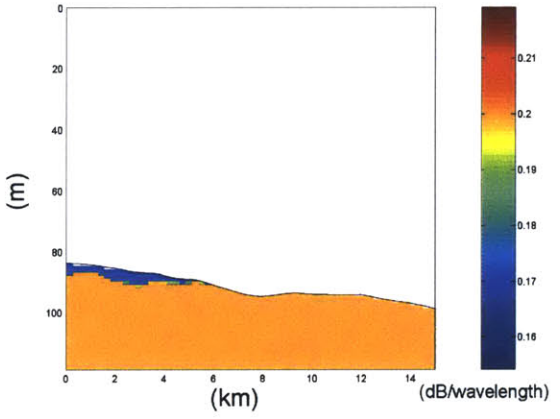
(a) Topography of Monterey bay, CA



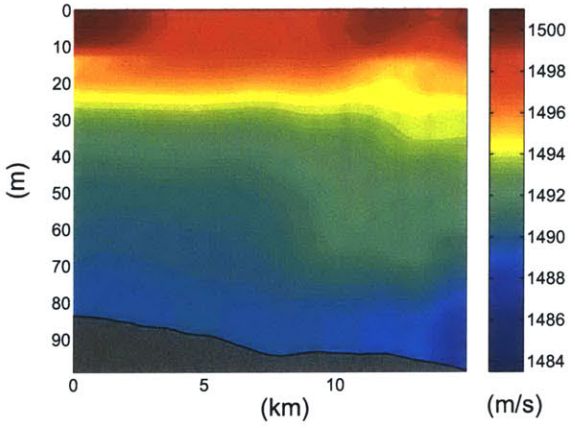
(b) SVP of seabed



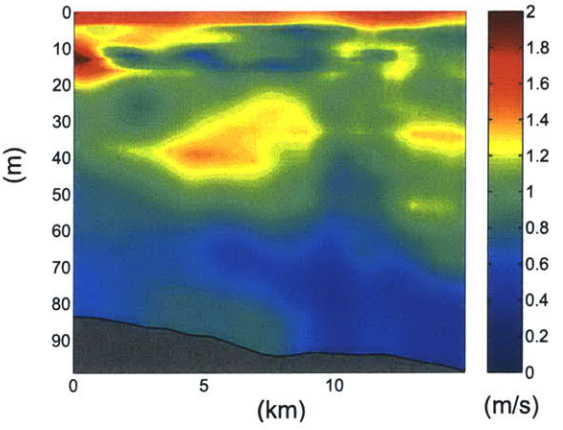
(c) Density profile of seabed



(d) Attenuation coefficient profile of seabed

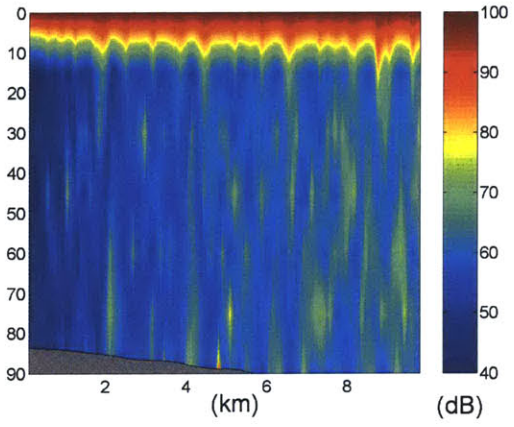


(e) Water column SVP principal estimation

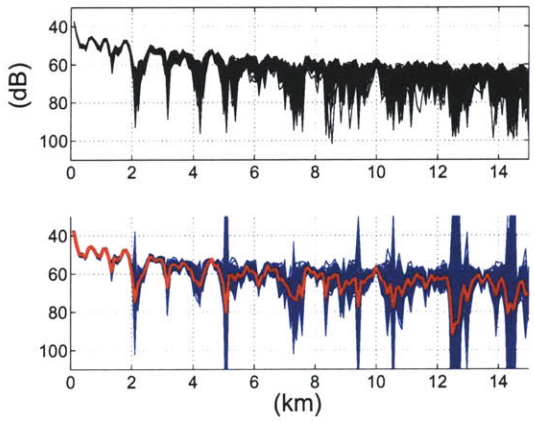


(f) Standard deviation of the SVP estimation error

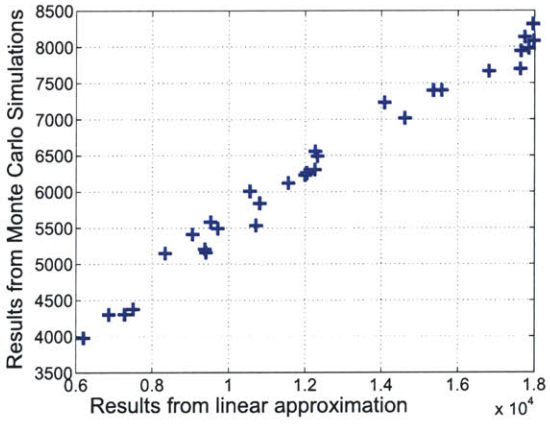
Figure 5-13: Profile of the bearing 4: parallel to the shelf break. $Lr = 2000\ m$ and $Lz = 3.5\ m$.



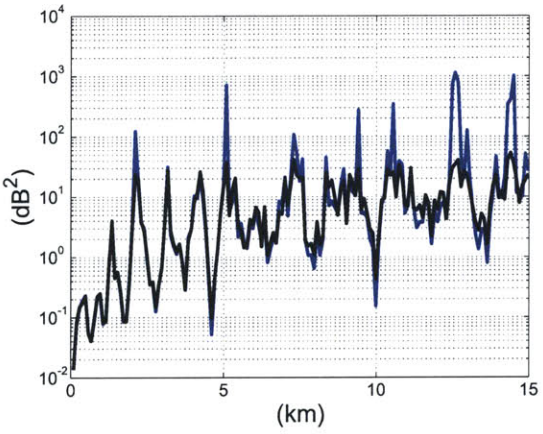
(a) TL_0



(b) TLs comparison at the depth of 75m

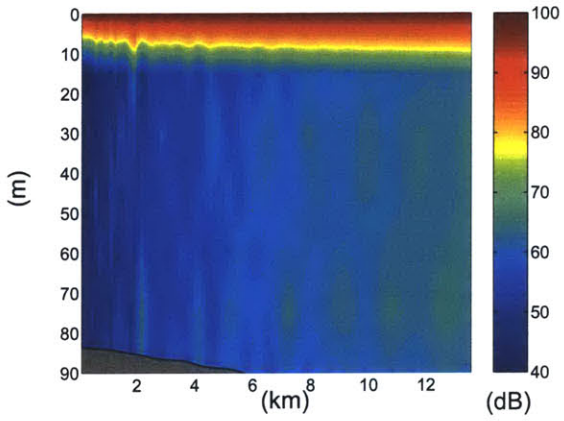


(c) Correlation test

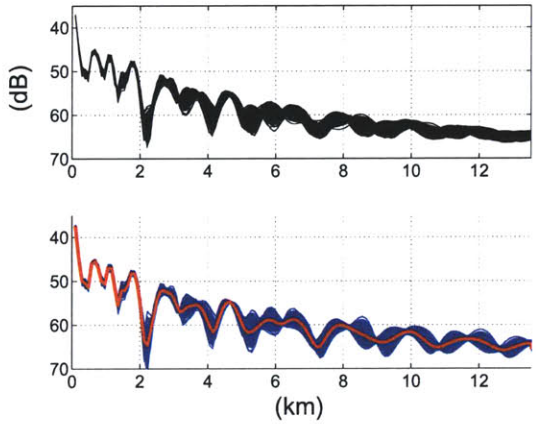


(d) TL sample variance comparison at the depth of 75m

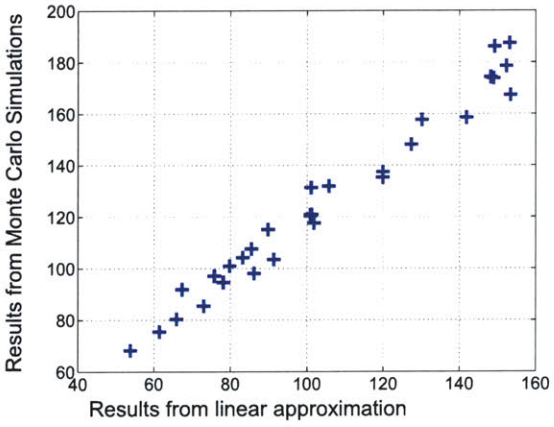
Figure 5-14: The 100Hz sound source is located at $(0\ km, 80m)$. $tr\left(var\left(TL\right)\right)$ from the linear approximation is equal to 13889.3 (dB^2). 300 Monte Carlo simulations give $tr\left(var\left(TL\right)\right)$ equal to 6813.9 (dB^2). The relative error is about 103.8%.



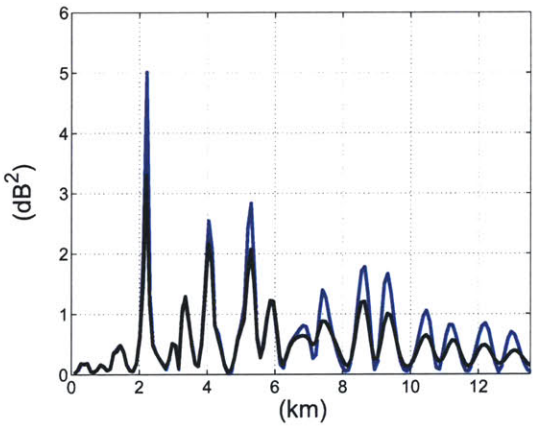
(a) TL_0



(b) TLs comparison at the depth of 75m

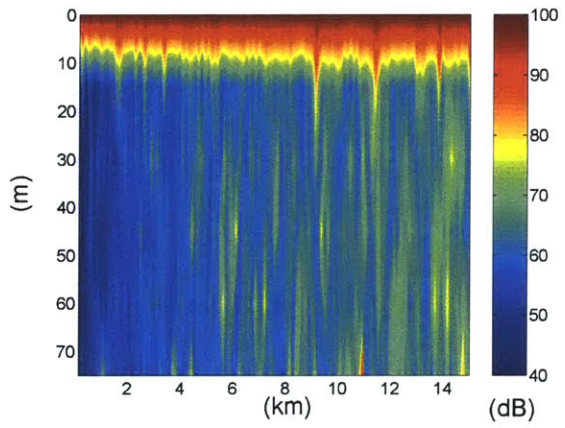


(c) Correlation test

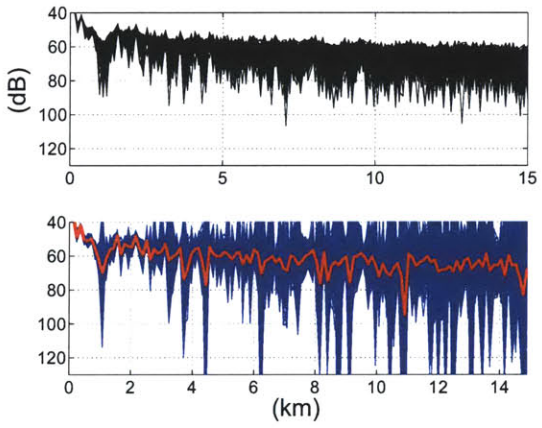


(d) TL sample variance comparison at the depth of 75m

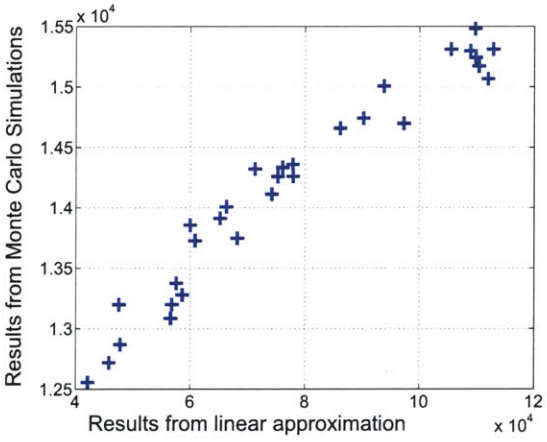
Figure 5-15: Using TL_r, $\alpha = 0.05$. $tr\,(var\,(TL))$ from the linear approximation is equal to 116.959 (dB^2). 300 Monte Carlo simulations give $tr\,(var\,(TL))$ equal to 135.464 (dB^2). The relative error is about 13.66%.



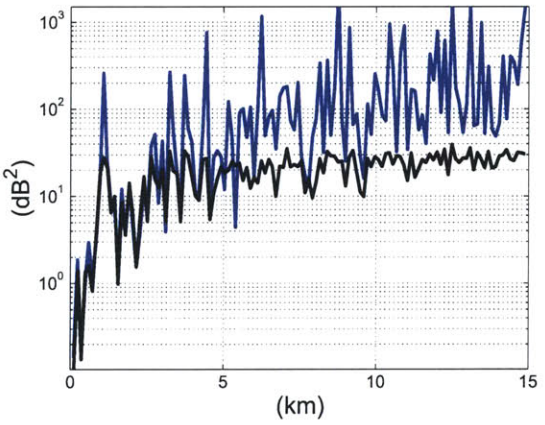
(a) TL_0



(b) TLs comparison at the depth of 75m

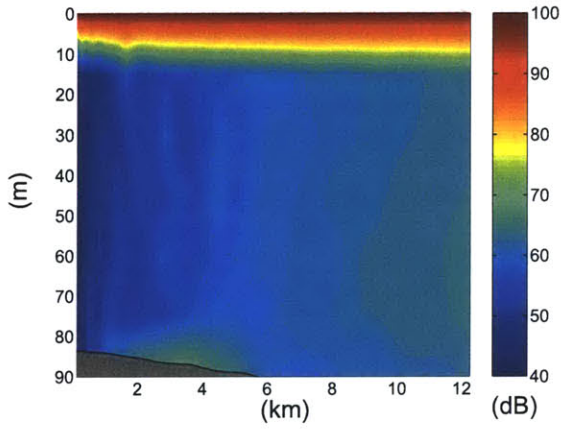


(c) Correlation test

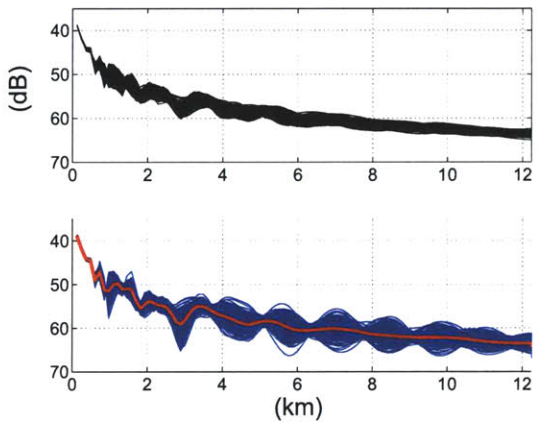


(d) TL sample variance comparison at the depth of 75m

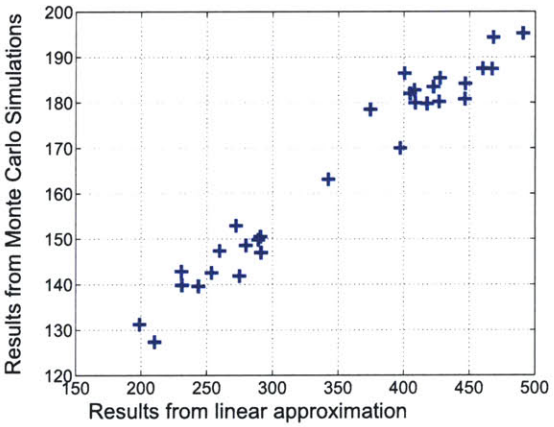
Figure 5-16: The 400Hz sound source is located at $(0\ km, 80m)$. $tr\,(var\,(TL))$ from the linear approximation is equal to 86030.2 (dB^2). 300 Monte Carlo simulations give $tr\,(var\,(TL))$ equal to 14572.7 (dB^2). The relative error is about 490.35%.



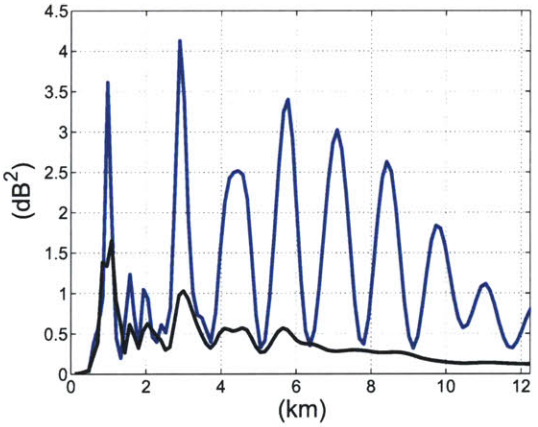
(a) TL_0



(b) TLs comparison at the depth of 30m



(c) 400Hz, $\alpha = 0.1$



(d) TL sample variance comparison at the depth of 30m

Figure 5-17: Using TL_r, $\alpha = 0.1$. $tr\,(var\,(TL))$ from the linear approximation is equal to 381.031 (dB^2). 300 Monte Carlo simulations give $tr\,(var\,(TL))$ equal to 171 (dB^2). The relative error is about 122.8%.

Chapter 6

Modeling the Adaptive On-board AUV Routing Problem

As mentioned before, the AUV path planning problem in regard to minimizing the posterior acoustic prediction uncertainty is a non-linear stochastic optimization problem, which can be solved by some non-linear programming methods, Genetic Algorithms (GA) etc. The optimized AUV path is a predetermined path. However in AREA, it is not necessary to fix the path before the AUV is launched, instead AUV can adaptively determine its waypoints on-board. For example, the $i+1$th waypoint can be determined in real-time when the AUV reaches the ith waypoint and all in-situ measurement data collected from the beginning to the ith waypoint can be used for the decision making. This adaptive on-board AUV routing problem can be viewed as a sequential decision making problem under stochastic disturbances. As discussed in Chapter 4.3, such kind of process can be formulated using Dynamic Programming (DP), in which what is to be optimized is not an AUV path (or say AUV route) but an AUV routing strategy. On the other hand, any AUV path can be viewed as a static routing strategy, which always makes the same decision no matter of changes in the in-situ measurement data collected. Therefore, the optimized AUV routing strategy is theoretically guaranteed to produce a better result than the optimized AUV path. However, the space of the admissible AUV routing strategy is much bigger than the space of the feasible AUV path. Solving a DP problem is usually much harder than

solving an ordinary optimization problem. So in the time limits in AREA, only a sub-optimal AUV routing strategy can be obtained in practice and the associated result may not be better than that of a sub-optimal predetermined AUV path.

It should be noticed that for the AREA problem in regard to minimizing the posterior SVP prediction uncertainty, the objective function is only dependent on the waypoints. For this problem, all AUV routing strategies are static and the associated DP problem is a deterministic DP problem [47]. Therefore, the optimized AUV routing strategy is actually identical to the optimized predetermined AUV path.

6.1 Modeling the AUV Routing Strategy Optimization using DP

Now let's think about how to model the adaptive on-board AUV routing problem in the DP frame. The first thing is to determine the state variable x_k (see Chapter 4.3 for more details). In principle, x_k must contain all informations required to determine the control u_k. In the adaptive on-board AUV routing problem, those informations include all the waypoints' locations and all in-situ measurement data collected so far. We have

$$x_k = \{z_0, z_1, \cdots, z_k;\ \omega_1, \cdots, \omega_k\}, \tag{6.1}$$

where z_i is the vertical coordinates of the ith waypoint and ω_i is all the in-situ measurement data collected between the $i-1$th waypoint and the ith waypoint.

$$\omega_i = \begin{bmatrix} r_i^1 & r_i^2 & \cdots & r_i^{n_i} \\ z_i^1 & z_i^2 & \cdots & z_i^{n_i} \\ c_i^1 & c_i^2 & \cdots & c_i^{n_i} \end{bmatrix} \tag{6.2}$$

where r_i^j and z_i^j are the coordinates of the jth in-situ measurement between waypoint $i-1$ and waypoint i; c_i^j is the jth in-situ measurement datum; n_i is total number of

in-situ measurements between waypoints $i-1$ and waypoint i.

After the waypoints $i-1$ and waypoint i are determined, r_i^j, z_i^j and c_i^j are not yet completely determined and many factors can disturb them. Because ω_k is part of x_k, ω_{k+1} will influence x_{k+1} — the state at the next stage. Thus ω_{k+1} is actually the stochastic disturbance at stage k. After the state variable and disturbance are determined, it is easy to see that at stage k, the control should be the next waypoint, i.e. $u_k = z_{k+1}$. The system dynamics is

$$x_{k+1} = x_k \oplus u_k \oplus \omega_{k+1}, \tag{6.3}$$

where $\oplus$ means the variable augmentation, like $\{x_1, x_2\} \oplus x_3 = \{x_1, x_2, x_3\}$.

The purpose in our problem is to minimize the posterior TL_r prediction uncertainty, so the cost only shows up at the termination stage and the cost per stage is zero.

$$g_N(x_N) = tr\{var(TL_r) \odot W\}, \tag{6.4}$$

$$g_k(x_k, u_k, \omega_k) = 0. \tag{6.5}$$

Eq. 6.4 and 6.5 show a non-separable cost function structure, which destroys the advantage of this DP algorithm and makes the DP problem much much more difficult to solve. How to resolve this difficulty, how to solve the DP problem quickly is one of the major contributions of this thesis and the discussions are in Chapter 7.

6.2 Summary

Based on the definition in Chapter 4.3, the adaptive on-board AUV routing problem can be written as follows.

$$\min \quad E[tr\{var(TL_r) \odot W\}] \tag{6.6}$$

$$\text{s.t.} \quad \{\mu_0, \mu_1, \cdots, \mu_{N-1}\} \text{ is admissible,} \tag{6.7}$$

where the expectation is over all possible a priori SVP realizations and measurement noises. $\{\mu_0, \mu_1, \cdots, \mu_{N-1}\}$ is the routing strategy, namely the adaptive sampling strategy, which is the decision variable in the DP problem.

6.3 Thermocline-oriented AUV Yoyo Control Optimization

One way to simplify the AUV routing strategy optimization is to restrict the routing strategies in a strategy pattern, in which the basis functions are carefully selected, while only some parameters are left to be optimized.

The AREA project focus more on shallow water region, where the variation of the thermocline depth often leads to the main SVP prediction uncertainties. Therefore, the adaptive sampling strategy that aims to capture the vertical variability of the thermocline due to fronts, eddies, internal waves, etc. can often capture the dominant SVP feature and its uncertainties and so also minimize the TL prediction uncertainty. To track the vertical variability of the thermocline, a thermocline-oriented AUV path control was researched, by which an AUV can be given guidance about the depths of the thermocline and move around these depths. Since the thermocline is the region where the sound speed changes rapidly with depth, a simple criterion determining the relative position between the AUV and the thermocline is to compare the absolute value of local vertical gradient of sound speed $\left|\frac{\partial c}{\partial z}\right|$ with a threshold. By doing so, the AUV can estimate whether it is above, inside of or below the thermocline.

It is assumed that at the beginning of the mission, the AUV stays on the surface. While it is diving, its CTD sensor collects data every second. The $\left|\frac{\partial c}{\partial z}\right|$ is estimated via *Linear Least Squares Fitting* method based on every p CTD data. If at the beginning, $\left|\frac{\partial c}{\partial z}\right| \leq \gamma$, where γ is the threshold, and then $\left|\frac{\partial c}{\partial z}\right|$ becomes greater than γ, and after that $\left|\frac{\partial c}{\partial z}\right|$ becomes lower than γ again, then the criterion will indicate that the AUV is now below the thermocline and it will turn around upwards. Thereafter while the AUV is going up, if $\left|\frac{\partial c}{\partial z}\right|$ becomes greater than and then lower than γ again,

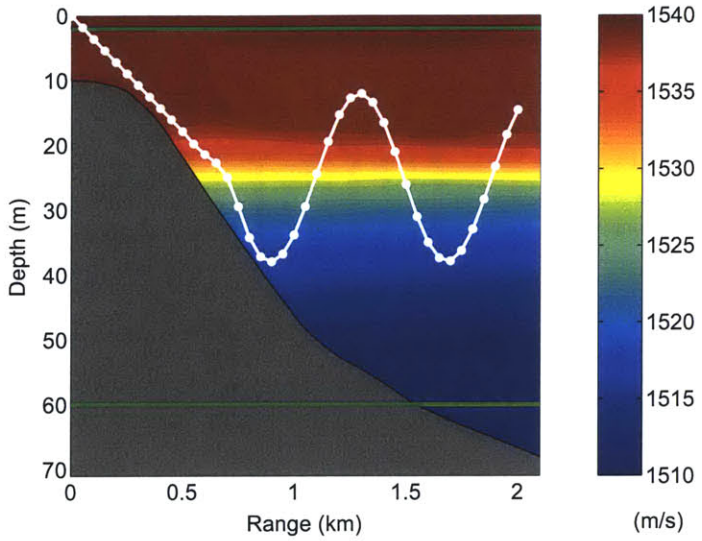


Figure 6-1: Illustration of the thermocline-oriented AUV yoyo track. The two green lines are the upper bound and lower bound respectively.

the criterion indicates that the AUV is now above the thermocline and it will turn around downwards. An upper bound and a lower bound on the depth range of the AUV were also set up. Should the AUV have crossed the thermocline or not, once the lower bound or upper bound is reached, the AUV has to turn around to avoid reaching too deep depths or getting off the surface. This path control will lead the AUV to carry an up-and-down yoyo track (Fig. 6-1).

The lower bound can be set even lower than seabed at some ranges and it is assumed that AUV will be forced to turn around at few meters above the seabed by collision avoidance device. There are several possible patterns for thermocline-oriented AUV path control (see Fig. 6-2). So far, the triangular wave pattern (pattern 2) is selected because of the following 2 reasons:

1. AUV prefers to go up or down at the maximum pitch angle and it is hard for AUV to follow a level line.

2. Following a level line doesn't help much to track the vertical variabilities of thermocline.

In this case, it is not necessary to discretize the ocean for AUV path and no waypoints

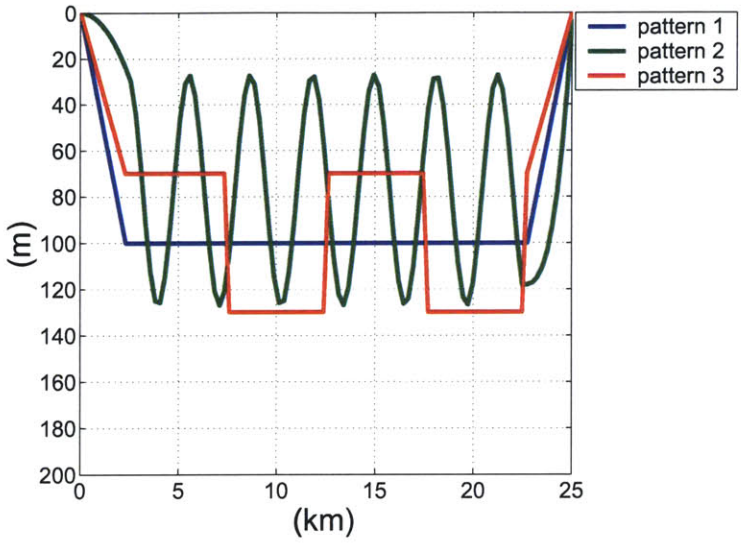


Figure 6-2: Illustration of 3 yoyo patterns.

will be determined by the AUV yoyo control. While in the control there are two parameters to be optimized: p and γ. p is the number of sampling points used to compute $\left|\frac{\partial c}{\partial z}\right|$; γ is the threshold used to compare with $\left|\frac{\partial c}{\partial z}\right|$. The γ defines how rapidly the sound speed changes with depth can be linked to the thermocline. The control parameters optimization problem can be formulated as:

$$\min \quad \mathbf{f}(p,\ \gamma) \tag{6.8}$$

$$\text{s.t.} \quad \gamma \geq 0,\ p \text{ is a positive integer}, \tag{6.9}$$

where the objective function is

$$\mathbf{f}(p,\ \gamma) = E\left[tr\left\{var\left(TL_r\right)\right\} \odot W\right], \tag{6.10}$$

which can be approximated using TL_r linear approximation. Again, here the *var* is taken with respect to all possible posterior SVP realizations and the E is taken with respect to all possible a priori SVP realizations and measurement noises. The objective functions in Eq. 5.23, Eq 6.6 and Eq 6.10 are actually identical. They are implicitly dependent on AUV path, AUV routing strategy and yoyo control parame-

ters respectively.

This optimization problem is essentially a mixed-integer non-linear stochastic programming problem. The objective function is only defined on integer-valued p, so it can't be solved by relaxation method. The advantage of the AUV yoyo control is that: the optimization result is not a predetermined AUV path but a yoyo sampling strategy; while compared with the associated adaptive on-board AUV routing problem, the search space is now just the space of the two control parameters and hence much smaller.

Chapter 7

Solving the Optimization Problems

In Chapter 5 and 6, the AUV path planning problem and the adaptive on-board AUV routing problem have been modeled in optimization problems. In this regard, the AREA project is an implementation of operational Research (OR) in ocean engineering. In this chapter, we will discuss how to quickly solve those optimization problems in detail.

7.1 Adaptive Rapid Environmental Assessment Simulation Framework

As shown in the precedence, the optimization models associated with AREA are all very complicated: the search spaces can be very huge and the objective functions don't have any helpful feature but require very intensive computations. Thus, none of them can be solved analytically or solved by numerical methods easily. To solve those optimization problems, find the optimal or sub-optimal AUV path or adaptive routing strategy and also test the optimization effects before doing very costly on-site experiments, an *Adaptive Rapid Environmental Assessment Simulation Framework* (AREASF) is really desired, by which we can also observe how AREA system will work and test if real-time feasibility. The AREASF can provide us a training and learning tool [39, 65].

An AREA simulator has been developed in C++ — an object-oriented language. Due to the object-oriented feature, every real object can have a corresponding simulated object in the computer, which can simulate all functions that the real one has. So referring to Fig. 3-5 and Fig. 7-1, basically speaking each component in the real AREA system has a corresponding module in the AREA simulation framework. However, the sonar system is divided into 2 modules: sonar array simulator and sonar signal processing center. The Control center is directly replaced with the observation database module, ocean predictor module and control agent module. For the control agent, several different sampling strategy algorithms have been embedded. In the end, a surveillance module and an output module were built to monitor the whole system and output results. In this way, the AREA simulation framework is upgradeable and flexible; and its structure is simpler and close to a real AREA system.

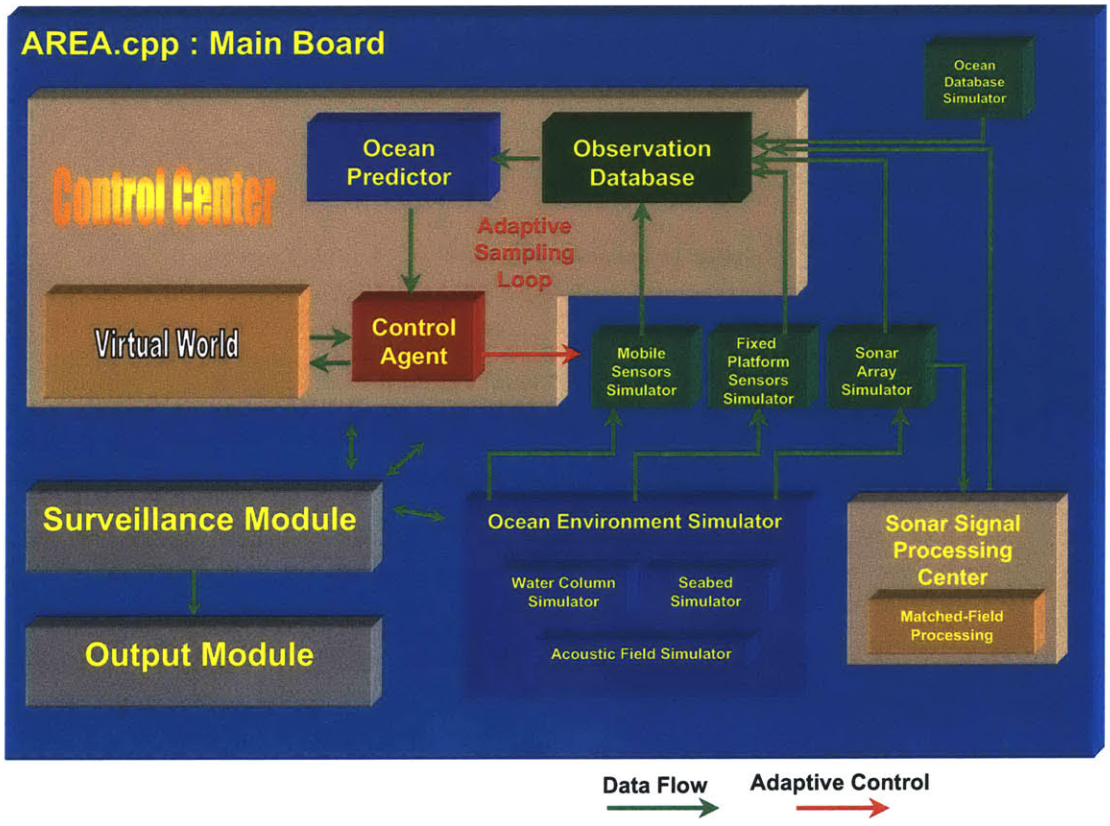


Figure 7-1: AREA simulation framework wiring diagram

As shown in Figure 7-1, the structure of AREA simulation framework is like an

integrated circuit board. AREA.cpp is the C++ main file containing the "main" function. It works like a human-computer interface where we can input almost all parameters for each module, select options and start running program (see Figure 7-2). AREA.cpp provides a working environment to all the other modules like the main board in PC to peripherals.

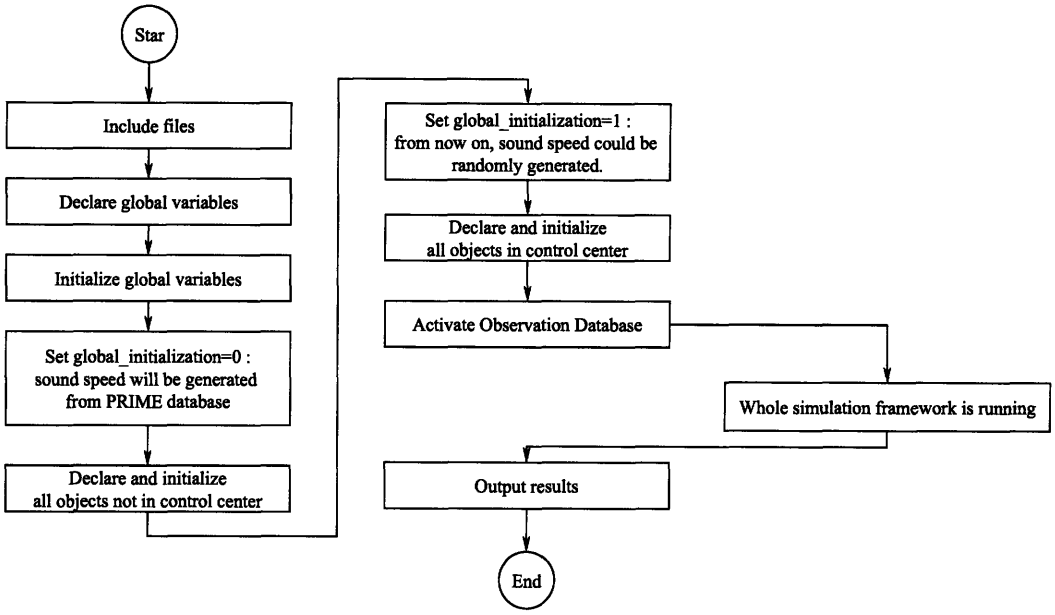


Figure 7-2: Flow chart of AREA.cpp

The Ocean Environment Simulator module is supposed to provide sensors and sonar arrays with oceanographic information, bathymetric information and acoustic signals. It includes 3 sub-modules: Water Column Simulator, Seabed Simulator, Acoustic Field Simulator (see Figure 7-1). Water Column Simulator and Seabed Simulator simulate the ocean environment in water column and seabed respectively. The Acoustic Field Simulator can generate the acoustic field according to water and seabed environment and sound source parameters input from AREA.cpp. The current acoustic model is RAM.

The Mobile Sensors Simulator module can be called and input controlling parameters by the Control Agent module. The Mobile Sensors Simulator can simulate how real mobile sensors move in the ocean and measure in situ by calling the Ocean Environment Simulator to output information at those measurement locations. By

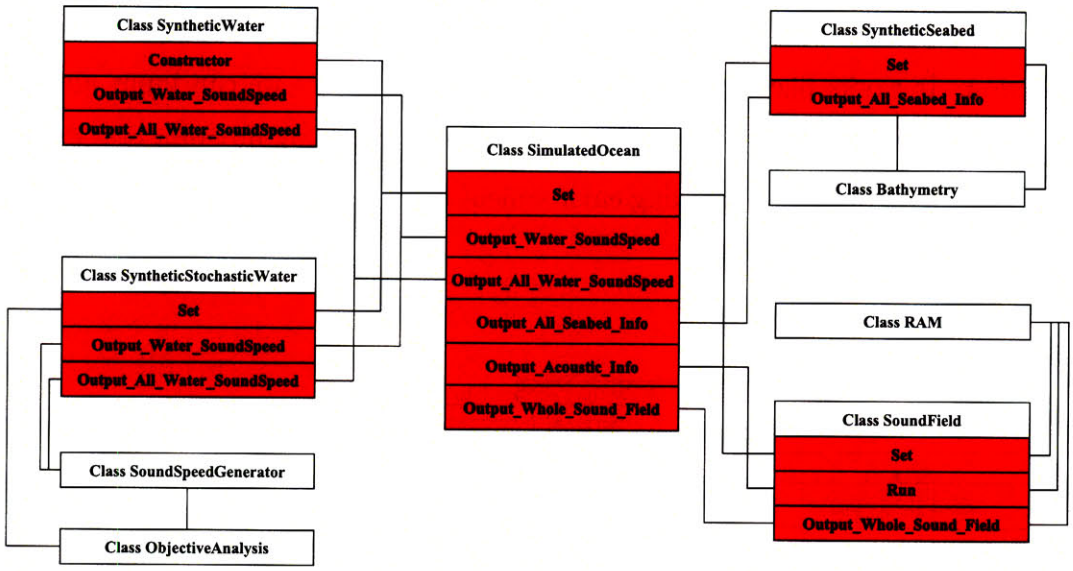


Figure 7-3: Ocean Environment Simulator wiring diagram

configuring the sensors and platforms differently, this module can simulate many sort of mobile sensors such as XBT carried on ship, CTD carried on AUV, hydrophones carried on AUV, or both of them carried on AUV.

The Fixed Platform Sensors Simulator module can retrieve oceanographic information and/or bathymetric information from the Ocean Environment Simulator as conventional oceanographic sensors do in ocean. The Fixed Platform Sensors Simulator may include several different objects, each of them corresponding to one particular sensor, which could be local CDT, satellite or acoustic remote sensing and a seabed mapping device. Because of the flexibility, this module can be quickly adapted according to requirement.

The Sonar Array Simulator module simulates a hydrophone array, which can call the Ocean Environment Simulator and retrieve data from the Acoustic Field Simulator. Acoustic signals received by the Sonar Array Simulator and signals received by the Mobile Sensors Simulator will be processed in the Sonar Signal Processing Center. The Sonar Signal Processing Center is a software package containing different sonar models and acoustic models; however, currently only *Matched-Field Processing* (MFP) method and RAM are included.

The Observation Database is the first module in the control center. Its function is

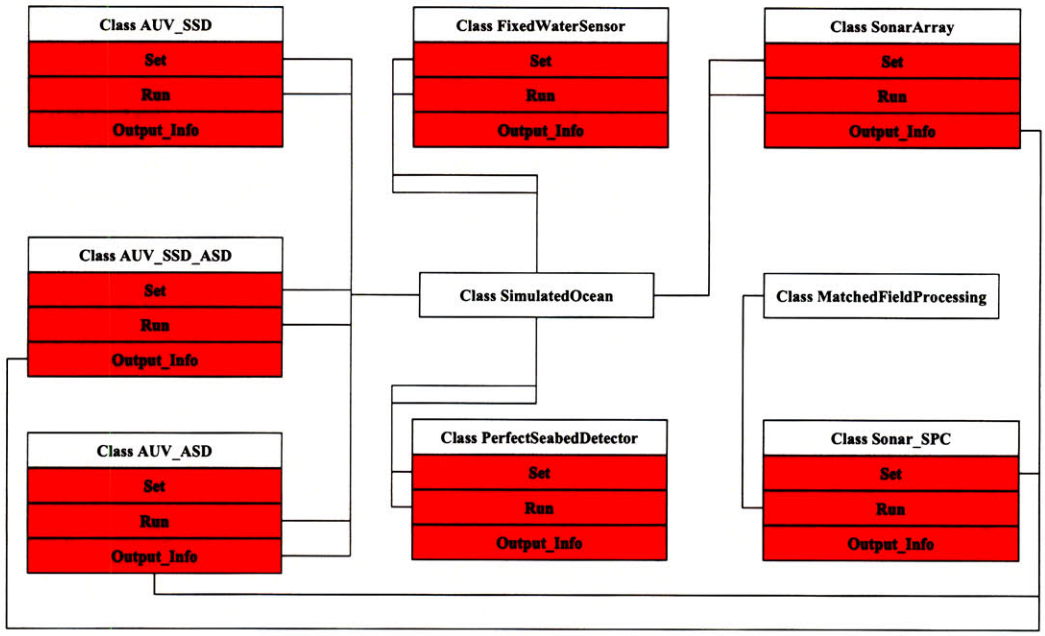


Figure 7-4: Wiring diagram of Mobile Sensors Simulator, Fixed Platform Sensors Simulator, Sonar Array Simulator and Sonar Signal Processing Center

to sequentially call and receive data output from the Ocean Database Simulator, Sonar Signal Processing Center, Sonar Array Simulator, Fixed Platform Sensors Simulator, Mobile Sensors Simulator and store the data. In fact, the whole simulation framework starts from the Observation Database calling and collecting initial information from those modules.

After the Observation Database finishes collecting all necessary initial information, it will call and activate module Ocean Predictor. This module uses some estimation algorithms such as an objective analysis technique to predict the ocean acoustic environment and simultaneously provide the error field.

At the end, the Control Agent will be called and passed those initial information and analysis results. Based on all information and according to adaptive sampling algorithm, the Control Agent may create a virtual world for trial purpose and determine optimal or sub-optimal commands through a complicated procedure. Details about the decision making procedure are out of the range of this thesis, but a major AREA research issue.

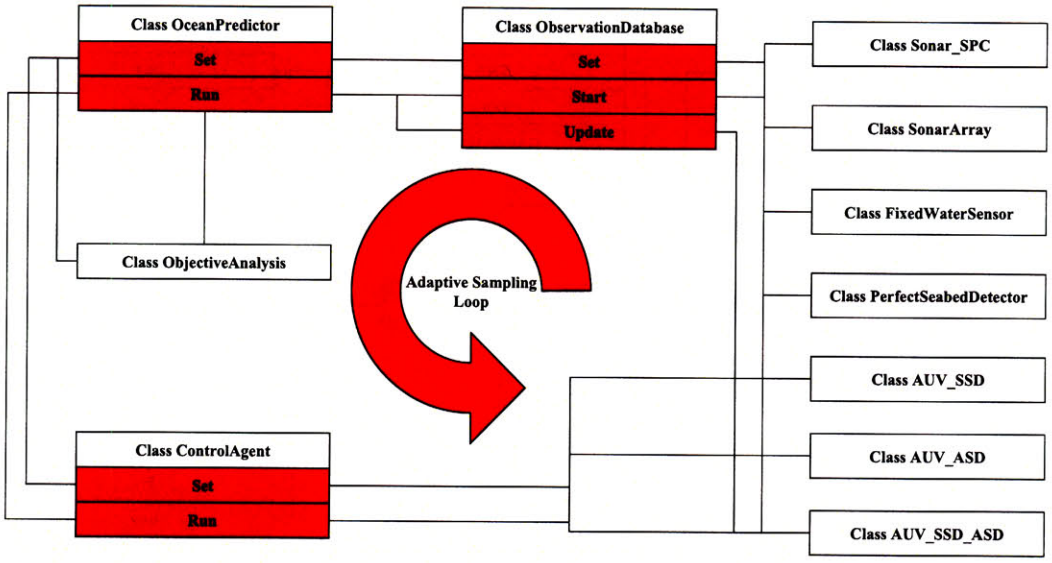


Figure 7-5: Wiring diagram of modules in control center

Once commands are determined, Mobile Sensors Simulator will be called and execute those commands to obtain the newest data. After that, Observation Database will be called and updated. Then, the adaptive sampling loop will be repeated again until the Mobile Sensors Simulator finishes all in situ measurements.

When all the above modules are running, a very special module - the Surveillance Module keeps watching all processes and records all interesting intermediate results. In the end, the Surveillance Module will send all records to Output Module through which results will be output into a file.

Note:

1. Since we don't have any ocean database for Ocean Predictor, currently there's no Ocean Database Simulator in the simulation framework. But it is easy to be added in this module later.

2. In this chapter, we simply introduced the structure and functions of the simulation framework. For more details, please refer to [39].

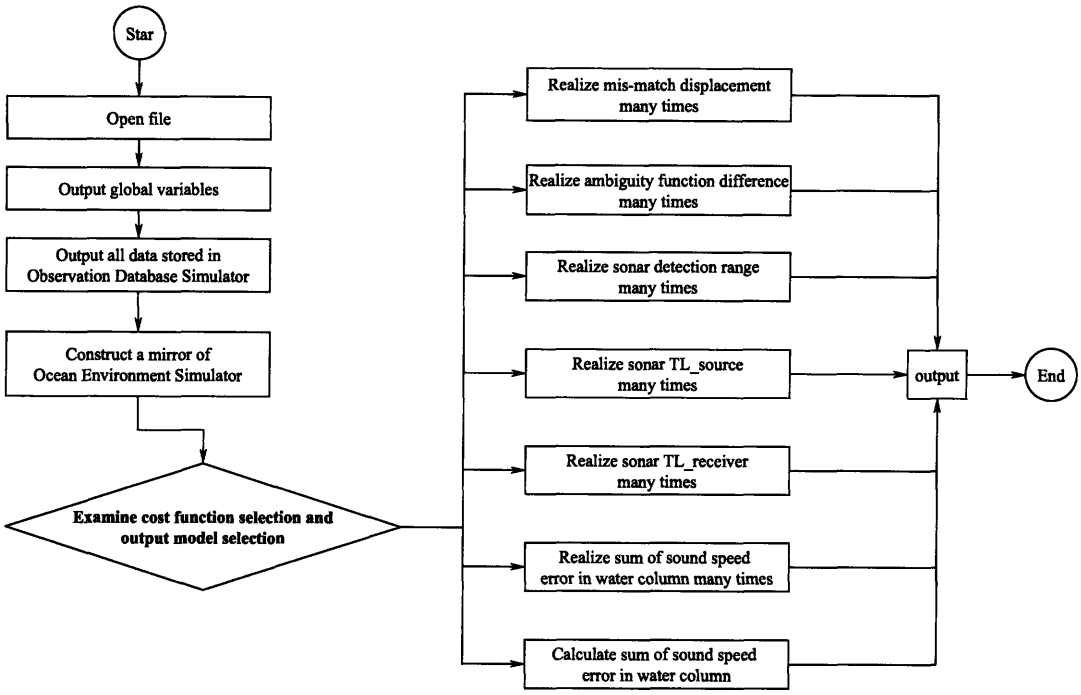


Figure 7-6: Simplified flow chart of the Surveillance Module

7.2 Solving the Thermocline-oriented AUV Yoyo Control Optimization Problem

In this section, we are going to discuss how to solve the AUV yoyo control parameters optimization problem. From Eq. 6.8, 6.9 and Eq. 6.10, it can be seen that the AUV yoyo control parameters optimization problem is a mixed-integer non-linear stochastic programming problem. It is very difficult to solve such kind of problem and many researches have been done in integer programming and non-linear programming. However, in our case, the objective function is only defined at those p with integer values, many major integer programming methods such as relaxation method can not be applied. Moreover, the objective function is neither convex nor concave, so no helpful features exist.

From the experience, in such kind of problem the random search methods such as genetic algorithm, simulated annealing etc. may work well and the corresponding formulas may be easily implemented. So let's check the feasibility of the random

search methods. First, let's rewrite the objective function Eq. 6.10 in the below,

$$\mathbf{f}(p,\ \gamma) = E\left[tr\left\{var\left(TL_r\right)\right\} \odot W\right].$$

The *var* is taken with respect to the posterior SVP realizations and the E is taken with respect to the a priori SVP realizations and measurement noises. If Eq. 6.10 is calculated by Monte Carlo simulations, based on our experiences it at least takes more than 10 minutes to obtain a convergent result. While in the AREA project, the real-time feasibility is always very crucial, which means that the optimization computation has to be finished in few hours. In fact, it's usually no more than 4 hours. Therefore, the objective function can not be computed for more than 24 times and convergence is hard to be guaranteed. This difficulty can not be solved so far and even if the PC could be several times faster in future, it would still be a big difficulty. To avoid this difficulty, exhaustive search in a small search space with respect to (p, γ) is adopted in current AREA. In practice, about 4 promising p values and about 4 promising γ values are empirically selected. In this 16-element space the exhaustive search can be finished within the time limit.

Another way to avoid the computation difficulty is to estimate the $var\left(TL_r\right)$ by Eq. 5.29, which usually takes only few seconds. The expectation with respect to a priori SVP realizations and measurement noises can be approximately obtained by Monte Carlo simulations. In this way, computing the objective function once takes about 2 minutes and thus the search space with respect to (p, γ) can be enlarged.

There are two things that should be noticed here:

1. Approximating the expectation with the SVP P.E. and zero measurement error is NOT proper in this case, since if doing so, the adaptivity of AUV yoyo control will be eliminated and the corresponding result will reflect nothing about the AUV routing strategy optimization but the AUV path optimization in the SVP P.E. scenario.

2. To reduce the number of Monte Carlo simulations and make the sub-optimal result better, for any two parameter pairs, (p_{i1}, γ_{i2}) and (p_{j1}, γ_{j2}), all random

variables used in Monte Carlo simulations for (p_{i1}, γ_{i2}) should be accordingly highly correlated with those of (p_{j1}, γ_{j2}). Therefore $\widetilde{\mathbf{f}}(p_{i1}, \gamma_{i2})$ will be highly correlated to $\widetilde{\mathbf{f}}(p_{j1}, \gamma_{j2})$, i.e. if $\widetilde{\mathbf{f}}(p_{i1}, \gamma_{i2})$ is above its true value $\mathbf{f}(p_{i1}, \gamma_{i2})$ then $\widetilde{\mathbf{f}}(p_{j1}, \gamma_{j2})$ is very possible above its true value $\mathbf{f}(p_{j1}, \gamma_{j2})$ too, and vice versa. This phenomenon is similar to that in estimating $var(TL_r)$ via linear approximation method. The accurate values can't be obtained, while the same descending / ascending order may be kept in the approximated results (as shown in Fig. 7-7). The true optimal solution is also very possibly the optimal one in the approximation. This method is popular in simulation based optimization problem [47].

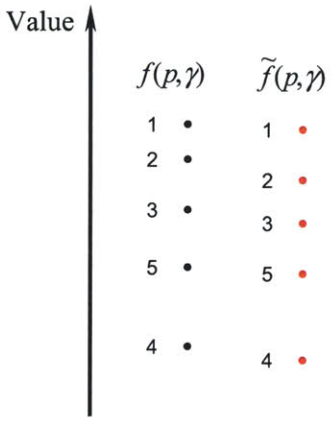


Figure 7-7: Illustration of the descending / ascending order in the true values $\mathbf{f}(p, \gamma)$ and in the approximate values $\widetilde{\mathbf{f}}(p, \gamma)$.

7.3 Solving the AUV Path Planning Problem

In this section, we will discuss how to solve the optimization problem defined in Eq. 5.44 and 5.45, which is to minimize the approximate posterior acoustic prediction uncertainty. When $A = I$, this optimization problem is the same as the one defined in Eq. 5.19 and 5.20, which is to minimize the posterior SVP prediction uncertainty. In practice, the Eq. 5.19 is often replaced with Eq. 5.44 and the matrix A is a diagonal weight matrix, in which each diagonal element is equal to the area associated with a

SVP grid point in the vertical plane.

This optimization problem is a non-linear programming problem. Some non-linear programming methods can be applied. It will be discussed later that in this problem the search space (i.e. all feasible AUV paths) can be very huge and exponentially grows as the resolution of the ocean discretization grows (see Chapter 5.2.1). So, those non-linear programming methods may not be efficient in this case.

Looking at the constraint in Eq. 5.45, the feasible solutions are vehicle paths and have network structure, therefore the problem may be solved by network optimization methods. In fact, the AUV path planning problem is easy to be represented as a shortest path problem introduced in Chapter 4.2.1. The graph will be like the one shown in Fig. 5-3. The starting point is usually at the origin and each node has one or more children except the nodes at the maximum range. An artificial ending node can be made as the child of all the nodes at the maximum range. This is an acyclic graph and for each node at a certain range, its children/child must be located at the next range. The total distance from the starting node to the ending node is the value of the objective function in Eq. 5.44, which is actually a non-separable function. While in a normal shortest path problem, the total distance is the summation of the length of arcs in a path from the starting node to the ending node, thus the objective function is additive. Without the additivity in the objective function, the optimization problem defined in Eq. 5.44 and 5.45 can not take any advantage from network optimization methods, which will actually degrade to normal non-linear programming methods in this case. The non-separability in the objective function is rooted at the correlation of SVP — the SVP prediction uncertainty is reduced at a single point, the SVP prediction uncertainty in the neighborhood will be reduced too. As a result, the objective analysis can only be carried for the whole AUV path. For a segment in the path, its contribution to the SVP prediction uncertainty reduction is affected by the prior and posterior segments. So to make the objective function in Eq. 5.44 more additive, we need to investigate the effects of the SVP correlation on the prediction uncertainty reduction.

7.3.1 Significance of SVP Correlation Lengths

In AREA, the SVP correlation lengths indicate the associated dominant oceanographic process scale. Fig. 7-8 shows the influence of horizontal and vertical correlation length on the acoustic prediction uncertainty. The scenario is the same as the 2nd example shown in Chapter 5.4.2. A single 100Hz sound source is located at $(0km, 80m)$. The acoustic prediction uncertainty is the a priori $tr\left(var\left(TL_r\right)\right)$ and $\alpha = 0.05$. 300 Monte Carlo simulations were run to compute $tr\left(var\left(TL_r\right)\right)$ for each Lr, Lz pair. Here, it is supposed that for all Lr, Lz pairs, the a priori SVP prediction uncertainty is the same as the one shown in Fig. 5-13(f).

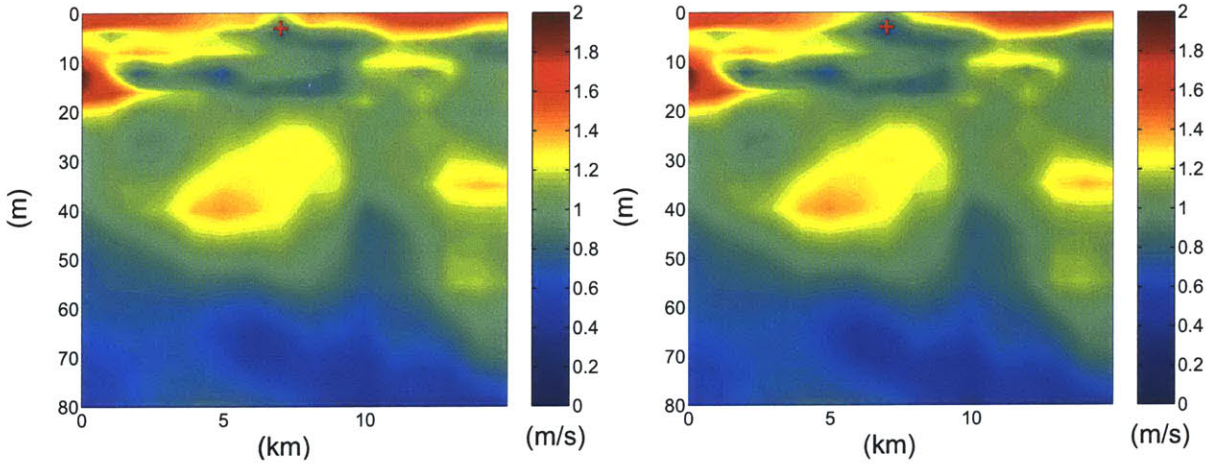
	Lr=0.1m	Lr=1m	Lr=10m	Lr=100m	Lr=1km	Lr=10km
Lz=0.1m	35.523	36.223	33.6714	35.6202	55.8852	87.3399
Lz=1m	37.0001	37.8883	37.1229	35.9475	56.7762	102.874
Lz=10m	98.1158	86.0759	93.5165	91.2069	104.792	191.569
Lz=100m	15.6137	14.67	15.0867	14.7656	8.24276	8.00563
Lz=1km	13.8058	14.1651	14.8768	14.7927	7.35397	8.98648

Figure 7-8: The 100Hz sound source is located at $(0km, 80m)$. The numbers are the values of $tr\left(var\left(TL_r\right)\right)$ estimated from 300 Monte Carlos simulations. The unit is dB^2.

In Fig. 7-8, it can seen that when Lz is very large, the $tr\left(var\left(TL_r\right)\right)$ becomes very small; while when Lz is about few meters, the $tr\left(var\left(TL_r\right)\right)$ gets the biggest. In shallow water, Lz is always about few meters. In this scenario, as Lr becomes larger, the $tr\left(var\left(TL_r\right)\right)$ will very possibly become larger too. In fact, when the dominant oceanographic process scale is as big as few kilometers, it will be easily captured by remote sensing techniques such as satellite sensing. So the real significance for different oceanographic process scales will be like the one shown in Fig. 3-1. The

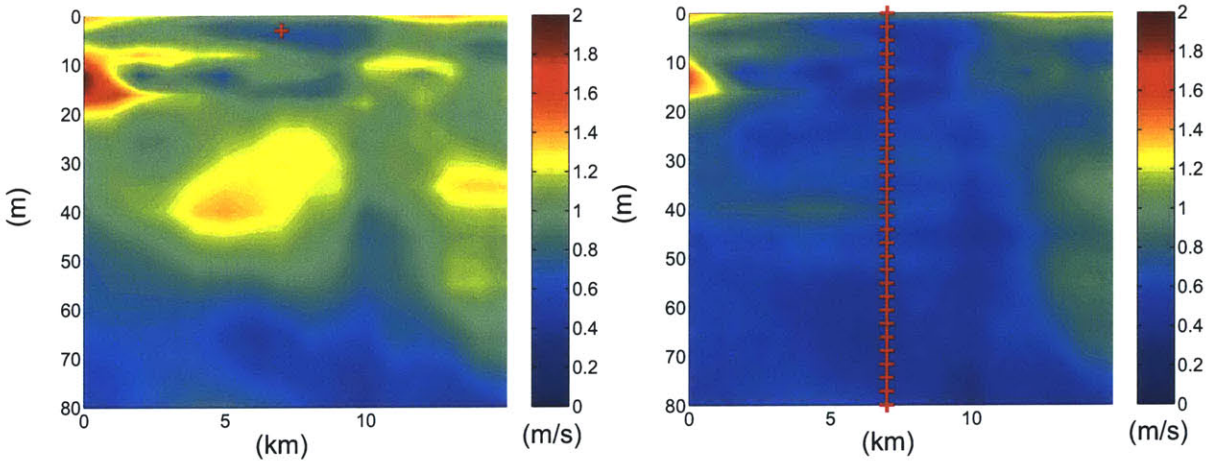
scales in few hundred meters will impact the acoustic prediction most strongly.

Now let's investigate how Lr will affect the prediction uncertainty reduction, if an in-situ measurement is made.



(a) $Lr = 100m$, $Lz = 3.5m$

(b) $Lr = 1000m$, $Lz = 3.5m$.



(c) $Lr = 10km$, $Lz = 3.5m$.

(d) $Lr = 10km$, $Lz = 3.5m$.

Figure 7-9: The posterior SVP prediction error standard deviation. In (a), (b) and (c), an in-situ measurement is made at $(7km, 3m)$. In (d), 30 in-situ measurements are made at the range of $7km$.

Fig. 7-9(a), 7-9(b) and 7-9(c) show the posterior SVP prediction uncertainty after an in-situ measurement is made at $(7km, 3m)$. The corresponding SVP prediction error variance reductions are 2.4×10^7 $(m/s)^2 \cdot m^2$, 6.26×10^7 $(m/s)^2 \cdot m^2$ and

5.1×10^8 $(m/s)^2 \cdot m^2$; the corresponding TL_r prediction error variance reductions are about 1.0×10^{-3} $(dB)^2$, 3.9×10^{-3} $(dB)^2$ and 0.23 $(dB)^2$. This means that when Lr is bigger, more SVP information around the in-situ measurement will be obtained. In Fig. 7-9(c) the in-situ measurement at $(7km, 3m)$ almost reduces all the SVP prediction uncertainty from the depth of $0m$ to about $7m$. So, when the Lr is really big, AREA system is not necessary. Dropping a CTD or XBT at the middle range will capture most uncertainties (see Fig. 7-9(d)). Where, the SVP prediction uncertainty reduction is 1.0×10^{10} $(m/s)^2 \cdot m^2$ and the TL_r prediction uncertainty reduction is about 114.4 $(dB)^2$; while the total a priori SVP prediction uncertainty is 1.86×10^{10} $(m/s)^2 \cdot m^2$ and the total a priori TL_r prediction uncertainty is about 182.1 $(dB)^2$.

In Fig. 7-9(c), it can be seen that once the in-situ measurement is made at the depth of $3m$, any other in-situ measurements around that depth will be redundant. This implies that the individual significance of an in-situ measurement to the SVP/acoustic prediction uncertainty reduction is coupled with locations of other in-situ measurements. To represent the original optimization problem as a shortest path problem with additive objective function, the representation of the "Individual Significance" of an in-situ measurement at a certain location is needed to be investigated.

7.3.2 Individual Significance and n-step Look-back Method

The individual significance can be represented with respect to the sound velocity at a certain location in the water column, the Empirical Orthogonal Functions (EOFs) or the System Orthogonal Functions (SOFs). Some researches have been done in [66, 63], where the objective is to find an uncoupled individual significance format, e.g. if the individual significance of an EOF is independent of the individual significances of other EOFs, the individual significance representation with respect to EOFs is an uncoupled format. Uncoupled individual significance format is very helpful to find the optimal AUV path. Unfortunately, due to the coupling in acoustic modeling and the SVP correlation, no uncoupled individual significance format has been found so far.

Decision Variable Augmentation

In the original optimization problem the decision variable is $\{z_1, z_2, \cdots, z_N\}$, now let's augment it as $\{s_1, s_2, \cdots, s_N\}$, where $s_1 = \{z_1\}$, $s_2 = \{z_1, z_2\}$, $s_N = \{z_1, \cdots, z_N\}$, i.e. all the past history is included in the current state. Fig. 7-10 illustrates an AUV path and the network graph associated with the original problem. The number of nodes is 17 and the number of arcs is 46. After the variable augmentation, the augmented graph will be like the one shown in Fig. 7-11. The size of the graph gets much bigger.

Let $R_{i,i+1,\cdots,i+k}$ denote the SVP/acoustic prediction uncertainty reduction, when the AUV moves from waypoint i to waypoint $i+k$ through waypoints $i+1, \cdots, i+k-1$. So for the AUV path in Fig. 7-10, $R_{0,1}$ is the SVP/acoustic prediction uncertainty reduction associated with the AUV moving from the start point to waypoint 1 and $R_{2,3,4}$ is the SVP/acoustic prediction uncertainty reduction associated with the AUV moving from waypoint 2 and to waypoint 4 via waypoint 3. The SVP/acoustic prediction uncertainty reduction associated with the whole path is $R_{0,1,2,3,4}$. In $R_{0,1,2,3,4}$, let $R'_{i,i+1,\cdots,i+k}$ denote the part contributed by the path segment (waypoint i, waypoint $i+1$, $\cdots$, waypoint $i+k$). Thus, we have $R'_{0,1} = R_{0,1}$, $R'_{1,2} = R_{0,1,2} - R_{0,1}$, $R'_{2,3} = R_{0,1,2,3} - R_{0,1,2}$ and $R'_{3,4} = R_{0,1,2,3,4} - R_{0,1,2,3}$. Those $R'_{i,i+1}$ reflect the SVP/acoustic prediction uncertainty reduction induced by each step on the AUV path, so they are actually the arc lengths in the augmented graph. After going through all possible AUV paths, lengths of all arcs in the augmented graph can be assigned. The total distance of a path in the augmented graph is equal to the SVP/acoustic prediction uncertainty reduction associated with the corresponding AUV path in the original graph. It seems that the original objective function is now converted to be additive and the original problem can be solved by some very efficient shortest path algorithms. This method is very much like the n-step look-ahead method used in chess playing algorithms [47] such as the one in Deep Blue; while in our case it's not look ahead but look back. So, let's call it **n-step look-back method**. However, now the n is equal to N and this method is actually trivial. In the current n-step look-back method, the SVP/acoustic prediction uncertainty reduction will have to be computed about

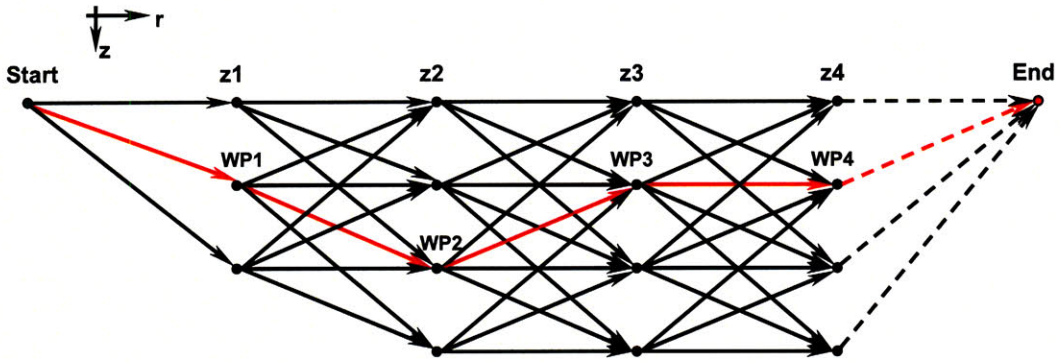


Figure 7-10: Illustration of a network graph and an AUV path. The AUV path is in red.

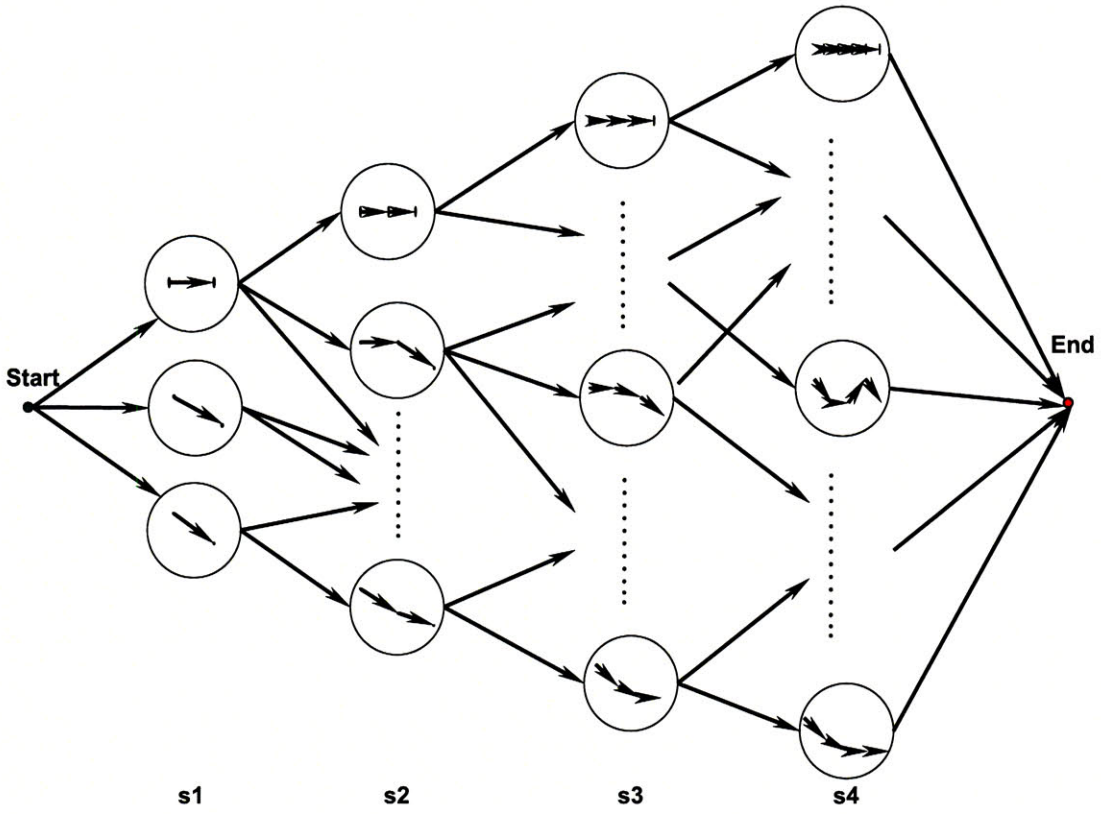


Figure 7-11: The graph associated with augmented decision variables.

200 times. This number is even bigger than the total number of possible paths in the original graph, which is equal to 139. So solving the AUV path planning problem by exhaustively searching all possible AUV paths is more efficient.

As discussed before, very large horizontal correlation lengths are not considered

in the AREA project. In fact, in shallow water area, due to the existence of many submeso-to-small-scale oceanographic processes such as internal waves, the Lr is often about 1 kilometer or less[14]. This fact implies that the individual significance of an in-situ measurement in the water column is only affected by other in-situ measurements nearby located within few kilometers. Out of the range, the influence will be weak (see Fig 7-12). Therefore, in decision variable augmentation, not all the preceding history $z_1, \cdots, z_{i-1}$ need to be included in s_i. For example, for the path segment (waypoint i, waypoint $i+1$) if we only take into account the influence from the preceding path segment (waypoint $i-1$, waypoint i) , i.e. only look back for 1 step, the augmented decision variable will be $\{s_1, s_2, \cdots, s_N\}$, where $s_1 = \{z_1\}$, $s_2 = \{z_1, z_2\}$, $s_3 = \{z_2, z_3\}$ and $s_N = \{z_{N-1}, z_N\}$. Also for the AUV path in Fig. 7-10, we have $R'_{2,3} \approx R_{1,2,3} - R_{1,2}$ and $R'_{3,4} \approx R_{2,3,4} - R_{2,3}$. The associated graph is shown in Fig. 7-13. The size of this graph is much smaller than that of the previous augmented graph. In fact, for a graph like Fig. 7-10, the number of total possible paths is $O\left(l^N\right)$; while the number of total nodes in Fig. 7-13 is $O\left(M \cdot N \cdot l^n\right)$ and the number of total arcs is $O\left(M \cdot N \cdot l^{n+1}\right)$, where the M is the average number of nodes at a range, the N is the number of total stages, l is the average number of outgoing arcs from one node and n is the number of steps looked back. To construct the partially augmented graph, the SVP/acoustic prediction uncertainty needs to be calculated for $O\left(N \cdot l^{n+1}\right)$ times. If $n < N$, then it can be $M \cdot N \cdot l^{n+1} << l^N$, which means that solving the shortest path problem defined in the partially augmented graph can be much faster than the exhaustive search.

The n-step look-back method gives us a way to convert the optimization problem defined in Eq. 5.44 and 5.45 into a shortest path problem with an approximate but additive objective function. This shortest path problem can be solved by some efficient algorithms such as the deterministic DP algorithm. It should be noticed that our problem is not a typical network optimization problem, in which the graph is given and the cost coefficients of all arcs are given too. While in our problem, the cost coefficients (i.e. lengths of arcs) are not given prior but must be calculated in a short time. So this problem is more like an engineering problem and in practice ,

we found that the bottleneck is just from the cost coefficients calculation in real-time. Compared with this, the time needed to solve the shortest path problem using deterministic DP algorithm can be ignored at all. So, N, l and n should be selected very carefully. The bigger the n, the approximation will be more accurate but the computation will be slower.

The above discussions are focused on the one way trip scenario, i.e. the AUV moves from the start point to the end point and then stops there. In practice, round trip scenario may be more realistic, in which the AUV moves from the start point to the end point and then makes a "U" turn to come back. In this scenario, both of the forward and backward AUV path should be optimized. The forward AUV path can be optimized using the n-step look-back method described above and the associated posterior SVP prediction uncertainty can be computed accordingly. Using this posterior SVP prediction uncertainty as the a prior SVP prediction uncertainty, the backward path can be optimized in the same way. Then the forward path can be refined again with taking into account the backward path. This process can be iterated for several times until it converges. A good sub-optimal result should be obtained then.

7.4 Solving the Adaptive On-board AUV Routing Problem

In this section, we are going to discuss how to solve the adaptive on-board AUV routing problem modeled in Chapter 6.1. For a graph such as Fig. 5-3 or a more complicated one, it is impossible to solve the AUV routing strategy optimization problem quickly and accurately on current PCs. In this DP problem, computation for the non-separable objective function is very intensive and the search space is now the AUV routing strategy space, which is huge. From another viewpoint, when the ocean and the sound velocity value are discretized in reasonably small resolutions, as defined in Chapter 6.1, the state space, the control space and the disturbance space

in the DP problem can be very big. The curse of dimensionality will be encountered.

Therefore, we wish some Approximate Dynamic Programming (ADP) methods could be helpful in this problem. Based on our experiences, however, neither of the temporal difference method and the Q-learning method etc. can solve this AREA DP problem in a way quick enough. In the adaptive on-board AUV routing problem, no closed mathematical form exists for the system dynamics and the objective function, hence only simulation-based methods [58] can be used. As mentioned before, Monte Carlo simulation-based estimation for the posterior $var\left(TL_r\right)$ usually takes more than 5 minutes for running once. Within about 4 hours, no more than 50 Monte Carlo simulations can be finished. This is often far away from convergence and thus simulation-based methods such as the temporal difference method and the Q-learning method are not practical in this problem. Some new faster methods have to be created.

The idea presented in this section is not to conquer the $\mathcal{NP}$-hard AREA DP problem by improving or speeding up any ADP algorithm. While, based on a static optimization method, a sub-optimal AUV path can usually be obtained prior for the SVP P.E. scenario. Since the sound velocity variation is usually no bigger than 1/100 of the mean value, it can thus be conjectured that the optimal AUV routing strategy may just lead the AUV to follow that sub-optimal AUV path in most scenarios but deviate a little bit in some specific cases. Therefore, all candidates of AUV path segment can be restricted in the neighborhood of the sub-optimal AUV path (see Fig. 7-15). The AREA DP problem can thus be much simplified and the dimensionality can be dramatically reduced so that a quick solution is possible. According to this idea, the result will be an sub-optimal AUV routing strategy that is theoretically guaranteed to be better than the preceding sub-optimal AUV path. In practice, however, due to the computation noise, the sub-optimal AUV routing strategy may not work as well as the expectation.

To explain the details in the idea, let's first investigate a very simple example to get some intuitions. In Fig. 7-14(a), the black line is an AUV path. Now let's suppose that the CTD only takes two in-situ measurements at $(0\ km, 0\ m)$ and $(7\ km, 100\ m)$.

When AUV finishes the second in-situ measurement, there are two possible paths to choose, the red curve and the blue curve. The decision should be made based on the two CTD data. We can simulate this process for many in the AREASF with a SVP prediction model generated by HOPS/ESSE and generate a lot of training data. Now, it is supposed that the simulation result is like the one shown in Fig. 7-14(b), in which when the 1st in-situ measurement value is bigger than the 2nd one, the red curve will most likely lead to a better result than the blue curve; while when the 2nd in-situ measurement value is bigger than the 1st one, then the blue curve will most likely be the better one. Therefore, in the real AREA operations, the AUV can take the red curve when the 1st in-situ measurement gives bigger value and take the blue curve in the other situation. Consequently, the final result will be better than sticking with either the red curve or the blue curve no matter of the in-situ measurement data. If the black line + the red curve is an optimized AUV path, then the above method provides a way to improve it.

Now let's analyze the above method using Q-factor approximation [47]. There are totally 2 stages in the problem. The stage 0 is at the origin, the stage 1 is at the range of 7 km and the stage 2 is at the range of 10 km. At stage 2, the $g_2(x_2)$ is equal to the posterior $tr(var(TL_r))$. From Eq. 4.46, we have

$$Q_1^*(x_1, u_1) = E_{\omega_2}\{g_2(x_2)\}, \tag{7.1}$$

where x_1 is a 2-dimension vector containing the two in-situ measurement data, u_1 is the selection between the red curve and the blue curve, the cost per stage is equal to 0. x_1 has a continuous space. In Fig. 7-14(b), it can be seen that for a certain value of x_1, there may not be enough training data in its neighborhood and thus the expectation in Eq. 7.1 is hard to be achieved. If the dimension of x_1 is higher, this problem will be more severe. In fact, we can replace the expectation with a single sample as used in the Q-learning method [58],

$$Q_1^*(x_1, u_1) \approx g_2(x_2). \tag{7.2}$$

From Eq. 4.47, the optimal AUV path selection is

$$\mu_1^*(x_1) = \arg\min_{u_1} Q_1^*(x_1, u_1) \tag{7.3}$$

$$= \arg\min_{u_1} g_2(x_2). \tag{7.4}$$

However for a certain value of x_1, there may be no any associated training data. In AREA operations, once a x_1 value is given, we can look around this value and check the k nearest points in its neighborhood in Fig. 7-14(b) and then make decision, i.e. if the red curve is a better choice for most points in the k-nearest-neighbor, then select the red curve, and vice versa. Doing so, the AUV will take the red curve in most scenarios when the 1st in-situ measurement value is bigger and take the blue curve in the opposite situations. This 2-stage DP problem can be viewed as an example of pattern recognition. In the first segment of AUV path, environment information is collected, then the environment pattern is recognized. For different pattern, different selection will be taken. The preceding method by comparing the values of the 1st and the 2nd in-situ measurement is a linear classification method and the k-nearest-neighbor method is a non-linear classification method [67].

Due to the non-linearity in the k-nearest-neighbor method, we can replace it with the following method,

$$\mu_1^*(x_1) = \arg\min_{u_1} \sum_{i=1,2,\cdots,N} f\left(D(x_1, x_1^i)\right) \cdot Q_1^*(x_1^i, u_1), \tag{7.5}$$

where, N is the number of training data; x_1^i is the realization of x_1 in the ith training datum; $D(x_1, x_1^i) = \|x_1 - x_1^i\|^2$ is the distance between x_1 and x_1^i; $f(d) = \exp(-d^2/L)$ is the weight function, which is exponentially decreasing as d is increasing. By increasing (decreasing) L, we can increase (decrease) the influence from the nearest points in the neighbor. This method can be implemented easily and quickly, and it is somehow like the kernel method in Data Mining (DM) [67]. Moreover, from some perspective, the summation in Eq. 7.5 can be viewed as doing expectation around the x_1.

In the real AREA operations, a sub-optimal AUV path can be firstly obtained via the n-step look-back method. Based on this path, some "branches" can be made empirically, as shown in Fig. 7-15. Those branches are usually better than the original path in some scenarios. The above method can then be used to construct a sub-optimal AUV routing strategy. It should be noticed that in the real situation the CTD will collect data every seconds. If all of those data collected in a path segment is put into x_i, the dimensionality of the state space will be extremely large. The practical way to reduce the dimensionality is to replace those original data with features, e.g. those original CTD data can be divided into n groups and then do average in each group, the n averaged values can be used as the features and put into x_i.

Furthermore, for N-stage AREA DP problem, we have

$$Q^*_{N-1}(x_{N-1}, u_{N-1}) \approx g_N(x_N), \tag{7.6}$$

$$Q^*_k(x_k, u_k) \approx \min_{u_{k+1}} Q^*_{k+1}(x_{k+1}, u_{k+1}), \tag{7.7}$$

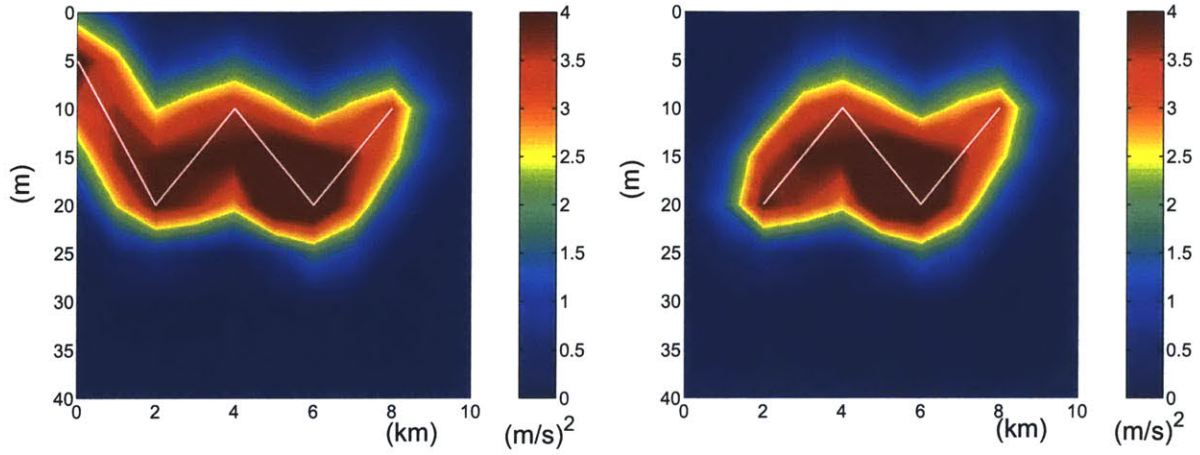
where the cost per stage $g_k(x_k, u_k, \omega_k)$ is equal to 0. In practice, the Eq. 7.5 can be adapted as

$$\mu^*_k(x_k) = \arg\min_{u_k} \left(\sum_{i=1,2,\cdots,N} f\left(D(x_k, x^i_k)\right) \cdot Q^*_k(x^i_k, u_k) + h(u_k) \right). \tag{7.8}$$

Here, if u_k is to take the original sub-optimal path, then $h(u_k) = 0$; otherwise $h(u_k) > 0$. By this we mean that we want to improve the original sub-optimal path conservatively. If a new path can not excess the original one over a certain threshold, the original sub-optimal path will still be selected. Doing so, we can avoid to select a really bad path segment, which however happens to look like very good due to the noises in the computation noise.

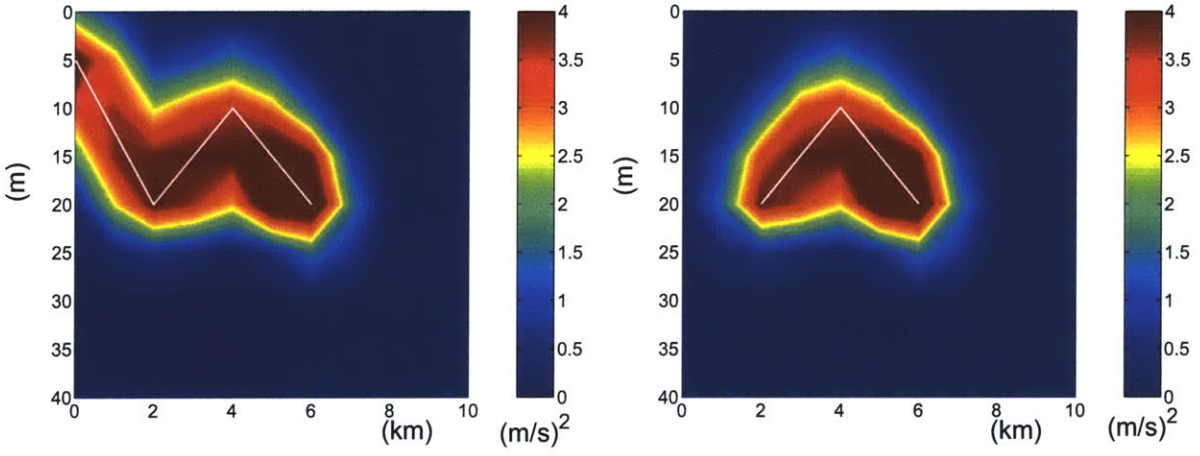
In AREA operations, those training data will be prepared off-line, i.e. generated before the AUV is launched. Then, the training data will be uploaded to the AUV. When it is running, the AUV will only do sound velocity sampling via the CTD

sensor, extracting features (the n averaged values) and solving the Eq. 7.8. This is a very fast on-line process.



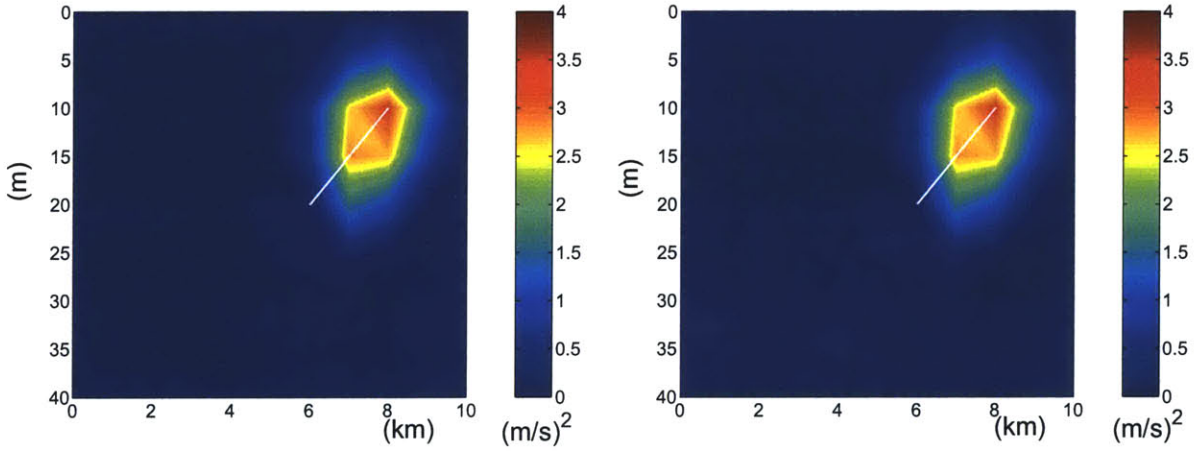
(a) SVP pred. unc. red. Path segment 1,2,3,4.

(b) SVP pred. unc. red. Path segment 2,3,4.



(c) SVP pred. unc. red. Path segment 1,2,3.

(d) SVP pred. unc. red. Path segment 2,3.



(e) Contribution from segment 4.

(f) Approximate contribution from segment 4.

Figure 7-12: (a) is the SVP prediction error variance reduction associated with path segment 1,2,3,4. (b) is the one associated with path segment 2,3,4. (c) is the one associated with path segment 1,2,3. (d) is the one associated with path segment 2,3. (e) is equal to (a)-(c). (f) is equal to (b)-(d). It can be seen that (e) is almost the same as (f), since the segment 1 is far away from the segment 4.

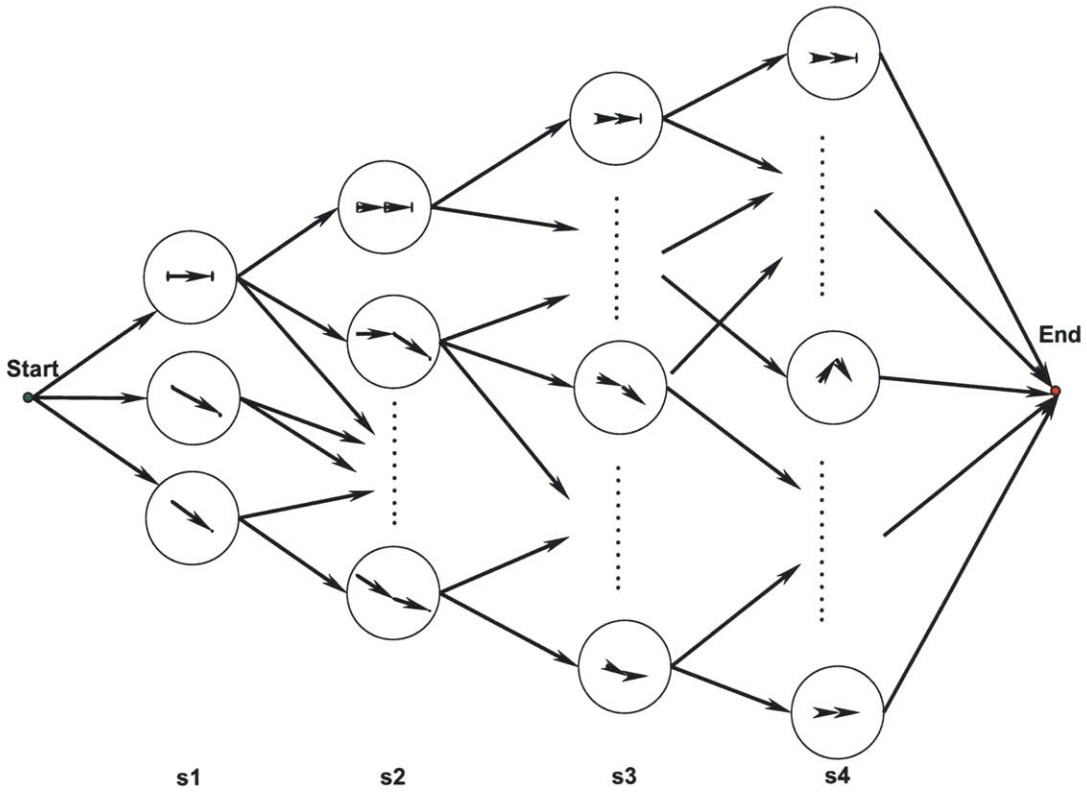
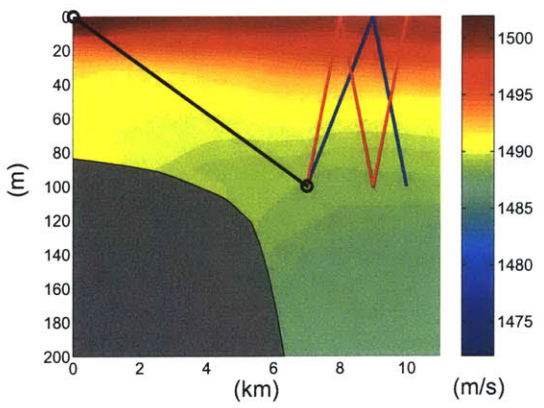
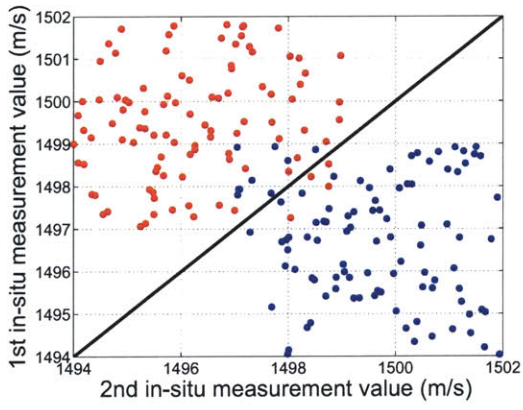


Figure 7-13: The graph associated with 1-step look-back method.



(a)



(b) One dot indicates on SVP scenario. Red dot means that in this situation the red curve gives a better result than the blue curve; blue dot indicates the opposite situation.

Figure 7-14: Illustration of a 2-stage adaptive on-board AUV routing problem.

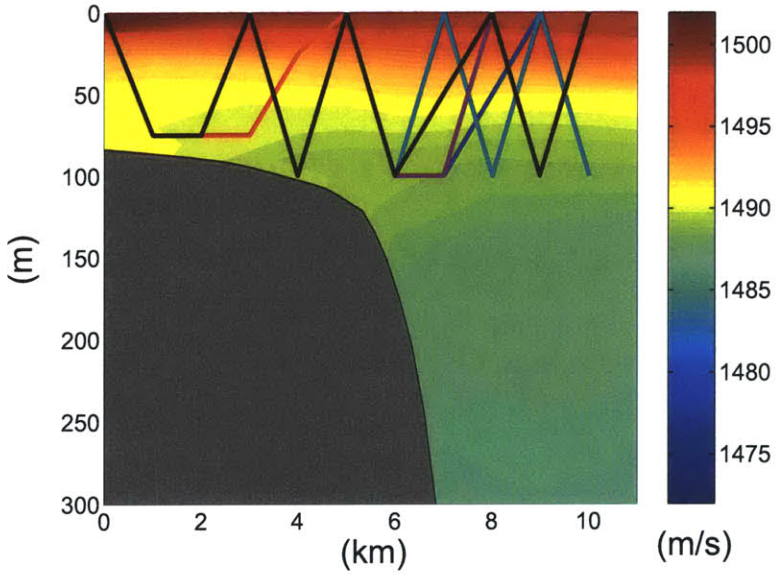


Figure 7-15: Illustration of adding branches to a sub-optimal AUV path. The black curve is the sub-optimal path and all the other curves are branches made empirically.

Chapter 8

Capturing Fronts

As introduced in Chapter 2.1.2, an oceanic front is the interface between two water masses of different properties. Usually, fronts show strong horizontal gradients of temperature, salinity and sound velocity, thus causing changes in acoustic propagation. Fronts can be predicted by HOPS/ESSE, but only in a coarse resolution. Capturing fronts locally is very meaningful for oceanographic research and underwater acoustic research. It can be viewed as an extension of the AREA project.

Fig. 8-1 shows a front prediction from HOPS/ESSE. To get more local information about the front on the surface, an AUV carrying a CTD sensor can go back and forth across the front with doing in-situ measurements. Those data can then be assimilated by HOPS/ESSE to get a better front estimation. Based on the front prediction, a predetermined AUV path can be made empirically such as the one shown in Fig. 8-1. This path may not be very efficient for capturing fronts. In next sections, several ways to develop more efficient adaptive AUV path crossing the front will be introduced.

8.1 Tracking the 2-D Temperature Gradient at A Constant Depth

One interesting way to track the front is to follow the azimuth of the temperature horizontal gradient at a constant depth. The AUV can stay on the surface and

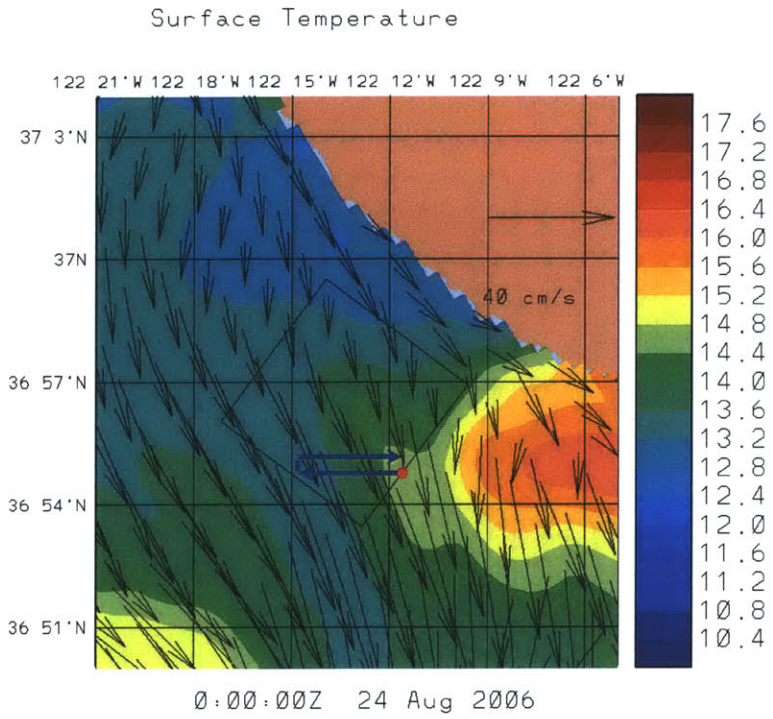


Figure 8-1: Illustration of a front prediction from HOPS/ESSE and a predetermined AUV path for capturing the front.

start moving in an initial azimuth, which could be the direction of the horizontal gradient of the surface temperature predicted by HOPS/ESSE. The CTD will collect temperature data every second. We can use every p data and the associated location information to estimate the horizontal gradient of temperature through the *Linear Least Square Fitting* method. This process can be done in real-time and once the new gradient azimuth is obtained, the AUV will turn to that direction. However, it should be noticed that the p points could happen to locate in a line, which is parallel to the front by accident. In this situation, the gradient estimation error will be very big. To avoid this, the AUV will actually move zigzag or sinusoidally in each small path segment but with the gross direction being the current gradient azimuth. In this way, the AUV will almost always move in the direction perpendicular to the front (see Fig. 8-2(a)). Thus, it is a more efficient way to go across the front.

8.2 Capturing Fronts by Horizontal Zig-zag Control

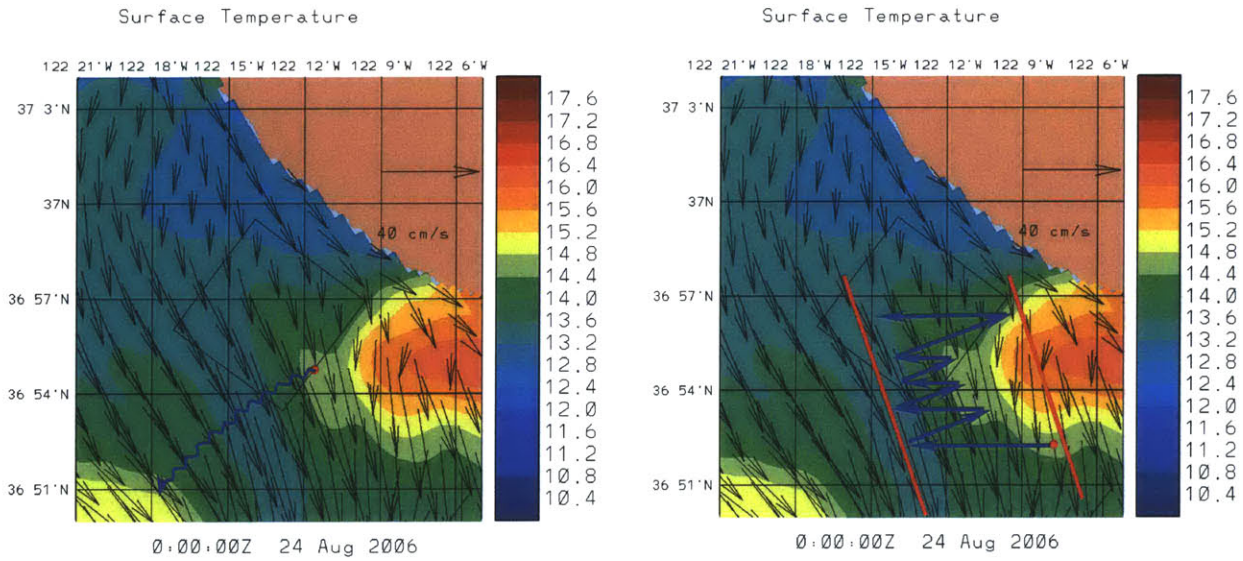
Similar to the adaptive AUV yoyo control in AREA, an adaptive horizontal AUV yoyo control with respect to temperature at a constant depth can be developed. This horizontal zig-zag control can make AUV move focusing on the front region. For example, in Fig. 8-2(b) the AUV starts from the red point, the two red lines are two boundaries and the AUV can only move in the region between them. Starting from the red point, the AUV will move in an predetermined azimuth. Once the temperature difference between the current location and the starting location is over a threshold (Eq. 8.1), the AUV will make a turn to another predetermined azimuth.

$$\Delta T = |T_s - T_c| > h, \tag{8.1}$$

where T_s is the temperature at the starting point, T_c is the temperature at the current point, h is a threshold. The turning point will then be updated as the new starting point. If the AUV hit any boundary, it will have to make a turn there. By repeating this process for a while, the AUV will move in a way keeping crossing the front.

8.3 Capturing Fronts by Horizontal Zig-zag and Vertical Yoyo Control

The preceding adaptive horizontal AUV zig-zag control is only for AUV moving at a constant depth. In fact, this horizontal zig-zag control can be easily combined with the adaptive vertical yoyo control introduced in Chapter 6.3. The AUV's motion will then be a 3-D yoyo track. Looking from up to down, the AUV's horizontal motion will be a horizontal zig-zag; while looking from side, the vertical motion will be a up-and-down yoyo. In this case, the temperature at many depths will be measured and thus in Eq. 8.1, ΔT can be the temperature difference at many depths.



(a) Tracking the 2-D temperature gradient.

(b) Adaptive horizontal zig-zag control.

Figure 8-2: Two adaptive AUV paths for capturing fronts.

8.4 Other Applications of AUV Yoyo Control

In this chapter and Chapter 6.3, several different AUV yoyo controls are discussed. Besides those applications, the vertical AUV yoyo control can be used in a non-adaptive mode by setting the threshold infinite, which will make the AUV just bounce between the upper and lower bound while moving forward. If the non-adaptive vertical yoyo control is combined with a circular horizontal motion, the AUV carrying CTD can then be used to capture internal waves.

Moreover, the non-adaptive yoyo mode can collaborate with the adaptive yoyo mode. For example, the AUV can firstly make several dives from the upper bound to the lower bound, during which the local vertical gradient of sound velocity can be measured and estimated. This result can be used to tune the control parameters in the adaptive yoyo mode. After those parameters get tuned, the AUV can work under the adaptive yoyo control. This mixed mode provides a completely automatic AUV control, which can be very useful in the scenario that the a priori ocean prediction is unavailable.

Chapter 9

Monterey Bay 2006 (MB06) Experiment

In this chapter, we are going to introduce the MB06 experiment. To test all methods, controls and ideas developed in this thesis and to integrate the whole AREA system, two entire at-sea experiments were carried out. The first one is the FAF05 experiment in the northern Tyrrhenian Sea in 2005 and the latest one is the MB06 experiment from Aug. 15 to Aug. 25, 2006, in the Monterey Bay, CA. This was a major field experiment sponsored by the Office of Naval Research. In next sections, the major achievements made in MB06 will be introduced in the order of precedence and the FAF05 real-time AREA simulations will be briefly introduced at the end.

9.1 Daily TL_r Forecasts

The topography of the experiment area is shown in Fig. 9-1. The center point of the experiment is located at latitude 36.9414^o, longitude -122.2232^o. The 3-D ocean environment was divided into 8 bearings. During MB06, on each day a SVP forecast ensemble for each bearing was generated by HOPS/ESSE for the next 36 hours and the associated seabed models were provided by Mike Porter. The example of Aug. 25, 2006 is shown in Fig. 9-2 to Fig. 9-9.

A SVP forecast ensemble includes not only a SVP principal estimation but also

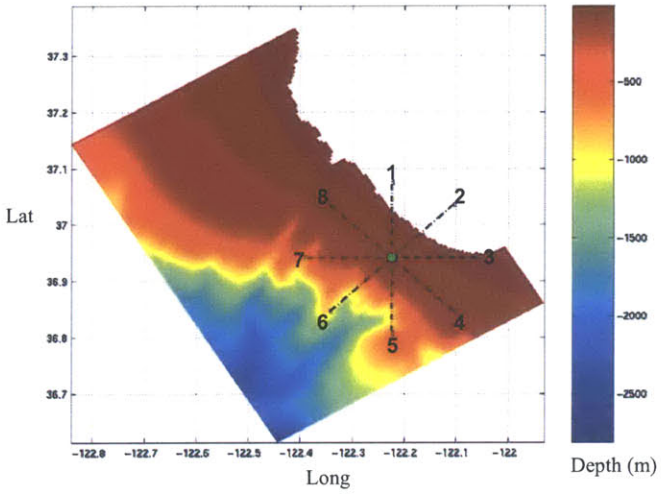
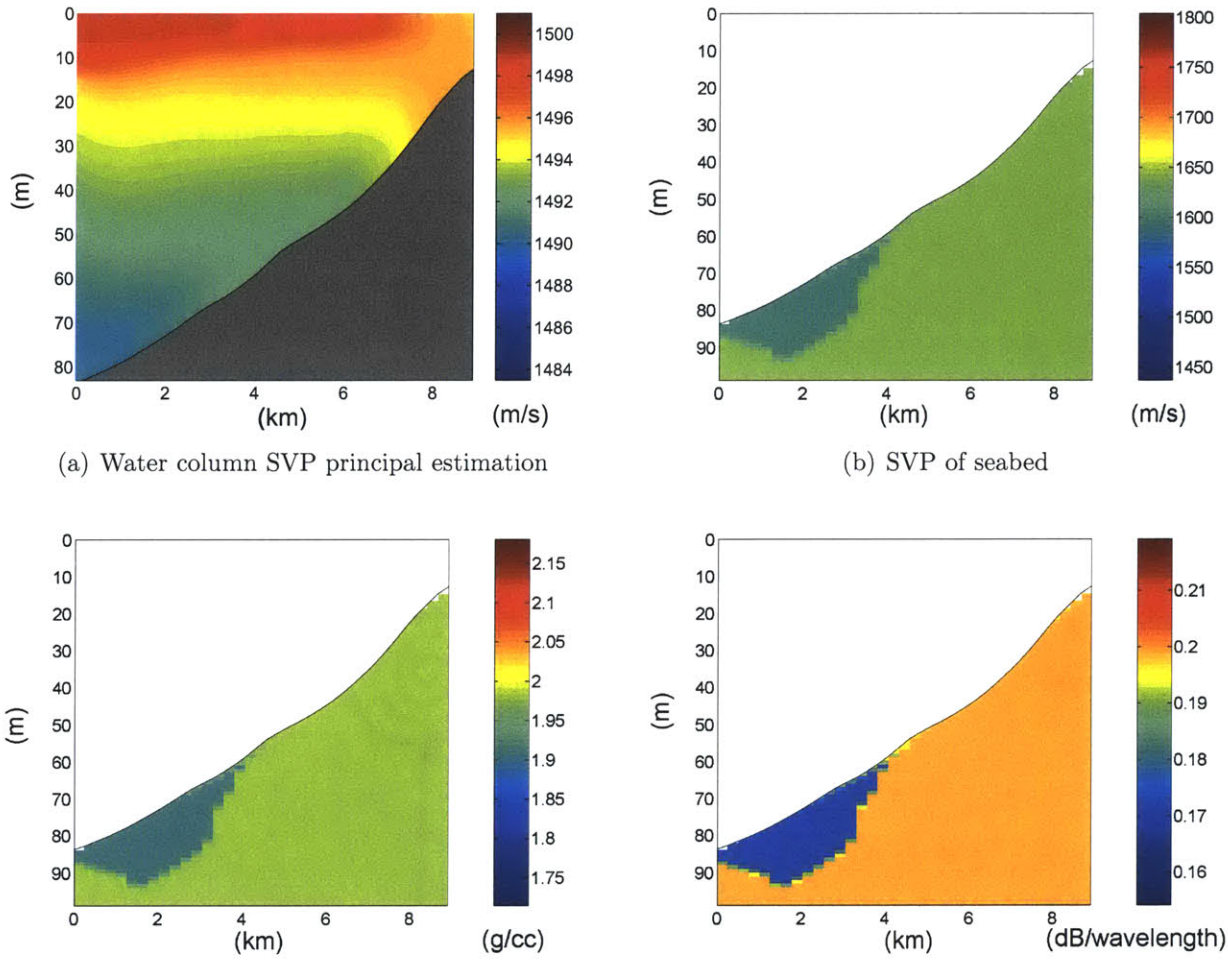


Figure 9-1: Topography of Monterey bay, CA.

around 20 SVP realizations for each bearing. On the basis of the SVP forecast ensembles, the TL_r forecasts were generated every day for the 8 bearings, 2 frequencies (100 Hz and 400 Hz), 3 sound source depths (5m, 40m, 80m) and 3 receiver depths (15m, 45m, 75m). The acoustic model was the RAM PE and the computations were finished in about 2 hours. To assure the real-time feasibility, the parameters in the RAM PE code must be selected very carefully. In our experiences, the parameter dz in RAM is the keyest one and it should be set small enough. Some examples of TL_r forecast on Aug. 25, 2006 are shown in Fig. 9-10 and Fig. 9-11, in which the frequency bandwidth coefficient $\alpha = 0.1$. It should be pointed out that **MB06 is the first time that the acoustic field forecasts have been linked with the ocean environment forecasts in real-time**.

Once the TL_r forecasts are obtained for all bearings, the one associated with the biggest TL_r prediction uncertainty can be known. In MB06, it was usually the bearing 5 or bearing 6. Both of them cross the shelf break.



(a) Water column SVP principal estimation

(b) SVP of seabed

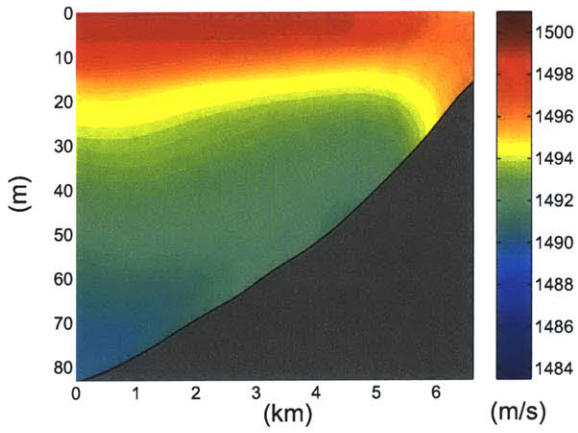
(c) Density profile of seabed

(d) Attenuation coefficient profile of seabed

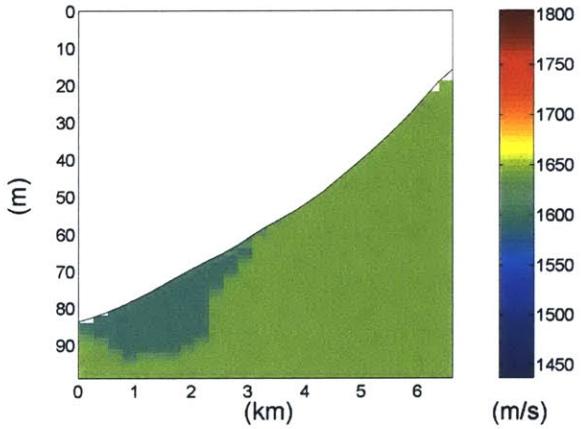
Figure 9-2: Bearing 1, Aug. 25, 2006.

9.2 Estimate the Lr and Lz

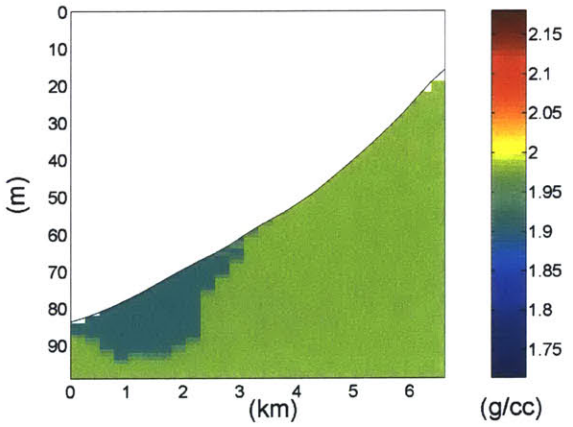
Lr and Lz are respectively the horizontal and vertical correlation length of the SVP. As aforementioned, Lr is very crucial to determine the way to capture the environment uncertainties. In MB06, Lr and Lz were determined semi-empirically in the following way. Firstly, a depth range, e.g. from 0m to 20m was taken and a non-linear optimization code was run to optimize the Lr and Lz for that depth range. The objective function is the summation of differences between the Λ_{c_1} estimated by HOPS/ESSE and the one given by Eq. 5.16. Then, some other depth ranges were



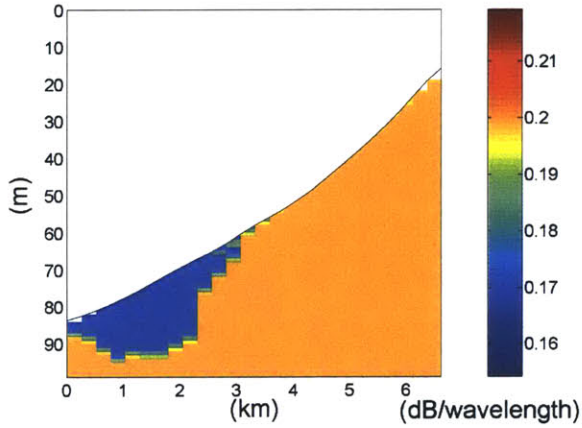
(a) Water column SVP principal estimation



(b) SVP of seabed



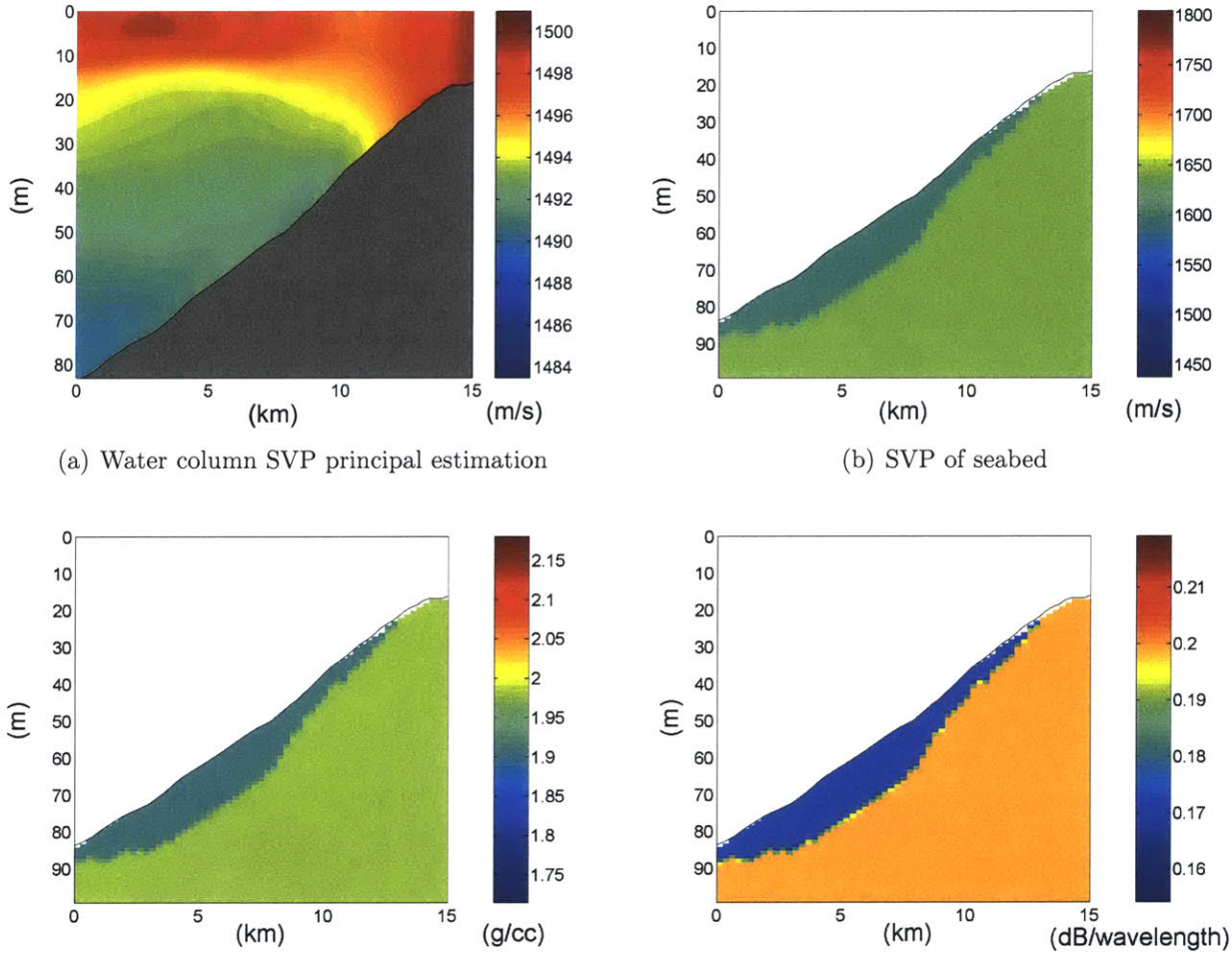
(c) Density profile of seabed



(d) Attenuation coefficient profile of seabed

Figure 9-3: Bearing 2, Aug. 25, 2006.

taken and the corresponding optimized Lr and Lz could be obtained. Among them, the smallest Lr and Lz would be the one used in operations, since they were associated with the smallest scale oceanographic process. Also, if taking into account the internal waves, the operational Lr used in MB06 might be even smaller. Generally, in shallow water area, internal waves can make the horizontal correlation length no longer than 1km.



(a) Water column SVP principal estimation

(b) SVP of seabed

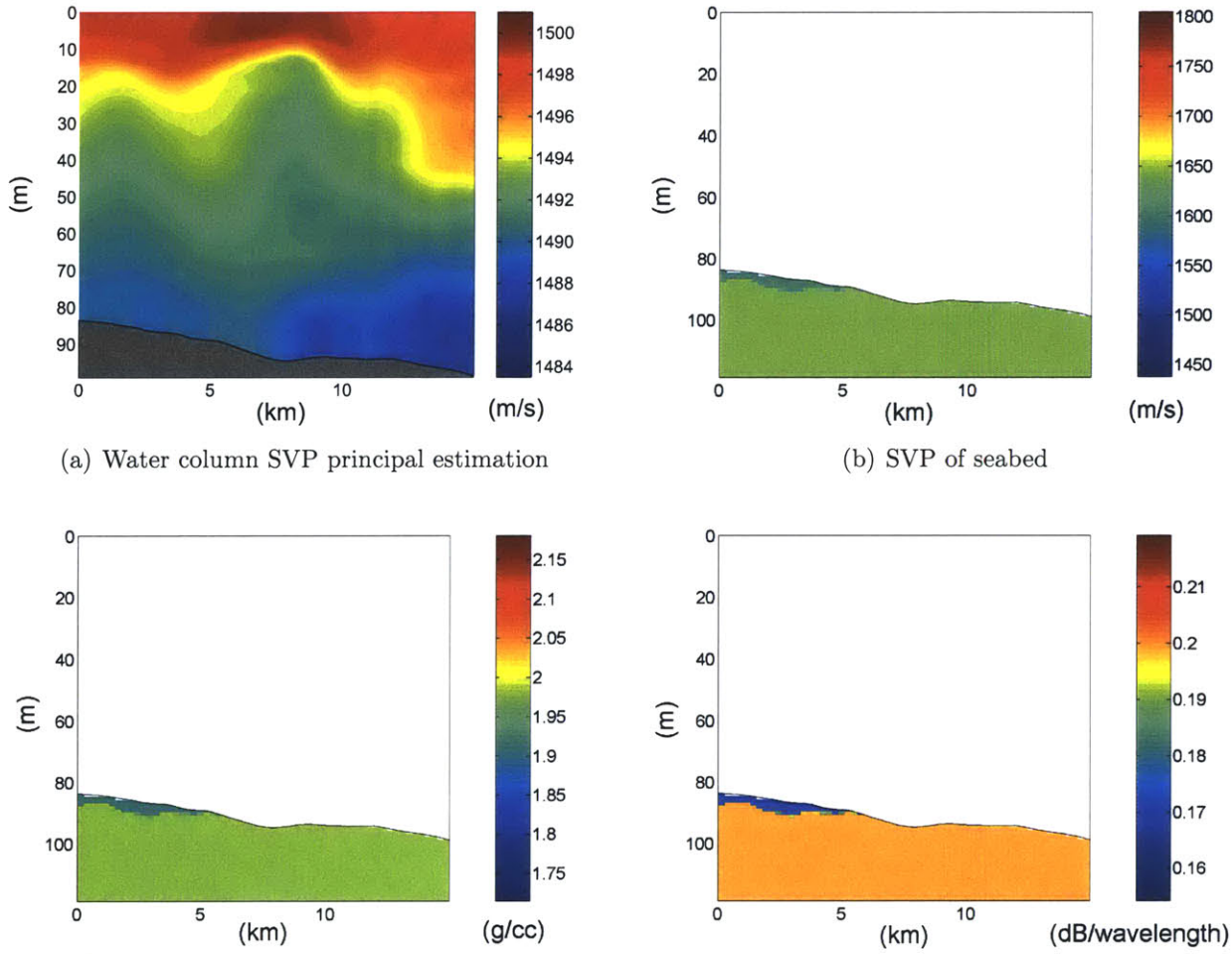
(c) Density profile of seabed

(d) Attenuation coefficient profile of seabed

Figure 9-4: Bearing 3, Aug. 25, 2006.

9.3 AUV Path Planning and Adaptive On-board AUV Routing

Sub-optimal AUV paths for capturing SVP/acoustic prediction uncertainty were generated almost everyday during MB06. The results of AUV path planning for capturing SVP prediction uncertainty on Aug. 21, 2006 are shown in Fig. 9-12 and Fig. 9-13. Fig. 9-12(a) is the principal estimation of SVP, Fig. 9-12(b) shows the associated error standard deviation field. Fig. 9-12(c) shows the ocean discretization grid for AUV path and the associated path graph. In this case, the maximum range for AUV



(a) Water column SVP principal estimation

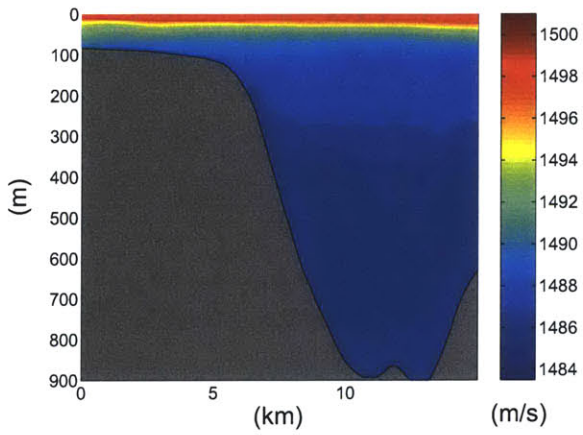
(b) SVP of seabed

(c) Density profile of seabed

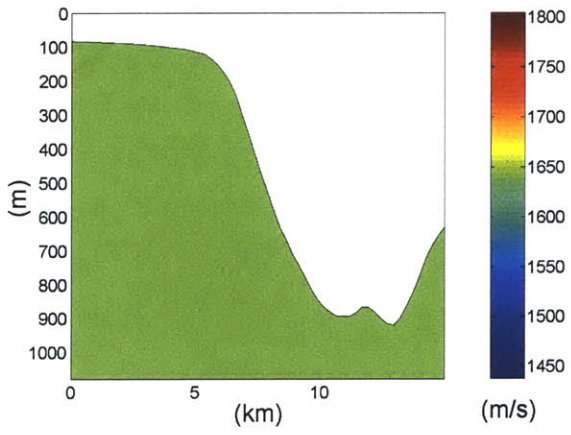
(d) Attenuation coefficient profile of seabed

Figure 9-5: Bearing 4, Aug. 25, 2006.

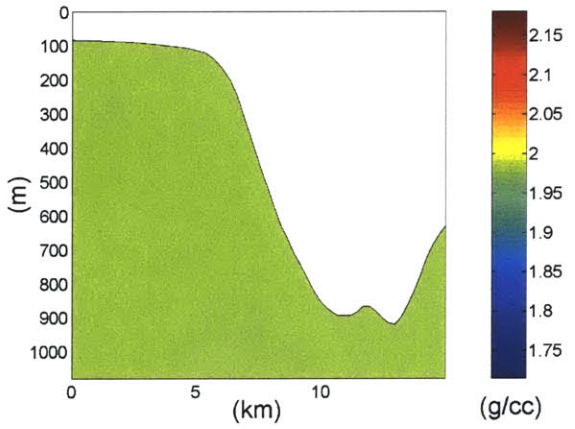
is 10km, the upper bound is 5m, the lower bound is 80m. The horizontal interval between two waypoints is 1km, the vertical interval is about 9m. Fig. 9-12(c) shows the sub-optimal AUV path produced by 1-step look-back method, which took about 30 seconds. Fig 9-12(e) shows the sub-optimal AUV path produced by 2-step and 3-step look-back methods. Both of them gave the same result. The 2-step look-back method took about 7 minutes, while the 3-step look-back method took about 2 hours. In this case the sub-optimal paths from n-step look-back method is very close to the non-adaptive up-and-down yoyo (Fig. 9-12(f)) constrained by the ocean discretization grid for AUV path. In this grid, the maximum pitch angle for AUV is about 4.3^o.



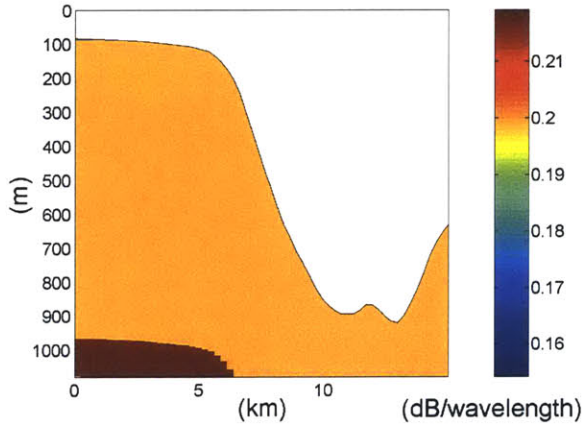
(a) Water column SVP principal estimation



(b) SVP of seabed



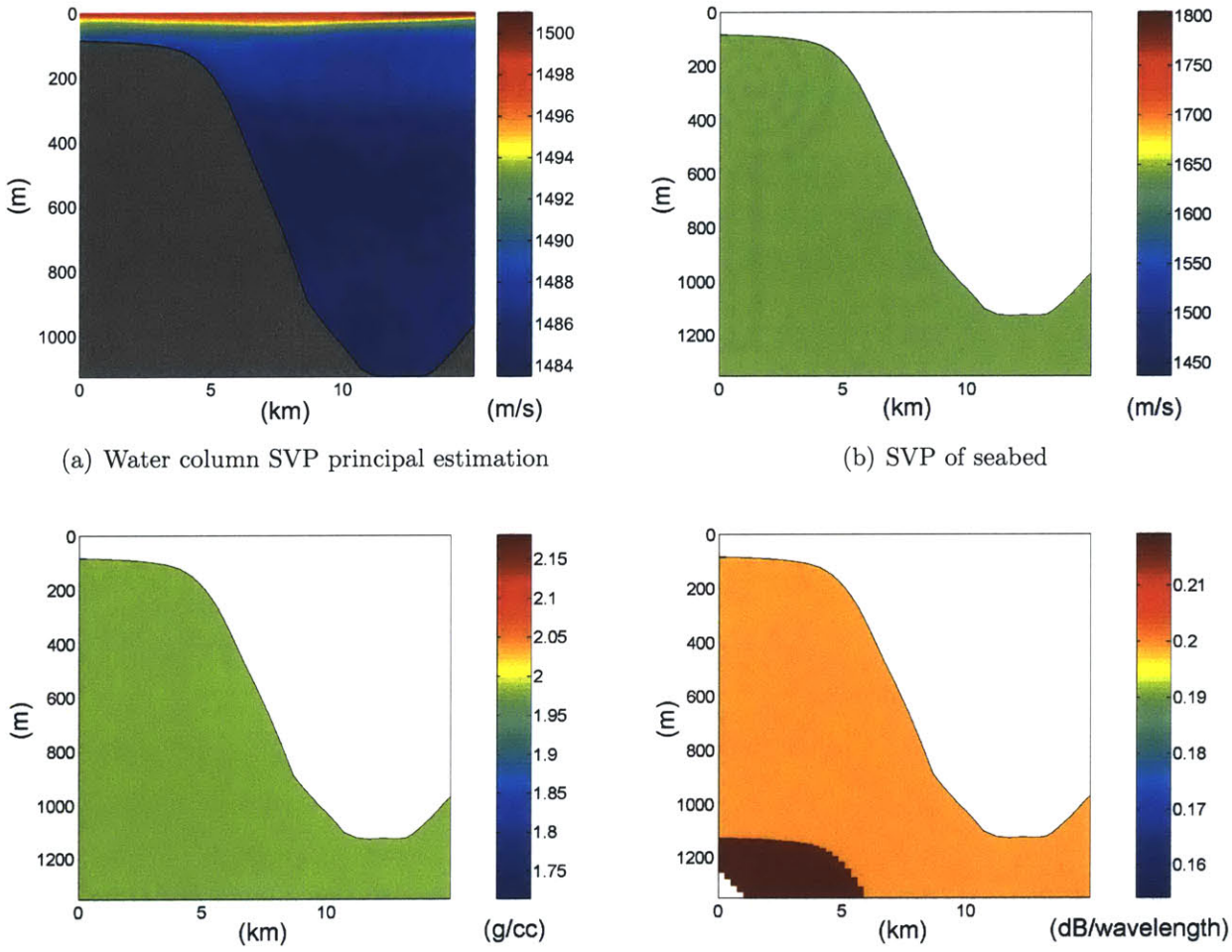
(c) Density profile of seabed



(d) Attenuation coefficient profile of seabed

Figure 9-6: Bearing 5, Aug. 25, 2006.

While in MB06, the AUV pitch angle can be as large as 10^o, i.e. due to the computation capability, the horizontal interval between two AUV waypoints can not be set very small, as a result the maximum AUV pitch angle has to be sacrificed for several degrees. Fig. 9-13(a) shows a predetermined path which is set empirically. Since in Fig. 9-12(b) the biggest error is around the depth of 25m, the predetermined path was set to stay in this region and was expected to capture a lot of SVP prediction uncertainty. The SVP prediction uncertainty reductions associated with all paths mentioned before are shown in Fig. 9-13(b). In MB06, the SVP in the bearing 5 is discretized in the way shown in Fig. 9-14(a) and there are totally 352 points. From



(a) Water column SVP principal estimation

(b) SVP of seabed

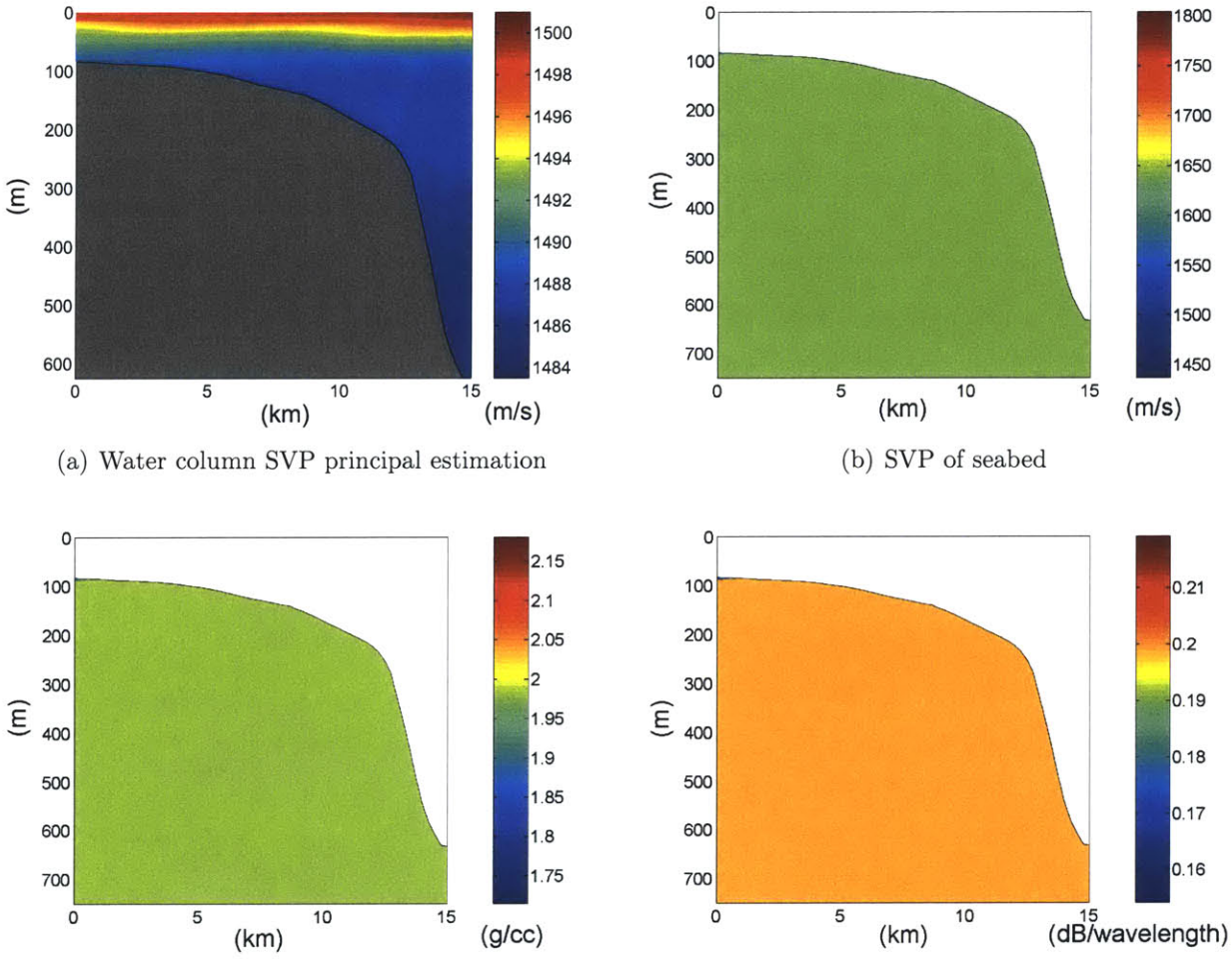
(c) Density profile of seabed

(d) Attenuation coefficient profile of seabed

Figure 9-7: Bearing 6, Aug. 25, 2006.

the results comparison, it can be seen that uncertainty reduction associated with the empirically predetermined path is the least. While the other three produce almost the same results.

The results of AUV path planning for capturing TL_r prediction uncertainty on Aug. 21, 2006 are shown in Fig. 9-15. Fig. 9-15(a) and Fig. 9-15(b) are the result of 1-step and 2-step look-back method respectively. Fig. 9-15(c) shows all the 6 path candidates for AUV routing strategy optimization. Those path candidates were constructed based on the path in Fig. 9-15(b). The TL_r prediction uncertainty reduction comparison is shown in Fig. 9-15(c), in which the green bars indicate the



(a) Water column SVP principal estimation

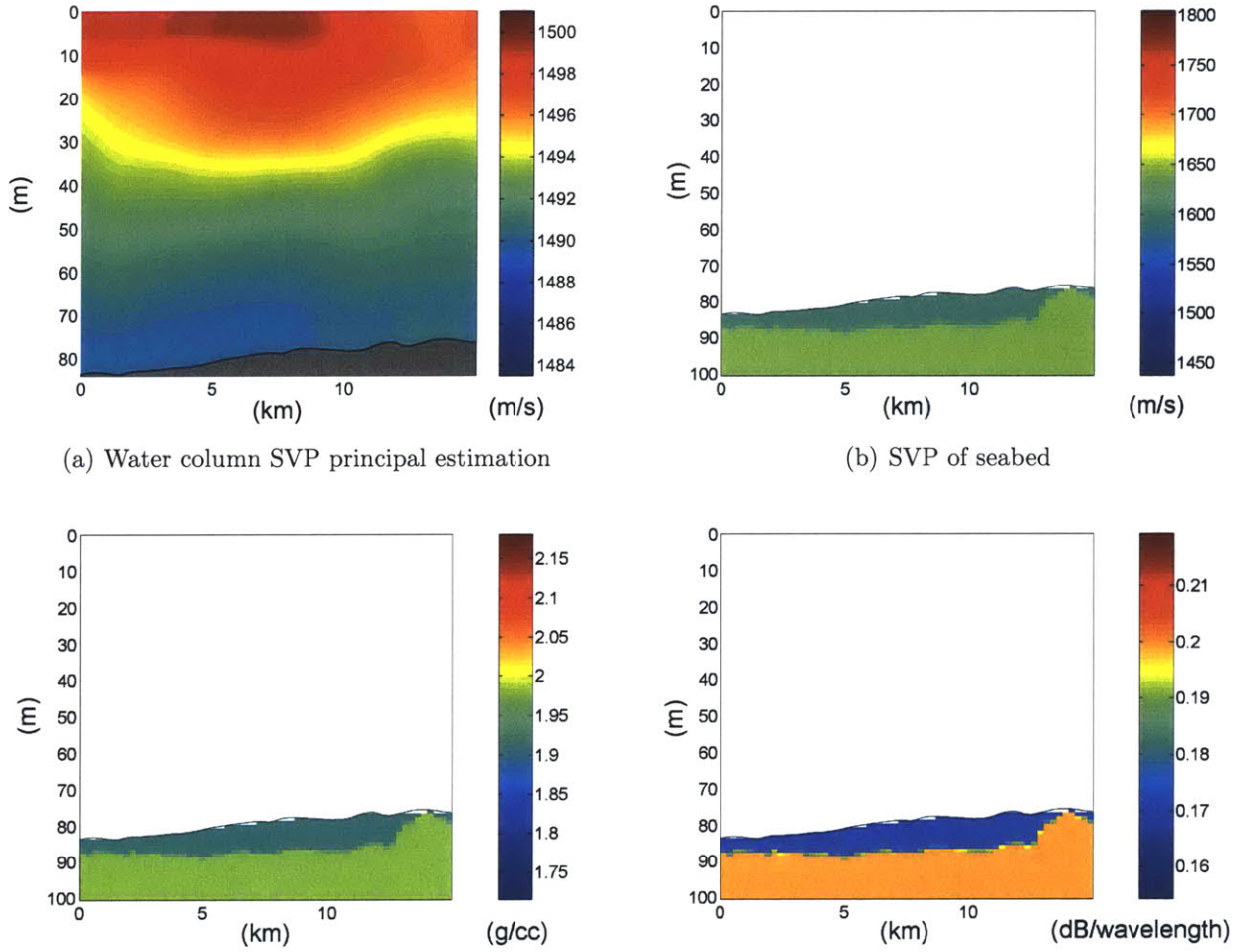
(b) SVP of seabed

(c) Density profile of seabed

(d) Attenuation coefficient profile of seabed

Figure 9-8: Bearing 7, Aug. 25, 2006.

results generated by the method based on TL_r linear approximation and the blue bars indicate the results generated by 200 Monte Carlo simulations. The ocean discretization grid for TL_r and OA are shown in Fig 9-14(b) and 9-14(c). In this case, the results from the n-step look-back method are also very close to the non-adaptive yoyo. In Fig. 9-15(d), it can be seen that the 1-step and 2-step look-back methods and the non-adaptive yoyo lead to very close results in both green bars and blue bars, but the 2-step look-back method is a little bit better than the other two. While, the predetermined path is the worst one again. The AUV routing strategy can only be tested by Monte Carlo simulation-based method and its result is even better than the



(a) Water column SVP principal estimation

(b) SVP of seabed

(c) Density profile of seabed

(d) Attenuation coefficient profile of seabed

Figure 9-9: Bearing 8, Aug. 25, 2006.

2-step look-back method's result. This is consistent to the preceding analysis. However in practice, it was found that in most Monte Carlo simulations the AUV routing strategy selected the same path as in Fig. 9-15(b). Only in a few cases, the AUV routing strategy selected different paths, which were better than the 2-step look-back result in those cases.

From the results comparison, there's another phenomenon can be observed: if the green bar is taller, the associated blue bar is taller too. This is consistent to the conclusion made about the TL linear approximation method — the TL uncertainty computed via the linear approximation method is highly positively correlated to the

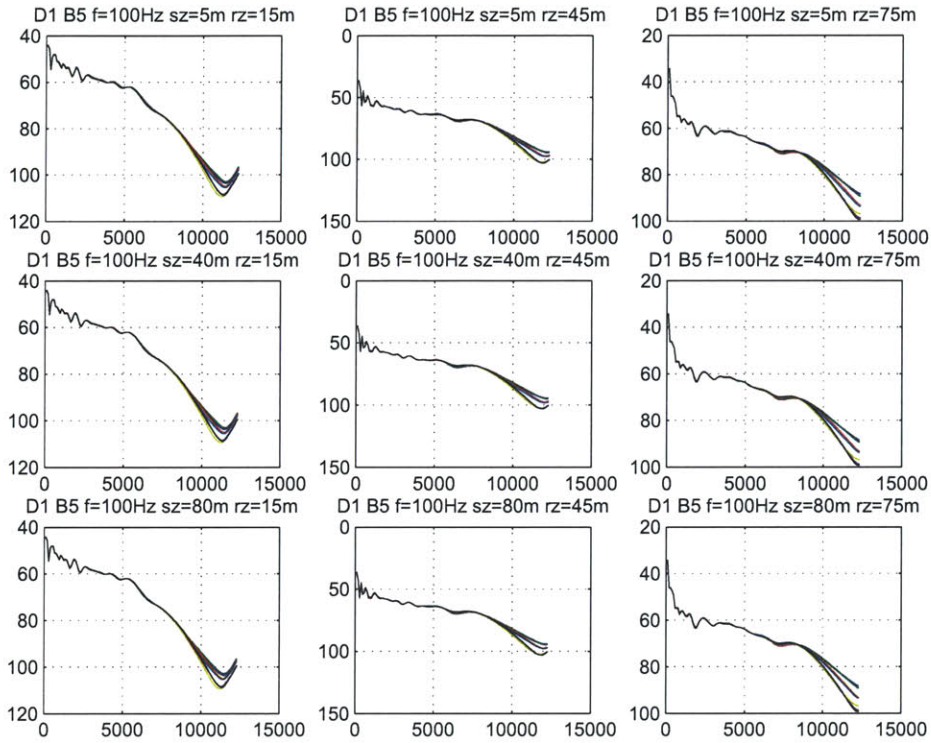


Figure 9-10: TL_r forecasts v.s. range on Aug. 25, 2006. B5 means the bearing 5; f, sz, rz are frequency, source depth and receiver depth respectively. D1 means that this forecast is for the first 12 hours.

one computed via Monte Carlo simulations.

For the bearing 5 on Aug. 21, 2006, if more internal waves existed, the horizontal correlation length could be as short as few hundred meters. Now, let's suppose $Lr = 400m$ and see what kind of results will be obtained. Fig. 9-16 shows the sub-optimal paths for capturing SVP prediction uncertainty and the results comparison. In this case, the optimized path is not like the non-adaptive yoyo any more, but focus more on the top layer, where the SVP prediction error is bigger. When Lr is very large, one in-situ measurement will reduce uncertainty in a large neighborhood. The shape of this area is like a very flat ellipse, long in range and short in depth. If we can move this ellipse up and down, then a lot of SVP prediction uncertainty will be reduced. Therefore, the optimal AUV path will be very possibly like a non-adpative yoyo. When Lr is shorter, the ellipse will be slimmer and then the optimal AUV

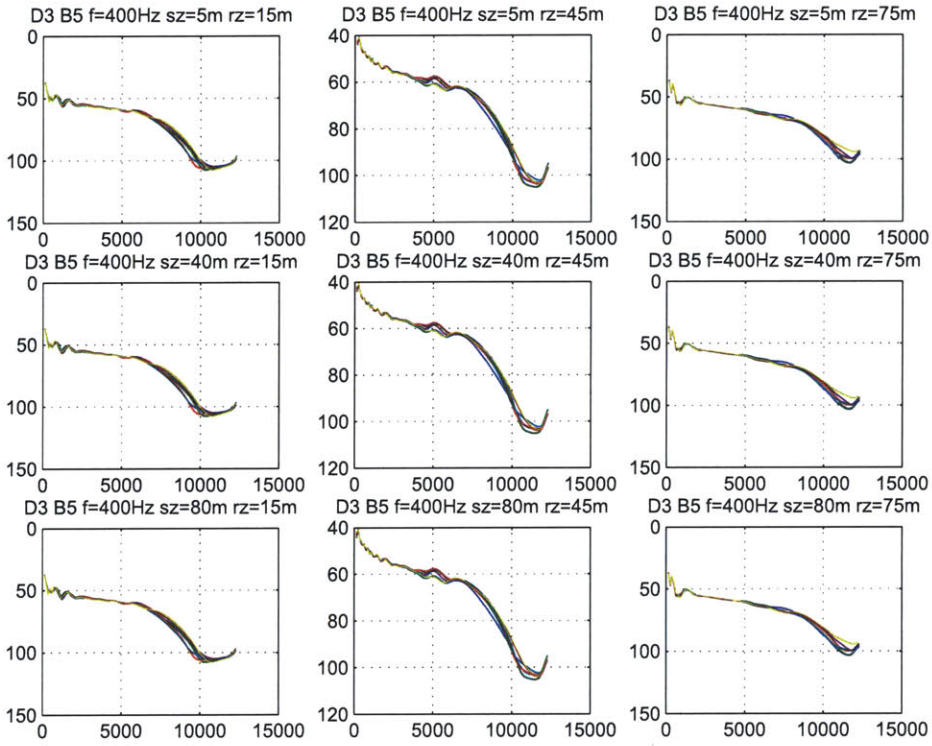
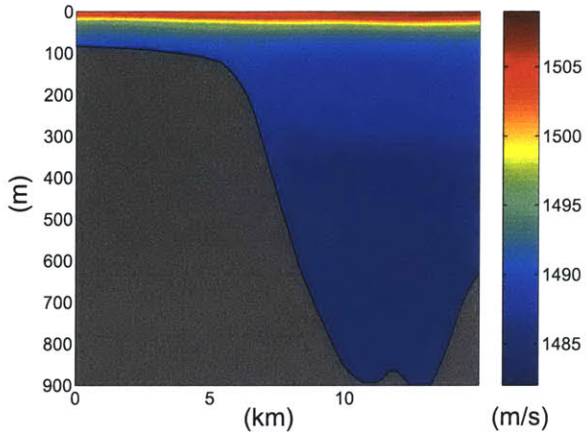


Figure 9-11: TL_r forecasts v.s. range on Aug. 25, 2006. B5 means the bearing 5; f, sz, rz are frequency, source depth and receiver depth respectively. D3 means that this forecast is for the third 12 hours.

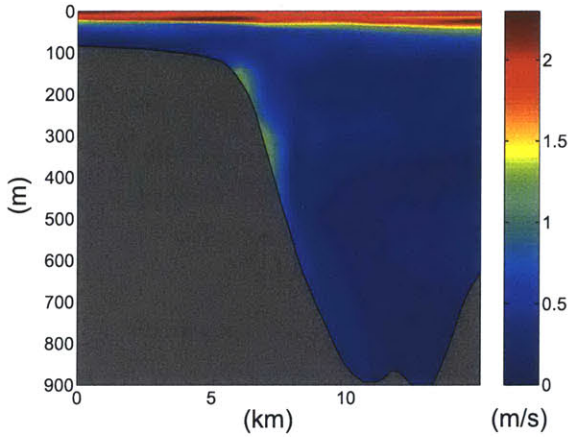
path will be less like a non-adaptive yoyo. The results for capturing TL_r prediction uncertainty whit $Lr = 400m$ are shown in Fig. 9-17.

9.4 Thermocline-oriented AUV Yoyo Control Optimization

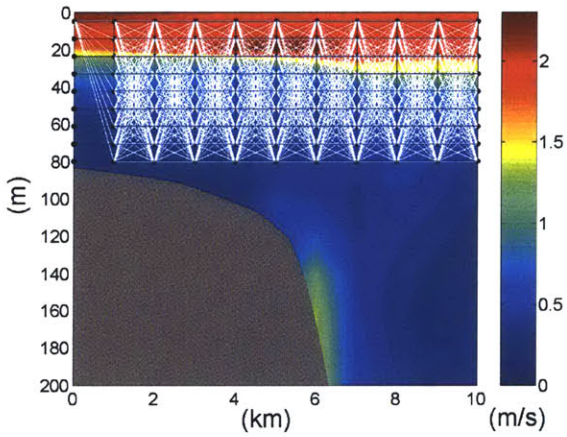
The results of the thermocline-oriented AUV yoyo control parameter optimization on Aug. 21 are shown in Fig. 9-18. Based on the Monte Carlo simulation method, the sub-optimal AUV yoyo control parameters are $p = 40, \gamma = 0.5$; while based on the TL linear approximation method, the sub-optimal AUV yoyo control parameters are $p = 20, \gamma = 0.9$. A predetermined parameters pair $p = 20, \gamma = 0.5$ was made



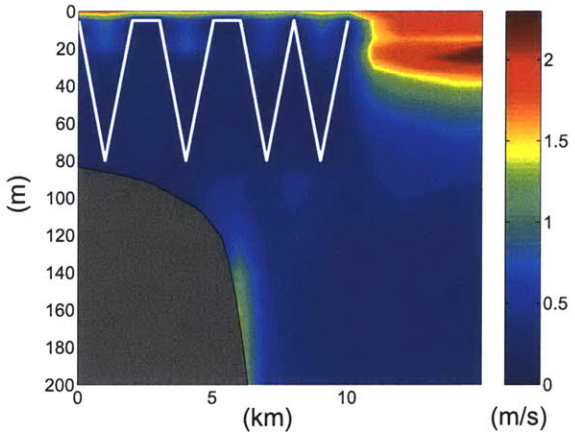
(a) SVP principal estimation



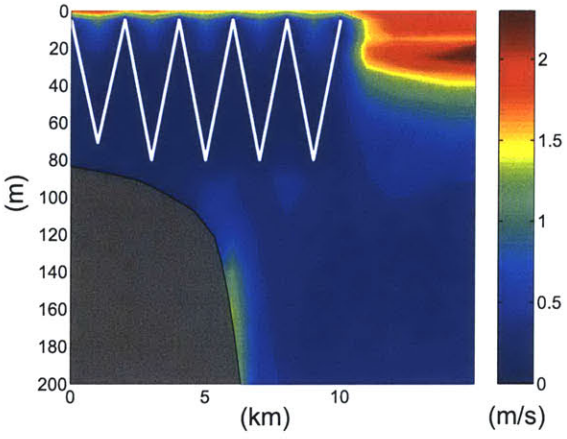
(b) A priori error std.



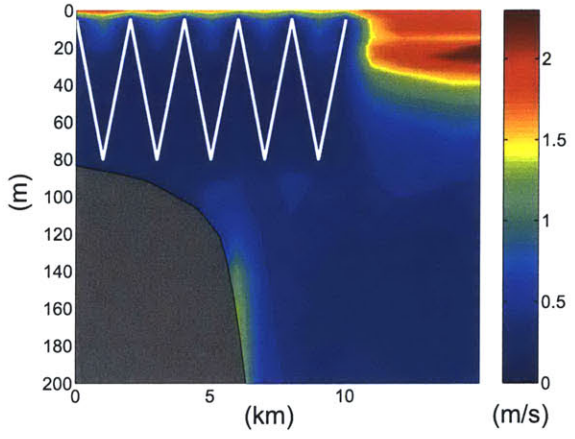
(c) Graph for all possible AUV paths



(d) 1-step look-back result

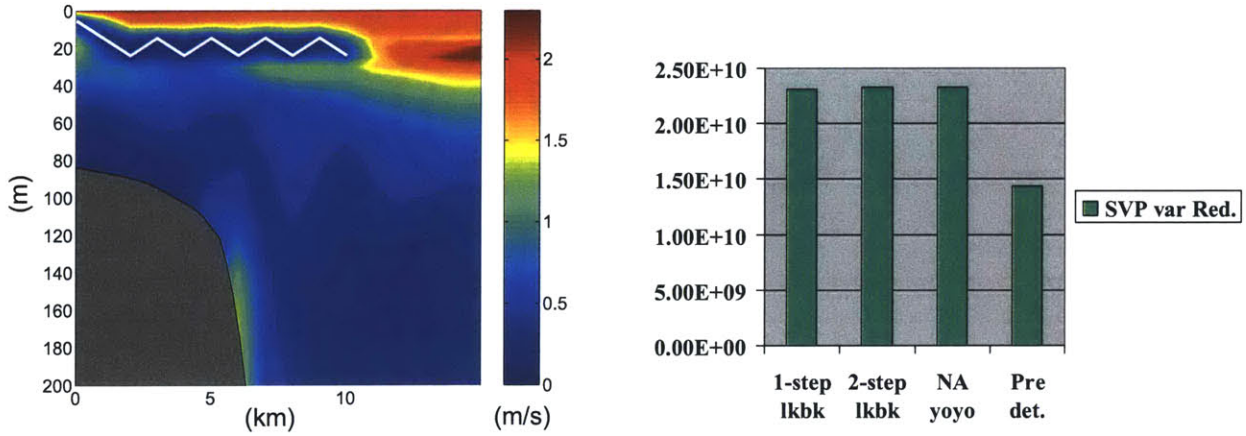


(e) 2-step and 3-step look-back result



(f) Non-adaptive yoyo

Figure 9-12: Aug. 21, 2006, bearing 5, AUV path planning for capturing SVP prediction uncertainty. The background in (c) is the a priori SVP prediction error std.; the backgrounds in (d), (e) and (f) are all the posterior SVP prediction error std.



(a) An empirically predetermined AUV path

(b) Results comparison

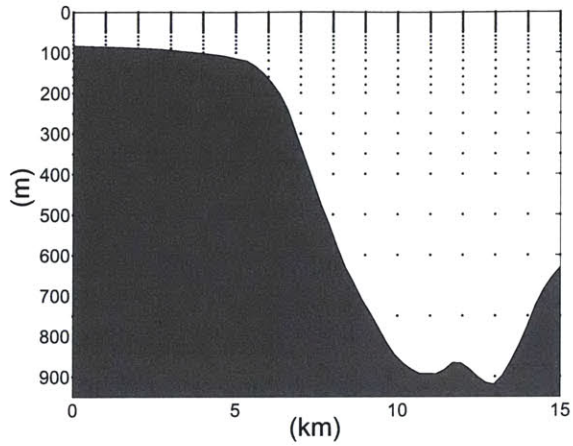
Figure 9-13: Aug. 21, 2006, bearing 5, AUV path planning for capturing SVP prediction uncertainty. $Lr = 1000m, Lz = 5m$. 1-step look-back: 2.30776e+10, 2-step look-back: 2.32952e+10, non-adaptive yoyo: 2.32836e+10, predetermined: 1.43376e+10. The unit is $(m/s)^2 \cdot m^2$.

empirically. The results comparison is shown in Fig. 9-20(a). The Monte Carlo simulation-based optimization takes about 7 hours, the TL linear approximation-based optimization takes about 5 hours. Note that in this case the maximum AUV pitch angle is 10^o.

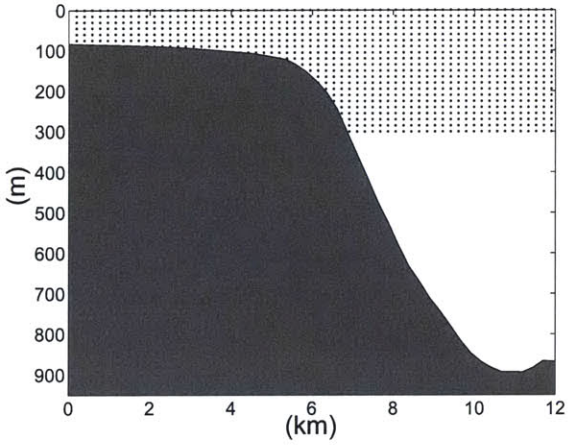
If $Lr = 400m$, the results of AUV yoyo control parameter optimization are shown in Fig. 9-19. The results comparison is shown in Fig. 9-20(b). The results comparison is shown in Fig. 9-20(b).

9.5 Capturing Fronts

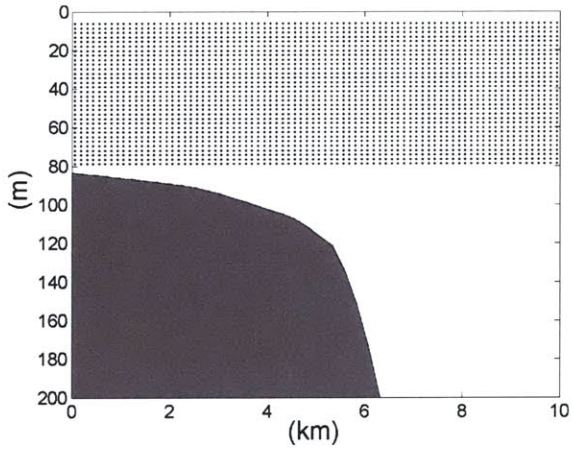
On Aug. 22 and 23, we planned AUV paths for capturing fronts on Aug. 24. They include predetermined path, horizontal AUV yoyo control and surface temperature gradient tracking etc. Some examples are shown in Fig. 9-21.



(a) SVP discretization grid



(b) Ocean discretization grid for TL_r

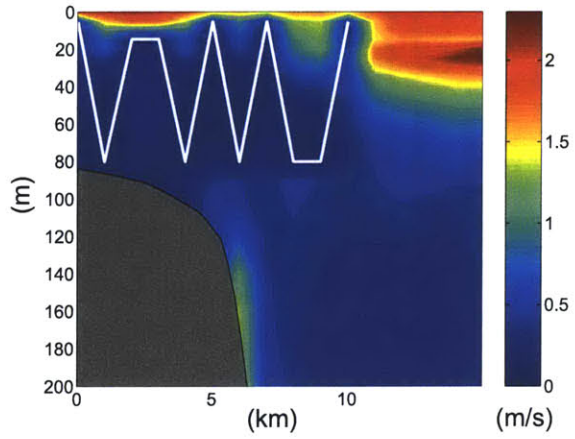


(c) Ocean discretization grid for OA

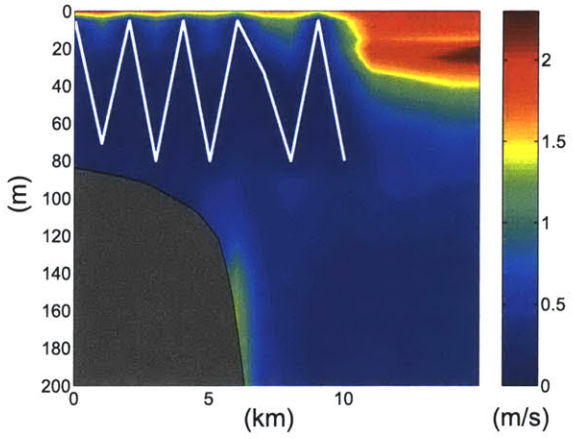
Figure 9-14: Ocean discretization grids for the bearing 5. In (b) the horizontal interval is 192m and the vertical interval is 15m. In (c) the horizontal interval is 125m and the vertical interval is about 2.34m.

9.6 FAF05

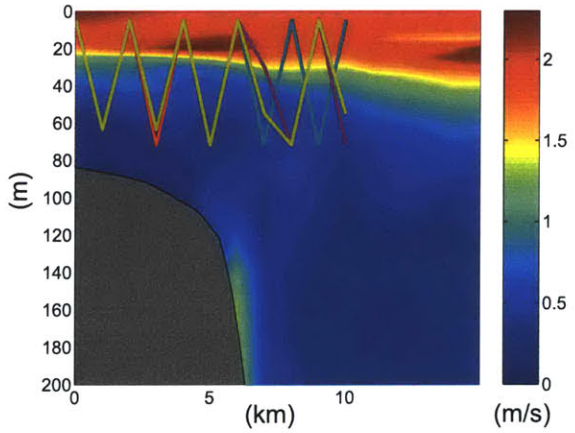
From 7/17/2005 to 7/26/2005, the Focused Acoustic Forecasting-05 (FAF05) experiment was held off Pianosa, Italy, within the northern Tyrrhenian sea, on the eastern side of the Corsican channel. The AREA concept was tested for two weeks within the AREASF, including connections to ocean and seabed models. AOSN techniques were tested at sea but the complete AREA framework was only carried out in simulations, based on real ocean data collected at sea within the AOSN exercises. Details about the FAF05 can be found in [68].



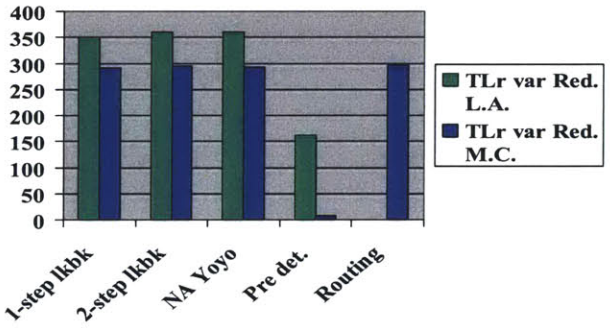
(a) 1-step look-back result



(b) 2-step look-back result

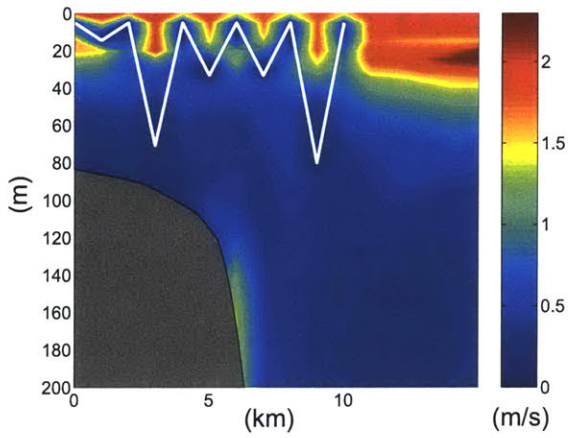


(c) Path candidates for AUV routing strategy optimization

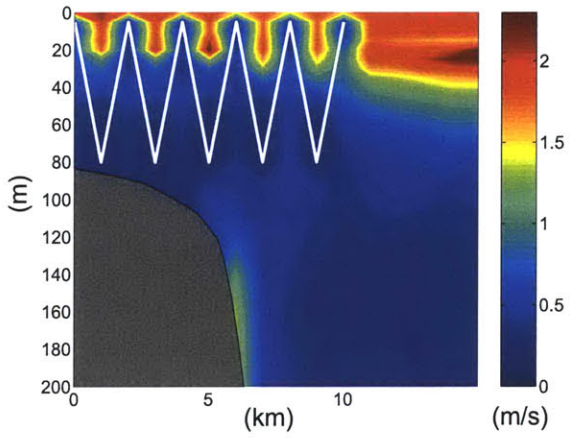


(d) Results comparison

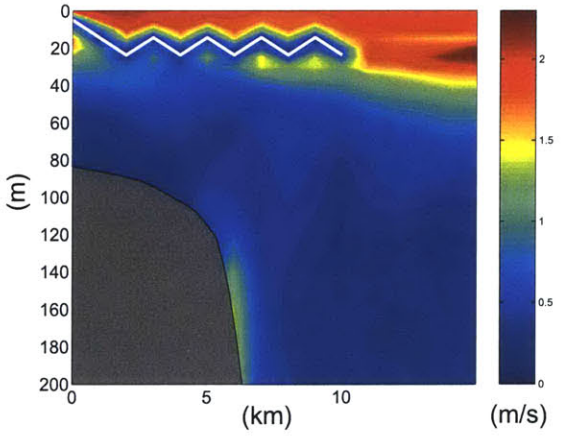
Figure 9-15: Aug. 21, 2006, bearing 5, AUV path planning for capturing TL_r prediction uncertainty. $Lr = 1000m, Lz = 5m$. 1-step look-back: green — 350.93, blue — 291.836; 2-step look-back: green — 359.475; blue — 294.9515; non-adaptive yoyo: green — 359.336, blue — 292.576; predetermined: green — 161.393, blue — 8.426; sub-optimal AUV routing strategy: blue — 295.286. The unit is dB^2.



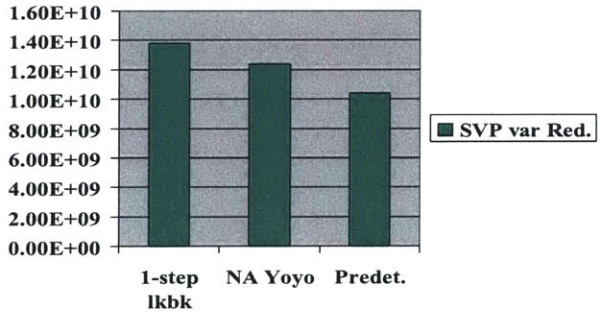
(a) 1-step and 2-step look-back results



(b) Non-adaptive yoyo

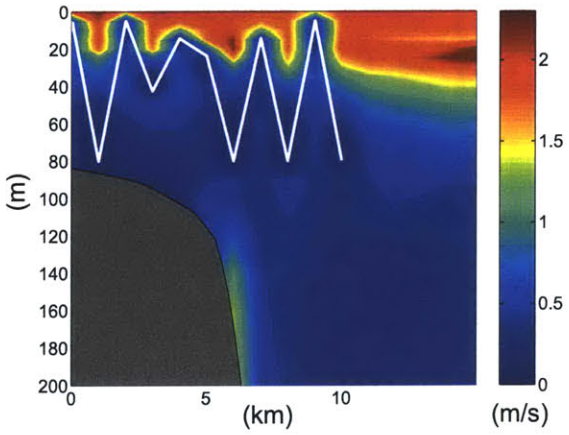


(c) An empirically predetermined AUV path

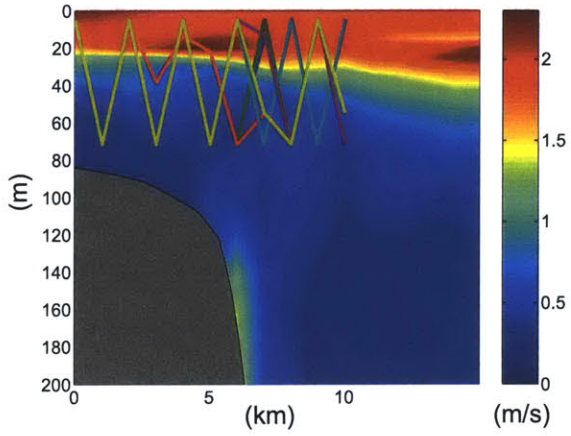


(d) Result comparison

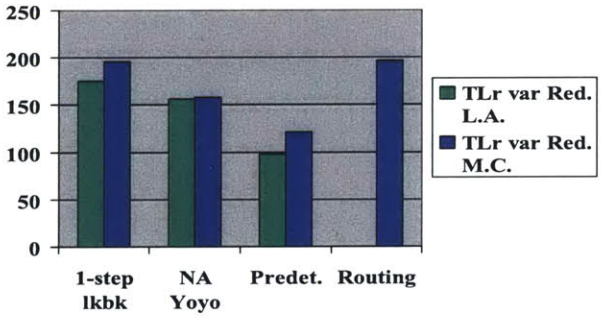
Figure 9-16: Aug. 21, 2006, bearing 5, AUV path planning for capturing SVP prediction uncertainty. $Lr = 400m, Lz = 5m$. 1-step look-back: 1.37786e+10, non-adaptive yoyo: 1.23973e+10, predetermined: 1.04063e+10. The unit is $(m/s)^2 \cdot m^2$.



(a) 1-step look-back

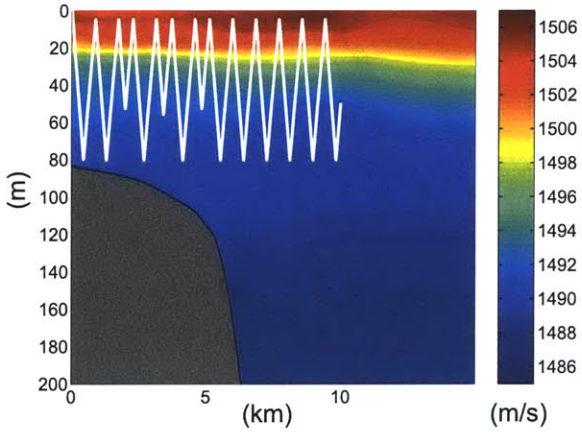


(b) Path candidates for AUV routing strategy optimization

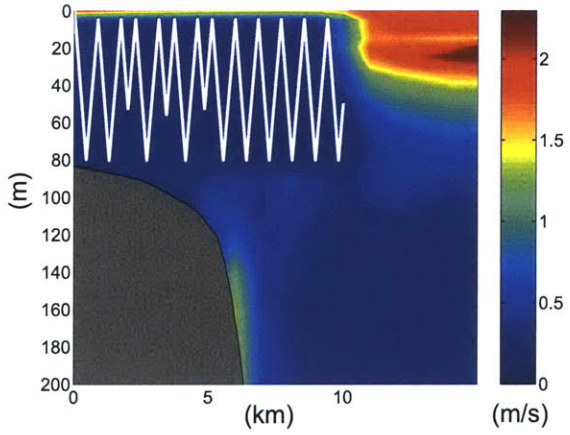


(c) Results comparison

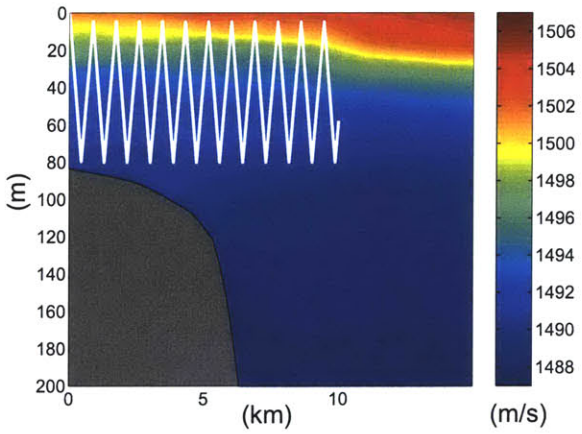
Figure 9-17: Aug. 21, 2006, bearing 5, AUV path planning for capturing TL_r prediction uncertainty. $Lr = 400m, Lz = 5m$. 1-step look-back: green — 175.676, blue — 195.751; non-adaptive yoyo: green — 156.352, blue — 158.181; predetermined: green — 98.0583, blue — 121.132; sub-optimal AUV routing strategy: blue — 196.828. The unit is dB^2.



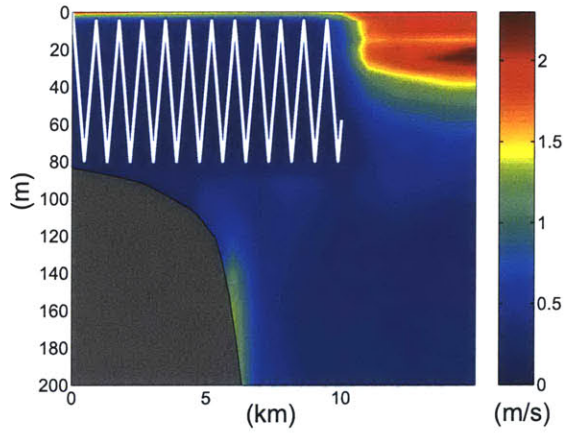
(a) $p = 40, \gamma = 0.5$, posterior SVP prediction



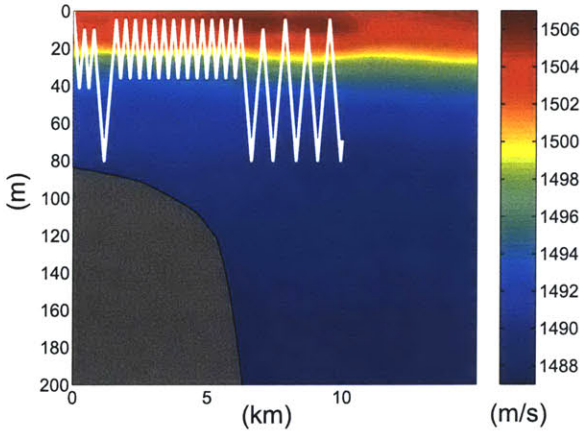
(b) $p = 40, \gamma = 0.5$, posterior SVP prediction error std.



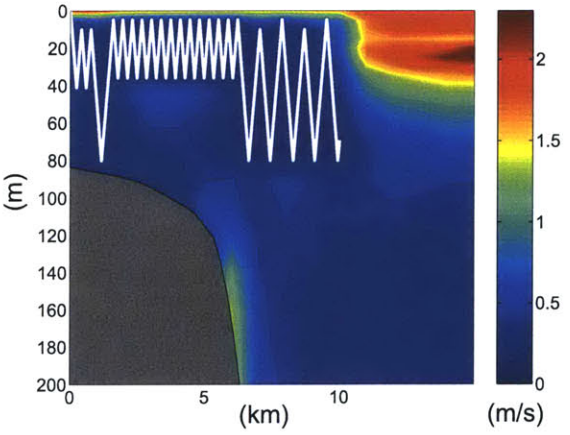
(c) $p = 20, \gamma = 0.9$, posterior SVP prediction



(d) $p = 20, \gamma = 0.9$, posterior SVP prediction error std.

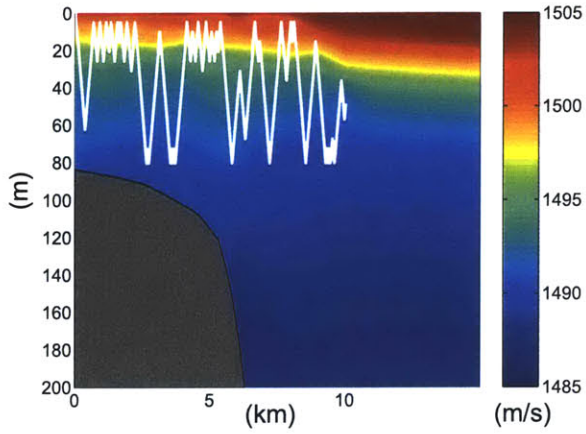


(e) $p = 20, \gamma = 0.5$, posterior SVP prediction

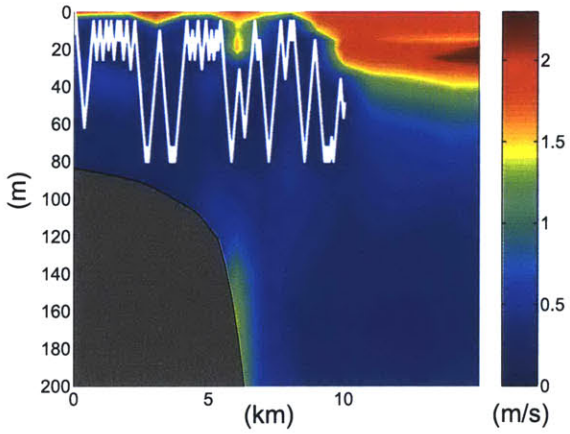


(f) $p = 20, \gamma = 0.5$, posterior SVP prediction error std.

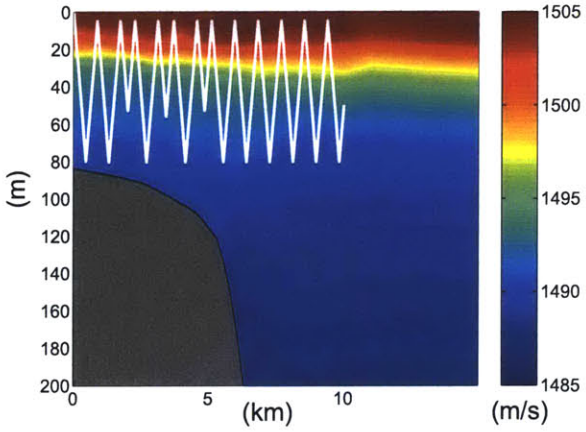
Figure 9-18: Aug. 21, 2006, bearing 5, $Lr = 1000m, Lz = 5m$, AUV yoyo control parameter optimization. Monte Carlo simulation-based: $p = 40, \gamma = 0.5$; TL linear approximation-based: $p = 20, \gamma = 0.9$; empirically predetermined: $p = 20, \gamma = 0.5$.



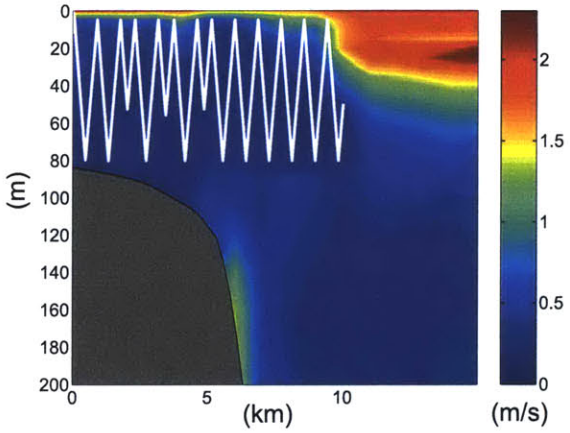
(a) $p = 10, \gamma = 0.5$, posterior SVP prediction



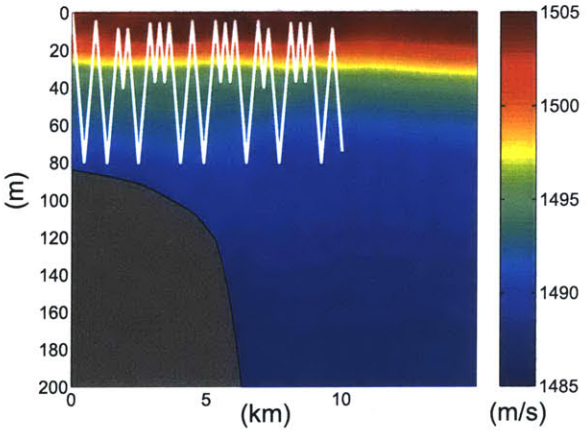
(b) $p = 10, \gamma = 0.5$, posterior SVP prediction error std.



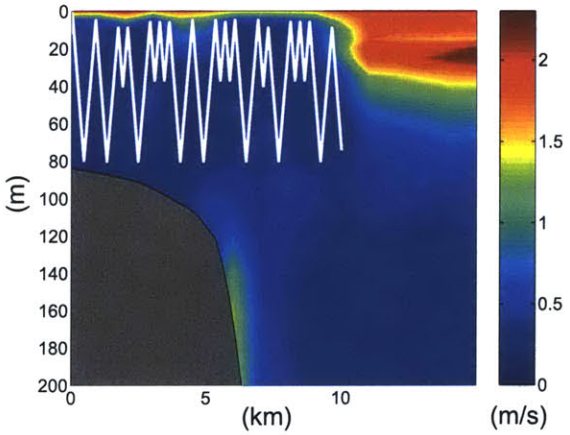
(c) $p = 40, \gamma = 0.5$, posterior SVP prediction



(d) $p = 40, \gamma = 0.5$, posterior SVP prediction error std.

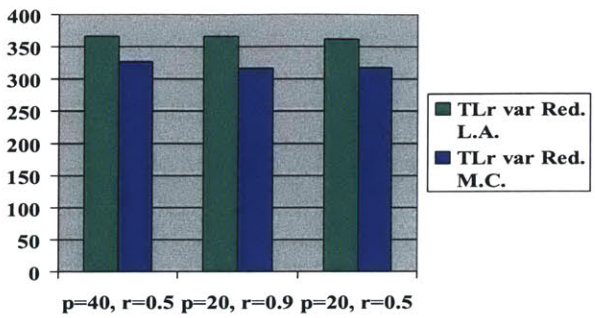


(e) $p = 20, \gamma = 0.5$, posterior SVP prediction

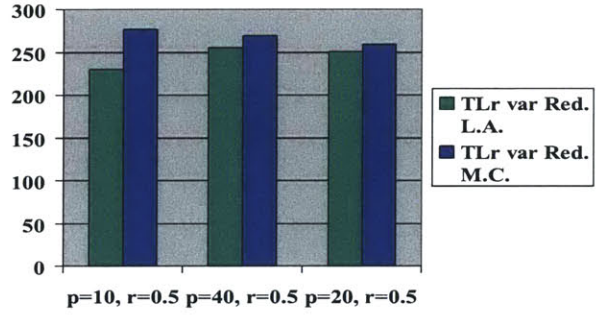


(f) $p = 20, \gamma = 0.5$, posterior SVP prediction error std.

Figure 9-19: Aug. 21, 2006, bearing 5, $Lr = 400m, Lz = 5m$, AUV yoyo control parameter optimization. Monte Carlo simulation-based: $p = 10, \gamma = 0.5$; TL linear approximation-based: $p = 40, \gamma = 0.5$; empirically predetermined: $p = 20, \gamma = 0.5$.

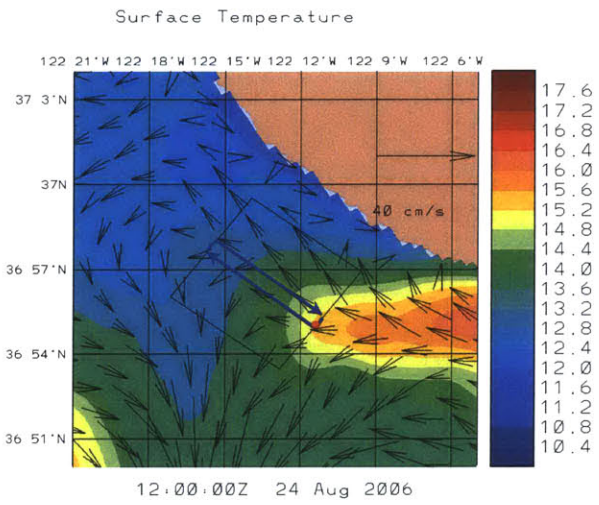


(a) $Lr = 1000m, Lz = 5m$

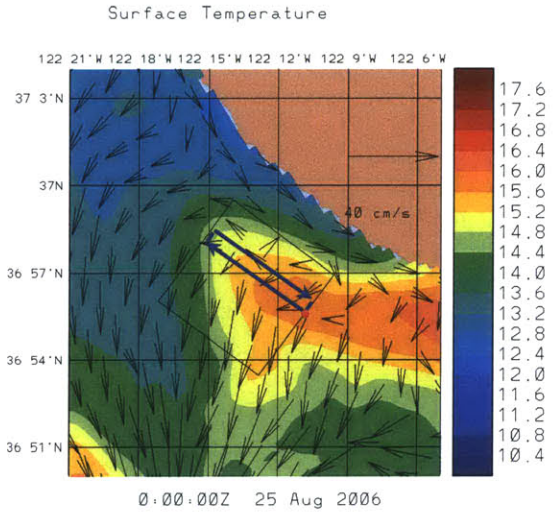


(b) $Lr = 400m, Lz = 5m$

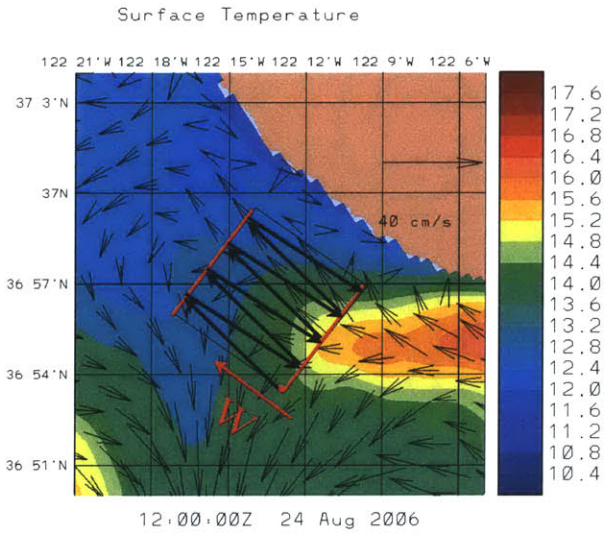
Figure 9-20: AUV yoyo control parameter optimization results comparison.



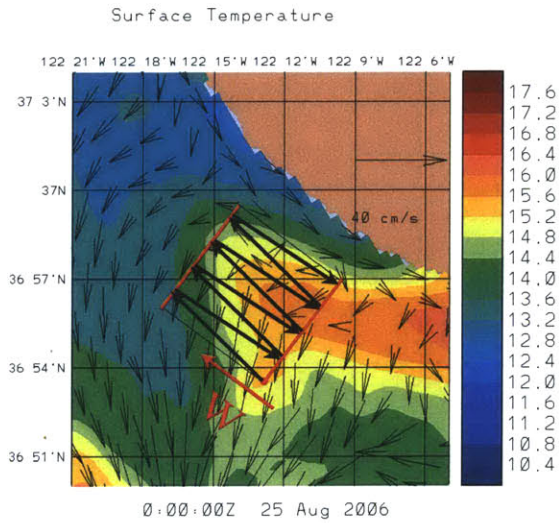
(a) For Aug. 24 morning



(b) For Aug. 24 afternoon



(c) For Aug. 24 morning



(d) For Aug. 24 afternoon

Figure 9-21: Capturing fronts.

Chapter 10

Conclusions

10.1 General Findings and Major Contributions

From the analysis and results described in this thesis, it can be concluded that to capture the SVP/acoustic prediction uncertainty in an ocean domain with size about 10~15km, it is proper to drop a CTD or XBT at the middle range instead of deploying AREA when the SVP horizontal correlation length Lr is very long, such as 5km or longer; when Lr is about 1km to 5km, it is proper to run a non-adaptive up-and-down yoyo; while when Lr is about few hundred meters, it is proper to run n-step look-back method to optimize the AUV path and then construct a sub-optimal adaptive on-board AUV routing strategy based on the result of n-step look-back method.

From the MB06 experiment results, it can be concluded that all ideas, methods developed in this thesis work well, at least for 100Hz or lower frequency scenario. The real-time feasibility of AREA in low frequency scenario is verified. By replacing RAM PE with ray methods, AREA can work in high frequency scenarios.

The major contributions made in this thesis include:

1. Some fundamental philosophies, ideas, concepts in AREA are analyzed and clarified. The principle and performance limit of AREA is very clear now.

2. An engineering model and a mathematical model are developed for AREA.

3. A modularized AREA simulator is developed in C++, which can test the real-

time feasibility of AREA and help solve the AUV path planning problem and the adaptive on-board AUV routing problem etc.

4. Linear approximation for TL is investigated and it is found that the value of $tr\,(var\,(TL))$ calculated via TL linear approximation is highly positively correlated to that of Monte Carlo simulations.

5. A fast AUV path planning method for AREA — the n-step look-back method is developed.

6. A fast adaptive on-board AUV routing method for AREA is developed.

7. A thermocline-oriented AUV yoyo control and control parameter optimization methods for AREA are developed.

8. Some AUV control algorithms for capturing fronts are developed, which includes horizontal AUV yoyo, 2-D temperature gradient tracking etc.

9. A framework for real-time TL forecasts is developed. This is the first time that TL forecasts have been linked with ocean forecasts in real-time.

10. All of the above ideas and methods were tested in the FAF05 and MB06 experiment.

10.2 Future Work Suggestion

1. For acoustic frequency equal to 1kHz or higher, the ray theory is the most proper one to compute acoustic field. In the future, the RAM PE code can be replaced by ray method and AREA can thus work for high frequency scenarios.

2. Currently, $\Lambda_{c_1}(-)$ — the a priori covariance matrix of c_1 is estimated by an approximate function and the Lr, Lz are currently set semi-empirically. In the future, some new and more accurate methods can be developed for estimating $\Lambda_{c_1}(-)$, Lr and Lz.

3. Now, the path candidates for adaptive on-board AUV routing are based on the sub-optimal AUV path generated by the n-step look-back method. In the future, that sub-optimal AUV path can be refined by some local search methods such as Genetic Algorithm (GA), Simulated Annealing (SA) etc. foremost and then the adaptive on-board AUV routing strategy can be developed based on the refined result.

4. In this thesis, we only discussed the path planning problem for a single AUV moving on a vertical plane. In the future, multi-vehicle moving in the 3-D space can be considered. The ideas, methods developed in this thesis may be combined with the methods developed by Namik [52].

Appendix A

Philosophy: Deterministic, Stochastic, Variability, Uncertainty

It's hard to argue that this world is essentially deterministic or stochastic. Based on the understanding of quantum theory, the world seems to be essentially stochastic, at least on microscale. While for those macroscale systems such as the ocean, some people (such as underwater acousticians) think it is essentially stochastic; some others (such as oceanographers) think it is essentially deterministic but extremely complex. Since in this project knowledge from underwater acoustics and knowledge from oceanography have to be combined together, how to view the nature of the ocean, how to bridge the gap between the deterministic hypothesis and the stochastic hypothesis inevitably becomes important.

In this project, the following self-justified hypothesises are presumed. The mere purpose is to build up a philosophical methodology to interpret the connection between variabilities in the ocean and uncertainties in ocean estimation. Otherwise, those hypothesises may be untenable or incomplete.

1. The macroscale world is essentially deterministic but very complicated.

2. In the macroscale world, any variable is essentially a deterministic variable, but

it could be **completely known**, **partially known** or **completely unknown** based on how much information about it is available.

3. Random variable is only a mathematical model used to approximate and represent a partially known deterministic variable. The true value of the deterministic variable can be viewed as a realization of the random variable.

4. In the macroscale world, any dynamic process is essentially a deterministic process. **Variability** is referred to changes of a deterministic process with respect to its arguments. If the variability is too complex and only partially known, the deterministic process can be modeled as a stochastic process. **Uncertainty** is referred to statistic characterizations, such as covariance function, of the stochastic process. The true deterministic process can be viewed as a sample path of the stochastic process.

Some supplemental explanations:

1. For a deterministic variable, if we know it exactly, then it is certain; if we completely know nothing about it, then it can only be viewed as deterministic but unknown; if we know something about it but not completely, then random variable is a very good mathematical model to represent it. However, if the PDF of the random variable is sharp enough, i.e. we know the deterministic variable more enough, then it can be approximately viewed as certain.

2. There exist several rather different concepts of probability, all covering a spectrum from subjective belief to objective frequencies [69]. In the philosophy used in this project, PDF of a random variable reflects people's beliefs or credibilities, or relative frequencies, or propensities, of some values being the true value of the deterministic variable. The PDF is constructed based on people's knowledge about the deterministic variable, logic reasoning, analysis and common sense etc. The PDF reflects people's knowledge about the deterministic variable and also people's ignorance. So the PDF sounds like a subjective term, but it is usually obtained based on objective data and objective methods.

3. For a partially known variable, different people may possess different information or use different analytical method and hence may construct random variables with different PDFs. For example, different ocean prediction systems may generate different PDFs for $c(\vec{r}_0,\ t_0)$, the sound velocity at a certain location and a certain time in the ocean.

4. If the ocean system is essentially stochastic, then $c(\vec{r}_0,\ t_0)$ is an essentially random variable. Different ocean prediction systems may generate different PDFs for $c(\vec{r}_0,\ t_0)$, so one may ask what is the true PDF. However, it is well known that ocean is nonstationary and strictly speaking, $c(\vec{r}_0,\ t_0)$ can be measured only once. Therefore, the true PDF of $c(\vec{r}_0,\ t_0)$ can't be known and its existence can't be proved. Considering $c(\vec{r}_0,\ t_0)$ to be an essentially random variable is not so proper in this scenario.

5. The partially known deterministic variable may reflect an event that will happen in the future, or an event that happened in the past but we only partially observed.

6. Data assimilation can be viewed as melding new information with old information and updating the PDF of a random variable.

Appendix B

Discussion About The Integration Path in The Fourier Transform in Acoustic Propagation Problem

Think about the simplest one-dimensional acoustic propagation problem. A unit source is located at x_0, the frequency is f_0 and sound velocity is c_0. The Helmholtz equation is:

$$\frac{d^2G}{dx^2} + k_0^2\, G = \delta\left(x - x_0\right) \tag{B.1}$$

where $k_0 = \frac{2\pi f_0}{c_0}$ and $G\left(x\right)$ is the wave field.

Figure B-1: One-dimensional acoustic propagation with a unit source at x_0

By Fourier transforms,

$$g\left(k\right) = \int_{-\infty}^{\infty} G\left(x\right) e^{-ikx} dx \tag{B.2}$$

$$G\left(x\right) = \frac{1}{2\pi} \int_{-\infty}^{\infty} g\left(k\right) e^{ikx} dk \tag{B.3}$$

Eq. B.1 can be transformed to be

$$-k^2\, g + k_0^2\, g = e^{-ikx_0} \tag{B.4}$$

so,

$$g = \frac{e^{-ikx_0}}{k_0^2 - k^2} \tag{B.5}$$

Base on Eq. B.3,

$$G(x) = \frac{1}{2\pi}\int_{-\infty}^{\infty} \frac{e^{-ikx_0}}{k_0^2 - k^2} e^{ikx} dk \tag{B.6}$$

For the integrand in Eq. B.6, $\frac{1}{2\pi}\frac{e^{-ikx_0}}{k_0^2-k^2}e^{ikx}$ has two poles at $k = \pm k_0$.

$$Res(k_0) = \frac{-e^{ik_0(x-x_0)}}{4\pi k_0} \tag{B.7}$$

$$Res(-k_0) = \frac{e^{-ik_0(x-x_0)}}{4\pi k_0} \tag{B.8}$$

According to the strict definition, the integration in Eq. B.6 doesn't exist [70]. The **Principal Value** [70] of the integration can be obtained by path integration as shown in Fig. B-2 and the *Cauchy theorem* [70], thus

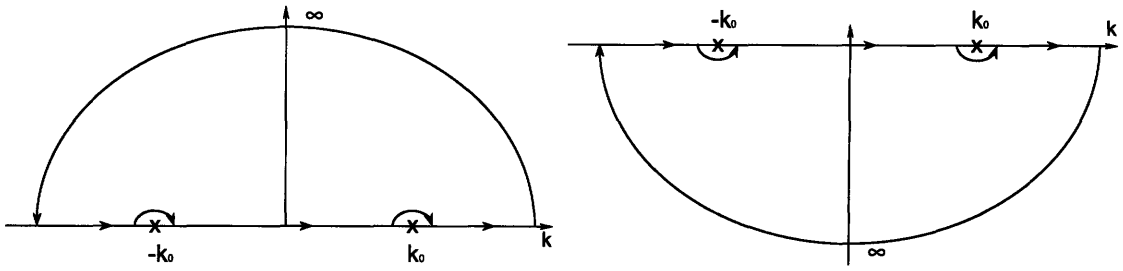
$$\begin{aligned} &\text{if } x > x_0 \\ &\quad G(x) = \pi i\left(\frac{1}{2\pi}\frac{-e^{ik_0(x-x_0)}}{2k_0} + \frac{1}{2\pi}\frac{e^{-ik_0(x-x_0)}}{2k_0}\right) \\ &\qquad\quad = \frac{sin\left[k_0\left(x-x_0\right)\right]}{2k_0} \end{aligned} \tag{B.9}$$

$$\begin{aligned} &\text{if } x < x_0 \\ &\quad G(x) = -\frac{sin\left[k_0\left(x-x_0\right)\right]}{2k_0} \end{aligned} \tag{B.10}$$

Note that the integration along the arc with ∞ radius is equal to 0.

However $G(x)$ in Eq. B.9, B.10 doesn't satisfy the *radiation condition*,

$$\frac{dG}{d\,|x|} - ikG = 0 \text{ for } x = \pm\infty \tag{B.11}$$



(a) $x > x_0$ (b) $x < x_0$

Figure B-2: Integration path for the Principal Value

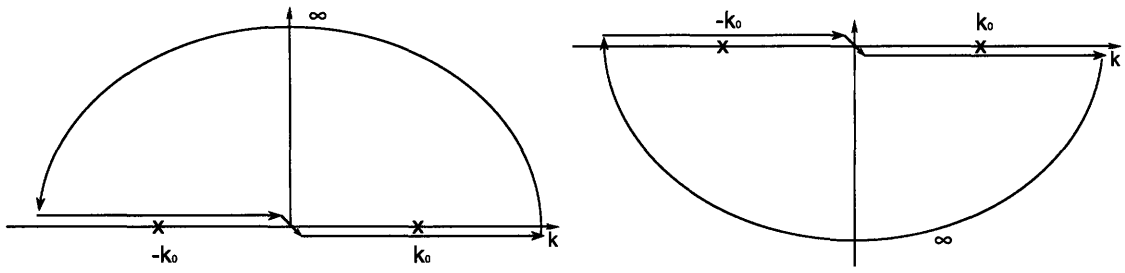
So it seems that in the above Fourier transforms, if poles exists on the real axis, the principal value of the integration is not proper. If now the integration path is deviated a little bit as shown in Fig. B-3, we can see that

$$\begin{aligned} &\text{if } x > x_0 \\ &\quad G(x) = 2\pi i \left(\frac{1}{2\pi} \frac{-e^{ik_0(x-x_0)}}{2k_0} \right) \\ &\quad\quad = \frac{e^{ik_0(x-x_0)}}{2ik_0} \end{aligned} \tag{B.12}$$

$$\begin{aligned} &\text{if } x < x_0 \\ &\quad G(x) = \frac{e^{-ik_0(x-x_0)}}{2ik_0} \end{aligned} \tag{B.13}$$

It can be shown that Eq. B.12, B.13 satisfy the *radiation condition* and the Helmholtz equation.

So, from the one dimensional case, it is clear that $\int_{-\infty}^{\infty}$, **the integration in the inverse Fourier transform is not really operated on the real axis, but on $k+i0^+$, when $k < 0$ and $k+i0^-$, when $k > 0$.** The basic reason of this phenomenon is related to the *radiation condition.* A strict explanation can be found on page 193-196 in [71]. In this book, Sommerfeld used the same 1-D example and let the *radiation*



(a) $x > x_0$ (b) $x < x_0$

Figure B-3: Deviated integration path

condition to be firstly

$$\frac{dG}{d|x|} - ikG = 0 \text{ for } |x| = l \tag{B.14}$$

and then let $l \to \infty$. When Eq. B.14 is used, $G(x)$ can be obtained in format of summing a infinite sequence of eigenfunctions (eq. 27.5 in [71]). In this case, no Fourier transform is used. Those eigenvalues of k (marked by x in Fig. B-4) are located at the origin and the 4th quadrant. Path W_1 is a curve passing through all eigenvalues. If now let $l \to \infty$, those eigenvalues will be everywhere dense on W_1, which is closing to the positive axis of k from below as $l \to \infty$. At the same time, the summation of eigenfunctions will become to an integration on $k + i0^-$ for $k > 0$. Utilizing the symmetry of the integrand, the integration can be extended to $k + i0^+$ for $k < 0$.

$$\begin{aligned} G(x) &= \frac{1}{\pi} \int_{W_1} \frac{\cos k(x - x_0)}{k_0^2 - k^2} dk \\ &= \frac{1}{2\pi} \int_{W_1+W_2} \frac{e^{ik(x-x_0)}}{k_0^2 - k^2} dk \end{aligned} \tag{B.15}$$

where, W_1 and W_2 are symmetric with respect to the origin.

Now, we can see that Eq. B.15 just gives us the same expression in Eq. B.6, the inverse Fourier transform, with the integration path slightly deviated from the real axis. This implicates that in solving acoustic propagation problems, the integration

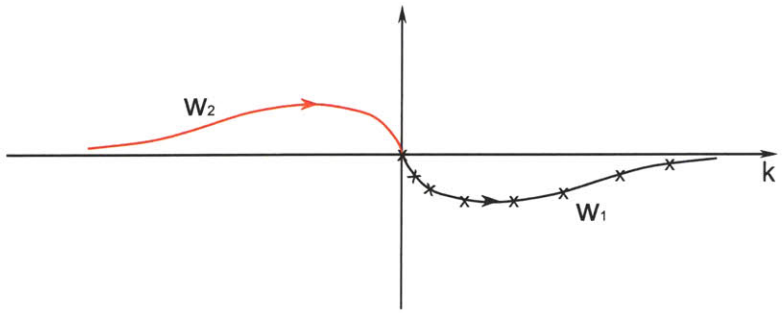


Figure B-4:

in the inverse Fourier transform is not exact on the real axis, but slightly deviated into the 2nd quadrant and the 4th quadrant. This result is also valid for the Hankel transform.

Appendix C

The Relation Between The Wavenumber Integration Method And The Normal Mode Method

The normal mode method can be derived from the wavenumber integration method. Let's first consider the idea waveguide problem. By the wavenumber integration method, the pressure field is

$$p(r,z) = \frac{1}{2} \int_{-\infty}^{\infty} p(k_r,z)\, H_0^{(1)}(k_r r)\, k_r dk_r \tag{C.1}$$

$$p(k_r,z) = -\frac{S_\omega}{4\pi} \begin{cases} \frac{\sin k_z z \sin k_z(D-z_s)}{k_z \sin k_z D}, & z < z_s \\ \frac{\sin k_z z_s \sin k_z(D-z)}{k_z \sin k_z D}, & z > z_s \end{cases} \tag{C.2}$$

where D is water depth.

$p(k_r,z)$ has poles for

$$k_z D = m\pi, \quad m = 1,2,\cdots \tag{C.3}$$

or, in terms of the horizontal wavenumber $k_r = \sqrt{k^2 - k_z^2}$,

$$k_{rm} = \sqrt{k^2 - \left(\frac{m\pi}{D}\right)^2}, \quad m = 1,2,\cdots \tag{C.4}$$

The integration path C and poles are shown in Fig. C-1. From Appendix. B, the

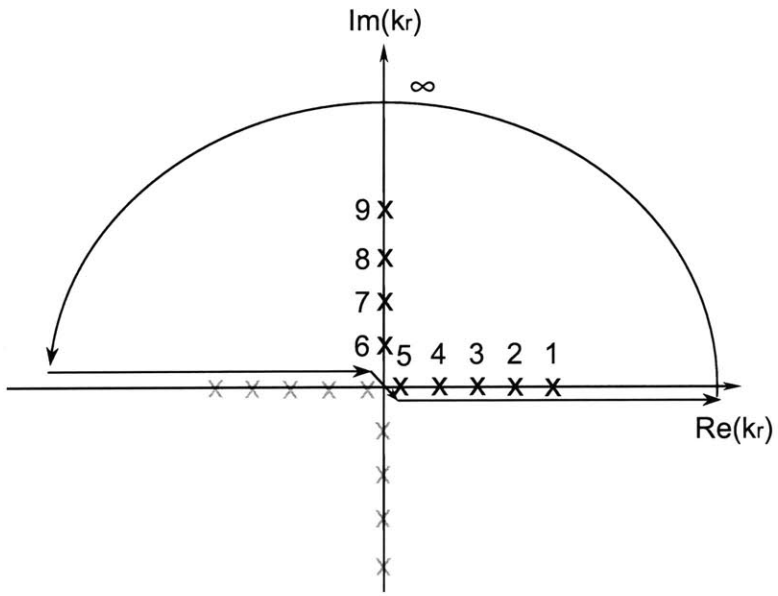


Figure C-1: Idea waveguide

integral path C is slightly deviated into the 2nd and 4th quadrant. Since no branch cut exists, the integral path can be closed by the semicircle with $r \to \infty$ in the upper half space. Due to the property of the Hankel function, the integration along the semicircle is equal to zero. So according to the Cauchy theorem,

$$p(r,z) = \frac{1}{2}\sum_{m=1}^{\infty} Res(k_{rm})\, H_0^{(1)}(k_{rm}r)\, k_{rm} \tag{C.5}$$

Eq. C.5 is actually the implementation of Eq. 2.44 in the idea waveguide scenario. It can be found that the total pressure field is equal to the summation of all modes, which include not only normal modes such as mode 1,2,3,4,5, but also all virtual modes such as mode 6,7,8,9. From this example, we can see that the wavenumber integration method and the normal mode method generate identical results. However, this conclusion is not always right.

Now, let's consider the Pekeris wave guide. Fig. C-2 shows poles, EJP branch cut and integral contour on the complex k_r-plane, where k_1 and k_2 are the wavenumber

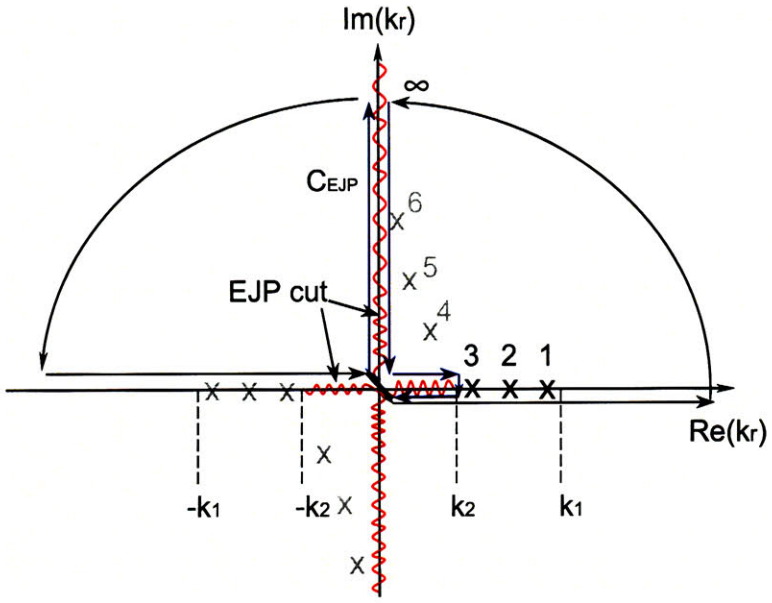


Figure C-2: Pekeris waveguide

in water column and seabed respectively. Normal modes such as mode 1,2,3 can only exist on the real axis between k_1 and k_2. Mode 4,5,6 are virtual modes. k_2 and $-k_2$ are two branch points. If the EJP branch cut (the red curves) is adopted, all virtual modes will exist on an unphysical Riemann sheet. However, all normal modes and the integral contour will be on another physical Riemann sheet, so it can be obtained that

$$p(r,z) = \frac{1}{2}\sum_{m=1}^{n_{max}} Res(k_{rm}) H_0^{(1)}(k_{rm}r) k_{rm} - \int_{C_{EJP}} \tag{C.6}$$

where, n_{max} is the max index of normal modes. From the above equation, we can see that in Pekeris waveguide scenario, $p(r,z)$ can't be completely expressed as the summation of all modes, but the summation of all normal modes plus contribution from a continuous spectrum. The reason of this phenomenon is that all mode functions don't form a complete set.

However, it can be found that the $\int_{C_{EJP}}$ (along the blue curve) corresponds to some decay waves, which only exist in the "near-field" and can be asymptotically decomposed in terms of virtual modes. So, in this case the acoustic field generated

by the normal mode method has increasing accuracy as the range increases.

It should be noticed that if another branch cut is used such as the Pekeris cut, the value of n_{max} may be changed and the summation may become to be the summation of all modes. Accordingly, the integration around the branch cut will change too [1].

Appendix D

Some Discussions For The Normal Mode Method Derivation In [1] And [2]

In cylindrical coordinates, the Dirac delta function is

$$\delta\left(\vec{x}-\vec{x_0}\right)=\frac{1}{\pi r}\delta\left(r-r_0\right)\delta\left(\varphi-\varphi_0\right)\delta\left(z-z_0\right) \tag{D.1}$$

except when $r_0=0$ when the function becomes

$$\delta\left(\vec{x}\right)=\frac{1}{\pi r}\delta\left(r\right)\delta\left(z-z_0\right). \tag{D.2}$$

The Helmholtz equation for the Green's function (as Eq. (2.58) in [1]) is

$$\left[\nabla^2+k^2\right]G_\omega\left(\vec{x},\vec{x_0}\right)=-\delta\left(\vec{x}-\vec{x_0}\right). \tag{D.3}$$

In cylindrical coordinates, Eq. D.3 is written as

$$\left[\frac{1}{r}\frac{\partial}{\partial r}\left(r\frac{\partial}{\partial r}\right)+\rho\left(z\right)\frac{\partial}{\partial z}\left(\frac{1}{\rho}\left(z\right)\frac{\partial}{\partial z}\right)+k^2\left(z\right)\right]G_\omega=-\frac{1}{\pi r}\delta\left(r\right)\delta\left(z-z_s\right). \tag{D.4}$$

So, the right side of Eq.(5.1) in [1] should be the same as that in Eq. D.5, i.e. the

Eq.(5.1) in [1] should be

$$\frac{1}{r}\frac{\partial}{\partial r}\left(r\frac{\partial p}{\partial r}\right)+\rho(z)\frac{\partial}{\partial z}\left(\frac{1}{\rho}(z)\frac{\partial p}{\partial z}\right)+k^2(z)p=-\frac{1}{\pi r}\delta(r)\delta(z-z_s). \tag{D.5}$$

Thus, the Eq.(5.11) in [1] should be

$$\frac{1}{r}\frac{d}{dr}\left[r\frac{d\Phi_n(r)}{dr}\right]+k_{rn}^2\Phi_n(r)=-\frac{\delta(r)\Psi_n(z_s)}{\pi r\rho(z_s)}. \tag{D.6}$$

However, the Eq.(5.12) in [1] is actually the solution of Eq. D.6.

Similarly, the Eq.(5.27) in [1] should be

$$\frac{1}{r}\frac{\partial}{\partial r}\left(r\frac{\partial p_0(r)}{\partial r}\right)+\frac{\omega^2}{c^2(z)}p_0=-\frac{\delta(r)}{r\pi}, \tag{D.7}$$

and the Eq.(5.28) in [1] is actually the solution of Eq. D.7.

In [2], the Eq. (5.5) can be directly obtained from Eq. D.3 and D.2 and its right side should be

$$-4\frac{\delta(r)\delta(z-z_0)}{r}.$$

Thus the Eq. (5.18) in [2] should be

$$\frac{1}{r}\frac{d}{dr}\left[r\frac{dR_n(r)}{dr}\right]+k_n^2R_n(r)=-4\frac{\delta(r)}{r}. \tag{D.8}$$

The Eq. (5.19) in [2] is actually the solution of the above equation.

Eq. D.8 can be transformed as

$$r^2\frac{d^2R_n}{dr^2}+r\frac{dR_n}{dr}+r^2k_n^2R_n=-4\delta(r)r. \tag{D.9}$$

set $x=k_nr$, the above equation will be

$$x^2\frac{d^2R_n}{dx^2}+x\frac{dR_n}{dx}+x^2R_n=-4\delta\left(\frac{x}{k_n}\right)\frac{x}{k_n}=-4\delta\left(x\right)x. \tag{D.10}$$

The solution of the above equation is

$$R_n=i\pi H_0^{(1)}\left(x\right)=i\pi H_0^{(1)}\left(k_nr\right) \tag{D.11}$$

Bibliography

[1] Finn B. Jensen, William A. Kuperman, Michael Porter, and Henrik Schmidt. *Computational Ocean Acoustics.* AIP Press, 1993.

[2] George V. Frisk. *Ocean and seabed acoustics: a theory of wave propagation.* Prentice-Hall, 1994.

[3] T. D. Dickey. Technology and related developments for interdisciplinary global studies. *Sea Technology*, pages 47–53, August 1993.

[4] Clarence S. Clay and Herman Medwin. *Acoustical Oceanography.* Wiley, 1977.

[5] A. R. Robinson, P. Abbot, P. F. J. Lermusiaux, and L. Dillman. Transfer of uncertainties through physical-acoustical-sonar end-to-end systems: A conceptual basis. In Pace and Jensen [10], pages 603–610.

[6] P. F. J. Lermusiaux, C.-S. Chiu, and A. R. Robinson. Modeling uncertainties in the prediction of the acoustic wavefield in a shelfbreak environment, May 2001.

[7] N. M. Patrikalakis, J. J. McCarthy, A. R. Robinson, H. Schmidt, C. Evangelinos, P.J. Haley, S. Lalis, P. F. J. Lermusiaux, R. Tian, W. G. Leslie, and W. Cho. Towards a dynamic data-driven system for rapid adaptive interdisciplinary ocean forecasting. In F. Darema, editor, *Refereed invited manuscript in Dynamic Data-Driven Application Systems.* Kluwer Acad. Pub., Amsterdam, 2006. Under review.

[8] Pierre F. J. Lermusiaux and Ching-Sang Chiu etc. Quantifying Uncertainties in Ocean Predictions. *Oceanography*, 19:92–105, March 2006.

[9] H. Schmidt. Area: Adaptive rapid environmental assessment. In Pace and Jensen [10], pages 587–594.

[10] N. G. Pace and F. B. Jensen, editors. *Impact of Littoral Environmental Variability on Acoustic Predictions and Sonar Performance.* Kluwer Academic Publishers, 2002.

[11] Herman Medwin and Clarence S. Clay. *Fundamentals of Acoustical Oceanography.* Academic Press, 1998.

[12] Hassan B. Ali. Oceanographic Variability in Shallow-Water Acoustics and the Dual Role of the Sea Bottom. *IEEE J. Ocean. Eng.*, 18:31–41, January 1993.

[13] Christopher Garrett and Walter Munk. Internal Waves in The Ocean. *Annual review of fluid mechanics*, 11:339–369, 1979.

[14] T. R. Chen, P. Ratilal, and N. C. Makris. Mean and variance of the forward field propagated through three-dimensional random internal waves in a continental-shelf waveguide. *J. Acoust. Soc. Am.*, 118:3532–3559, 2005.

[15] William A. Kuperman and James F. Lynch. Shallow-Water Acoustics. *Physics Today*, November 2004.

[16] Dirk Tielbürger, Steven Finette, and Stephen Wolf. Acoustic propagation through an internal wave field in a shallow water waveguide. *J. Acoust. Soc. Am.*, 101:789–808, 1997.

[17] Terry E. Ewart. A numerical simulation of the effects of oceanic finestructure on acoustic transmission. *J. Acoust. Soc. Am*, 67:496–503, Feb 1980.

[18] Santha Unni and Charles Kaufman. Acoustic fluctuations due to the temperature fine structure of the ocean. *J. Acoust. Soc. Am*, 69:676–680, March 1981.

[19] John A. Knauss. *Introduction to physical oceanography.* Prentice Hall, 2nd edition, 1996.

[20] Stephen Pond and George L Pickard. *Introductory Dynamical Oceanography.* BH, 2nd edition, 1983.

[21] Bruce A. Warren and Carl Wunsch, editors. *Evolution of Physical Oceanography.* The MIT Press, 1981.

[22] G. Neumann. *Ocean Currents.* Elsevier, 1968.

[23] Richard R. Hobbs. *Marine Navigation: Piloting and Celestial and Electronic Navigation.* Naval Institute Press, 3rd edition, 1990.

[24] Vladimir E Ostashev. *Acoustics in Moving Inhomogeneous Media.* E & FN SPON, 1997.

[25] Dmitry Yu Mikhin, Oleg A. Godin, Olaf Boebel, and Walter Zenk. simulations of Acoustic Imprints of Meddies in the Iberian Basin: Toward Acoustic Detection of Meddies. *Journal of Atmospheric and Oceanic Technology*, 14:938–949, 1997.

[26] PETER F. WORCESTER, ROBERT C. SPINDEL, and BRUCE M. HOWE. Reciprocal Acoustic Transmissions: Instrumentation for Mesoscale Monitoring of Ocean Currents. *IEEE JOURNAL OF OCEANIC ENGINEERING*, OE-10:123–137, 1985.

[27] N. S. Grigor'eva and M. I. Yavor. Influence on the sound field in the ocean of a large-scale ocean current that qualitatively alters the nature of guided-wave sound propagation. *Sov. Phys. Acoust.*, 32:482–485, 1986.

[28] Robert H. Headrick, John L. Spiesberger, and Paul J. Bushong. Tidal signals in basin-scale acoustic transmission. *J. Acoust. Soc. Am*, 93:790–802, February 1993.

[29] R. F. Henrick, W. L. Siegmann, and M. J. Jacobson. General analysis of ocean eddy effects for sound transmission applications. *J. Acoust. Soc. Am*, 86:1465–1477, October 1989.

[30] Marshall V. Hall and Mark A. Irving. Application of adiabatic mode theory to the calculation of horizontal refraction through a mesoscale ocean eddy. *J. Acoust. Soc. Am*, 62:860–870, October 1977.

[31] A. D. Heathershaw and C. E. Stretch. Coupled ocean-acoustic model studies of sound propagation through a front. *J. Acoust. Soc. Am*, 89:145–155, January 1991.

[32] G. Brooke Farquhar. Biological sound scattering in the oceans: A review. *J. Acoust. Soc. Am*, 59:S73, April 1976.

[33] S. D. Rajan and G. V. Frisk. Seasonal variations of the sediment compressional wave speed profile in the Gulf of Mexico. *J. Acoust. Soc. Am*, 91:127–135, 1992.

[34] S. D. Rajan and G. V. Frisk. The effect of seasonal temperature fluctuations in the water column on sediment compressional wave speed profiles in shallow water. In J. Potter and A. Warn-Varnas, editors, *Ocean Variability and Acoustic Propagation*. Kluwar Academic Publishers, 1990.

[35] Pierre F. J. Lermusiaux. *Error Subspace Data Assimilation Methods for Ocean Field Estimation: Theory, Validation and Applications*. PhD thesis, Harvard University, May 1997.

[36] A. R. Robinson and P. F. J. Lermusiaux. Overview of Data Assimilation. Harvard University Reports in Physical / Interdisciplinary Ocean Science #62, 2000.

[37] P.F.J. Lermusiaux and A.R. Robinson. Data assimilation via error subspace statistical estimation, part i: Theory and schemes. *Monthly Weather Review*, 127(8):1385–1407, 1999.

[38] P.F.J. Lermusiaux. Data assimilation via error subspace statistical estimation, part ii: Mid-atlantic bight shelfbreak front simulations, and esse validation. *Monthly Weather Review*, 127(8):1408–1432, 1999.

[39] Ding Wang. Adaptive rapid environmental assessment system simulation framework, December 2004.

[40] A. R. Robinson, P. F. J. Lermusiaux, and N. Q. Sloan. Data assimilation. *The Sea: The Global Coastal Ocean*, 10:541–594, 1998.

[41] A. R. Robinson. Forecasting and simulating coastal ocean processes and variabilities with the harvard ocean prediction system. In *Coastal Ocean Prediction*, AGU Coastal and Estuarine Studies Series, pages 77–100. AGU, 1999.

[42] A.B. Baggeroer, W.A. Kuperman, and H. Schmidt. Matched field processing: Source localization in correlated noise as an optimum parameter estimation problem. *J. Acoust. Soc. Am.*, 83:571–587, 1988.

[43] T. Curtin, J.G. Bellingham, J. Catipovic, and D. Webb. Autonomous oceanographic sampling networks. *Oceanography*, 6(3):86–94, 1993.

[44] H. Schmidt, J. G. Bellingham, and P. Elisseef. Acoustically focused oceanographic sampling in coastal environments. In E. Pouliquen, A. D. Kirwan, and R. T. Pearson, editors, *Rapid Environmental Assessment. SACLANTCEN Conference Proc. Series CP-44*, 1997.

[45] Nicholas C. Makris, Purnima Ratilal, Deanelle T. Symonds, Srinivasan Jagannathan, Sunwoong Lee, and Redwood W. Nero. Fish Population and Behavior Revealed by Instantaneous Continental Shelf-Scale Imaging . *Science*, 311:660–663, February 2006.

[46] P. Elisseeff, H. Schmidt, and W. Xu. Ocean acoustic tomography as a data assimilation problem. *IEEE Journal of Oceanic Engineering*, 27(2):275–282, 2002.

[47] Dimitri P. Bertsekas. *Dynamic Programming and Optimal Control.* Athena Scientific, 2000.

[48] A.F. Shchepetkin and J.C. McWilliams. The regional oceanic modeling system: A split-explicit, free-surface, topography-following-coordinate ocean model. *Ocean Modelling*, 9:347–404, 2004.

[49] Paolo Toth and Daniele Vigo. *The Vehicle Routing Problem.* SIAM, 2002.

[50] Teodor Gabriel Crainic and Gilbert Laporte. *Fleet Management and Logistics.* Kluwer Academic Publishers, 2004.

[51] M.O. Ball, T.L. Magnanti, C.L. Monma, and G.L. Nemhauser, editors. *HANDBOOKS IN OPERATIONS RESEARCH AND MANAGEMENT SCIENCE, 8: NETWORK ROUTING.* ELSEVIER, 1995.

[52] Namik Kemal Yilmaz. *Path Planning of Autonomous Underwater Vehicles for Adaptive Sampling.* PhD thesis, MIT, September 2005.

[53] N. K. Yilmaz, C. Evangelinos, N. M. Patrikalakis, P. F. J. Lermusiaux, P. J. Haley, W. G. Leslie, A. R. Robinson, D. Wang, and H. Schmidt. Path Planning Methods for Adaptive Sampling of Environmental and Acoustical Ocean Fields. In *Proceedings of IEEE/MTS Oceans'06 Conference*, Boston, MA, September 2006.

[54] T. Schouwenaars, B. DeMoor, E. Feron, and J. How. Mixed integer programming for safe multi-vehicle cooperative path planning, September 2001.

[55] http://en.wikipedia.org/.

[56] Dimitris Bertsimas and John N. Tsitsiklis. *Introduction to Linear Optimization.* Athena Scientific, 1997.

[57] Dimitri P. Bertsekas. *Network Optimization: Continuous and Discrete Models.* Athena Scientific, 1998.

[58] Dimitri P. Bertsekas and John N. Tsitsiklis. *Neuro-Dynamic Programming.* Athena Scientific, 1996.

[59] Richard S. Sutton and Andrew G. Barto. *Reinforcement Learning: An Introduction.* The MIT Press, 1998.

[60] W. Luo. A three-dimensional coupled mode solution for range-dependent waveguides, Feb 2005.

[61] N. C. Makris. The effect of saturated transmission scintillation on ocean acoustic intensity measurements. *J. Acoust. Soc. Am.*, 100(2):3532–3559, 1996.

[62] C. H. Harrison and J. A. Harrison. A simple relationship between frequency and range averages for broadband sonar. *J. Acoust. Soc. Am.*, 97(2):1314–1317, 1995.

[63] Wen Xu and Henrik Schmidt. System-Orthogonal Functions for Sound Speed Profile Perturbation. *IEEE JOURNAL OF OCEANIC ENGINEERING*, 31(1), January 2006.

[64] P.F.J. Lermusiaux and C.-S. Chiu. Four-dimensional data assimilation for coupled physical-acoustical fields. In Pace and Jensen [10], pages 417–424.

[65] Ding Wang and Henrik Schmidt. Adaptive rapid environmental assessment simulation framework. *J. Acoust. Soc. Am.*, 117:2624–2625, 2005.

[66] Ding Wang, Wen Xu, and Henrik Schmidt. Investigation of the relative significance of individual environmental parameters to sonar performance prediction uncertainty. *J. Acoust. Soc. Am.*, 112:2425, 2002.

[67] Trevor Hastie, Robert Tibshirani, and Jerome Friedman. *The Elements of Statistical Learning.* Springer, 2001.

[68] Ding Wang, Pierre F. J. Lermusiaux, Patrick J. Haley, Wayne G. Leslie, and Henrik Schmidt. Adaptive Acoustical-Environmental Assessment for the Focused Acoustic Field-05 At-sea Exercise. In *Proceedings of IEEE/MTS Oceans'06 Conference*, Boston, MA, September 2006.

[69] Mario Bunge. Two faces and three masks of probability. In Evandro Agazzi, editor, *Probability in The Sciences.* Kluwar Academic Publishers, 1988.

[70] Francis Begnaud Hildebrand. *Advanced calculus for applications.* Prentice-Hall, 1976.

[71] Arnold Sommerfeld. *Partial Differential Equations in Physics.* Academic Press, 1964.