The design and implementation of dynamic fish model in South Florida on parallel architecture\textsuperscript{1}

Dali Wang *

The Institute for Environmental Modeling, Department of Ecology and Evolutionary Biology, 569 Dabney Hall, University of Tennessee, Knoxville TN 37996
Email: wang@tiem.utk.edu Tel: +1 8659742773

Louis J. Gross

The Institute for Environmental Modeling, Department of Ecology and Evolutionary Biology, 569 Dabney Hall, University of Tennessee, Knoxville TN 37996
Email: gross@tiem.utk.edu Tel: +1 8659743065

Eric A. Carr

The Institute for Environmental Modeling, Department of Ecology and Evolutionary Biology, 569 Dabney Hall, University of Tennessee, Knoxville TN 37996
Email: carr@tiem.utk.edu Tel: +1 8659742003

Michael W. Berry

Department of Computer Science, 203 Claxton Complex, University of Tennessee, Knoxville TN 37996, Email berry@cs.utk.edu Tel: +1 8659743838

\textsuperscript{1} This research has been supported by the National Science Foundation under Grant No. DEB-02-19269.

* Corresponding author.
Abstract

A parallel, spatially explicit landscape fish population model (ALFISH) was presented here, which is a part of the Across Tropic Level System Simulation (ATLSS) project for freshwater wetlands of the Everglades and Big Cypress Swamp. ALFISH models the impacts of different water management strategies in the South Florida region on the fresh water fish population, which in turn provides the information on the food resource available to wading birds. Structurally, ALFISH model contains four basic components: landscape component, hydrology component, lower trophic components, and fish component. Adapting a static computational domain partitioning, each of the computational processors in the ALFISH model only simulates the fish behaviors in one row-size block-striped partition of the landscape. Also using message-passing, the parallel ALFISH mimics the basic behaviors of fresh water fish based on the interaction of those four components – landscape, hydrology, lower trophic biomass, and fish, over a time span up to several decades. Parallel ALFISH model delivers faithful results in simulation: Similar outputs of fish population were obtained from the serial and parallel ALFISH models. Compared with the averaged simulation time of sequential model, which is about 35 hours, the performance evaluation of the parallel model reveals substantial speed improvement: On a symmetric multiprocessor (SMP), the execution of parallel ALFISH model using 13 processors is less than 4 hours, the speedup factor in this case is approximately 9.

Key words: computational ecology, parallel implementation, message passing model, spatial-explicit simulation, symmetric multiprocessor
1. Introduction

The landscape of South Florida is a complex environment that has been subjected to years of environmental stress. Disruptions in the natural water flows have been the catalyst for profound changes in the vegetation and animal life in the region. Attempts are now being made to repair the devastating effects of these changes in the water flow on the ecosystem of the South Florida region [1]. The effects of these corrections must be modeled to ensure that these new changes do not further harm the fragile region. The most effective way to evaluate the effects of this complex environment is through computer modeling [2, 3]. Across Trophic Level System Simulation (ATLSS), a family of models, was developed to address this regional environmental problems which span a wide variety of spatial, temporal and organismal scales. The ATLSS Landscape Fish Model (ALFISH) [4,5] is one component of ATLSS. One objective of ALFISH model is to compare, in a spatially explicit manner, the relative effects of alternative hydrologic scenarios on freshwater fish densities across South Florida. Another objective is to provide a measure of dynamic, spatially-explicit food resources available to wading birds.

1.1 Notation

In the following sections, \( m \) denotes meter, and notation \( C_{i,j} \) is used to refer to a grid cell contains information about an area that is 500m by 500 m, and subscripts \( i \) and \( j \), of course, indicate the location of the gird cell in landscape map. The notation \( C_i \) is used to refer to a row of grid cell in landscape map. Since a two layer communication model was adapted in parallel model, \( P_i \) is used to represent the processor ID number, and \( P_0 \) always refers to the master processor which only used to collect data from other processors and to control and output operations.

1.2 Computational Environment

The computational environment used in this research is a Sun Enterprise 4500, configured with 14 400MHz Sun Ultra Sparc II processors, 10 GB memory and 3GB/s interconnections. This SMP is one of the main build blocks of the Scalable Intracompus Research Grid (or SinRG) [6] at the University of Tennessee, Knoxville. An implementation of the MPI standard library, LAM-MPI [7], was selected to support the message-passing communication in parallel ALFISH model.

2. ALFISH Model Components

The study area for ALFISH modeling contains 26 regions as determined by the South Florida Water Management Model [8,9]. A complete list of these regions is provided in Figure 1. As described in this

Figure 1: Subregions used by the ALFISH model.
section, ALFISH model contains four basic components: landscape component, hydrology and lower trophic components, and the fish component.

2.1 Landscape Component

The total area of Everglades modeled in ALFISH contains approximately 111,000 landscape cells, with each cell 500m a side. Each landscape cell in the ALFISH model represents has two basic landscape types: marsh and pond. The difference between marsh and pond areas is that the latter is always considered wet (contains water) regardless of any available water data. With the marsh area of each cell, there is a distribution of elevations based upon a hypsograph [4]. This hypsograph is used to determine the fraction of the marsh area that is still under water at a given water depth. The pond refers to permanently wet areas of small size, such as ponds or alligator holes, which are a maximum 50 m² or 0.02% of the cell.

2.2 Hydrology Component

The hydrology component models one of most important external forces on fish ecology. The landscape hydrology data for ALFISH model is produced by the South Florida Water Management Model (SFWMM), which is developed by the South Florida Water Management District (SFWMD). This is a large-scale mathematical model of the network of canals, structures, levees, and pumps that make up the highly managed system controlling water levels and flows through out the areas [8]. These hydrology data are created at two-mile by two-mile grid cell size. An auxiliary model is used to create 500m by 500m topography, which is used to provide an estimate for hydrology data at a 500m-resolution grid [2, 9].

2.3 Lower Trophic Components

The appropriate lower trophic level values for each cell are determined using a time series or a presumed constant value according to the experience of field scientists.

2.4 Fish Component

The fish population model simulated by ALFISH is size-structured and is divided into two functional groups: small and large fishes. Both of these groups are used in each of the marsh and pond areas. Each functional group in each area is further divided into several fish categories according to age. The fish population that occupies cell area is represented as the fish density (bio-mass) within that cell. Basic behaviors of fish are simulated in the model, including movement within cells, movement between cells, mortality, growth and reproduction.

2.4.1 Movement Within Cells

This function is designed to simulate the movement of fish between the marsh and the pond areas within each cell. First, the model calculates the largest size of fish, \( \text{minAge} \), that can survive in the new water depth and compares this value to the old value from the previous timestep. At current timestep, if water depth of the cell has increased, and the cell has a pond, an appropriate fraction of fish of the sizes previously too large is move from the pond areas into the marsh area. If the cell water has decreased, some of the large fish is will not survive in that cell. In this case, an appropriate fraction of those fish are moved from marsh area into the pond area, another fraction of those fish are moved to adjacent cells, and remaining portion of those fish are eliminated. The left plot of Figure 2 demonstrates the intra-cell
movement of fish using 3 age classes instead of several dozen classes in ALFISH model. The color of each cell represents water depth. White color shows no water in mesh areas of the cell. Light/medium/dark gray color means that the water depth in the cell is suitable for small/medium/large fish to survive. Green spots represent pond area in cells. For the purpose of demonstration, let us assume that there are three kinds of fish (small, medium and large fish) that exist in the central cell \(C_{i,j}\). At present timestep, water depth decreases to a medium gray level, so that a fraction of large fish in this cell should move to pond area in the cell. Another fraction of large fish moves to adjacent cell, \(C_{i-1,j}\) (colored with heavy gray) where water depth is suitable for large fish to survive. Thin black arrows in left plot of Figure 2 represent these movements. All remaining large fish in this central cell die at this timestep. Since water depth in the central cell is still suitable for those fish of small and medium size, those fish remain in the cell. It is important to note that this kind of movement is independent of fish-density. Similarly, we assume that there are small and medium fish exist in top-left cell, \(C_{i-1,j-1}\), water depth drops at present timestep, only small fish can survive. Therefore, the medium fish must move to either cell \(C_{i-1,j}\) or cell \(C_{i,j}\). In the left plot of Figure 2, bold black arrows represent these movements.

### 2.4.2 Movement Between Cells

This function is designed to simulate the movement of fish between adjacent cells mainly due to the relative differences in water depth and fish densities. Movement into or out of a cell is assumed to only occur when the cell is more than 50% flooded. The mathematical formula used to determine the number of fish can be moved is not presented in this paper (see Gaff [4,10]. The right plot of Figure 2 demonstrates fish movement between cells. After fish movement within cells, only small and medium size fish exist in the central cell, \(C_{i,j}\). Those fish have different movements according to the difference of water depth and fish densities. Small fish can move to five adjacent cells with light/medium/dark gray color. Green arrows in right plot of Figure 2 represent those movements. Due to the limit of water depth, medium fish in the central cell can only move to two adjacent cells with medium/dark gray color. For the reason of simplicity, those movements are not shown in the right plot of Figure 2. Similarly, small fish in cell \(C_{i-1,j-1}\) can move to the adjacent cells according to the relative difference of density and water depth. Black arrows in right plot of Figure 2 represent those movements. The density dependent decision process relies upon pre-movement densities and updates the fish landscape matrix after all movement calculations are complete, to remove any order based bias. In this phase, only a small fraction (3%) of the total density in cells is allowed to move due to density levels.

### 2.4.3 Fish Mortality, Aging, Reproduction, and Growth

In ALFISH model, the fish mortalities depend on age and prey availability. The food-based mortality is calculated as the ratio of prey biomass in the cell to the amount needed times a parameter.
representing the amount of the prey in the cell, which is available to the fish. The age mortality is directly related to the fish size and age. The food-based mortality is compared to the age-dependent mortality for each age class, and the greater of the two is applied. The age classes for the fish functional groups are 30-days. If the timestep is at the end of that 30-day age class, all of the fish are moved to the next age class. For each functional group, if it is the appropriate time of year, the number of offspring is calculated using 0.5 times the number of fish of reproductive age for that functional group times the number of offspring per female per reproductive event. To prevent the population from producing too many new fish in a reproductive event, a constant maximum reproduction density is used. Each fish can be divided into six class, all fish in an age class are within a single size class at any timestep. Every 5 day, fish in each age class will advance to next size class.

3. Parallelization Methodology

To make the ALFISH model for multiprocessor execution, several modifications to the sequential model were required. A two-layer communication model was deployed to provide message-passing functions between all processors. Considering the sequential output operations are needed in each timestep, one processor ($P_0$) was dedicated to collect fish data from other processors, referred to as computational processors, and store those data into disk. Each computational processor only simulates fish behaviors in a row-size block-striped partition of landscape.

3.1 Landscape Partition

In the parallel ALFISH model, a mask is used to remove unstudied area from the landscape map, and this mask is duplicated in all processors in order to reduce data immigration time. The entire landscape map is statically partitioned among all computational processors. In order to enable inter-processor data communication between adjacent processors, a ghost row is attached to the upper and down side of each partitioned landscape. Figure 3 shows a partition of the landscape map on three computational processors. Part of the landscape map, an internal region (row $C_0$ to row $C_{i-1}$) plus a ghostzone (row $C_i$), is assigned to processor 1, another part of the landscape map, an internal region (row $C_i$ to row $C_{j+1}$) plus two ghostzones (row $C_{i-1}$ and row $C_j$), is assigned to processor 2, the remaining part of the landscape map, an internal region (row $C_j$ to row $C_n$) plus a ghostzone (row $C_{j-1}$), is assigned to processor 3. Ghostzones in Figure 3 are represented in gray, which are used to mimic fish movements in sequential code.

Figure 3. Landscape partition on 3 computational processors

Figure 4. Calculations of fish movement within cells in parallel model. Green dots refer to ponds. The colors of cells represent different water depths. The movements of large/medium size fish are represent by those thin black/ bold black arrows
3.2 Parallel Fish Movement Within Cells

Fish movement within cells is not density-dependent, so that the order of computation is not an influent factor on the simulation. In parallel ALFISH model, the internal region of landscape partition on each processor is treated as those in sequential model. In order to enable the same computations implied in serial model, partial results be calculated simultaneously, and data exchanges are required between adjacent processors. Once again, we assume that configuration is the same as that described in Section 2.4.1. Figure 4 demonstrates the calculation of fish movement within cells in parallel ALFISH model. On processor \( i \), row \( C_{i,j} \) is an internal row, and row \( C_i \) is in ghostzone. At the present timestep, water depth in cell \( C_{i,j} \) drops, so medium size fish must move to adjacent cell \( C_{i,j-1} \) and cell \( C_{i,j} \). The black arrows in the left plot of Figure 4 represent those movements. Since row \( C_i \) is in a ghostzone that processor \( i \) will not simulate the fish movement from \( C_{i,j} \). On the other hand, on processor \( i+1 \), row \( C_{i+1,j} \) is in a ghostzone, and row \( C_i \) is an internal row. Therefore, processor \( i+1 \) will calculate those fish movements from \( C_{i,j} \). The greens arrows in the right plot of Figure 4 represent those movements. At each timestep, the changes of fish densities in ghostzones are stored. After the completion of fish movements on each processor, the values of fish densities in a ghostzone and its adjacent internal row are passed to the adjacent processor to mimic the computations in sequential model.

3.3 Parallel Fish Movement Between Cells

Similar to calculations of fish movement within cells, each processor only simulates the fish movement from internal cells. This kind of fish movement depends on fish densities, so that the changes of fish density are held in a temporal storage until all calculations are complete, and then added to the original landscape matrix to create a new fish landscape matrix for the next calculation. We reuse the same configuration as described in Section 2.4.2. Figure 5 demonstrates the calculation of fish movement between cells (which are adjacent to a ghostzone), in the parallel ALFISH model. On processor \( i \), row \( C_{i,j} \) is an internal row, and row \( C_i \) is in a ghostzone. The black arrows in left plot of Figure 5 represent those fish movements from cell \( C_{i,j-1} \). Since row \( C_i \) is in a ghostzone that processor \( i \) will not simulate the fish movement from \( C_{i,j} \). On the other hand, on processor \( i+1 \), row \( C_{i+1,j} \) is in a ghostzone, and row \( C_i \) is an internal row. Therefore, processor \( i+1 \) will calculate those fish movements from \( C_{i,j} \). The greens arrows in the right plot of Figure 5 represent those movements. This kind of fish movement depends on fish densities, so that the changes of fish density in internal cells are held until all calculations are complete, and then added to the original matrix to create new fish landscape matrix for the next calculation. At each timestep, the changes of fish densities are stored in a ghostzone. After the calculation of fish movements on each processor, the values of fish densities in a ghostzone and its adjacent row are passed to the adjacent processor to mimic the computations in sequential model.

![Figure 4. Calculations of fish movement within cells in the parallel model.](image1)

![Figure 5. Calculations of fish movement between cells in the parallel model.](image2)
3.4 Parallel Fish Mortality, Aging, Reproduction, and Growth

The calculations of fish mortality, aging, reproduction and growth depend on local information of cells where fish exists. This local information includes the values of lower trophic data, age status of fish, etc. In parallel ALFISH model, those computations are basically the same as those in the sequential model, except that those computations are only executed on cells in the internal region of landscape map (on each processor).

4. Verification and Performance

4.1 Scenarios

The ALFISH models are mainly used to determine the pattern of fish density on the landscape for a variety of hydrology scenarios. The motivation for the particular scenarios chosen was the Restudy process for the selection of a plan for Everglades restoration [11]. In this paper, one scenario, called as F2050, was applied. F2050 is a standard base scenario, which uses water data based on a 31-year time serial of historical rainfall from 1965 through 1995, as well as sea level, population level and socioeconomic conditions projected for the year 2050. It also includes all of the previously legislated structural changes for the water control measures. Therefore, the simulation time of both sequential and parallel ALFISH models are 31 years, from 1965 to 1995 and the timestep is 5 days.

4.2 Comparison of Selected Outputs

To verify the parallel model’s correctness and accuracy, that is, its ability to produce results similar to those of the sequential model, we compared outputs of both the sequential model and parallel model. We produced and analysis several outputs, but choose one set of data for comparison, it is a 31-year mean fish density and distribution on April 1. Figure 6 shows mean fish density map comparison on April 1. The left graph represents the output from the parallel ALFISH model, the right graph is the output from sequential ALFISH model. There are no observable differences between in the outputs of parallel and sequential models.

Figure 6: Spatial 31-year average fish density map comparison in Everglades on April 1
4.3 Parallel Performance Results

In order to measure the scalability of the parallel ALFISH model, we executed both the sequential and parallel program on a Sun Enterprise 4500. We first ran the sequential model and recorded the execution time, followed by a series of parallel simulations using different numbers of processors (ranging from 3 to 13). Figure 7 shows the speedup factor for the parallel ALFISH model over the sequential ALFISH model. We note, that even using static partitioning of the landscape, the parallel ALFISH model demonstrates exceptional scalability. The average execution time of the sequential model is about 35 hours, while the execution of the parallel ALFISH model with 13 processes is about 4 hours (the speedup factor being about 9). Considering that one process is always dedicated to output operations, the scalability of this parallel model is particularly noteworthy.

Figure 7: The scalability of the parallel ALFISH model

6 Conclusions and Future Works

The nearly exact outputs and excellent speed improvement obtained from the parallel ALFISH model, as compared with the sequential model, provide strong evidence that grid-based partition can be highly effectively for age- and size-structure based explicit spatial landscape fish ecological models. Our simulation proves that that even using static landscape partitioning, the parallel ALFISH model demonstrates remarkable scalability. Comparing to the averaged execution time of sequential model, which is about 35 hours, less than 4 hours of parallel ALFISH model (using 13 processors) turnaround time defines substantial runtime improvement.

In the parallel ALFISH model, the landscape component is statically partitioned, so one might expect that the computational workloads will not be well balanced all computational processors, especially since explicit synchronizations are enforced at each timestep. These two factors yield tremendous computational overloads into the simulation. Since the computational intensity of each landscape cell varies from time to time, dynamic loading balance techniques can be adapted to further improve the parallel model performance.

Further work on the parallel ALFISH model will include introducing a computation index (CI) according to the water depth, fish density, and so on, and repartitioning the landscape map according to the CI, so that computational workload can be dynamically balanced among the processors during simulation. In the message-passing parallel ALFISH model, the explicit data synchronization and data communication take significant time, due to the complex age- and size-structure of fish component. This certainly indicates that the ALFISH model is more suitable for parallelization on shared-memory multiprocessor system using multi-threads. Further plans for the sequential ALFISH model include adapting finer grid resolution and taking account of the effects of features such as canals.
Acknowledgement

This research has been supported by the National Science Foundation under grant No. DEB-0219269. The serial implementation was developed with support from the U.S. Geological Survey, Biological Resource Division, through Cooperative Agreement No. 1445-CA09-95-0094

References

[7] Indiana University, LAM/MPI: Parallel Computing, Bloomington, IN, 2002