USING GENETIC ALGORITHMS TO OPTIMIZE BATHYMETRIC SURVEYS FOR HYDRODYNAMIC MODEL INPUT

A Thesis

by

DINESH MANIAN

Submitted to the Office of Graduate Studies of
Texas A&M University
in partial fulfillment of the requirements for the degree of

MASTER OF SCIENCE

December 2009

Major Subject: Ocean Engineering

USING GENETIC ALGORITHMS TO OPTIMIZE BATHYMETRIC SURVEYS FOR HYDRODYNAMIC MODEL INPUT

A Thesis

by

DINESH MANIAN

Submitted to the Office of Graduate Studies of
Texas A&M University
in partial fulfillment of the requirements for the degree of

MASTER OF SCIENCE

Approved by:

Chair of Committee,	James M. Kaihatu
Committee Members,	Jennifer L. Irish
	Emily M. Zechman
	Steven F. DiMarco
Head of Department,	David V. Rosowsky

December 2009

Major Subject: Ocean Engineering

ABSTRACT

Using Genetic Algorithms to Optimize Bathymetric Surveys for Hydrodynamic Model Input. (December 2009)

Dinesh Manian, B.Tech., Indian Institute of Technology Madras,

Chennai, India

Chair of Advisory Committee: Dr. James M. Kaihatu

The first part of this thesis deals with studying the effect of the specified bathymetric resolution and ideal bathymetric form parameters on the output from the wave and hydrodynamic modules of Delft-3D. This thesis then describes the use of an optimization to effectively reduce the required bathymetric sampling for input to a numerical forecast model, by using the model's sensitivity to this input. A genetic algorithm is developed to gradually evolve the survey path for a ship, AUV, or other measurement platform to an optimum, with the resulting effect of the corresponding measured bathymetry on the model, used as a metric. Starting from an initial simulated set of possible random or heuristic sampling paths over the given bathymetry using certain constraints like limited length of track, the algorithm can be used to arrive at the path that would provide the best possible input to the model under those constraints. This suitability is tested by a comparison of the model results obtained by using these new simulated observations, with the results obtained using the best available bathymetry. Two test study areas were considered, and the algorithm was found to consistently converge to a sampling pattern that best captured the bathymetric variability critical to the model prediction.

To my parents

ACKNOWLEDGEMENTS

I would first like to express my heartfelt gratitude to my advisor, Dr. James Kaihatu, for providing me constant support throughout my research and for his guidance and patience while allowing me to work in my own way. I would also like to thank him greatly for giving me opportunities to present my work in scientific gatherings, which helped make me more aware of other developments in the field and build confidence in my own presenting skills. I feel that I could not have wished for a better advisor. I am also indebted to my committee members, Dr. Steven DiMarco, Dr. Jennifer Irish, and Dr. Emily Zechman, for consenting to be on my committee and for providing guidance at various points throughout the course of this research. I must especially thank Dr. DiMarco and his Ocean Observing Systems class for providing the motivation to pursue this line of research and Dr. Zechman for her guidance on the subject of Genetic algorithms and her invaluable help in working out a strategy to approach my thesis problem.

Thanks also go to all my friends and colleagues and the Ocean Engineering faculty and staff for making my time at Texas A&M University a memorable experience.

This work would not have been possible without funding from Award number N00014-07-1-0169, granted by the Coastal Geosciences Program (Code 322 CG), Office of Naval Research, which provided the necessary financial support for my graduate study and research.

Finally, I thank my parents, for always standing by me providing encouragement and for being a strong source of moral support.

TABLE OF CONTENTS

CHAPTER Page

LIST OF FIGURES

FIGURE Page

LIST OF TABLES

CHAPTER I

INTRODUCTION

Hydrodynamic models are of significant utility for many naval operations where knowledge of waves and currents in a region near the coast is required. Among the inputs required for these models are estimates of wind, wave spectra at open boundaries, and good bathymetric data (i.e. underwater topography) at the modeled locations. However, significant knowledge gaps exist in the coastal bathymetric record in many potentially strategic areas, requiring supplemental field surveys. Moreover, weather, security and sovereignty issues make coastal seabed mapping potentially difficult and dangerous. Under these circumstances, where timely information is required and highly valued, it becomes significantly important to know the dominant spatial scales of resolution that determine model accuracy, so that no more time is spent on collecting field data than what is absolutely necessary. Autonomous underwater vehicles (AUVs) have been shown to be a tenable platform for bathymetric data collection in such situations. Thus, to make the best possible use of available resources and data, techniques that make use of existing datasets to optimize the application of such platforms would be beneficial.

The Delft-3D model (Lesser et al. 2004) is a commercial hydrodynamic software package widely used in many engineering and defense applications for simulations of waves, flow, sediment transport, morphology, etc. The Flow-Wave module of the Delft-3D in particular is of significant interest to the US Navy, who have in recent years acquired a Navy-wide license for the software. This thesis therefore aims to study the behavior of this model and its sensitivity

---

This thesis follows the style of Coastal Engineering.

to bathymetric input, and develop a method to optimize the required data collection.

There has been some related prior work on this topic. Plant et al. (2002) developed an interpolation technique for bathymetric data processing which allowed for control of the spatial scale of various bottom features in the interpolated result. The resulting technique was also useful in determining the relative amount of smoothing smaller scale features would undergo. The effect of this smoothing on the Delft3D-Wave (SWAN) and Delft3D-Flow models was analyzed by Plant et al. (2009). They determined that the wave and flow models responded to changes in spatial scales in the bathymetry.

1. OUTLINE

The first part of this study (Chapter II) analyses flow and wave propagation over ideal bottom features - elliptical shoal, continuous sandbar, and sandbars with rip channels to quantify the sensitivity of the model to the defined characteristic parameters of each form. These features have been selected for study, as they are among the most commonly occurring in the coastal zone, and the results serve as a good first step towards a more general picture of how the model responds to changes or errors in input bathymetry. However, extending the same approach to develop sampling criteria for practical applications could be rather cumbersome as it would involve among other things, identifying these features in the naturally occurring bathymetry, applying the appropriate sampling conditions for each, and coupling them to a realistic range of forcing conditions.

To develop a feasible way to determine an optimal sampling – optimized for use as a model input to Delft-3D - for practical purposes, the spatial variation in the sensitivity of model predicted wave-heights and currents to bathymetry particular to a given study area, needs to be examined. The goal of this study is then to identify the critical bottom features that need to be

captured in sampling so as to obtain an interpolated bathymetry that helps the model to produce the most accurate results of wave-heights and currents. The use of a global optimization scheme for this purpose is proposed in Chapter III, which first compares different approaches to the sampling problem.

Computer generated random survey paths are then evaluated for their ability to capture the areas critical to the model function and a genetic algorithm is developed and applied to evolve the fittest possible path. This fitness is ascertained by comparing the model results produced from the path interpolated bathymetry with those from the best available bathymetry. A general background of the concept of genetic algorithms is laid out in Chapter III, and Chapter IV describes the procedure followed to adapt and apply this technique to the current problem. Chapters V and VI present the results and conclusions respectively, along with possible improvements and alternatives to the present approach.

CHAPTER II

EFFECT OF ERRORS IN BATHYMETRIC INPUT

The effect of errors in bathymetric input on the Delft-3D model results was studied for three ideal-form bathymetries. A brief introduction to the Delft-3D model, which was the hydrodynamic model used throughout this work follows.

1. DELFT-3D MODEL DESCRIPTION

1-A Delft-3D Wave

SWAN or Simulating Waves Near-shore (Booij et al., 1999), the wave module of the Delft-3D hydrodynamic package used in this study, is a third generation wave model based on the Eulerian formulation of the conservation of wave action density. This equation in Cartesian coordinates takes the form:

$$\frac{\delta}{\delta t}N+\frac{\delta}{\delta x}C_x N+\frac{\delta}{\delta y}C_y N+\frac{\delta}{\delta \sigma}C_\sigma N+\frac{\delta}{\delta \theta}C_\theta N=\frac{S}{\sigma}. \quad (1)$$

The first three terms of this equation represent the local rate of change of action density and the propagation of action over geographical space, respectively. The fourth term represents the shifts in relative frequency due to bottom depth variation and currents, while the fifth term represents the depth and current induced refraction. The source term *S* on the right hand side accounts for wave generation, dissipation and non-linear – quadruplet and triad - wave-wave interactions, which are all represented explicitly. The model is driven by wave boundary conditions and local winds. The SWAN wave model has been extensively validated with field data, most recently by Rogers et al. (2007). However, the majority of the analyses of model performance have focused on the model's source terms (e.g. Ris et al. 1999; Rogers et al. 2003).

An analysis of model behavior with respect to large scale bathymetric smoothing was performed by Kaihatu and O'Reilly (2002), who studied the impact of the level of detail present in two bathymetric surveys on the model predictions. However, this analysis was not further developed.

The wave induced set-up in SWAN is computed using the vertically integrated momentum balance equation which, in the 2D case, incorporates the observation of Dingemans et al. (1987) that the rotation-free part of the wave force drives the set-up, while the divergence-free part drives the wave induced currents. This leads to the use of the following approximation:

$$\nabla.(\vec{F} + gd\nabla\eta) = 0 \,. \qquad (2)$$

where, η is the water surface elevation including the wave induced set-up,

d is the total water depth, and

F represents the wave force vector.

SWAN can also estimate wave transmission through obstacles, though diffraction around its ends is not well modeled. Some developments in this regard (Holthuijsen et al. 2003) have been put forth, but are not presently included in the version incorporated in Delft3D.

1-B Delft-3D Flow

The Delft-3D Flow module simulates 2-D depth averaged or 3-D unsteady flow and transport phenomena by solving the unsteady non-linear shallow water equations for an incompressible fluid. It makes use of the hydrostatic pressure assumption, as the time and horizontal length scales of the modeled flow phenomena are significantly larger than the vertical scales. The governing system of equations also consists of the equation of continuity to compute vertical velocity in 3-D models, and the transport equations for the conserved constituents. The equations are formulated in orthogonal curvilinear coordinates. The flow is forced by water-level

or tides at the open boundaries, wind stress at free surface, density gradients or water-level induced pressure gradients.

2. PARAMETERIZATION OF BOTTOM MORPHOLOGY

2-A Elliptical shoal

The elliptical shoal was represented in terms of the following non-dimensional parameters:

a) R_y / L

b) R_x / L

c) h_s / h = (Shoal height) / (Water depth)

where, L = characteristic wavelength, x = long-shore direction, y = cross-shore

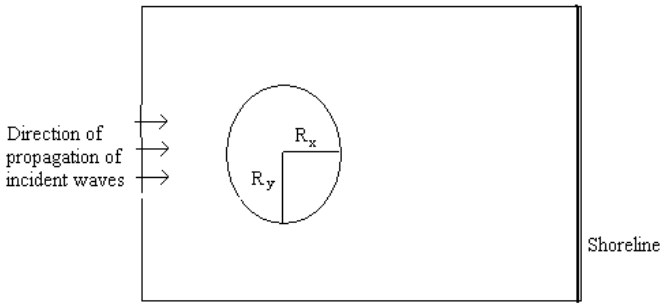


Fig. 2.1: Sketch of the form of elliptical shoal domain.

The equation for the depth over the elliptical shoal (Fig. 2.1) was adapted from the one used for depth over a circular shoal by Chawla et al. (1996), in such a way that it can be applied to any elliptical shoal of specified radii and height. This derivation is shown in the Appendix.

2-B Sandbars with rip channels (on a quadratic beach)

The sandbar forms (Fig. 2.2) were represented in terms of the following parameters:

a) L_s / L_c

b) L_b / L_c

c) h_s / h

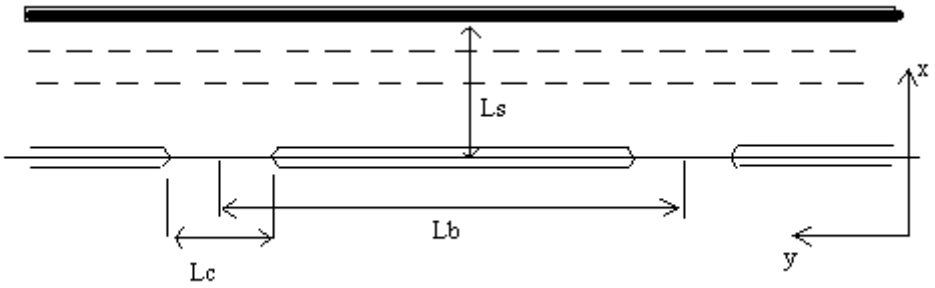


Fig. 2.2: Sketch of sandbars with rip channels on a quadratic beach.

The ideal-form bathymetry described in Haas (2000) - where it was used for SHORECIRC simulations - was used. The offshore distance of the bars, the length and crest depth of the bars, and the width of the rip channels were thus used as parameters, characteristic of the system.

2-C Continuous sandbars of Gaussian profile on a sloping beach (Reniers-Battjes set up)

The sandbar forms (Fig. 2.3) were represented in terms of the following parameters normalized by wavelength (except slope):

a) L_s – distance of bar from shore-line

b) h_c – height of the sandbar

c) w_b – width of the bar

d) S – slope of the beach

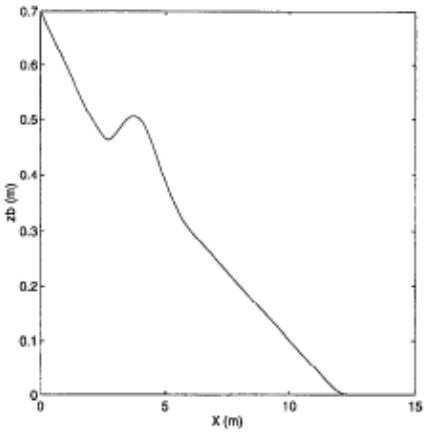


Fig. 2.3: Profile sketch of continuous sandbar on sloping beach.

The bathymetry from the Reniers and Battjes (1997) set-up for the study of long-shore currents was used here. Waves of oblique incidence were used to drive long-shore current behind the sandbar.

3. METHOD

The goal is to study two fundamental aspects of the model response to variations in bathymetry. The first is the effect of insufficient detail in the given bathymetric dataset on the performance of the model. This was achieved by examining the model behavior at different bottom grid resolutions. The second is the effect of errors in given bathymetric feature size and position on the model results. This was done by studying the response of the model to changes in the feature form parameters mentioned in the previous section.

To obtain the sensitivity of the model results to each of these parameters, the length scales of interest must first be defined. For the elliptical shoal, the analysis results over a test grid of resolution R_x by R_y (Fig. 2.1) were considered, while for the sandbars study, a scale of L_c by b was used, where b is the width of the sandbar. The results for one set of form parameters at highest bottom grid resolution were assumed as standard, with respect to which the magnitudes

of the relative error in the other results were computed. The region of maximum error, i.e., the region most sensitive to the changes in input was then identified. These errors were plotted as a function of each parameter value, in an attempt to establish a general trend.

4. RESULTS

4-A Elliptical shoal

A domain of size 18 m by 18 m, with a maximum bottom grid resolution of 0.1 m by 0.1 m was used. The boundary conditions specified are as follows:

- JONSWAP spectrum (Hasselmann et al., 1973) with peak enhancement factor, $\gamma = 7$, at open boundary
- Significant wave-height, Hs = 0.024 m
- Peak period, $T_p = 0.73$ s

The estimated wave-length of the incident waves was about 0.7 m, which is much higher than the bottom grid resolution. The wave shoaling and subsequent breaking behind the circular shoal of radius 2.57 m can be seen in Fig. 2.4.

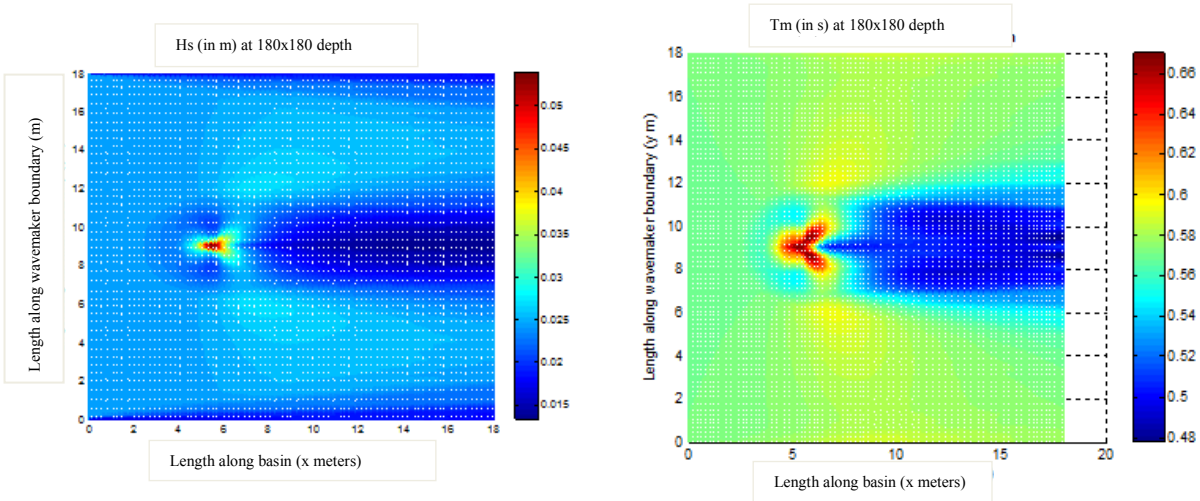


Fig. 2.4: Wave transformation over a circular shoal (wave-height and mean period).

(a) Sensitivity to bottom grid resolution

The bottom resolutions along cross-shore and long-shore were varied separately, while keeping the other fixed at a maximum. On comparing the variation of model wave-heights and wave periods with changing bottom resolution, the wave-heights were found to be far more sensitive, and thereby the critical factors that determine the required resolution for a given level of model performance.

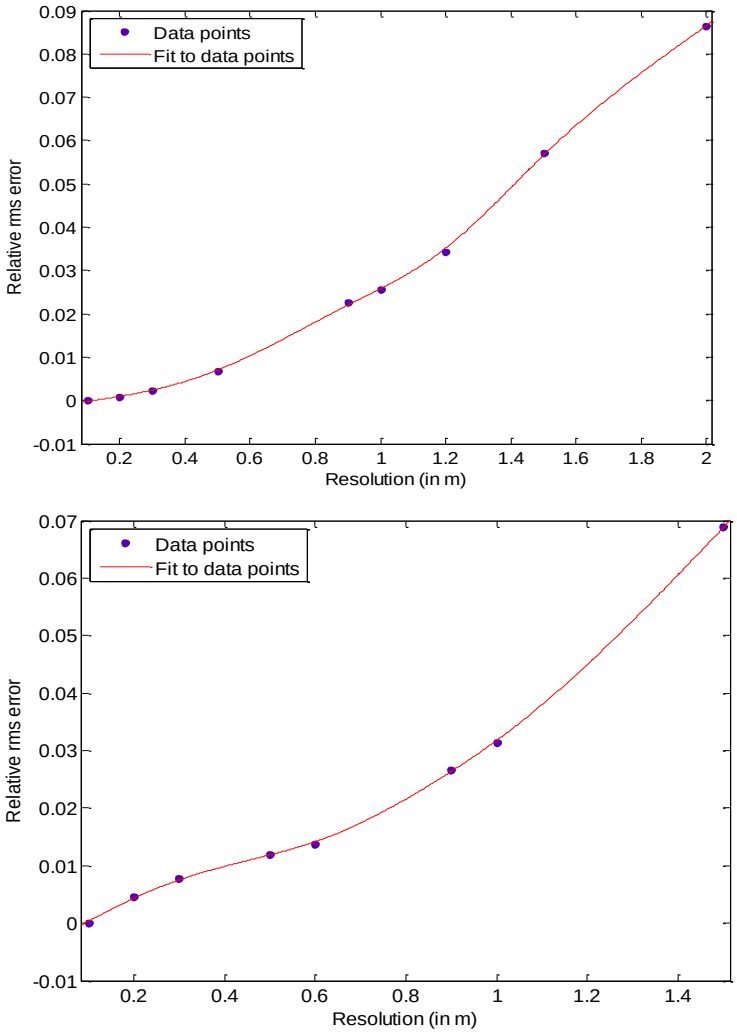
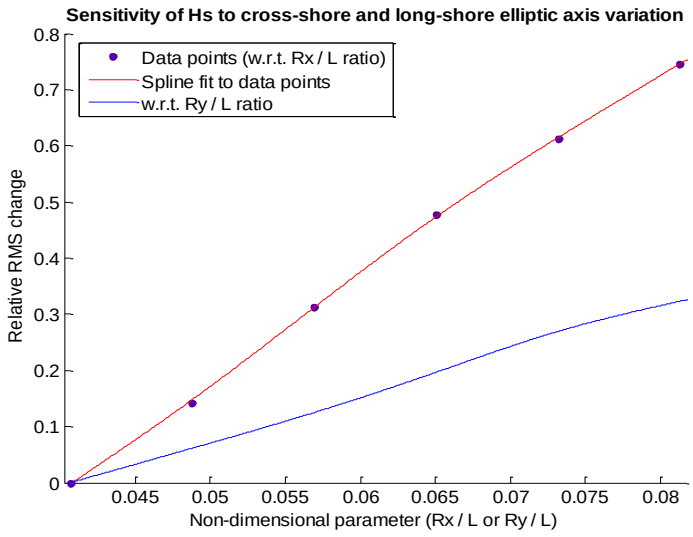


Fig. 2.5: Relative RMS error in H_s as a function of cross-shore (top) and long-shore (bottom) bathymetric resolutions.

As can be seen from Fig. 2.5, the relative RMS error shows a non-linear increase with coarser resolution. However, the model appears to be slightly more sensitive to the long-shore resolution of the shoal than the cross-shore.

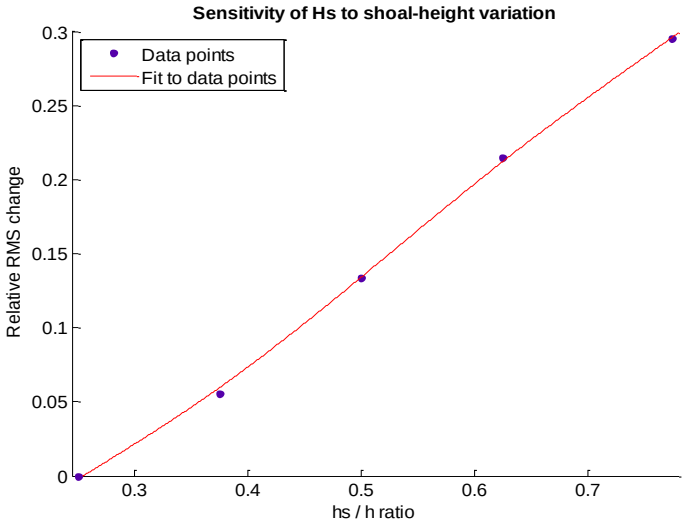
(b) Sensitivity of the model results to form parameters of the shoal

There was found to be a linear variation of the relative error with each of the three form parameters. The model was found to be most sensitive to the variation in the long-shore axis parameter, R_x / L of the shoal. The slope of the plots (Fig. 2.6) of the relative error in wave-heights against the parameters R_x / L, R_y / L, and h_s / h defined in section 2, were found to be 12, 8 and 6, respectively.



(a)

Fig. 2.6: Sensitivity of model results to shoal form parameters: (a) R_x / L, and R_y / L, (b) h_s / h.



(b)

Fig. 2.6 (contd.)

4-B Sandbars with rip channels

A domain of size 32 m by 14 m, with a maximum bottom grid resolution of 0.2 m by 0.2 m was used. It was chosen to be reasonably higher than the estimated wave-length of the incident waves, which was about 0.92 m. The boundary conditions were selected from the set-up of Haas (2000) as was the bathymetry. They were specified as follows:

- JONSWAP spectrum with peak enhancement factor, $\gamma = 7$, at open boundary
- Hs = 0.07 m
- Peak period, $T_p = 1.0$ s

The width of channel between sandbars, L_c , was kept fixed at 2 m through all the simulations, while the sandbar length, L_b, and its offshore distance, L_s were varied. Fig. 2.7 shows the wave-height and period results over the region for the setting L_b = 12 m and L_s = 5 m, and the corresponding depth-averaged flow velocity is shown in Fig. 2.8. These depict reduced wave breaking, and presence of rip currents and increased set up at the locations of the gaps between sandbars.

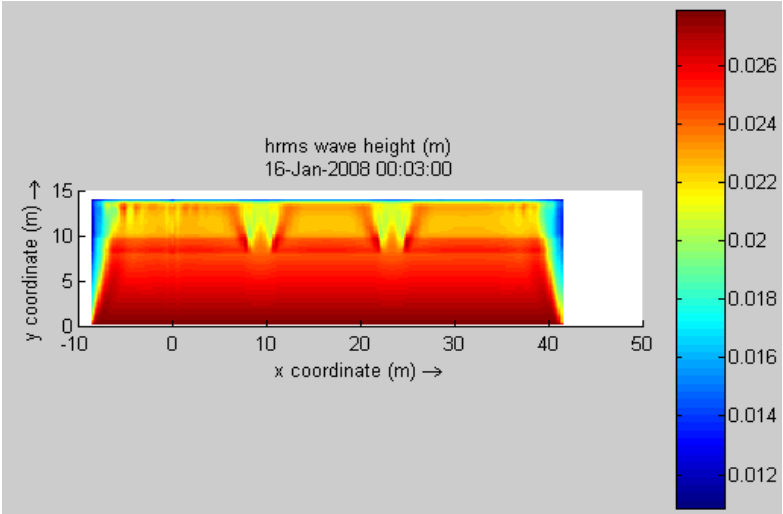


Fig. 2.7: Delft-3D Wave results for L_b = 12 m and L_s = 5 m.

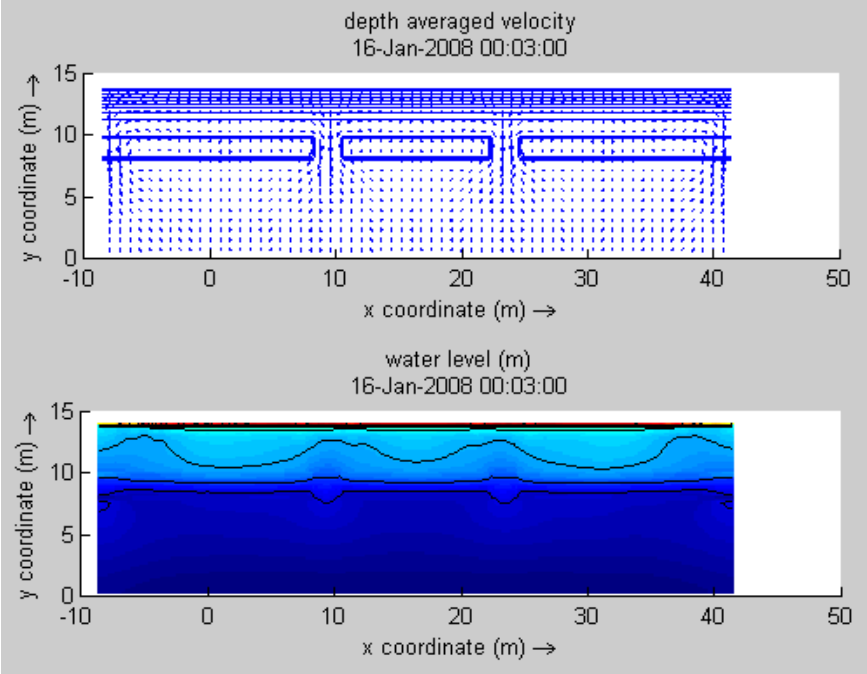


Fig. 2.8: Delft-3D Flow results for L_b = 12 m and L_s = 5 m.

(a) Sensitivity to bottom grid resolution

In this case too, the wave heights being more sensitive to changes in input, tended to determine the required resolution for a given level of model performance. Since the bathymetric contours in

the input are parallel to shore, the study of the model sensitivity to its long-shore resolution is a trivial case that would only depend on resolving the gap between the sandbars. Also, the change in model response with cross-shore resolution (Fig. 2.9) is due to the quadratic beach only, as the sandbars of constant height are fully captured at all resolutions. The small magnitude of the errors in model results indicate, that interpolated bathymetric data at computational grid points serves as a good proxy for the actual data in this case.

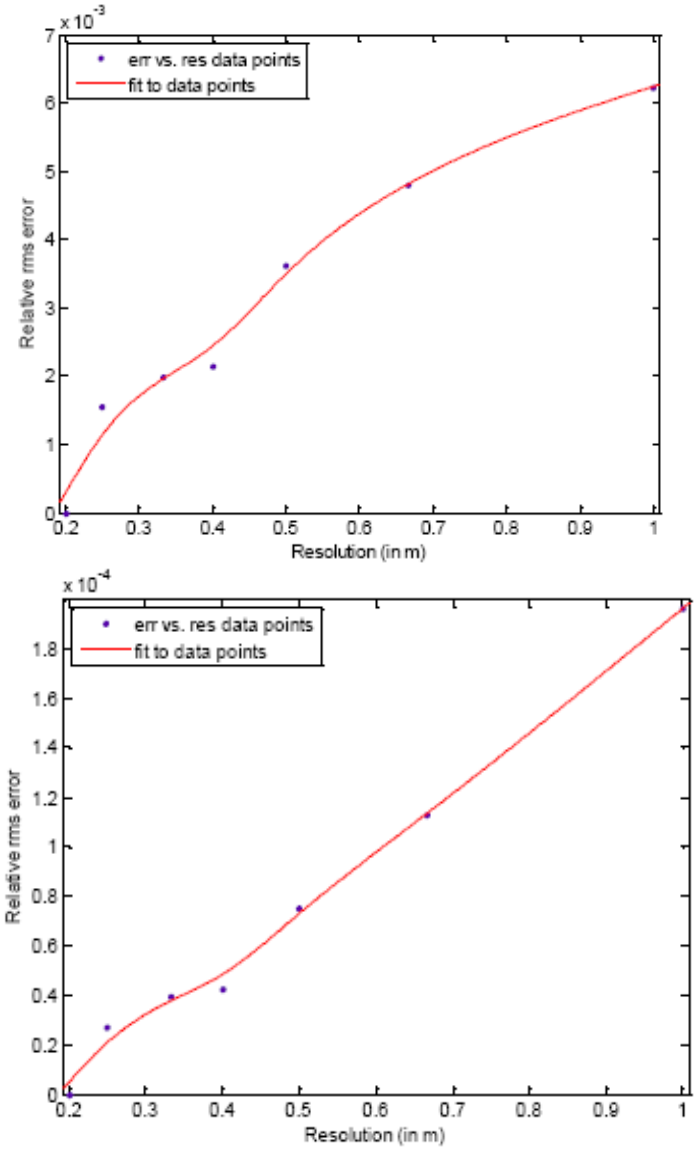


Fig. 2.9: Relative RMS error in wave-height and vorticity as a function of cross-shore resolution.

(b) *Sensitivity of the model to form parameters of the bar*

While moving the bar away from the shore, i.e. increasing the parameter L_s / L_c has some impact on the model prediction of RMS wave-height (Fig. 2.10; smoothing spline fitted curve) as might be expected, increasing the bar length or L_b / L_c only causes the gaps to shift and therefore the maximum relative error simply levels off.

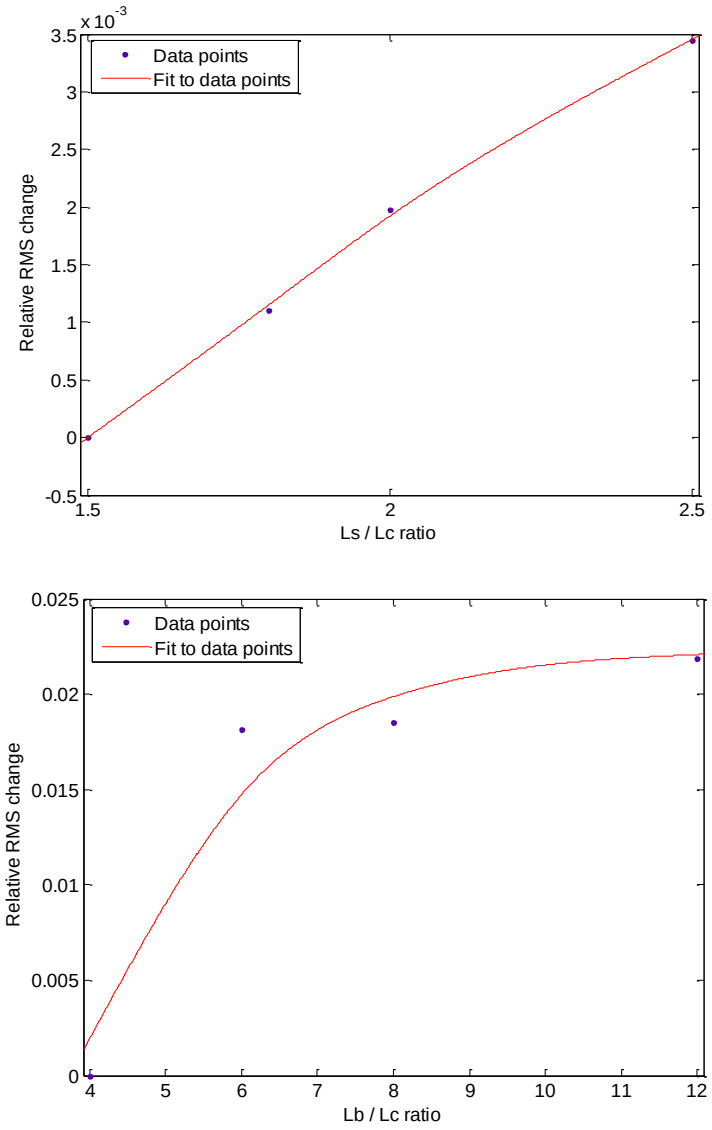


Fig. 2.10: Relative RMS error in wave-height as a function of bottom-form parameters.

Moreover, a fair correspondence between the absolute errors in the depth, RMS wave height and cross-shore current velocity over the whole beach, was found; for instance, at the rip channel where the depth is under-predicted by interpolation, an over-prediction of the RMS height and cross-shore current is observed. Fig. 2.11 shows the plot of the correlation of predicted H_{rms} error at any given computational point to the errors in depth at points situated in the same cross-shore section.

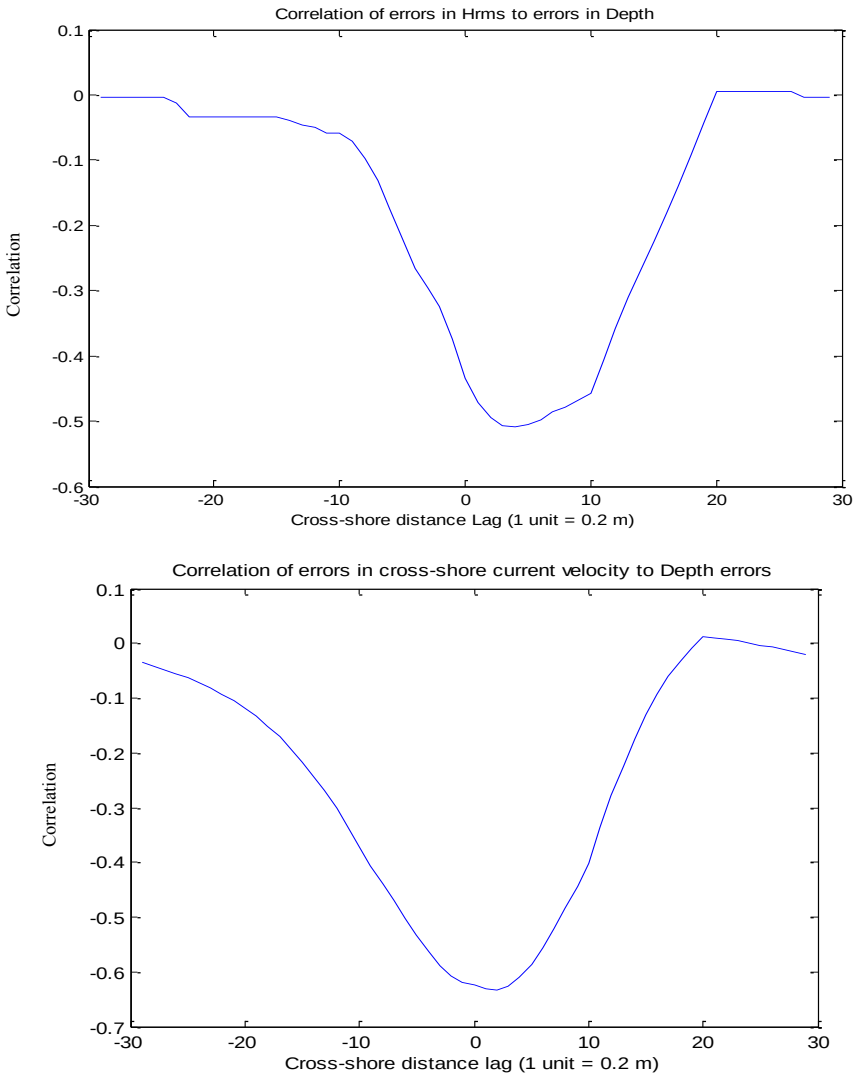


Fig. 2.11: Autocorrelation of wave-height and current velocity errors along a cross-shore transect.

4-C Continuous sandbars on sloping beach

The numerical simulations were conducted on a 93 m by 26 m domain, with a maximum bottom grid resolution of 0.2 m by 0.2 m. The boundary conditions – which were selected from the set-up of Reniers and Battjes (1997), just as the bathymetry – were specified as follows:

- JONSWAP spectrum with peak enhancement factor, $\gamma = 7$, at open boundary
- Hs = 0.08 m, $\theta = 340^{o}$ (nautical)
- Peak period, $T_p = 1.0$ s

The estimated wave-length of incoming waves was about 0.92 m, higher than the maximum specified bottom grid resolution. The simulations were carried out for different offshore bar distances (L_s). Fig. 2.12 shows the wave-height and long-shore current magnitude fields for L_s = 4 m. The oblique wave incidence at the offshore boundary was responsible for the asymmetric flow field that grew steadily stronger moving longshore.

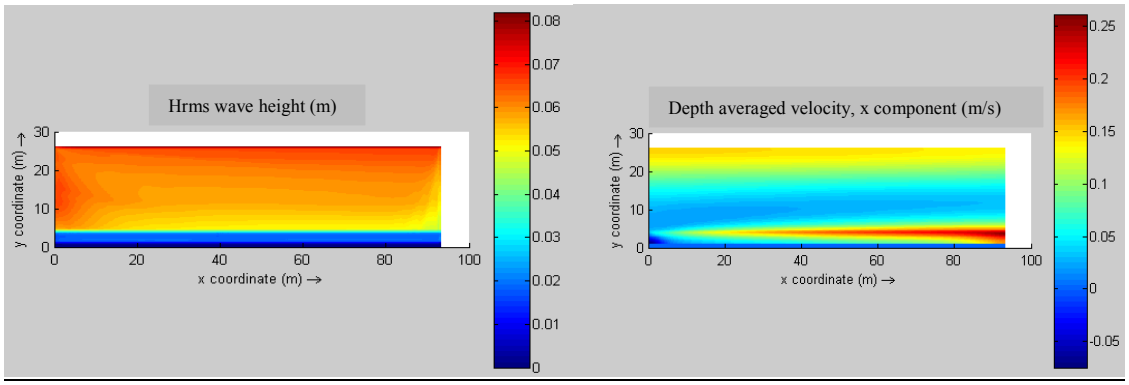
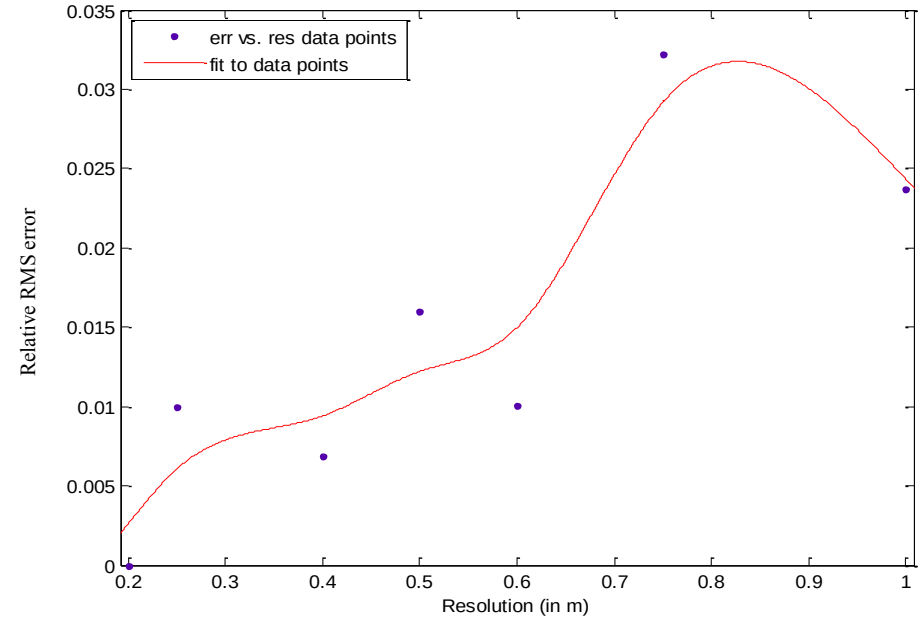


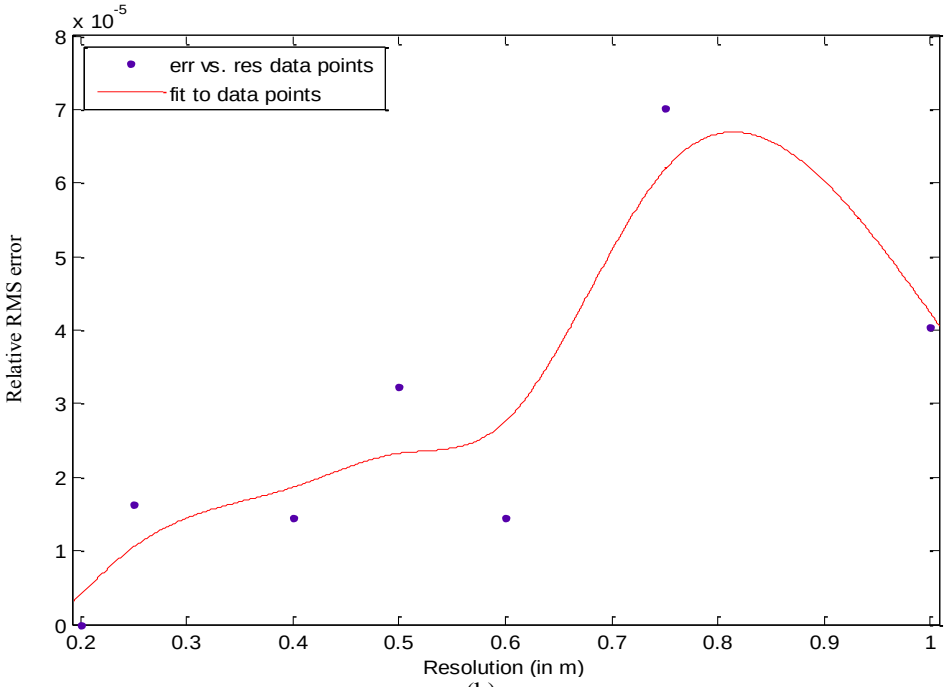
Fig. 2.12: Model wave-heights and long-shore currents.

(a) Sensitivity to bottom grid resolution

The errors in the wave-height were found to approximately mirror the relative errors in the interpolated bathymetry (Fig. 2.13), though currents were found to be more sensitive than wave-height to changes in resolution, with a higher relative error magnitude at coarse resolutions.

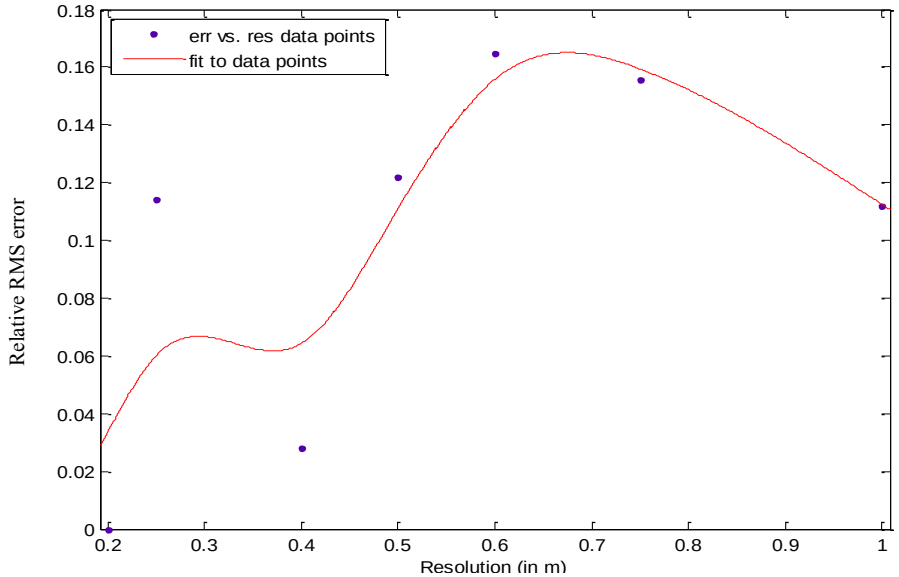


(a)

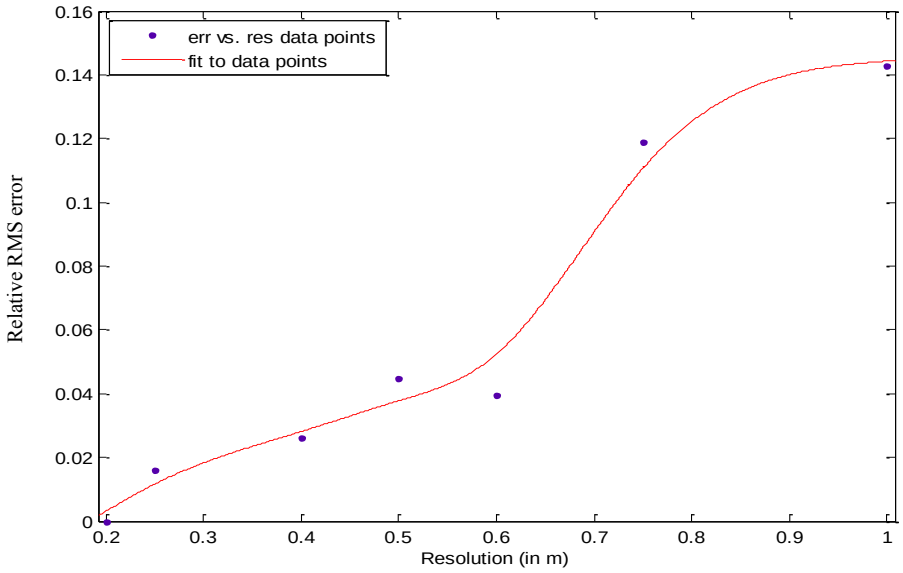


(b)

Fig. 2.13: Sensitivity of (a) interpolated depth, (b) sig. wave height, (c) cross-shore velocity, and (d) long-shore velocity - to bottom depth resolution.



(c)



(d)

Fig. 2.13 (contd.)

The errors observed at a discrete set of bottom resolutions, and a smoothing spline fitted to the observations, are shown in Fig.2.13. The 'coefficients of determination' (r^2) of the fits are 0.78, 0.76, 0.72 and 0.94 respectively. The absence of any apparent trends in the pattern of variation of

these errors points to the importance of capturing the sandbar crest. In cases where the peak of the bar was not resolved correctly, the maximum errors in wave-height occur at that position. Where the peak was resolved, the maximum errors in wave-height were smaller, and occurred just behind the bar. Fig. 2.14 shows the cross-shore depth profile as seen at different resolutions.

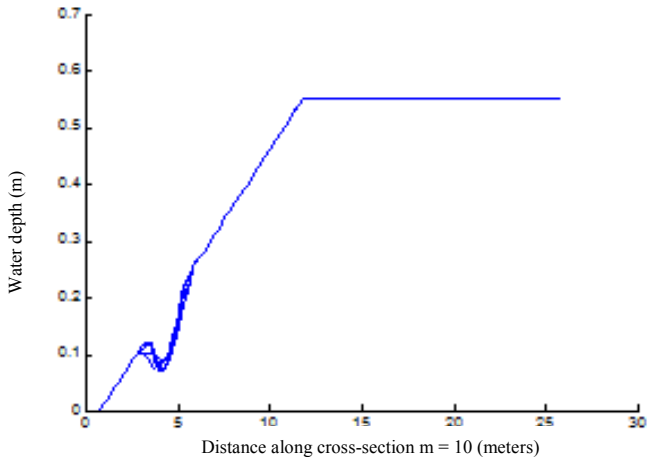
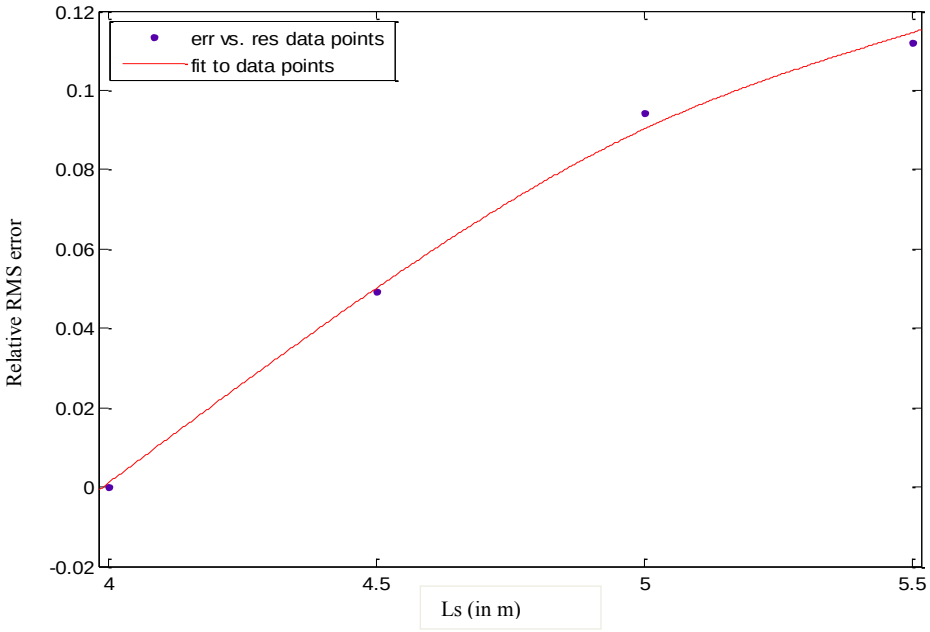


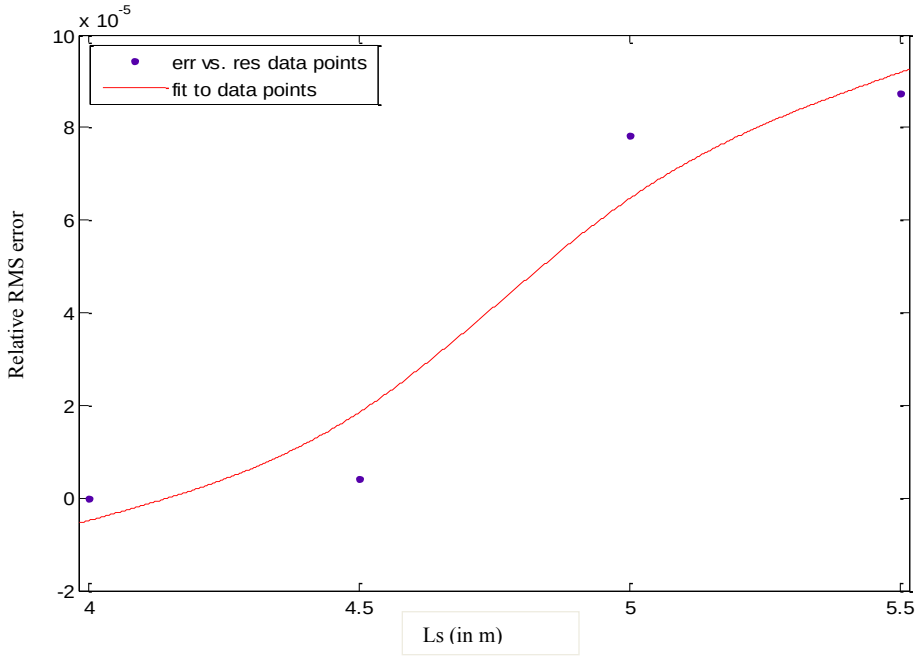
Fig. 2.14: Evolution of the interpolated depth profile with changing bottom resolution.

(b) Sensitivity to bottom parameters

The offshore distance of sandbar, L_s, was varied while keeping the boundary conditions the same. A smoothing spline was fitted to the error observations at different parameter values (Fig. 2.15). The currents were found to be most sensitive to errors in bottom parameters, while the effect on wave-heights was relatively very small.

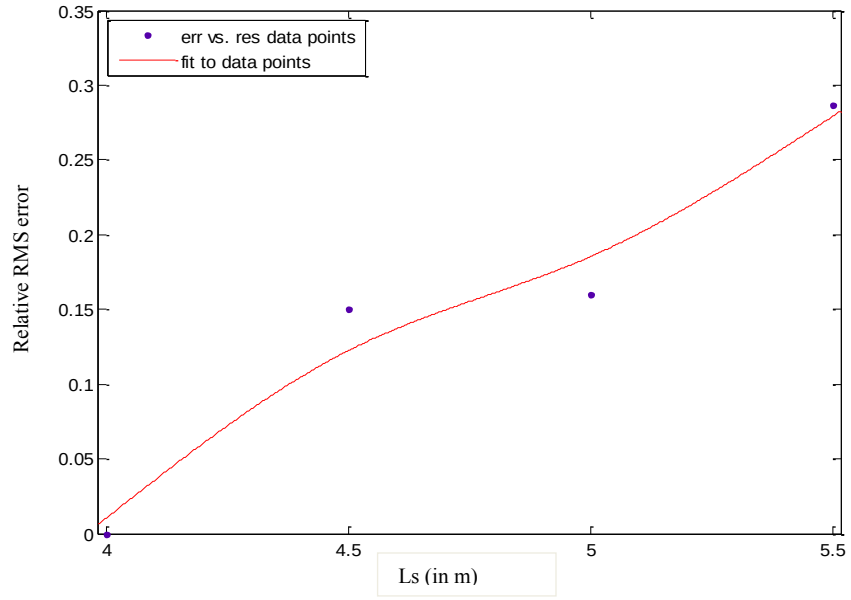


(a)

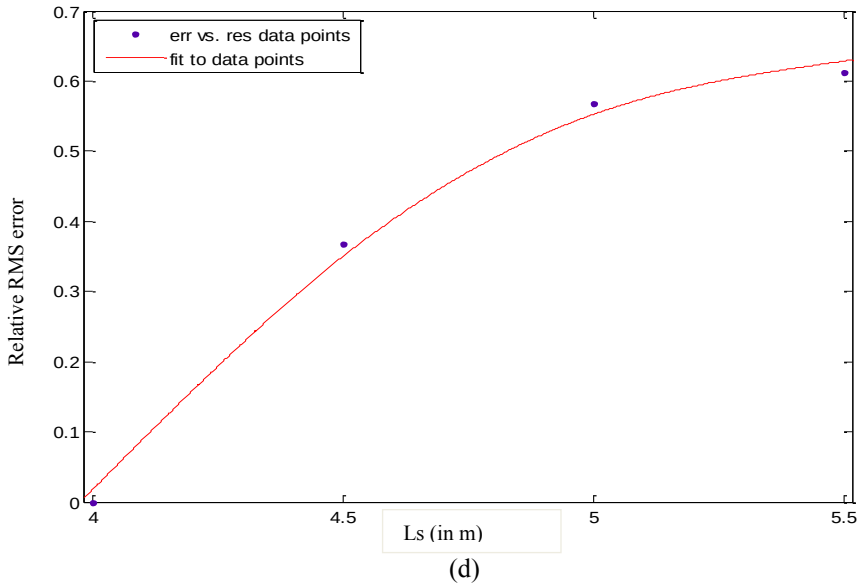


(b)

Fig. 2.15: Sensitivity of (a) interpolated depth, (b) sig. wave height, (c) cross-shore velocity, (d) long-shore velocity - to offshore distance (Ls) of sandbar.



(c)



(d)

Fig. 2.15 (contd.)

4-D Results over natural bathymetry

To get an indication of how the quantity of bathymetric data provided affects the model results in a field scenario that includes at least one of the ideal bottom features studied here, a

site off the coast of La Jolla, California was considered. Shepard and Inman (1950) indicated that the submarine canyons in this area cause a large spatial variation of the incident waveheight, which can lead to the presence of rip currents. A large scale field study of this region, called the Near-shore Canyon Experiment (NCEX) was conducted in 2003, with many different agencies contributing to the data collection. The updated bathymetric data and offshore wave spectra from the NCEX program were used for this model study. The dimensions of the area modeled were about 12 km by 14.6 km. The corresponding bathymetry and model wave-heights are shown in Fig. 2.16. The resolution in each direction of the rectilinear grid was varied from 40 m (compared to an incident wavelength of 80 m) to 200 m. On reducing the y-direction (long-shore) resolution, significant variation was found in the model cross-shore and long-shore velocities averaged over the entire domain. Fig. 2.17 shows this variation along with a smoothing spline fit to the error observations, of coefficient of determination (r^2) 0.94 and 0.98.

5. SUMMARY

While the sensitivity analysis study of the model to the parameters of the elliptical shoal feature produced some linear trends in the errors in model results, the study on ideal-form barred beaches showed that errors in the sandbar profile lead to errors in the wave-generated long-shore current behind the bar. It was not always possible, however, to establish a generic relationship between the bottom resolution and the model results. Also, there are practical difficulties in the application of the results of even much more comprehensive sensitivity studies over different ideal bathymetries such as this, to answering the questions of how much bathymetric information is needed, and where, for a given flow or wave model to produce acceptable results.

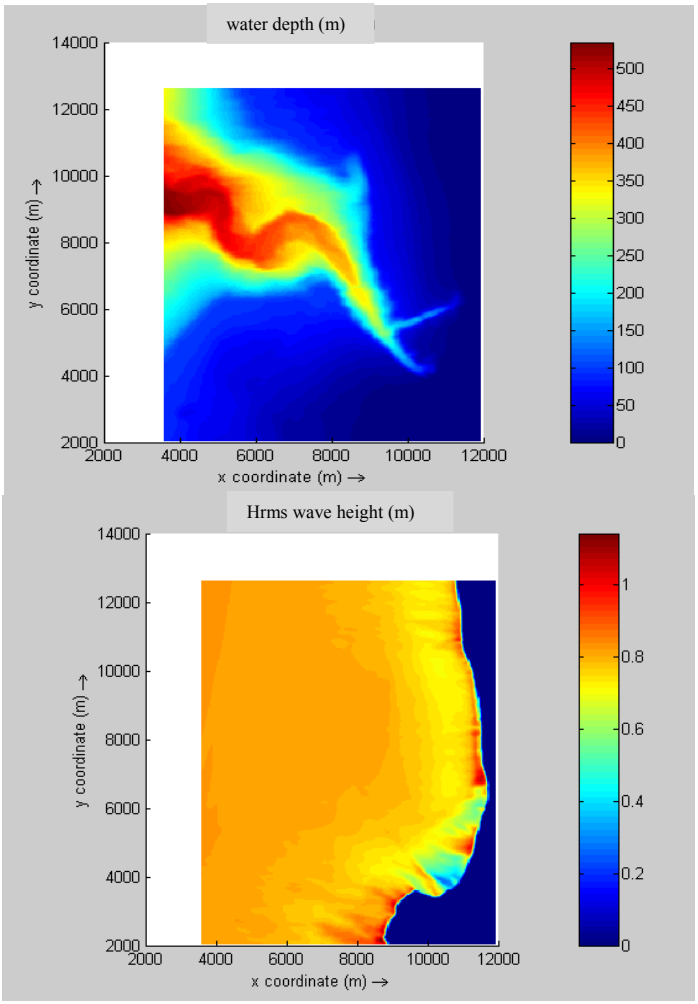


Fig. 2.16: Bathymetry and modeled wave-height off La Jolla, California coast.

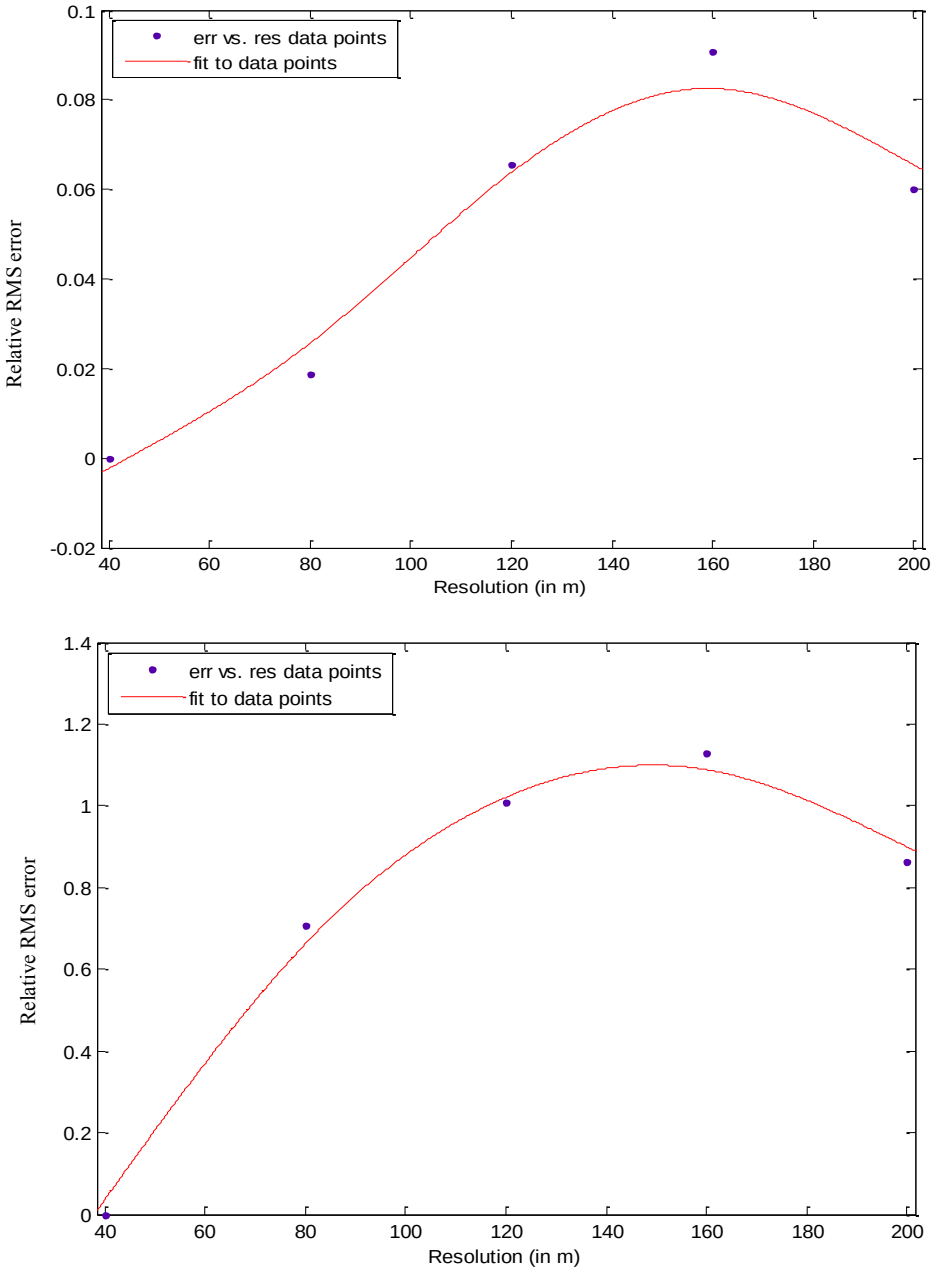


Fig. 2.17: Sensitivity of interpolated depth (top) and cross-shore velocity (bottom) to bottom long-shore resolution.

CHAPTER III

GENETIC ALGORITHMS

1. APPROACHES TO THE SAMPLING PROBLEM

If the objective of the bathymetric sampling is defined as the generation of the most cost effective input for the Delft-3D hydrodynamic model, one possible method of determining the sampling strategy could be by studying the theoretical sensitivity of the predicted wave heights and currents at the required location(s) to the bathymetry. This sensitivity would vary in space, and a high sensitivity at any given set of locations would indicate that greater importance must be given to capturing those locations during sampling. However, the ubiquity of depth dependent terms in the governing equations of both the Wave and Flow modules as seen in Chapter II, and the nonlinear dependence of predicted waves and currents on depth, precludes this. Therefore, any effort to quantify the effect of the input local bathymetry on the model results would have to use nonlinear or stochastic methods.

Genetic algorithms (GA) are just such a method. GA's are widely used to solve problems in optimization and machine learning. The general theme of this approach is to evolve a set of candidate solutions to a problem using operators inspired from natural selection and genetic variation.

2. GENETIC ALGORITHMS

The stochastic approach of GA's gives them an advantage over other traditional methods of optimization, especially when dealing with multi-modal problems (i.e. when there is more than one extremum). While traditional methods like gradient descent might not be able to differentiate between a global and local extremum, the GA ensures that the whole search space is

covered, therefore increasing the probability of convergence to a global extremum. GA are also well suited for application to multi-criteria optimization problems, where the optimum solution is required in the presence of several conflicting constraints.

Natural processes and phenomena have historically inspired many inventions and problem solving techniques (Liu and Tsui, 2006). The robustness and efficiency of natural phenomena has led to many attempts to replicate these concepts in artificial systems. The idea of evolutionary computation, of which genetic algorithms are a sub-class, came about as a result of efforts to emulate nature to solve optimization problems (Davis, 1991). In nature, there exist variations within species which are brought about by reproduction. Of the many offspring produced by the individuals of a population, only a few survive to adulthood and reproduce in turn. This is known as 'natural selection', where the survival of an offspring is dependent on its suitability and adaptability to its environmental conditions. Similarly, a genetic algorithm maintains a population of possible solutions to the given problem, to which the concept of 'survival of the fittest' is applied. Reproduction is achieved by the 'crossover' of parent chromosomes, and the mutation operator is responsible for introducing diversity by randomly changing some genes. Holland (1975) was the first to successfully model the mixing of chromosomes that occurs in nature, to create a functioning genetic algorithm. The following section further describes this method, and brings out the meaning of a few of these technical terms in the context of the optimization problem.

3. METHOD OUTLINE

A basic genetic algorithm has the following steps (Goldberg, 1989):

1) Representing the problem;
2) Generating initial population of candidate solutions;

3) Evaluating the fitness parameter for each;
4) Selecting individuals to produce new offspring;
5) Recombining the selected individuals using genetic operators for crossover and mutation;
6) Reinserting the new solutions to the population;
7) Repeating steps 3 to 6 till convergence.

The following is a brief general description of each of these steps. Chapter IV relates these to the problem at hand.

Problem representation: A set of candidate solutions to the problem, maintained by the GA is referred to as a population. Each individual of the population is called a chromosome. Each chromosome is composed of a string of genes. These genes could typically be binary valued representations or real valued numbers or an array composed of individual members, referred to as alleles.

Population initialization: To produce the initial population, random number generators are typically used to produce the gene values for each individual, to ensure sufficient spread over the entire search space. The size of the population depends on the nature of the problem, or (more specifically), the nature of the search space. However, a larger population size, though requiring larger number of computations to produce offspring, translates to greater diversity which usually helps in faster convergence in a fewer number of generations.

Fitness evaluation: The suitability of a particular solution is characterized by computing the 'fitness value'. This fitness corresponding to the specified optimization criterion is calculated for each individual, and the objective of the algorithm is to maximize this in subsequent generations.

Selection: In the genetic algorithm framework, the fitness value also acts as a tool with which to select individuals for mating, which is done so that individuals with higher fitness have a greater

effect on the properties of the next generation of solutions. There are various methods of selection documented in the literature (Goldberg and Deb, 1991). The 'Roulette selection' method is one such, where the probability of a member being selected to reproduce is directly proportional to its calculated 'fitness value' relative to the other members.

Reproduction: New individuals are produced from the selected pairs in each generation using the genetic operations of crossover and mutation. The crossover operation combines the characteristics of the two parent chromosomes to form new offspring. 'Single point crossover' technique selects a random r^{th} gene in the n gene sequence, and genes 1 to r of the first chromosome and $(r + 1)$ to n of the second chromosome are copied to the new offspring. The probability of a crossover operation is specified by the crossover rate. Similarly, the probability of mutation of a gene is specified by the mutation rate, and is carried by inserting a new randomly created gene in place of the selected gene.

Reinsertion: The new offspring are then immediately inserted into the population in place of the least fit individual, which are discarded. This variation, which is known as elitism, ensures that the best performing individuals are always retained, and helps in faster convergence.

CHAPTER IV

METHODOLOGY

Genetic algorithms (GA) have been used as a global optimization scheme in a variety of applications. They were first used in the area of oceanographic experiment design by Barth (1992), who considered a time dependent design problem for an idealized experiment. Baehr et al. (2004) first employed it for pre-deployment array design in optimizing an observing system for the North Atlantic meridional overturning circulation. They compared this technique with both the simulated annealing method and a heuristic approach, and found genetic algorithms to be a significantly faster method than simulated annealing and at the same time more successful in finding the optimum solution than the heuristic approach.

On the use of AUVs for oceanographic surveys, Bellingham and Wilcox (1996) described a method to determine the optimum resolution and extent of survey that would minimize the energy cost and the total survey error. However, no investigations of methods to increase the efficiency of non-uniform oceanographic surveys could be found in the scientific literature. The parameter to be optimized in such surveys would be the path of motion of the survey vehicle, rather than a 1-D point distribution as would be the case when optimizing a typical observing system. The problem thus becomes one of path planning to minimize the total survey error at an acceptable energy cost. The problem of path planning in general and for AUVs in particular, has been quite extensively studied before in the context of getting from one point to another in the ocean with a given current field, at minimum energy cost (Alvarez et al., 2004, Fox et al., 1999). However, the survey path planning problem has somewhat conflicting requirements of maximizing survey extent to achieve minimum survey error, and doing so with minimum path distance, which calls for a different GA scheme.

1. OBJECTIVE

The broad objective of this study is to estimate the required amount of bathymetric information for the hydrodynamic model, and the sampling strategy needed to minimize this amount. This is done by setting ourselves two goals.

The first goal is to estimate the spatial variation in the optimum cross-shore and long-shore resolution requirements of bathymetry for the model for a given study area. This is hence referred to as the 'optimum resolution problem'. This is equivalent to the problem of designing the spatial distribution of parallel long-shore or cross-shore bathymetry survey tracks for best possible bathymetric input to the model.

The next goal is to design an optimum continuous path for a bathymetric survey vessel such as an AUV, focusing on the utility of the thus sampled, model-grid interpolated data as a model input for Delft-3D (hence referred to as the 'survey design problem'). Two different schemes for the same goal are developed and evaluated.

Under scheme 1, the problem is somewhat simplified by constraining the number of degrees of freedom of the path such that it can only take perpendicular turns, and is of the form of a line by line sweep of the coverage area as shown in Fig. 4.1. The sweep direction shown in the figure is long-shore. This approach ensures at the outset a more or less uniform coverage of the area of interest. However in this case, the basic structure of the path to be followed is pre-determined.

Under scheme 2, the GA is given more leeway to decide the structure of the path to be followed, by using a strategy that only sets an upper limit of the length of the path, and allows it to move in any of eight possible directions. Starting from an initial randomly chosen point, the step-length is chosen as the minimum unit of distance that must be travelled in a straight line, before changing direction. Thus, the path has to pass through adjacent points on a grid of

resolution equal to the step-length. Though a smaller step-length might be desirable in order to make the algorithm select and converge on the best possible path, the step-length must be set high enough so that there is no excessive concentration or looping in a small area at the expense of wider coverage of the complete area of interest.

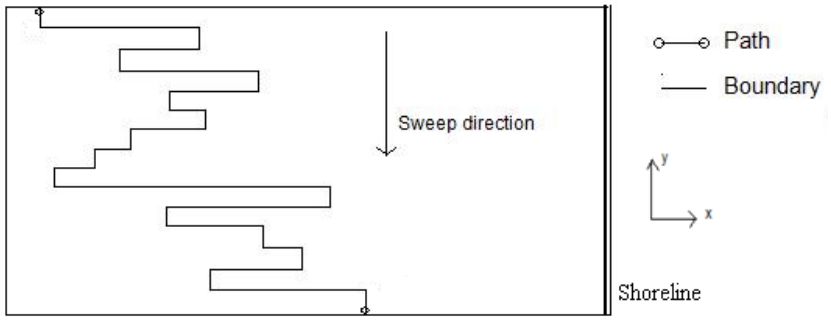


Fig. 4.1: A typical AUV path for bottom sampling as modeled in the study (Scheme 1).

2. APPLICATION OF THE GA

The general blueprint of the GA followed in both these problems mentioned in the objective, remains the same as described in the previous chapter. However, the problem representation schemes and implementations vary. The following sub-sections provide a detailed description of the implementation of these steps for the stated problems in the previous section.

2-A Optimum spatial resolution problem

(a) Encoding scheme

The 'encoding scheme', which describes how each individual solution is encoded as a chromosome, needs to be carefully selected, as it determines the effectiveness of the genetic operations performed on the chromosomes. In this case, each individual solution was encoded as a string of a fixed number of real valued numbers (genes) that correspond to the y-coordinates of the long-shore tracks, or x-coordinates of the cross-shore tracks. The positive x-direction was

assumed to point towards the shoreline. The number of genes in the encoded solution varied according to the number of tracks desired.

(b) Initial population generation

The individuals of the initial population were generated by using the built-in random number generating function 'rand' in *Matlab* to simulate a real numbered value (x or y coordinate in the given domain) corresponding to each gene. Different population sizes were tried, and their effect on the performance of the GA was studied, with the goal of using a population size leading to the best convergence characteristics in the solution.

(c) Fitness calculation

The model was first run with the best available, highest-resolved bathymetry and the results thus obtained were used as the standard of comparison (a so-called “golden standard”) for evaluating the fitness of the sub-sampling schemes. The inverse of the mean absolute difference between the 'golden standard' results and the individual results over the entire area of interest was then defined as the fitness parameter of that individual.

(d) Selection

The Roulette selection method described in the previous chapter was used to select individuals to reproduce based on the 'fitness value' calculations.

(e) Genetic operations

Single point crossover was used, which means that the parent chromosomes were truncated at a single randomly selected gene, before being recombined to form new offspring. A 100 percent crossover rate was used so that this operation is always performed on the selected individuals. Then a mutation operation was used in order to introduce new properties (random genes) to the offspring solution at a specified mutation rate, which was selected for best

convergence on the basis of the results of the application of the GA to a test problem, described in the section IV-3.

(f) Replacement

The new offspring was immediately inserted into the population in place of the individual with the least fitness. This variation, referred to as elitism, ensures that the fittest available individuals are always retained at every stage of the process, which ultimately helps in faster convergence.

2-B Survey design problem: Scheme 1

(a) Encoding scheme

In this case, each individual is a random path of fixed length. Since spatial location is the primary property characteristic of these various alternate solutions, the encoding scheme must be such that each gene has a uniquely identifiable geographical area associated with it. Only then can crossover and other genetic operations be used to create progressively better solutions. Therefore, the path (chromosome) was divided into sections (genes) of equal length, and the study area divided along the y-axis (see Fig. 4.1) at equal distances, so that each cross-shore strip of area had just one associated gene.

By constraining the structure of the solution path to the form of a line by line sweep with a fixed jump in the y-direction after each sweep line, each gene could then be encoded as a string of numbers representing the starting x-coordinates and the directions of each line associated with that gene. To make sure that the length of the section of path corresponding to each gene is a constant '*l*', '*n*' sub-sections (parallel tracks) were created by selecting '*n-1*' random points between 0 and '*l*', where '*n*' is the number of equidistant parallel tracks (could be long-shore or cross-shore) associated with that gene. This produced a set of *n* real numbers 'd_i', obeying the following constraint:

$\Sigma_1^n d_i = l.$ (3)

The x-coordinate of the starting point of the section was set equal to the x-coordinate of the end point of the previous adjacent section, and the x-coordinate of the start of each new sub-section was decided by:

$x_{j,i} = x_{j,i-1} + k_{i-1} d_{i-1}$, (4)

where, $k_i = \pm 1$,is a random number,

j is the gene index, and

i is the allele index.

The next step was to check whether the path thus described obeys the constraints of the given boundary. If not, the current sequence was discarded and a new one created until the constraints were met. The number of such iterations required was found not to be large enough to significantly increase computational costs. Thus the encoded solution was of the form:

$[\{x_{1,1}, \ldots, x_{1,n}\}, \{x_{2,1}, \ldots, x_{2,n}\}, \ldots, \{x_{m,1}, \ldots, x_{m,n}\}]$,

where, m is the number of genes, and n, the number of alleles of a gene.

(b) Initial population generation

The genes of all the chromosomes of the initial population were initialized with the help of the *Matlab* pseudorandom number generating function '*rand*', while making sure that each member satisfies the set constraints of the domain boundaries and the constant length.

(c) Fitness calculation and selection

The fitness parameter was calculated here as in the previous case, by calculating the inverse of the mean absolute total error, and the Roulette selection method was used.

(d) Genetic operations

The crossover rate used here was 100 percent, and a single-point crossover at a randomly selected r^{th} gene was performed. However, a simple crossover would not produce an

acceptable offspring in this case as the resultant path would most likely be discontinuous. Therefore, while the remaining genes of either parent were carried over to the offspring during the crossover in a normal fashion, a random mutated gene was inserted in place of the r^{th} gene, such that the resultant individual was acceptable, and the conditions of continuity and length of the path maintained.

Also, to make sure that the length of the section of path corresponding to each gene is constant at '*l*', and at the same time continuous with the adjacent sections, '*n-2*' sub-sections of the form $[(0, l_1), (l_1, l_2), \ldots, (l_{n-3}, l_{n-2})]$ were created, by selecting '*n-2*' random points between 0 and '*l*'. The first '*n-2*' alleles were then encoded and tested as described in the encoding scheme. The $(n-1)^{th}$ allele was chosen to satisfy:

$d_{n-1} + d_n = l - \Sigma_i^{(n-2)} d_i$, and

$$k_{n-1}d_{n-1} + k_n d_n = x_{j,n} - x_{j,n-2} \qquad (5)$$

In the absence of real positive solutions for d_{n-1} and d_{n-2}, the procedure was repeated until one was found. The new chromosome was then inserted into the population in place of the least fit member.

2-C Survey design problem: Scheme 2

The following steps in this scheme differ from those in scheme 1:

(a) Encoding scheme

A grid of a specified step-length was defined. The start point of the vehicle path was picked at random from the nodes of the defined grid. The integer index numbers of this grid point constituted the first two genes of the candidate solution. One of the adjacent grid points was picked as the next decision point on the path. The relative position of these grid points to the current location was encoded according to the numbering convention shown in Fig. 4.2, where

the center represents the current location. This string of relative directions encoded as integers formed the remaining genes of the individual.

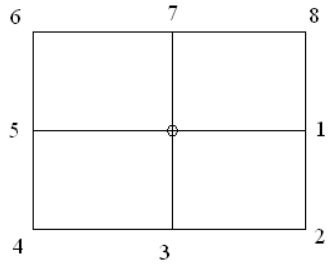


Fig. 4.2: Relative direction encoding scheme.

(b) Genetic operations

A single-point crossover was used, and mutation of individual genes carried out at the specified rate. Thus in this case, the geographic location information of one of the two parent paths was lost, and only its structure passed on to the offspring.

3. MATHEMATICAL TEST FUNCTIONS

The GA developed was tested by applying it to standard mathematical test functions for global optimization techniques. The first such function used was the Rastrigin function (DeJong, 1975), which is defined for 2 independent variables as:

$$Ras(x) = 20 + X_1^2 + X_2^2 - 10(\cos 2\pi X_1 + \cos 2\pi X_2)\,. \qquad (6)$$

The minimization of the Rastrigin function is an unconstrained optimization problem, as there are no additional constraints on the unknown variables X_i. The GA was used to minimize this function for *n* number of independent variables, where *n* represents the number of genes in the problem representation. While this function has multiple local minima, it has a known unique global minimum of Ras(x) = 0 at $[X_1, X_2, \ldots, X_n] = [0,0,\ldots, 0]$.

In the intended AUV path problem however, the solution path is constrained by the maximum allowed length of path and the regional boundaries. To test for an *n*-dimensional constrained optimization problem, the Griewank G2 function (Griewank, 1981) was minimized. The G2 problem is defined as:

$$\max_x f(x) = \left| \frac{\sum_{i=1}^{n} \cos^4 x_i - 2\prod_{i=1}^{n} \cos^2 x_i}{\sqrt{\sum_{i=1}^{n} i x_i^2}} \right|, \tag{7}$$

such that: $g_1(x) = -\prod_{i=1}^{n} x_i + 0.75 \le 0$,

$$g_2(x) = \sum_{i=1}^{n} x_i - 7.5n \le 0 .$$

The results of the Rastrigin function minimization tests are shown in Figs. 4.3 and 4.4. In Fig. 4.3, the large spread in the function values seen in the initial population disappears in the final population, where almost all the members of the population seem to have reached an equal minimum value. This indicates convergence, as any further evolution is unlikely with a predominantly homogeneous population. On varying the population size (Fig. 4.4), it was found that at smaller population size, the number of iterations for initial convergence is generally smaller even though the actual global minimum of zero is not always reached. This could be attributed to the fact that for a smaller population size, the same number of cost function computations (iterations) correspond to a larger number of generations, which means more opportunities to crossover and mutate. The algorithm here was found to take about 12-15 generations to converge. Thus the population size may be selected by taking into account the acceptable error criteria for the given problem, and the value placed on quicker computation. Table 5.1 summarizes the performance of the algorithm on the Rastrigin function minimization

problem, with different settings. The average of the results of 40 runs of each case was used to create this table. The last two columns of the table show the average final solution obtained after convergence, and the number of iterations required for the convergence.

The algorithm was then applied to the problem of G2 function maximization for 17 independent variables. For each combination of settings of the population size and mutation rate, the algorithm was executed 40 times, and the average results obtained were compared with the similarly averaged results for the same settings with the *Matlab GA toolbox*. Table 5.2 summarizes this comparison. As seen in Fig. 4.5, the variance in the fitness value distribution of the population was found to decrease substantially with subsequent generations, implying convergence. In this case, an average of 15-20 generations was required for convergence. The effect of the specified mutation rate on the performance of the algorithm was then studied, and relatively higher rates were found to be most effective for quickest convergence to a low enough cost function.

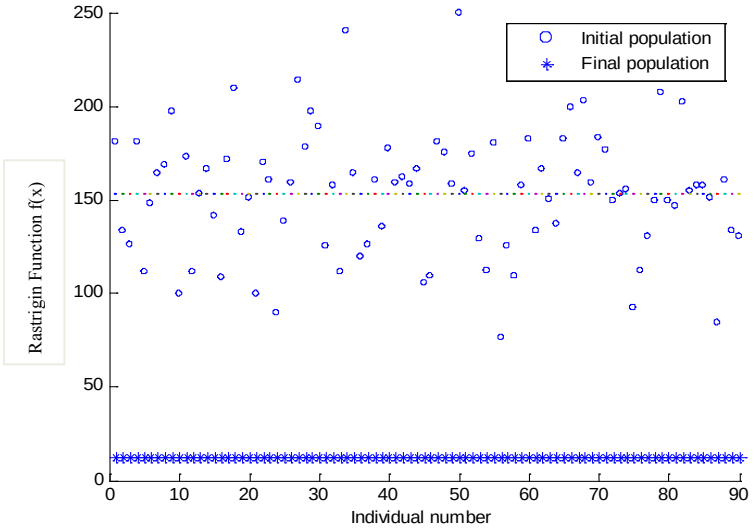


Fig. 4.3: Rastrigin function distribution of members of the initial and final population (after 6000 iterations).

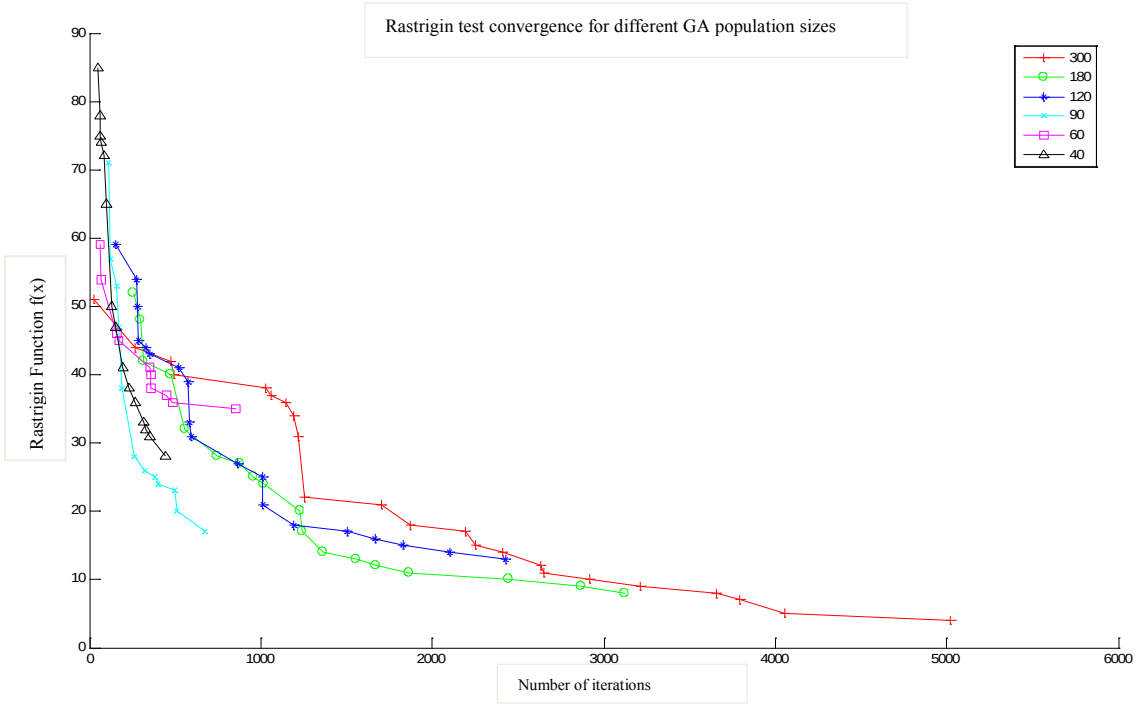


Fig. 4.4: 20-D Rastrigin function minimization: Best solution at each iteration (Max. iterations = 6000).

TABLE 5.1: Rastrigin function minimization (Global minimum known to be zero)

Population size	Mutation %	Best solution	Cost func. evals.
	Developed GA		
50	7	28.57	740
100	11	15.97	1413
100	7	15.27	1328
200	11	8.92	3031
200	9	8.77	3020
200	7	8.37	2923
200	5	9.15	2983
300	11	6.67	4788
300	7	6.12	4740

TABLE 5.2: G2 function maximization

Population size	Mutation %	Best solution	Cost func. evals.
	Developed GA		
50	11	0.2517	7871
50	9	0.2410	7798
50	7	0.2440	7144
50	5	0.2440	7521
100	11	0.2628	14592
100	7	0.2646	24623
200	11	0.2628	22491
200	7	0.2630	19888
	GA Toolbox		
50	11	0.3382	7885
50	9	0.3611	7736
50	7	0.3267	7925
50	5	0.3257	7951
100	11	0.3650	15522
100	7	0.3165	15713
200	11	0.4457	30592
200	7	0.3333	31640

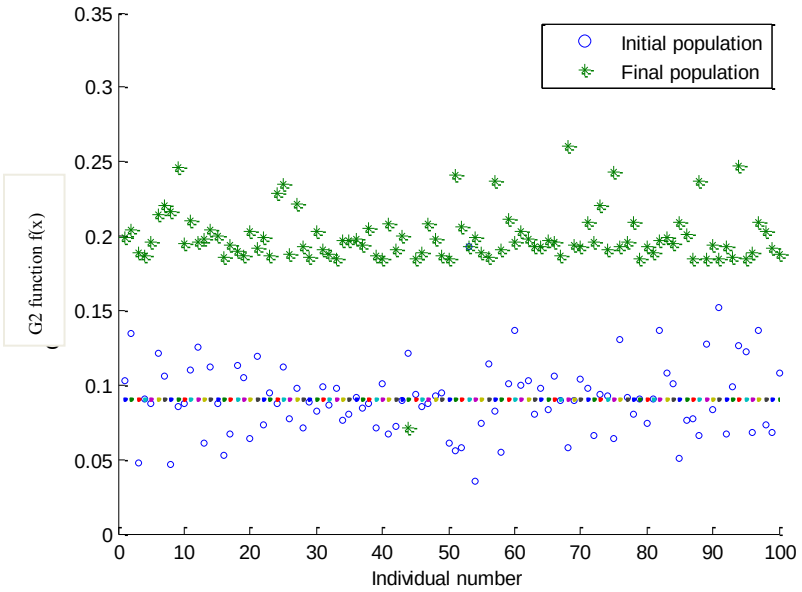


Fig. 4.5: G2 function maximization: Initial and final population distributions (after 15000 iterations, pop. size = 100).

CHAPTER V

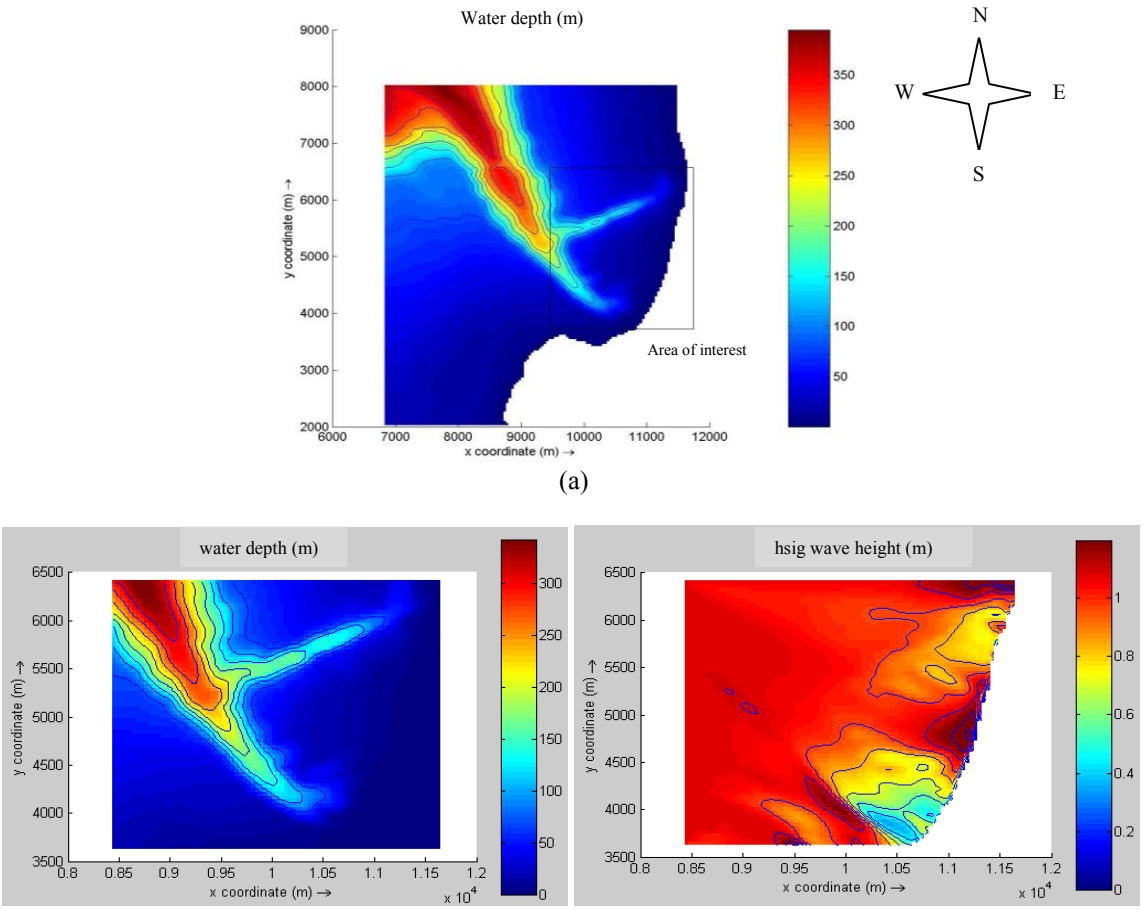
RESULTS

1. STUDY AREA DESCRIPTION

The data from the site of the Nearshore Canyon Experiment (NCEX), conducted near the Scripps and La Jolla canyons off the coast of La Jolla, California in 2003, was used to test the effectiveness of the GA method. The bathymetric data and the offshore wave spectrum forcing for the model was made available for this region from the surveys conducted during this period. The steep topography at the canyons could be expected to cause significant changes in wave energy longshore, and this expected spatial variation in sensitivity of the overall model results to bathymetry was one of the reasons for selecting this as the study area.

Due to the offshore wave spectrum data being about 7 km offshore, the SWAN model was first run over a larger domain (shown in Fig. 5.1a), and the wave spectrum results along the offshore boundary of an approximately 3 km by 3 km grid (shown in Fig. 5.1b) was written out to be used as the boundary condition for the SWAN runs over this smaller area of interest. A computational grid resolution of 40 m both in the cross-shore and longshore was used. This was chosen to be sufficiently high to reduce the effect of numerical artifacts on the model results, and thereby maximize the impact of the input conditions.

An area off the coast of Camp Lejeune, North Carolina served as a second test study region. This area of about 22 km by 18 km has much more small scale irregularity in the bathymetry (Fig. 5.2) relative to the La Jolla region bathymetry which is dominated just by the two canyons. Therefore, one might expect a significant difference in the product bathymetry and thereby model results depending on where the sampling is done. This might provide a greater incentive for performing the optimization and would be a test of the usefulness of the method.



(a)

(b)

Fig. 5.1: (a) Bathymetry over the extent of the larger Wave domain (La Jolla), and selected area of interest, (b) Selected bathymetric input field, and corresponding wave-height results.

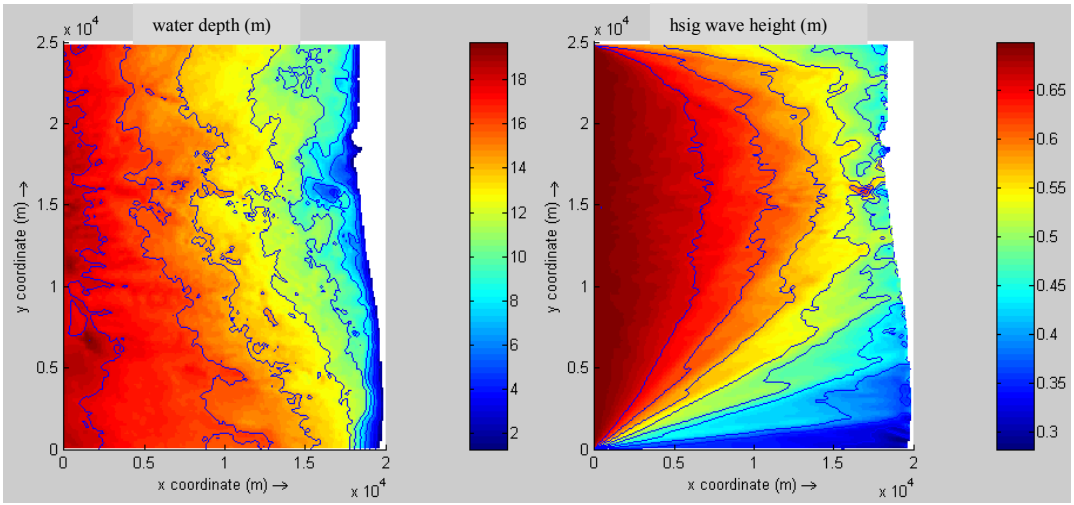


Fig. 5.2: Camp Lejeune bathymetric field (left) and corresponding wave-height results (in m).

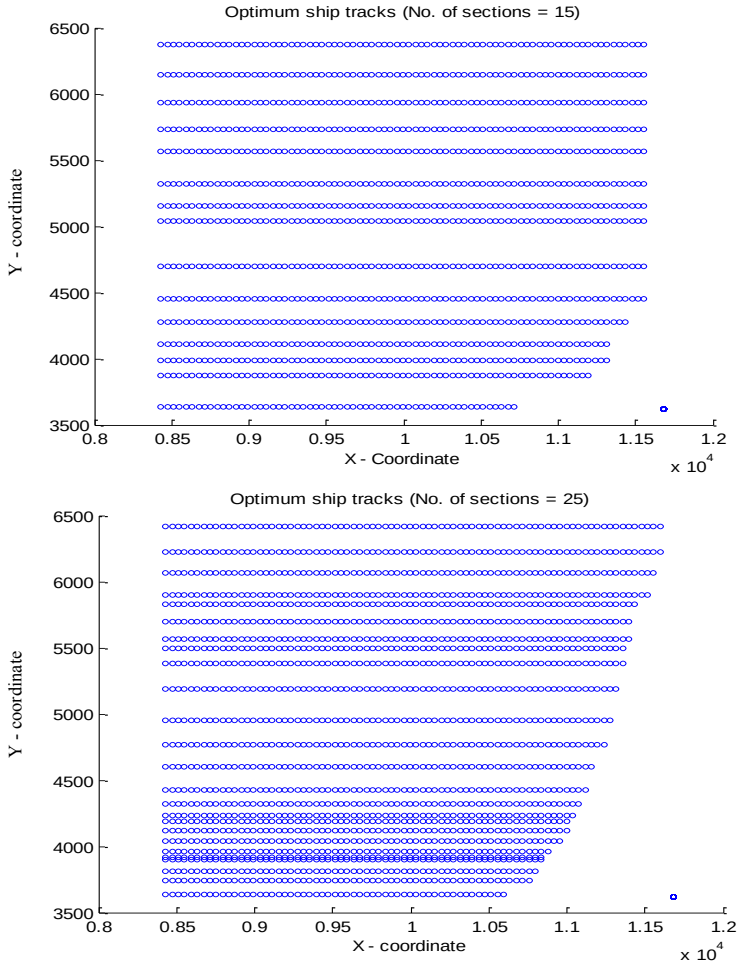
The 'golden standard' model wave-heights corresponding to the best bathymetric input is also shown in Fig. 5.2. This served as a benchmark for comparing the model results from sparsely sampled input. The typical wave boundary conditions for this region were obtained from an NDBC (National Data Buoy Center) buoy located at 34.476^{o} N, 77.280^{o} W (Station 41035, Onslow Bay, NC).

2. RESULTS FOR OPTIMUM SPATIAL RESOLUTION

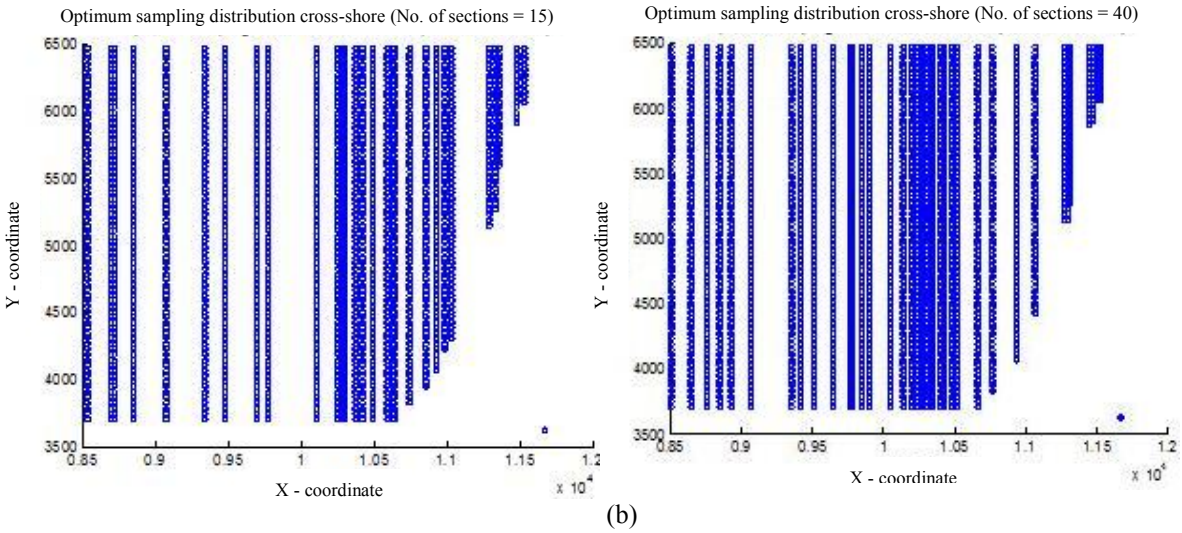
The performance of the method was measured in terms of the relative percentage mean absolute difference between the computed wave-heights (with sub-sampled bathymetry as input) over the entire study area, and the expected wave-heights - predicted using all available bathymetric data as input. Figs. 5.3*a* and *b* show the results for the case when the longshore and cross-shore sampling were optimized to a minimum cost of 5.5 % and 3% error, respectively.

As can be seen from the figures, the resultant sampling scheme is denser near the south canyon, near which the wave-height gradient is also at its greatest (Fig. 5.3a). Also the area closer to the shore seems to require greater sampling (Fig. 5.3b), which is consistent with what one might expect given the bathymetric complexity in this area.

Though the average discrepancy in the newly computed wave-height over the entire study area was used as an indicator of the performance of the method, the performance at any given location might be very different, and Fig. 5.4 shows this spatial variation in the error computed at different locations. Fig. 5.5a shows a convergence plot for this average performance metric against the number of model iterations. Fig. 5.5b shows the convergence plot when only the wave-heights of the nearshore short waves (with wavelength < 60m) are considered for optimization. The effect of sampling extent on the best achievable solution in a given maximum computational time is shown in Fig. 5.6.



(a)



(b)

Fig. 5.3: (a) Longshore and (b) Cross-shore variation in model sensitivity to the bottom.

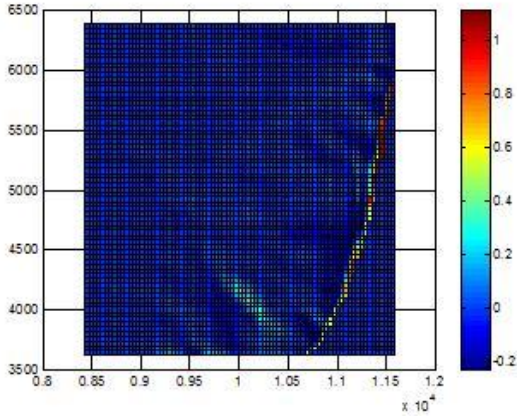
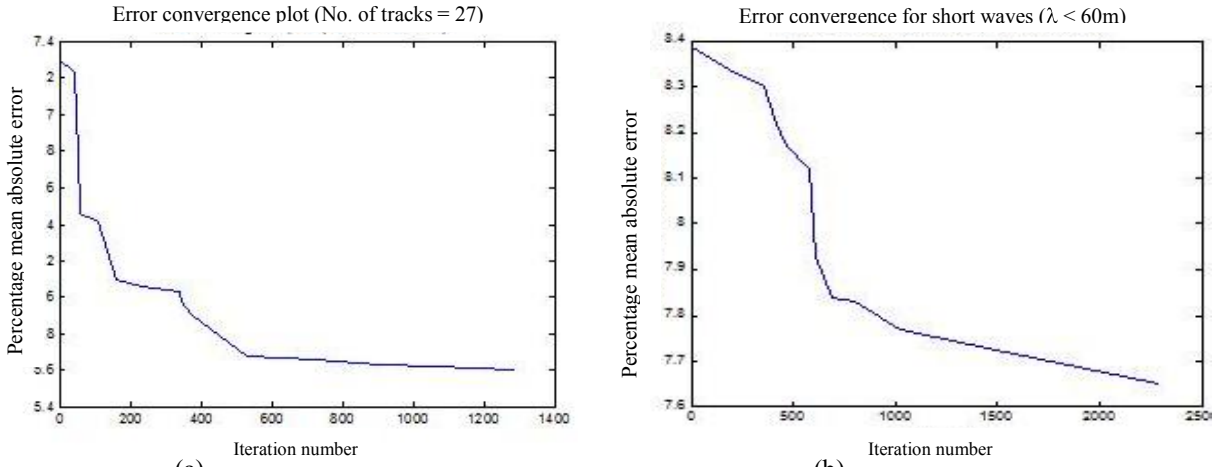


Fig. 5.4: Spatial variation of wave-height error (in m).



(a) (b)

Fig. 5.5 Error convergence plot for (a) all waves in the domain, (b) short waves only.

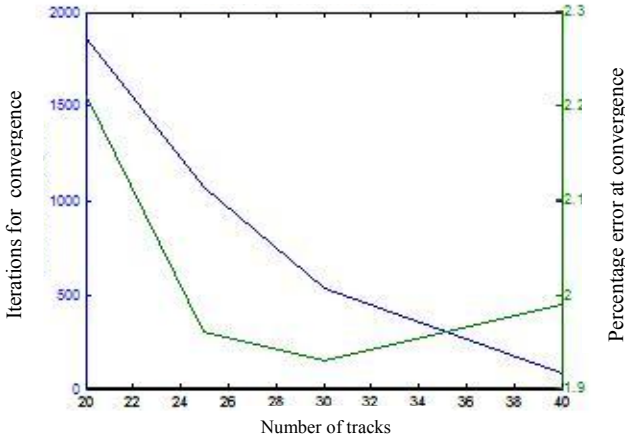


Fig. 5.6: Effect of extent of sampling on the time and level of convergence.

3. RESULTS FOR SURVEY DESIGN: SCHEME 1

Fig. 5.7 shows a 27 km length of survey track, optimized to serve the model to produce the best possible wave-heights over the selected study area. The corresponding grid interpolated bathymetry and the resultant wave-heights from the model are shown in Figs. 5.8 a and b respectively. The performance metric used was the same as described in the previous section, and Fig. 5.9 shows the convergence plot. The computation time was about 12 seconds per iteration of the GA on a 2.66 GHz dual core processor of 2 GB memory. Increasing the extent of sampling has a definite but relatively small effect in improving the performance of the algorithm, as seen in Figure 5.10.

For different specified lengths of path, the optimized paths always converged to a pattern that sampled the deep north-west canyon, and the trend-line lying in the north-west – south-east direction, so that the south canyon was also mostly captured.

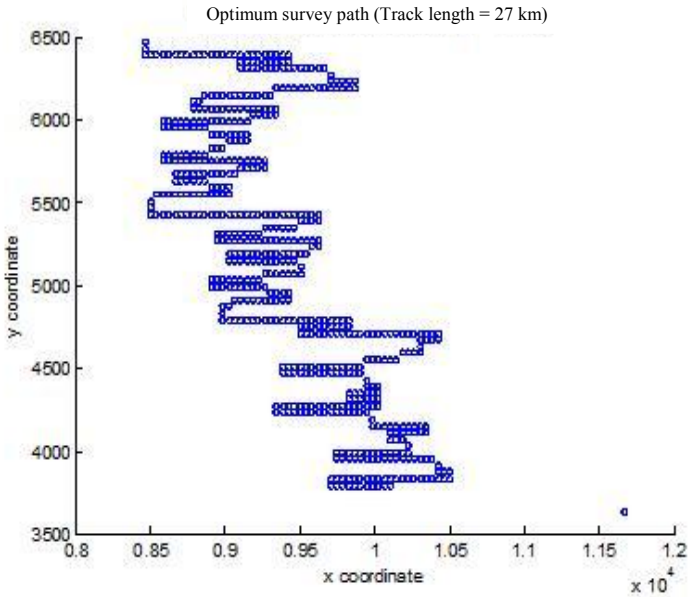
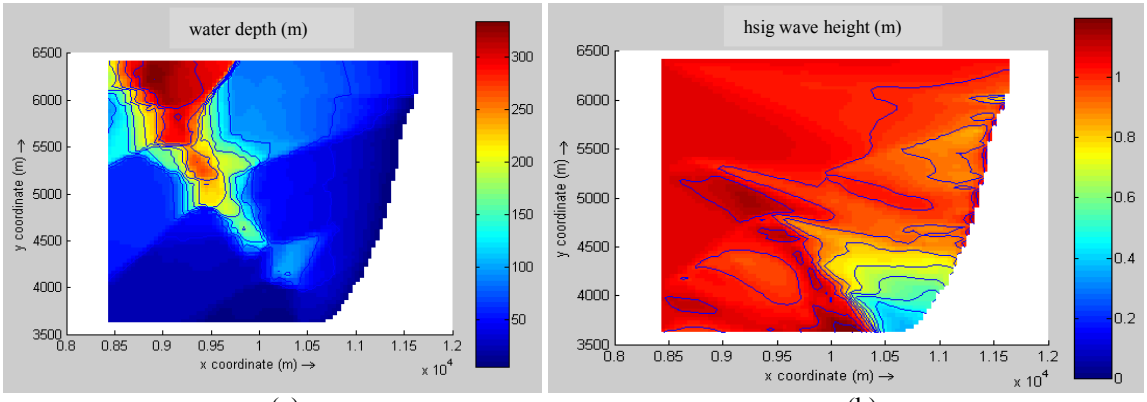


Fig. 5.7: An optimized survey track of length 27 km over the study area.



(a) (b)

Fig. 5.8 (a) Bathymetry interpolated from the sampling, and (b) Corresponding model wave-height results.

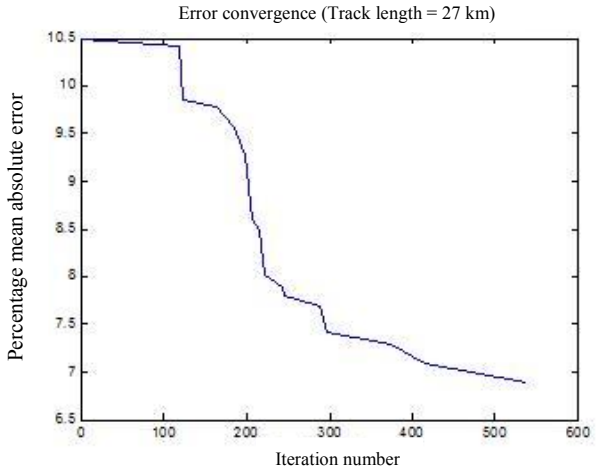


Fig. 5.9: Scheme 1: Convergence plot.

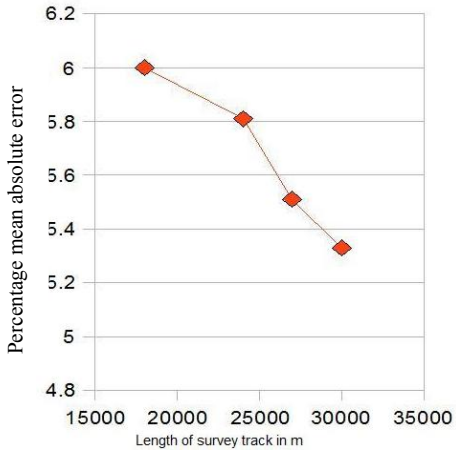


Fig. 5.10: Convergence vs. sampling extent.

4. RESULTS FOR SURVEY DESIGN: SCHEME 2

In this scheme, there are fewer constraints on the structure of the desired path. In the study over the La Jolla region, a relatively low maximum survey path length of 8 km was specified. An optimized solution path after 15 generations is shown in Fig. 5.11, while the derived bathymetry and wave-heights, and convergence are shown in Figs. 5.12 and 5.13 respectively. The computation time for this study region was of the order of 7 seconds per iteration of the GA. A step-length (see section IV-2-c) of 400 m was used here, which means every decision point on the path was followed by a straight line section of at least 400 m. This was chosen such that excessive local looping that would impede possible wider coverage is avoided, and at the same time enough decision points are allowed to give room for the solution path to evolve.

As model wave-height is the only criterion considered for optimization of the bottom sampling, the possibility of many distinct bottom configurations producing similar looking wave-height fields arises, especially so when the sampling extent is small. Also, since the cost function used considers only the spatially averaged wave-height errors with respect to the 'golden standard' (see Fig. 5.1b), there could be considerable difference in the actual spatial variation of wave-heights as seen in Fig. 5.12. As seen in the figure, the optimum path sampled the south canyon well, while the north canyon was almost completely ignored. The effect of this is evident in the wave-height fields produced, which show a better match with the 'golden standard' in the southern part of the modeled area.

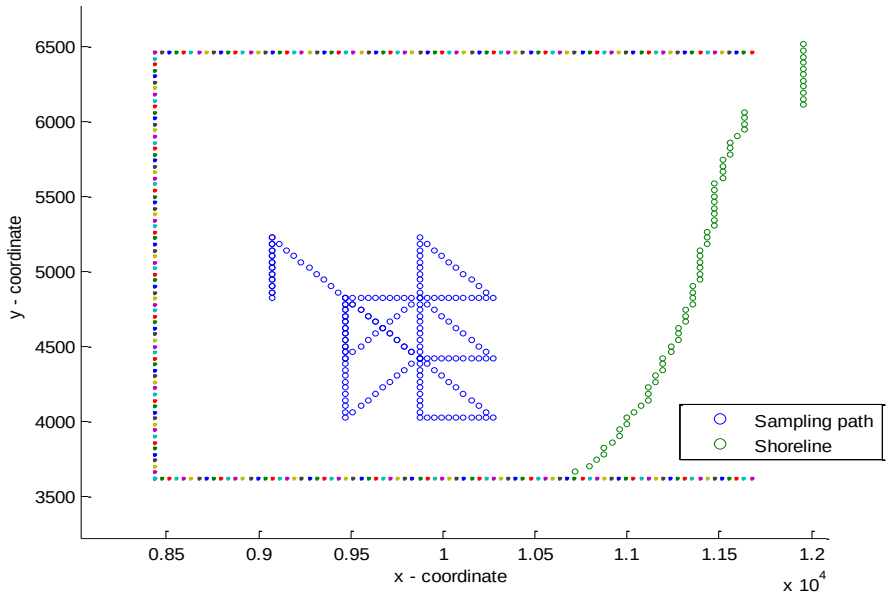


Fig. 5.11: An optimized AUV path for specified max. survey length of 8 km (La Jolla).

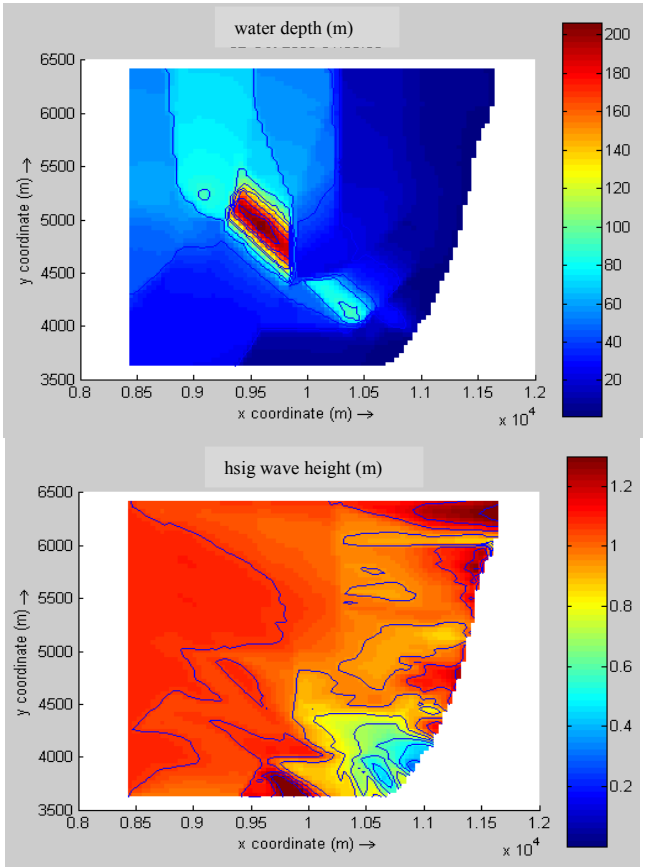


Fig. 5.12: Derived bathymetry and resultant model wave-height from optimized survey.

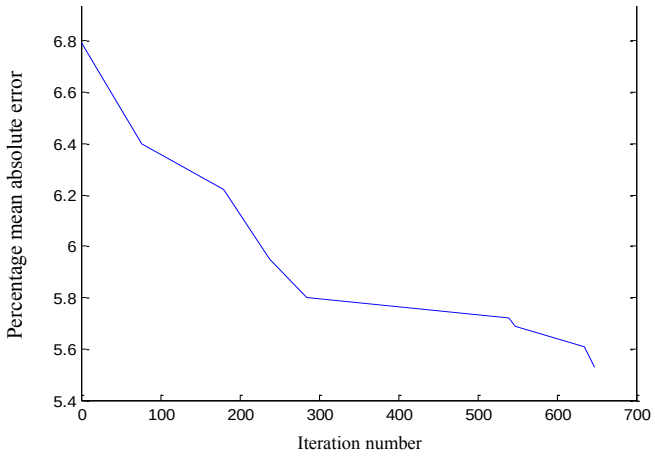


Fig. 5.13: Scheme 2: Convergence plot.

The results of the study over the Camp Lejeune, NC bathymetry – the optimized path, and the derived bathymetry and model wave-heights - are shown in Figs. 5.14 and 5.15 respectively. The upper limit used for the length of survey was 55 km on an area of about 18 km by 20 km. This length was chosen to be relatively small, in order to assess one of the worst possible scenarios. A step-length of 10 times the computational grid resolution of 92.5 m was used. In contrast with the study over the La Jolla region, the spatially averaged relative error in modeled wave-height for such a small length of survey path was found to be relatively very high - of the order of 23 percent. Moreover, due to the large size of the study area, the computation time required was of the order of 2 minutes per iteration.

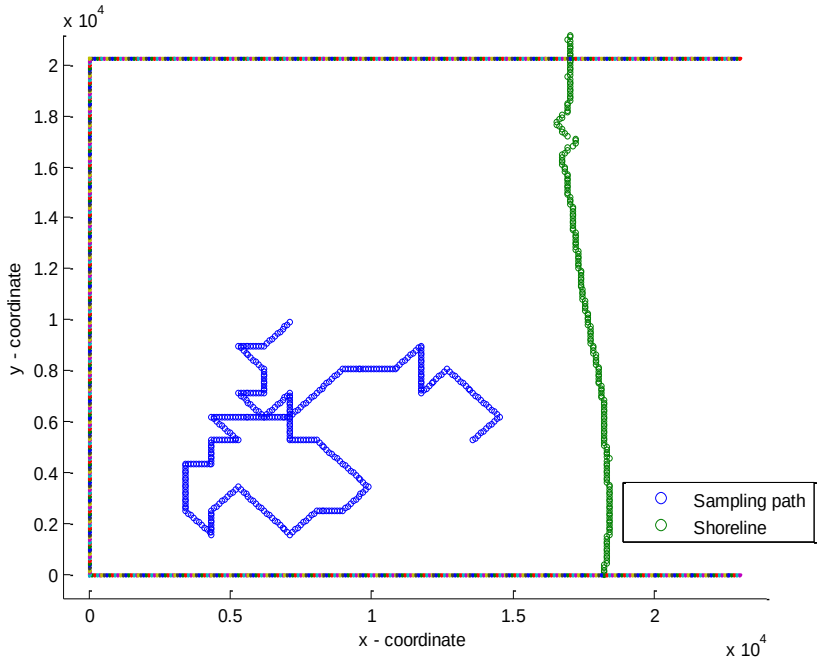


Fig. 5.14: An optimized AUV path for specified maximum survey length of 55 km (Camp Lejeune).

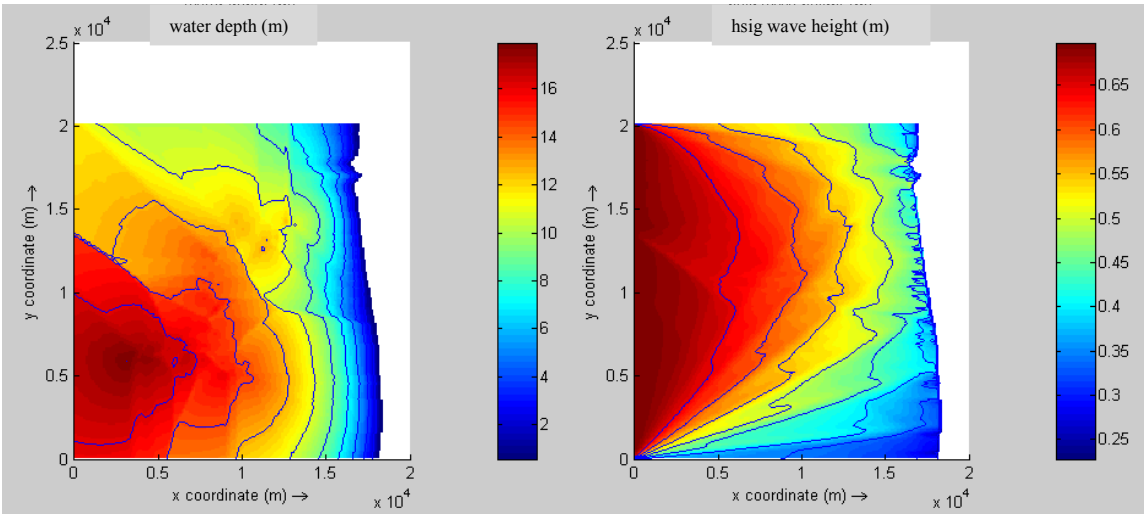


Fig. 5.15: Derived bathymetry and model wave-heights from optimized survey.

CHAPTER VI

CONCLUSIONS AND FURTHER WORK

Single beam bathymetric survey sonars provide less than 100 percent coverage of the area of interest, depending on survey density. They are still popularly used by virtue of being relatively inexpensive compared to multi-beam survey sonars which are capable of providing much better coverage of 200 percent or more. The genetic algorithm approach developed here attempts to make more efficient use of the former, for the task of providing better hydrodynamic model input.

First, the sensitivity of the Delft-3D model wave-heights and currents to the non-dimensional form parameters of several ideal bathymetric features was studied. The errors resulting from insufficient bottom resolutions were estimated and plotted. All these serve as a measure of the level of bathymetric detail required by the model to efficiently model the waves and currents at the required spatial scale.

Then, a Genetic Algorithm (GA) was developed and tuned, by applying it to standard test functions – the Rastrigin function and the G2 function. The results of these tests – the rate and level of convergence - were used to determine GA parameters such as the population size and mutation rate. Two test study areas were considered, and the GA was then applied to answer the questions of how much bathymetric information is needed, and where, for the Delft-3D Wave model to produce acceptable wave-height results. Only one set of offshore wave conditions were considered though, which might be acceptable if they are known to be relatively invariant during the period of interest. Otherwise, a more general approach would be to execute the GA for different sets of possible wave conditions, and consider the weighted results of the sampling strategies to design a new sampling path.

The large number of cost function evaluations involved, meant a large number of Delft-3D model runs, and running the combined Flow and Wave modules was found to be to take an infeasible amount of time with the given resources. Currents and waves could be expected to have different sensitivities to bottom variations, and considering both for the purpose of optimization of the sampling path would have been a challenge as such.

The interpolation scheme used to arrive at the model input bathymetry from the sampled one, could have a large impact on the solution path, and needs to be carefully selected. Inverse distance weighted triangular interpolation was used for the purpose of this study. Using the techniques of Plant et al. (2009) could potentially help improve the interpolation process.

The objective of the optimization schemes used here was to minimize the spatial average of the error in wave-height, which was effectively used as a cost function. This, however, may not always be a good measure of performance for the model, and was used in this study only as a rough indicator. The objective function can then be easily redefined to be location specific, or to suit a different output parameter, or be more representative of the spatial variance of the error.

One potential issue in the practical applicability of the proposed methods is the possible redundancy in data collection that would occur by not integrating and acknowledging the acceptable data from previous surveys. To this end, a future study could employ the sediment and morphology modules of Delft-3D to predict the expected changes in bathymetry over the given area during the interim period between surveys, with the help of an older bathymetric dataset and long-term model forcing. This would make it possible to concentrate the search for an optimal solution path in areas that are classified as having high expected bathymetric change.

While uniform sampling of bathymetry along parallel survey tracks is today mostly the norm, well informed non-uniform sampling schemes offer the potential of the same level of coverage at smaller time and energy costs.

REFERENCES

Alvarez, A., Caiti, A., and Onken, R., 2004. Evolutionary path planning for Autonomous Underwater Vehicles in a variable ocean. IEEE J. Oceanic Engineering 29 (2), 418-429.

Baehr, J., Hirschi, J., Beismann, J.O., and Marotzke, J., 2004. Monitoring the meridional overturning circulation in the North Atlantic: A model-based array design study. J. Mar. Res. 62, 283-312.

Barth, N.H., 1992. Oceanographic experiment design II: Genetic Algorithms. J. Atmos. and Oceanic Tech. 9, 434-443.

Bellingham, J.G., and Wilcox, J.S., 1996. Optimizing AUV Oceanographic Surveys. Proc. IEEE AUV'96 Conf., Monterey, CA, pp. 391-398.

Booij, N., Ris, R.C., and Holthuijsen, L.H., 1999. A third-generation wave model for coastal regions 1. Model description and validation. J. Geophys. Res., 104 (C4), 7649–7666.

Chawla, A., Özkan-Haller, H.T., and Kirby, J.T., 1996. Experimental study of breaking waves over a shoal. Proc. 25th Intl. Conf. Coastal Engineering, Orlando, FL, pp. 2-15.

Davis, T., Ed. 1991. The Handbook of Genetic Algorithms. Van Nostrand Reinhold Co., New York, NY.

DeJong, K., 1975. An analysis of the behavior of a class of Genetic Adaptive Systems. PhD dissertation, U. Michigan, Ann Arbor, MI.

Dingemans, M.W., Radder, A.C. and de Vriend, H.J., 1987. Computations of the driving forces of wave-induced currents. Coastal Engineering 11, 539-563.

Fox, R., Garcia, A., and Nelson, M.L., 1999. A three dimensional path planning algorithm for autonomous vehicles. Proc. 11th Intl. Symp. Unmanned Untethered Submersible Tech., Durham, NH, pp. 546-556.

Goldberg, D.E., 1989. Genetic Algorithms in Search, Optimization and Machine Learning. Addison-Wesley Longman Publishing Co., Inc., Boston, MA.

Goldberg, D.E., and Deb, K., 1991. A comparative analysis of selection schemes used in Genetic Algorithms. In: Foundations of Genetic Algorithms. Morgan Kauffman, San Mateo, CA, pp. 69-93

Griewank, A.O., 1981. Generalized descent for global optimization. J. Optimization Theory and Applications 34, 11-39.

Haas, K.A., 2000. Three dimensional modeling of rip current systems. PhD dissertation, U. Delaware, Newark, DE.

Hasselmann, K., Barnett, T.P., Bouws, E., Carlson, H., Cartwright, D.E., Enke, K., Ewing, J.A., Gienapp, H., Hasselmann, D.E., Kruseman, P., Meerburg, A., Müller, P., Olbers, D.J., Richter, K., Sell, W., and Walden, H., 1973. Measurements of wind-wave growth and swell decay during the Joint North Sea Wave Project (JONSWAP). Dtsch. Hydrogr. Z. A8 (12), 95 pp.

Holland, J.H., 1975. Adaptation in natural and artificial systems. PhD dissertation, U. Michigan Press, Ann Arbor, MI.

Holthuijsen, L.H., Herman, A., and Booij, N., 2003. Phase-decoupled refraction-diffraction for spectral wave models. Coastal Engineering 49, 291-305.

Kaihatu, J.M., and O'Reilly, W.C., 2002. Model predictions and sensitivity analysis of nearshore processes near complex bathymetry. Proc. 7th Intl. Workshop on Wave Hindcasting and Forecasting, Banff, Alberta, Canada, pp. 353-360.

Lesser, G.R., Roelvink, J.A., van Kester, J.A.T.M., and Stelling G.S., 2004. Development and validation of a three dimensional morphological model. Coastal Engineering 51, 883-915.

Liu, J., and Tsui, K.C., 2006. Toward nature-inspired computing. Communications of the ACM 49 (10), 59-64.

Plant, N.G., Holland, K.T., and Puleo, J.A., 2002. Analysis of the scale of errors in nearshore bathymetric data. Marine Geology 191, 71-86.

Plant, N.G., Edwards, K.L., Kaihatu, J.M., Veeramony, J., Hsu, L., and Holland, K.T., 2009. The

effect of bathymetric filtering on nearshore process model results. Coastal Engineering 56, 484-493.

Reniers, A.J.H.M., and Battjes, J.A., 1997. A laboratory study of longshore currents over barred and non-barred beaches. Coastal Engineering 30, 1–22.

Ris, R.C., Holthuijsen, L.H., and Booij, N., 1999. A third generation wave model for coastal regions. 2. Verification. J.Geophys.Res.104, 7667-7681.

Rogers, W.E., Hwang, P.A., and Wang, D.W., 2003. Investigation of wave growth and decay in the SWAN model: Three regional-scale applications. J. Phys. Oceanography 33, 366-389.

Rogers, W.E., Kaihatu, J.M., Hsu, L., Jensen, R.E., and Holland, K.T., 2007. Forecasting and hindcasting waves with the SWAN model in the Southern California Bight. Coastal Engineering 54, 1-15.

Shepard, F.P., and Inman, D.L., 1950. Nearshore circulation. Proc. 1st Conf. on Coastal Engineering, Council on Wave Res., Univ. of California, Long Beach, CA, pp. 50-59.

APPENDIX

1. Elliptic Shoal Bathymetry

An equation for bathymetry over an elliptic shoal was developed by modifying the equation for the bathymetry over a circular shoal used by Chawla et al. (1996):

$$depth(x,y) = h - [h-d1(x,y)] * [S_{max} / (h-d1(x_{cen},y_{cen}))]$$

(Scaling the depth so that peak ht. = S_{max})

where,

$$d1(x,y) = h + b_o - ((r_x^2 + r_y^2)/2)^{1/2} *[(c_o/r_x)^2 + (c_o /r_y)^2 – shoal(x,y)^2]^{1/2}$$

$$shoal(x,y) = [\{(x-x_{cen})/r_x\}^2 + \{(y-y_{cen})/r_y\}^2] ^{1/2}$$

$$c_o = 9.1 / \sqrt{2}$$

$$b_o = [(r_x^2 + r_y^2)/2]^{1/2} *[(c_o /r_x)^2 + (c_o /r_y)^2 – 1] ^{1/2}$$

S_{max} = *maximum shoal height*

r_x , r_y = *major, minor axis of ellipse*

x_{cen}, y_{cen} = *coordinates of centre of the ellipse*

h = *surrounding water depth*

VITA

Name:	Dinesh Manian
Address:	2 F, MSD Flats, Minto Road, New Delhi, Delhi – 110002, INDIA
Email Address:	dineshm@tamu.edu
Education:	B.Tech., Naval Architecture and Ocean Engineering, Indian Institute of Technology – Madras, July 2007
	M.S., Ocean Engineering, Texas A&M University – College Station, December 2009