

## INTEGRATION OF SIMULATION AND GEOGRAPHIC INFORMATION SYSTEMS: MODELING TRAFFIC FLOW IN INLAND WATERWAYS

William E. Biles  
Daniel Sasso

Department of Industrial Engineering  
University of Louisville  
Louisville, KY 40292, U.S.A.

Jerry K. Billbrey

Department of Computer Science  
Francis Marion University  
Florence, SC 29501, U.S.A.

### ABSTRACT

This paper describes the integration of Geographic Information Systems (GIS) with simulation modeling of traffic flow on inland waterways. Two separate modeling efforts are described: (a) GIS/AutoMod modeling of barge traffic on the Ohio River, and (b) GIS/Arena modeling of the transit of ocean-going vessels through the Panama Canal. These modeling efforts demonstrate the benefits that accrue both to modeling realism and to the initialization process with discrete-event models of traffic flow on these waterways.

### 1 AN AUTOMOD MODEL OF BARGE TRAFFIC ON THE OHIO RIVER

An AutoMod 11.0 model was developed of barge traffic on the Ohio River. The focus of this modeling effort was to

investigate the delays incurred at the numerous control structures along the river. Figure 1 illustrates the locks and dams on the Ohio River as it borders Kentucky. The main resources used within the model are the control structures (locks and dams) on the inland waterway system. These control structures delay each of the tows as they journey from their origin to their destination. However, they are necessary for the safe and consistent travel of these tows. The locks and dams provide a means of maintaining depth within the pools of water as well as keeping any rapids from creating disastrous conditions for the tows.

#### 1.1 Modeling Methodology

The model is created using a path-mover system in AutoMod version 11.0 build 1942.7. There are two paths that follow the contour of the river. A non-accumulating sys-

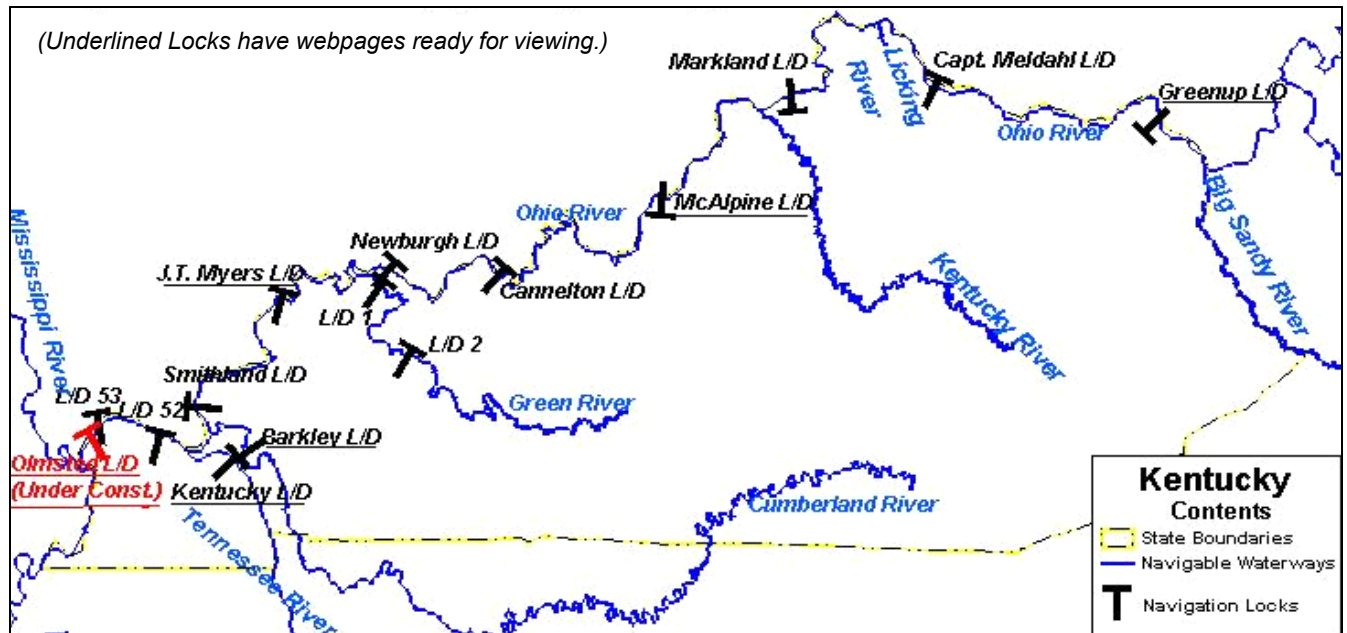


Figure 1: Control Structures along the Ohio River

tem is used to keep the tows from having collisions and for model efficiency (Moon 2003). Geographic data for the Ohio River is obtained from the Water Resource Information System Website (2003). The data comes in the form of AutoCAD format *.dwg* files. These files are imported into AutoMod and then used to draw the paths to scale based on the boundaries of the river. Also, geographically referenced initialization data is downloaded from the US Army Corps of Engineers website.

A “tow” is one of the basic entities flowing along the river system. It is modeled within AutoMod as a vehicle carrying a load. A “tow” consists of a tow boat and a set of barges (e.g., 15 barges). Entities flow either upstream or downstream, and can stop at any of several ports along the river to drop off and/or pick up barges. The simulation is initialized by placing entities at typical points along the river based on geographically referenced input data.

Due to the relatively small number of tows in a given area on the river at a given time, it is assumed that passing is not a problem within the model. Along this same line of logic, tows may start or continue a pass (Brewer 2004) in the bend of a river. Also, using a non-accumulating system it is assumed that as many tows as is necessary may concurrently be side by side. In actuality it is not likely that more than three tows would be in the same area at the same time. So, it is also for this reason that it is assumed that passing is not a significant problem within the model system. One end is the Greenup L & D another entity utilized within the AutoMod model is the control point. Control points are placed throughout the model to represent each of the logical locations where a tow would be required to stop. They represent geographically referenced locations that are used to direct the tows to their next location. As well as references for the initial locations of the tows and logical stopping points, control points are used to represent the locations of the ports between the locks. While stopped at these control points (ports) the system generates new data for the tows upcoming trip.

After model initialization, arrivals enter the system through the two ends of the system. One end is the Greenup L & D and the other end is the Lock 52 L & D. The rest of the system is self-contained accordingly. Once entering the system a tow only leaves the system by traveling to one of the two ends. At the beginning of a tow’s trip its parameters are determined. Parameters that are calculated are tonnage, number of barges empty, number of barges full, destination, speed, and the number of the port where the tow is scheduled to stop. At model initialization and again after every stop a tow updates its parameters and follows this newly prescribed route until it is finished. Then the process repeats. This process is illustrated in Figure 2.

There are three points between each of the locks at which a tow may stop. These are evenly distributed between the locks to indicate “close travel”, “medium travel”, and “distant travel” from the previous lock. These points are derived in an attempt to model the many different ports

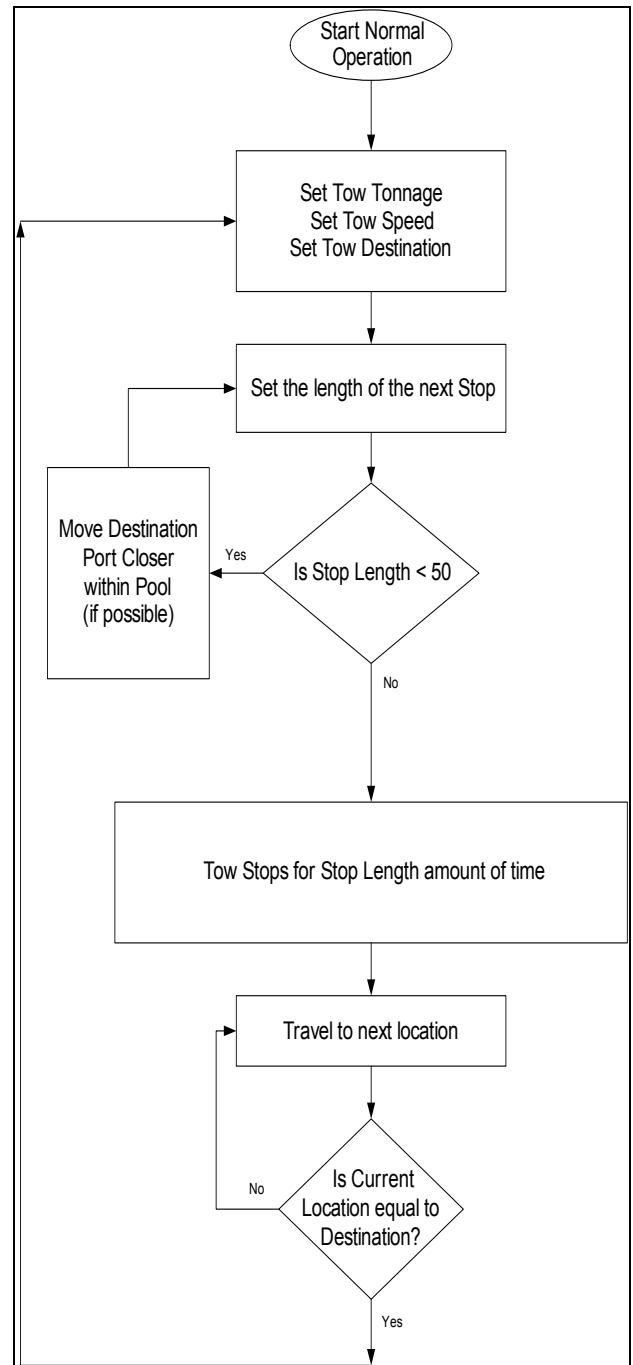


Figure 2: Looping Process of the Ohio River Model

on the river system and closely represent the actual ports on the Ohio River. Within the model it is assumed that one third of the tows stopping in the current pool will stop at the closest point, one third will stop at the middle point, and the last third will stop at the further point. This is an attempt to demonstrate the tows traveling to different ports to load and unload their cargo while utilizing each of the ports evenly over a long period of time.

## 1.2 Model Parameters

There are several parameters in the model, most of which are based on the database provided by the United States Army Corps of Engineers (Lisney 2002). This data is from the year 2001 and pertains to all of the lockages that happened on the Ohio River. It includes all arrival times beginning lockage times and ending lockage times. These times are used for various calculations to determine queue waiting distributions, resource (locking) distributions, and travel speed distributions. Other parameters included within the database are boolean stop flags, tonnage, number of loaded and unloaded barges within a tow, vessel number, lock name, and river mile of lock.

Within the model, tows have attributes including vessel number, tonnage, array of locks for the trip, speed, number of empty barges, number of loaded barges, stop flag, destination, and stop time length. There are many other standard simulation parameters that are calculated for the entire system such as lock utilization, queue delays, number in queue, etc. There are both upstream and downstream distributions for speeds derived from historical data. There are also upstream and downstream locking time distributions used at each lock.

There is a stopping time distribution calculated for the tows on the river. At the end of each trip a tow stops for a length of time calculated from this distribution before proceeding through the model. It is assumed that each tow will stop for a period of time no longer than 50 minutes. In reality, this is not what happens at all. Tows may stop for several days. However, as some tows stop for several days and other tows start after several days of being stopped the natural system reaches somewhat of a constant state. For

routine short stops, 50 minutes is considered to be a reasonable amount of time (Gardner 2003).

Behavior of the tows is also converted into a distribution which in turn is given to the behavior attribute of the tow. The upstream behavior is illustrated in Figure 3 as the number of pairs of locks a tow passes through before stopping. For instance, about 70 percent of the tows on the river pass through only one lock (hence zero pairs) before stopping at a port. This function is used to determine the future behavior of all of the tows in the model.

## 1.3 Model Inputs

Inputs are crucial to the success of the model. Also, some desired inputs and associated expected outputs are not presently available. These have to be derived using reasonable assumptions. Model initialization data and arrival data fall into this category.

One assumption that has to be made concerns the destination. It has to be estimated because the data is not available. There is a large range of potential errors in the system. For instance, there is no data for precise starting or stopping time. In fact, the stopping time has to be generally estimated using basic assumptions about the flow of traffic. Tributaries allowing tows to enter and exit the system are yet another major source of potential error. There are four significant rivers that enter the Ohio River as it borders Kentucky. These are the Tennessee, Kentucky, Green and Cumberland. However, data was not available to add to this model. With these consequent entries and exits into and from the system it could easily change some of the other parameters.

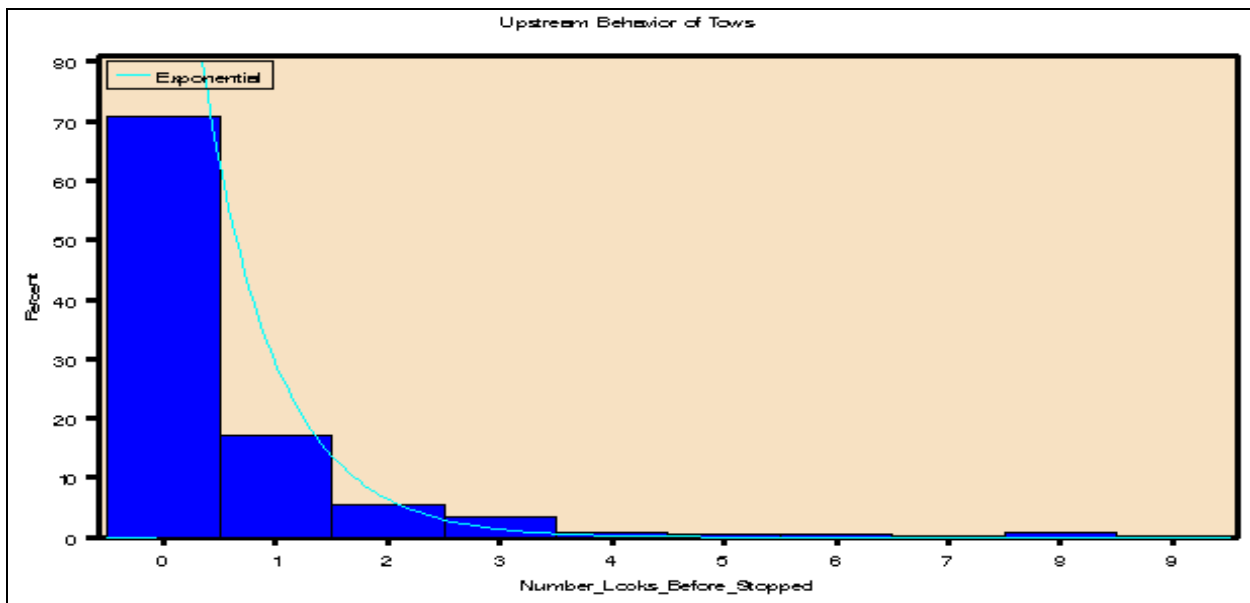


Figure 3: Distribution of the Number of Pairs of Locks the Tows Typically Travel through before Stopping

## 1.4 Results

The model warm-up time is determined using graphs of the queue utilizations. It is determined that the model takes between 90 and 120 days of runtime to reach steady-state from an empty and idle startup. Consequently, using initialization data there is no warm-up time necessary. This saves computer time as this warm-up period takes approximately 5 minutes on a Pentium IV 2.2 GHz machine.

It is found that the model gives good indicators for the short-term future. The measurement tested is the number of tows in each pool. Although this number is not as precise as is desired, it is the best measure with the available data. The model is initialized with the input data and ran for periods of 1-25 hours. Averages are taken of the output of the model and compared with the actual results for the time period. The results are shown in Figure 4. These results are based on the formula  $\text{abs}[(\text{observed} - \text{expected})/\text{expected}]$ .

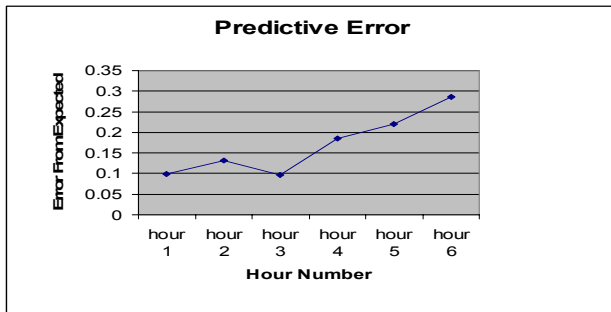


Figure 4: Results of Number of Tows in the Pool in Pool

The results are quite promising. The average error for the three runs for the first hour is 9.7%. The average error for the three runs for the second hour is 13.3%. Lastly, the average error for the three runs for the third hour is 9.7%. This suggests that the model could be used to predict the state of the model with an error of less than 15% for the first three hours. It is seen in the graph that starting at hour 4, the variability of the observed from expected number in each pool approaches 20% and continues to climb each hour thereafter. It is noted that some of the values in the obtained data files are obviously wrong. Several tows are found to have traveled more than 90 miles in about an hour. This is very inaccurate as the tows have a calculated maximum speed of 15 mph and a top actual speed of around 7 mph. This inaccuracy within the data gives reason to believe that the predictability would be much stronger if the input data set were more accurate. Also, the input data is far from detailed about the destination of the individual tows. Obviously, if the destinations were known for the initialization data the short-term accuracy would improve.

## 2 A GIS/ARENA SIMULATION MODEL OF THE PANAMA CANAL

The Panama Canal is a 50-mile long waterway that joins the Atlantic and Pacific Oceans through the Isthmus of Panama. It consists of three sets of locks, Miraflores, Pedro Miguel, and Gatun. Arriving vessels are classified in different categories based on their beam, draft, length, type, and priority. These categories determine their scheduling scheme and the time required to go through the Canal.

The objective of this model is to populate a sample database with different kinds of vessels with their coordinates and randomly re-sample them placing them at their corresponding positions in a geo-referenced map to start the simulation. This is accomplished by gathering the plotting data from Global Positioning System (GPS) readings and loading it into a Microsoft Access database. A Geographic Information System (GIS) reads these coordinates and plots them in a geo-referenced map of the Panama Canal. The database will also contain the vessel categories as their attributes which will be used in the Arena simulation model to determine their transit process through the Canal.

### 2.1 Description of the Geographic Information System (GIS)

One of the key elements of this research is the GIS that must be integrated into the simulation model. The importance of this implementation is based on the capacity of having a vessel management tool to start the simulation with the actual number of vessels in the system with their corresponding positions. The integration of this interface will allow the daily scheduling of vessels ready to transit and currently on transit taking into account waiting time in queue, booking status, navigation restrictions and their times through the locks.

The interface development was initiated with the configuration of the GIS software, ArcView 8.3, from Environmental Systems Research Institute (ESRI). The mapping scheme that the software uses consists of layers. A layer references the data contained in the data source which can be coverages, shapefiles, CAD files, grids, images and others. Once the data is referenced, it allows the layers on a map to automatically reflect the most up to date information in the GIS database (Murad-al-Shaikh et al. 2003).

Using the layer principle, the next step consisted on constructing the map for the Panama Canal. First, the political data layer had to be loaded, which is the country's shape and boundaries. On top of the political data layer came the lakes and rivers layers that constitute Panama's hydrography (Geo Community Website 2004). It is important to have every layer with the same geographic coordinate system (GSC) in order to have an accurate map. After adding all the layers we have our starting map, shown in Figure 5.



Figure 5: A Starting GIS Map of the Isthmus of Panama

With the map in place, the following procedure was to define the route for the vessels through the Canal and location of the locks in the map with their corresponding coordinates (latitude and longitude). Figure 6 shows the Panama Canal with the positions of routes and locks.



Figure 6: Panama Canal with Routes and Locks

## 2.2 Description of the Database

To start the experiment, it is required to populate a database with the vessel coordinates at different points along the route. For this purpose, historical data was gathered from the Panama Canal Authority (Panama Canal Authority Website 2004), using vessel transits for the fiscal years 2002 through 2003 to estimate the number of transits per day.

Vessels are classified in different categories that can be configured by the database manager. These categories are beam, draft, length, market segment/type, and priority.

Table 1 shows a summary of the number of transits by market segment for the years 2002 and 2003.

Table 1: Panama Canal Transits per Market Segment

Market Segment	No. of Transits		
	2003	2002	Avg %
Container Vessels	2,369	2,012	19
Dry Bulk Carriers	2,347	2,676	21
Refrigerated Cargo	2,207	2,135	18
Tankers	1,581	1,802	14
General Cargo	827	985	8
Vehicle Carriers	816	773	7
Passenger Ships	203	206	2
Others	1,375	1,271	11
<b>Total</b>	<b>11,725</b>	<b>11,860</b>	<b>100%</b>

This mix of vessels was entered into a Microsoft Access database with their coordinates and categories as the sample to choose from to initialize the simulation which will be explained in the next section. One of the advantages of using Microsoft Access is that it easily interfaces with ArcView, allowing it to read the coordinates for each vessel plotting them on the map.

## 2.3 Description of the Simulation Model

Figure 7 shows a simplified Arena model of vessel transit of the Panama Canal. Note that there is a model segment for the Atlantic-to-Pacific route and another for the Pacific-to-Atlantic route. A previous approach to the development of a simulation model for the Panama Canal was undertaken by Golkar, Shekhar, and Buddhavarapu (1998) who divided the

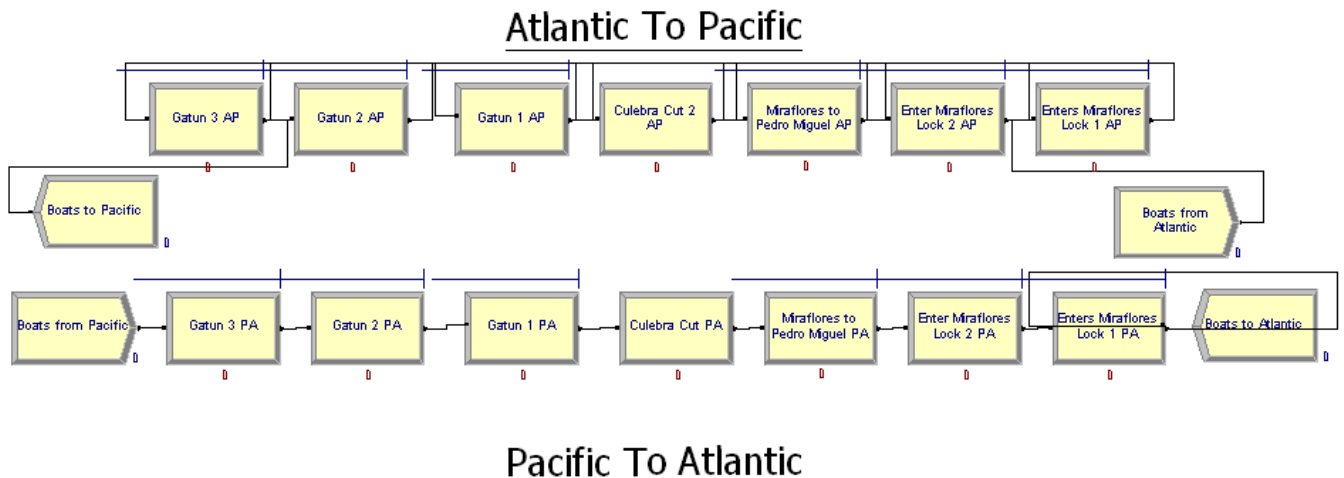


Figure 7: An Arena Model of Traffic Flow on the Panama Canal

development process in two phases. In this paper we make reference to operational elements of the simulation engine described in phase two with a few modifications.

The simulation model is better developed and understood by dividing its design in modules. The modules represent key components that will simulate operating scenarios with assigned rules, strategies and restrictions (Panama Canal Authority RFP 2003). The modules are the following:

- *Locks Module*: This module includes all the activities in the Gatun, Pedro Miguel and Miraflores Locks. The activities include: the utilization of the water chambers, locomotives and linehandlers.
- *Navigation Channels Module*: The channels are: Pacific Entrance, Miraflores Lake, Culebra Cut, Gatun Lake, and Atlantic Entrance. The design of this module consists on specifying the navigation speeds and times per vessel category for each channel.
- *Tie-up Stations and Moorings Module*: These are: Miraflores Moorings, Paraiso tie-up station, Gatun Anchorage.
- *Tugboat Module*: The module represents the availability of the tugboat fleet to assist the vessels.
- *Pacific and Atlantic Anchorages Module*: These are the locations where vessels arrive and need to be cleared by authorities before they start transit. This is the queue prior to entering the Canal. The generated vessels are scheduled based on their categories. The schedule may consist on a first come first serve discipline, preference, booking, or selected reordering. For this module a vessel arrival mix is generated based on historical data or probabilistic distribution and selected from the GIS database.
- *Restrictions and Weather Module*: In this module the user specifies all the nature conditions that may influence the transit through the Canal such as daylight hours, tides, lake levels, water availability, and fog.

## 2.4 Model Initialization

The initialization consists on having a Visual Basic macro within Microsoft Access randomly re-sample from the database a certain number of vessels to be plotted on the Canal map. The vessels plot represents a day's traffic through the Canal. The length of a day's run may be extended to more days by editing the database.

Before the simulation begins, all the data from the GIS is loaded into Arena so that the software interprets the information and proceeds according to the specifications of each vessel in transit and in queue. Each vessel will carry its attributes which include: coordinate, direction, and category. This means that if a vessel is currently located at a certain coordinate on the Gatun Lake going into the Atlantic, Arena would automatically determine the remaining

time before it enters the locks based on the vessel category and traffic. This would be the same case with every vessel chosen from the database for a simulation run.

## 3 SUMMARY AND CONCLUSIONS

It has been shown that using geographically referenced data adds much value to a simulation model of an inland waterway system. Using a geographically referenced model allows the model to be initialized with current Global Positioning System (GPS) data and gives an indication of the short-term future state of the actual system. There are other initialization and sampling issues that remain to be investigated, but in both models initialization with real-time data is feasible and greatly reduces simulation runtime.

## REFERENCES

- Brewer, R. 2004. Personal phone conversation, January 29, 2004. Boat Captain, Crouse Corporation, Paducah, Kentucky.
- Gardner, Capt. R. 2003. Personal phone conversation, December 30, 2003, Boat Captain, Barge America, Louisville, Kentucky.
- Geo Community Website. 2004. Available online via <<http://data.geocomm.com>>, (accessed February 15, 2004).
- Golkar, J., A. Shekhar and S. Buddhavarapu. 1998. Panama Canal Simulation Model. In *Proceedings of the 1998 Winter Simulation Conference*, eds. D. J. Medeiros, E. F. Watson, J. S. Carson and M. S. Manivannan, 1229—1237, The Society for Computer Simulation International, Washington, DC.
- Lisney, M. W. 2002. Personal Correspondence, May 2002, Civil Engineer, Louisville District, United States Army Corps of Engineers.
- Moon, W. 2003. Accumulating vs. Non-Accumulating Vehicles, *Autoflash*, Brooks Automation, Vol. 16, No. 5.
- Murad-al-shaikh, M., K. Page, M. Stewart, and M. Taggart. 2003. *Introduction to ArcGIS (for ArcView 8, ArcEditor 8, and ArcInfo 8)*, Course Lectures and Exercises, ESRI Educational Services, Redlands, California.
- Panama Canal Authority. 2003. Request for Proposal No. CDO-150552.
- Panama Canal Authority Website. 2004. Available online via <<http://www.pancanal.com>>, (accessed 04-30-04).
- Water Resource Information System Website. 2003. Available online via <<http://wris.state.ky.us/wris/mapgis/data.htm>>, [accessed March 9, 2003].

## AUTHOR BIOGRAPHIES

**WILLIAM E. BILES**, Ph.D., P.E. is the Edward R. Clark Professor of Computer Aided Engineering in the Depart-

ment of Industrial Engineering of the University of Louisville. He received the B.S.Ch.E. from Auburn University, the M.S.I.E. from the University of Alabama in Huntsville, and the Ph.D. in Industrial Engineering and Operations Research from Virginia Polytechnic Institute and State University. Dr. Biles has served on the faculties of the University of Notre Dame, The Pennsylvania State University, and Louisiana State University. He is a Fellow of the Institute of Industrial Engineers and a registered Professional Engineer in Indiana and Kentucky. His email address is <webile01@gwise.louisville.edu>.

**JERRY K. BILBREY, JR.** is an Assistant Professor of Computer Science at Francis Marion University. He received the B.S. degree in Industrial Engineering from Tennessee Technological University, the M.E. degree in Computer Engineering from the University of South Carolina, and the Ph.D. degree in Industrial Engineering at the University of Louisville. His research interests include simulation using geographic data, object-oriented simulation modeling and real-time simulation. His research is motivated by logistics problems and systems inefficiencies. His email address is <jbilbrey@fmarion.edu>.

**DANIEL SASSO** is a Ph.D. student in Industrial Engineering at the University of Louisville. He has a B.S. degree in Electronics Engineering from Santa Maria La Antigua University in Panama City, Panama and the M.Eng. in Engineering Management from University of Louisville. He has worked as a contractor for commercial airlines and the U.S. Air Force. His current interests are in simulation and transportation logistics. He is a member of IIE and INFORMS. His email address is <daniel@sasso.com>.