Design and Implementation of a Parallel Fish Model for South Florida

Dali Wang

The Institute for Environmental Modeling, 569 Dabney Hall, University of Tennessee, Knoxville TN 37996 Email: wang@tiem.utk.edu

Eric A. Carr

The Institute for Environmental Modeling, 569 Dabney Hall, University of Tennessee, Knoxville TN 37996 Email: carr@tiem.utk.edu

Michael W. Berry

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

Louis J. Gross

The Institute for Environmental Modeling, 569 Dabney Hall, University of Tennessee, Knoxville TN 37996 Email: gross@tiem.utk.edu

Abstract

A parallel, spatially explicit landscape fish population model (ALFISH) is presented to model the impacts of different water management strategies in the South Florida region on the fresh water fish population, which in turn provides information on the food resource available to wading birds. Adopting a static domain partitioning scheme and using message-passing, the parallel ALFISH model mimics the basic behaviors of fresh water fish based on the interaction of four components – landscape, hydrology, lower trophic biomass, and fish, over a time span up to several decades. The parallel ALFISH model delivers accurate results in simulations. Compared to the average simulation time of the sequential model, which is about 35 hours, the parallel model yields substantial speed improvement. On a symmetric multiprocessor (SMP), the execution time of the parallel ALFISH model on 13 processors is less than 4 hours – a speedup factor of near 9.

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 problem which spans 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 the ALFISH model is to compare, in a spatially explicit manner, the relative effects of alternative hydrologic scenarios on fresh-water 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 containing information about an area that is $500m$ by $500m$, and subscripts i and j, indicate the location of the grid cell in the landscape map. The notation C_i is used to refer to a row of grid cells in the landscape map. Since a two-layer communication model was adapted in the parallel model, P_i is used to represent the processor ID number, and P_0 always refers to the master

processor which is used to collect data from other processors and to control input 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 Symmetric Multiprocessor (SMP) serves as one of the many computational clusters defining the Scalable Intracampus 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 message-passing communication in the 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 section, the ALFISH model contains four basic components: landscape component, hydrology and lower trophic components, and the fish component.
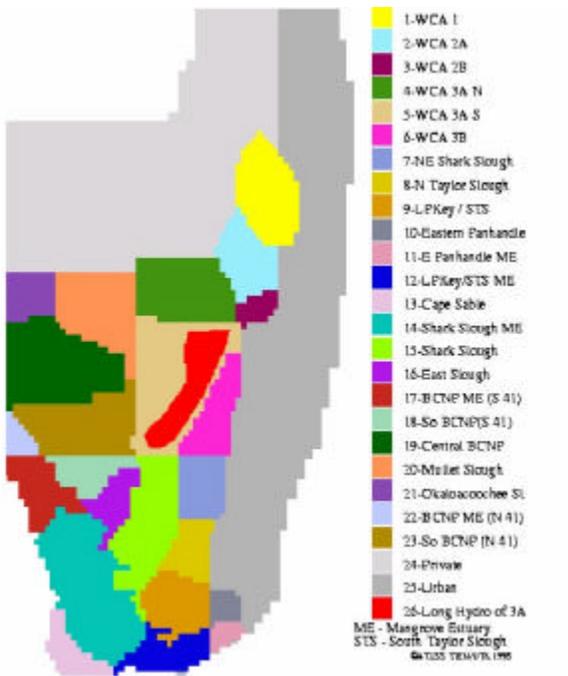

Figure 1. Subregions used by the ALFISH model.

2.1. Landscape component

The total area of the Everglades modeled in ALFISH contains approximately 111,000 landscape cells, with each square cell representing $500m$ on a side. Each landscape cell in the ALFISH model includes 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. For 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. Pond areas refer to permanently wet areas of small size, such as ponds or alligator holes, which are a maximum $50m^2$ or 0.02% of the cell area.

2.2. Hydrology component

The hydrology component models one of the 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 resolution. 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 fish* and *large fish*. 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 size. The fish population that occupies a cell area is represented as the fish density (biomass) within that cell. Basic behaviors of fish are simulated in the model, including *density-independent fish movement*,

diffusive fish movement, mortality, growth and reproduction. The general data flow of ALFISH model simulation is shown in Figure 2.
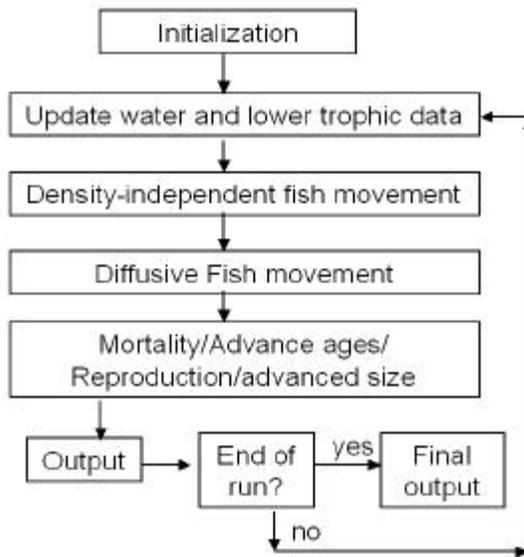

Figure 2. Flowchart diagram for the ALFISH model.

2.4.1. Density-independent fish movement. This function is designed to simulate the fish movement between marsh and pond areas within each cell caused by drying out or reflooding. First, the model calculates the largest size of fish, *minSize*, that can survive in the new water depth and compares this value to the old value from the previous timestep. At the current timestep, if the 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 moved from pond areas into marsh areas. If the cell water has decreased, some of large size fish will not survive in that cell. In this case, an appropriate fraction of those fish are moved from marsh area into pond area, another fraction of those fish are moved to

adjacent cells, and the remaining portion of those fish are eliminated. For the purpose of demonstration, the left graph of Figure 3 is presented to demonstrate this kind of fish movement using 3 size classes (small, medium and large fish) instead of the several dozen size classes in the ALFISH model. The shade of each cell represents water depth. White indicates no water in the cell. Light/medium/dark gray color means that the water depth in the cell is suitable for small/medium/large fish to survive, respectively. Circular regions represent pond area. Assume that all 3 sizes of fish exist in the central cell $C_{i,j}$ at previous tiemstep. At the present timestep, water depth decreases to a medium gray level, so that a fraction of the large size fish in this cell must move to pond area within the cell. Another fraction of the large size fish moves to an adjacent cell, $C_{i-1,j}$ (colored with dark gray) where water depth is suitable for the large size fish to survive. Thin black arrows in the left plot of Figure 3 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. Similarly, we assume that there are small and medium size fish in the top-left cell, $C_{i-1,j-1}$, and that the water depth drops at the present timestep so that only the small size fish can survive. Therefore, the medium size fish must move to either cell $C_{i-1,j}$ or cell $C_{i,j}$. In the left plot of Figure 3, bold black arrows represent these movements.

2.4.2. Diffusive fish movement. 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. Also, only a small fraction (3%) of the total fish density in cells is allowed to move. The mathematical formula used to determine the number of fish to be moved is not presented here (see Gaff [4,10]). The right plot of Figure 3
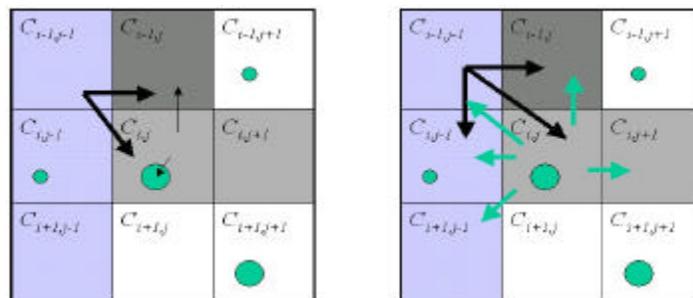

Figure 3. Calculation of fish movements.

(Left: density-independent movement; Right: diffusive movement)

demonstrates the diffusive fish movement between cells. After *density-independent movement* (see Section 2.4.1), only the small and medium size fish exist in the central cell, $C_{i,j}$. Those fish have different movements according to the differences in water depth and fish densities. The small size fish can move to five adjacent cells of light/medium/dark gray color. The green arrows in the right graph of Figure 3 represent those movements. Due to the limit of water depth, the medium size fish in the central cell can only move to the two adjacent cells of medium/dark gray color. For simplicity, those movements are not shown in the right plot of Figure 3. Similarly, the small size fish in cell $C_{i-1,j-1}$ can move to the adjacent cells according to the relative difference of fish density and water depth. Black arrows in the right graph of Figure 3 represent those movements. Since this kind of fish movement is density dependent, the fish landscape matrix is only updated after all movement calculations are complete, to remove any order-based bias.

2.4.3. Fish mortality, aging, reproduction, and growth. In the ALFISH model, 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 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. Fish that survive through the mortality phase will grow and age. The age classes for the fish functional groups are 30-days. Every 30 days (6 timesteps of 5-days each), 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.

3. Parallelization methodology

To redesign 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 that sequential output operations are needed at each timestep, one processor (P_0) was dedicated to collect fish data from other processors (referred to as computational processors) and to write those data to disk. Each computational processor only simulates fish behaviors in a row-size block-striped partition of the 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 on all processors in order to reduce data migration 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 lower segment of each partitioned landscape. Figure 4 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
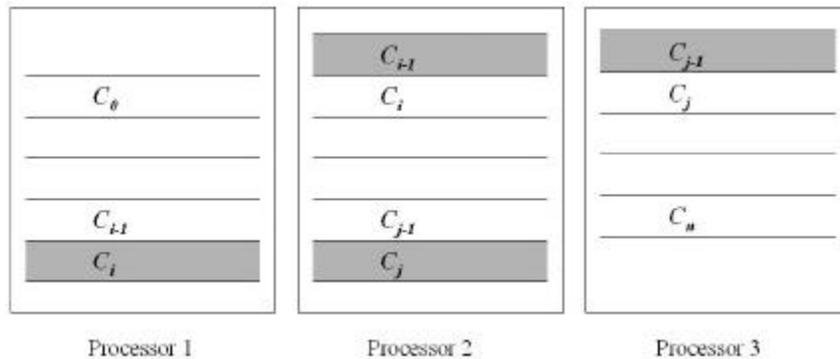

Figure 4. Landscape partition on three computational processors.

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 4 are represented in gray, which are used to mimic fish movements in the sequential code.

3.2. Parallel computational model

In the parallel ALFISH model, explicit message-passing is used to mimic the same fish dynamics described in the Section 2.4. The computational model is shown in Figure 5. Compared to the flow chart shown in Figure 2, the parallel ALFISH model contains explicit data exchange and computational synchronizations (represented by dash lines). At each timestep, 3 explicit synchronizations are required. The synchronization before the *density-independent fish movement* guarantees that all processors enter the computation phase at the same time. Since the movement of fish depends and has influence on the adjacent cells, an explicit data exchange is required after the computation of *density-independent fish movement* and after the computation of *diffusive fish movement*.

In the parallel ALFISH model, blocking LAM-MPI primitives (i.e., MPI_Send, MPI_Recv, and MPI_Sendrecv) are used. The main reason for
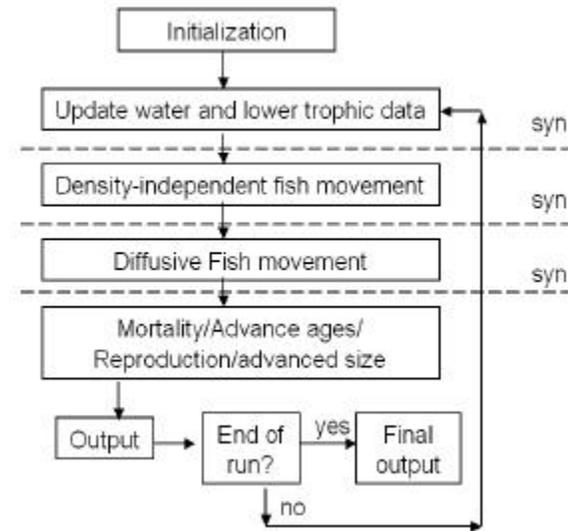

Figure 5. Flowchart diagram for the ALFISH model.

choosing blocking communication primitive is that such primitives eliminate the code segment for explicit synchronizations, which are required several times within each timestep. In addition, MPI_Barrier is used to guarantee synchronization during the simulation.
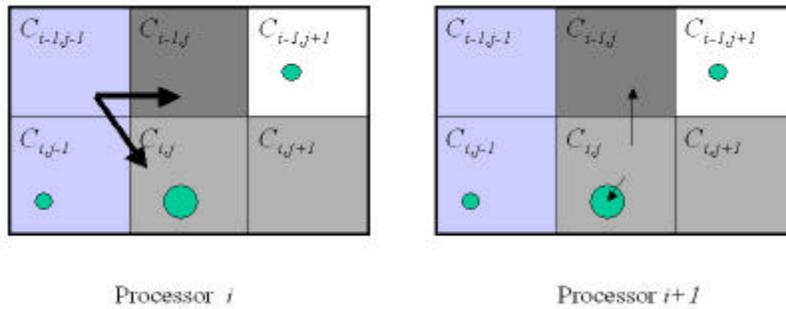
3.3. Parallel density-independent fish movement

Figure 6. Calculations of density-independent fish movement in the parallel model.
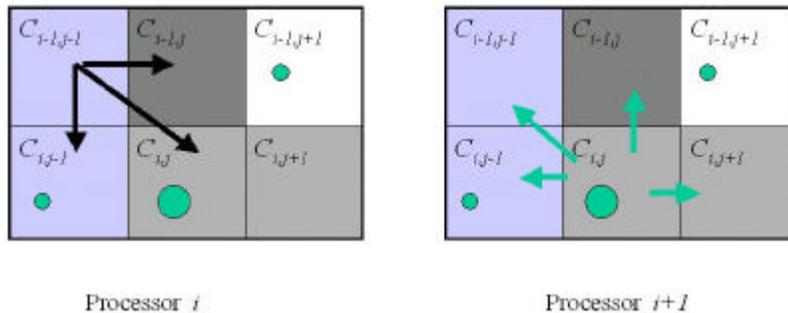

Figure 7: Calculations of diffusive fish movement in the parallel model.

In the parallel ALFISH model, the computation on each internal landscape cell is the same as that in the sequential model. In order to enable the same computations implied in the sequential model, partial results must be calculated simultaneously, and data exchanges are required between two adjacent processors. Once again, we assume that the configuration is the same as that described in Section 2.4.1. Figure 6 demonstrates the calculation of *density-independent fish movement* in the parallel ALFISH model. On processor i, row C_{i-1} is an internal row, and row C_i is in a *ghostzone*. At the present timestep, water depth in cell $C_{i-1,j-1}$ drops, so the medium size fish must move to two adjacent cells ($C_{i-1,j}$ and $C_{i,j}$). The black arrows in the left graph of Figure 6 represent those movements. Since row C_i is in a *ghostzone*, that processor i will not simulate the fish movements from $C_{i,j}$. On the other hand, on processor $i+1$, row C_{i-1} 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 black arrows in the right graph of Figure 6 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 the sequential model.

3.4. Parallel diffusive fish movement

We reuse the same configuration as described in Section 2.4.2. Figure 7 demonstrates the calculation of *diffusive fish movement* (which are adjacent to a *ghostzone*) in the parallel ALFISH model. On processor i, row C_{i-1} is an internal row, and row C_i is in a ghostzone. The black arrows in the left graph of Figure 7 represent those fish movements from cell $C_{i-1,j-1}$. 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} 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 green arrows in the right plot of Figure 7 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 a 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 the sequential model.

3.5. Parallel fish mortality, aging, reproduction, and growth

The calculations of fish mortality, aging, reproduction and growth depends on local information of cells where fish exist. This local information includes the values of lower trophic data, age status of fish, and so on. In the 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 the 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 referred to as F2050 was applied. F2050 is a standard base scenario, which uses water data based on a 31-year time series 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 the sequential and parallel ALFISH models is 31 years, from 1965 to 1995 using a timestep of 5 days.

4.2. Comparison of selected outputs

To verify the parallel model's *correctness*, that is, its ability to produce results similar to those of the sequential model, we compared outputs of both the sequential and parallel models. We analyzed several outputs and selected one set of data for comparison – the 31-year mean fish density and distribution on April 1. Figure 8 shows the mean fish density map comparison on April 1. The left graph represents the output from the parallel ALFISH model, and the right graph is the output from the sequential ALFISH model. There are no observable differences between the outputs of these two models.
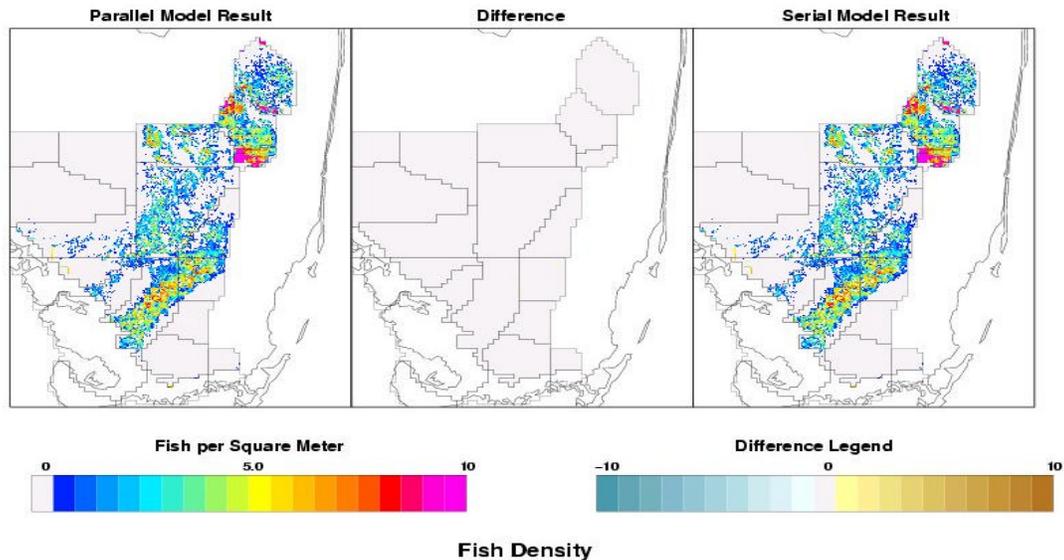
Figure 8. Spatial 31-year average fish density map comparison in Everglades on April 1.
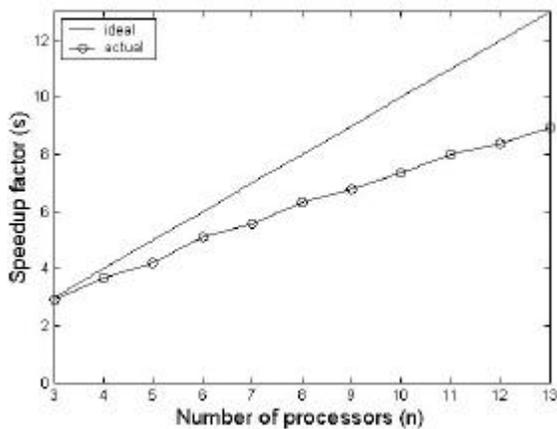

Figure 9. The scalability of the parallel ALFISH model.

4.3. Parallel performance results

In order to measure the scalability of the parallel ALFISH model, we executed both the sequential and parallel programs 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 9 shows the speedup factor for the parallel ALFISH model over the sequential ALFISH model. Although static partitioning is applied to 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 on 13 computational processors is about 4 hours (the speedup factor being about 9).

5. Conclusions and future work

The nearly identical outputs and excellent speed improvement obtained from the parallel ALFISH model, as compared with the sequential model, provide strong evidence that grid-based partitioning can be highly effectively applied for age- and size-structured explicit spatial landscape fish ecological models. Our simulation demonstrates that even though static landscape partitioning is used, the parallel ALFISH model demonstrates remarkable scalability. The average execution time was reduced from 35 to less than 4 hours using 13 processors.

In the parallel ALFISH model, the landscape component is statically partitioned, so one might expect that the computational workloads will not be well-balanced for all computational processors, especially since explicit synchronizations are enforced at each timestep. These two factors produce 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 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 the 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 the fish component. This certainly indicates that the ALFISH model is more suitable for parallelization on shared-memory multiprocessor systems 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.

6. Acknowledgement

7. References

[1] D. DeAngelis, S. Bellmund, W. Mooij, M. Noot, E. Comiskey, L. Gross, M. Huston, and W. Wolff, "Modeling Ecosystem and Population Dynamics on the South Florida Hydroscope", in The Everglades, Florida Bay, and Coral Reefs of the Florida Keys: An Ecosystem Sourcebook, edited by J. Porter and K. Porter, CRC Press, 2002.

[2] D. DeAngelis, L.Gross, M. Huston, W. Wolff, D. Fleming, E. Comkskey, and S. Sylvester, "Landscape modeling for Everglades ecosystem restoration", Ecosystems, 1998, No.1, pp.64-75.

[3] L. Gross and D. DeAngelis, "Multimodeling: New Approaches for Linking Ecological Models", in Predicting Species Occurrences: Issues of Accuracy and Scale, edited by J. M. Scott, P.J. Heglund, M.L. Morrison, Island Press, 2002.

[4] H.Gaff, D. DeAngelis, L. Gross, R. Salinas, and M. Shorrosh, "A Dynamic Landscape Model for Fish in the Everglades and its Application to Restoration", Ecological Modeling, Vol. 127, pp33-53, 2000.

[5] A. Immanuel, "A parallel Implementation of ALFISH: Compartmentalization Effects on Fish Dynamics in the Florida Everglades", MS Thesis, Department of Computer Science, The University of Tennessee, Knoxville, TN, 2002.

[6] Innovative Computing Laboratory at the University of Tennessee, SInRG: Scalable Intracampus Research Grid, Knoxville, TN, 2002.

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

[8] R.L.Fennema, C.J. Neidrauer, R.A. Johnson, T.K. MacVicar, and W.A. Perkins, "A Computer Model to Simulate Everglades Hydrology", In Everglades: The Ecosystem and Its Restoration, S.M. Davis and J.C. Ogden, eds., St. Lucie Press, 1994, pp249-289.

[9] Duke-Sylvester, S. and L. J. Gross. Integrating spatial data into an agent-based modeling system: ideas and lessons from the development of the Across Trophic Level System Simulation (ATLSS). Chapter 6 in: Integrating Geographic Information Systems and Agent-Based Modeling Techniques for Stimulating Social and Ecological Processes, H. R. Gimblett (ed.), Oxford University Press. 2002.

[10] Gaff, H., J. Chick, J. Trexler, D. L. DeAngelis, L. J. Gross, and R. Salinas. 2003. Evaluation of and insights from ALFISH: a spatially-explicit, landscape-level simulation of fish populations in the Everglades. Hydrobiologia (to appear)

[11] U.S. Army Corps of Engineers, C & SF Restudy Draft Feasibility Report. http://www.restudy.org/overview.pdf. (Accessed 17 November 1998.)