Informing water policy with large-scale, high fidelity simulation

Gerard P. Learmonth Sr., Ph.D.
Ryan Bobko
Department of Systems and Information Engineering
University of Virginia
jl5c@virginia.edu, rpb6eg@virginia.edu

ABSTRACT

As the world experiences rapid growth and urban development, the management of our water resource is of increasing importance. Policy-makers and regulators are faced with the prospect of making decisions today to assure adequate supply of clean water for tomorrow’s world. Water supply, however, is a local problem. And, issues of adequate water supply are the result of intertwined human and natural forces. Humans require, consume, and waste precious water. Nature of its own, and owing to increased pressures of climate change, adds an element of unpredictability. Here we describe two approaches to better understanding the “water problem.” One is an existing participatory simulation enabling human players to engage dynamically in decision-making about water. The other is a large-scale, high fidelity simulation-only model that will provide greater depth of understanding into this complex human and natural system. Preliminary results are presented.

Presented at Association for Public Policy Analysis and Management (APPAM) Annual Meeting, December 2011.

INTRODUCTION

Providing access to adequate quantities of clean water is an increasingly important issue faced by policymakers at all levels. With reduced budgets and competing demands for public services, hard policy choices aimed at solving even the short-term “water problem” are easily set aside. And, being widely perceived as a “free” commodity, the public generally does not appreciate the magnitude of the problem.

The dimensions of the problem of water policy, both quantity and quality, are temporal, spatial, and behavioral. Implementing new water initiatives involves long-term planning and infrastructure projects. The historical pattern of development, driven largely by population growth, often does not consider the availability of an adequate and sustainable water supply. And, increasing demand for water creates a classic tragedy of the commons situation. Add to this the looming problem of climate change and the problem of future water supply becomes more uncertain.

This paper describes how this problem is being examined in the Chesapeake Bay Watershed in the United States. The Chesapeake Bay Watershed is the largest estuary in U.S. covering large parts of six Mid-Atlantic States and the District of Columbia, an area of some 64,000 square miles with a population of 17.2M persons in 2010 (Figure 1). This watershed does not suffer from a lack of water, but rather from the deleterious effects of human activity on the Bay’s environment. Previous attempts have been made to organize collective action to
mitigate these impacts without success (*Chesapeake 2000, 2000*). In May 2009, the Obama Administration issued Executive Order 13508 mandating the restoration and sustainability of this precious national resource (*Executive Order 13508, 2011*).

To gain deeper insight into this coupled natural and human system, a large-scale simulation model of the Chesapeake Bay was developed. The model will project the state of the Bay's health over twenty years from 2010 to 2030 on a monthly scale. Spatially, the model has fidelity to the level of an acre and incorporates river flow with over 1,000 river segments. The population of the watershed is captured at the level of the household.

The model was selected for massive parallel execution on the IBM World Community Grid (*World Community Grid, 2011*). This will provide the opportunity to explore a large number of scenarios to better understand the potential impacts of policy choices as the individual states implement their respective plans to achieve Bay health goals as postulated by the President's Executive Order. Preliminary results are presented below.

![Figure 1: The Chesapeake Bay Watershed (Source: Chesapeake Bay Program Office)](image-url)
BACKGROUND

Modeling coupled natural and human systems requires an approach that captures the relevant spatial and temporal dynamics of the underlying natural system while adequately integrating the effects of human activity. Our approach begins by framing the problem as one of a complex system in contrast to a merely complicated system (Ottino, 2004).

Among the hallmarks of a complex system are that it is:

- **Nonlinear**, i.e., macro-level behavior is not simply the sum of micro-level behaviors;
- **Nonstationary**, i.e., it changes continuously;
- **Nonequilibrium**, i.e., it displays statistically improbable dynamics and patterns; and
- It does not have a single scale, i.e., it often displays fractal, or tree-like, organization.

Because of these characteristics, complex systems often display unpredictable and often surprising behavior. Such outcomes are called emergent outcomes. Consequently, complex systems do not lend themselves to closed-form solutions or traditional analytic methods. Rather, to gain insight into complex system behavior, simulation models are used. These models are not predictive of future behavior but rather produce possible outcomes, even emergent outcomes, depending on the characterization of the underlying system's dynamics.

THE UVa BAY GAME®

A multidisciplinary team of researchers at the University of Virginia took first steps at modeling the Chesapeake Bay Watershed with human participation through the development of a participatory simulation – the UVa Bay Game® (Learmonth et al., 2011).

This participatory simulation, developed and launched in April 2009, consists of a high-level simulation model of the Chesapeake Bay developed as a systems dynamics model with a game interface allowing human players to interact with the simulation model as it moves through ten, two-year rounds simulating twenty years—2000 - 2020. The first five rounds (ten years) of the participatory simulation cover historical years where key data are known; the last five two-year rounds use randomly projected values for key model variables.

Participants take on an assortment of roles including crop farmers, livestock farmers, land developers, watermen, and their corresponding regulators. Players are distributed into seven sub-watersheds. At each round of game play, participants review information about their decisions and individual performance in previous rounds; they also see regional (sub-watershed) results and limited global (watershed-wide) results. Based on their analysis of this information, they then possibly revise their choices and decisions entering the next round. The UVa Bay Game can accommodate a total of 166 live players but typically has 16 to 40 live players.

The UVa Bay Game is regularly played in university classes at UVa and six other universities as part of a research consortium. Additionally, regulators and policy-makers from the National Oceanographic and Atmospheric Administration, the U.S. House of Representatives, and the Institute for Environmental Negotiation have played the UVa Bay Game. More interestingly, game plays with industry executives have proven quite
successful as they seek to better understand the economic risk associated with the “water problem.”

The underlying simulation model in the UVa Bay Game is at a very high level of aggregation—spatial representation is at the level of seven major sub-watershed; temporal representation is at two-year intervals (although the underlying uses a quarter-year step); and with human agents (players) each representing on the order of hundreds or thousands of actual persons. The results produced by the model are, however, accurate—in the parlance of simulation modeling, the results have been verified against empirically reported data through 2010.

The goal of the UVa Bay Game model is not to maintain deep fidelity to the natural processes happening in the Chesapeake Bay, but rather for game players, in their various roles, to gain understanding of the complexity of this human and natural system and to realize that their human decisions and actions can lead to positive outcomes, but ultimately, uncontrollable Nature rules.

**UVa Bay Game®/Analytics — A Large-Scale, High Fidelity Simulation Model**

With the high-level of aggregation in the UVa Bay Game model, there is limited ability to examine the impact of proposed policy choices such as the effectiveness of agricultural and urban BMPs (Best Management Practices); land use changes over time; and population growth. Consequently, a more detailed model with greater fidelity to spatial and temporal dimensions would be required.

A model with finer spatial detail; a monthly versus quarterly time step; and the ability to represent the impact of human activities was proposed. This simulation model would not have live human players. Also, at this level of fidelity, it would not only consist of a simulation program and database of extraordinary size, but with varying only a few parameters, each over a limited range of choices, would require a (combinatorially) large number of computer runs. This computational challenge being, as computer scientists would say, *embarrassingly parallel*, led to the search for computational resources on which to run large-scale, high fidelity simulation-model-only experiments.

The World Community Grid, sponsored by IBM, is such a computational resource. The World Community Grid consists of over 1.5M computers donating their unused computer time to problems of social importance. The volunteered computers have screen savers installed which, when started during idle time, run computationally intensive programs in the background. When the owner of the computer resumes his or her work and thus interrupts the screen saver, the background research program stops and waits until the next idle time when it then resumes.

In 2010, the World Community Grid chose three water-related projects to place on the grid for execution (IBM, 2010). One of these is UVa Bay Game/Analytics—a name chosen to differentiate the high fidelity, simulation-only model of the Chesapeake Bay Watershed from the participatory simulation version. This simulation-only model, described below, is currently being prepared for launch on the World Community Grid with a planned date in late December 2011 or January 2012. Prior to preparation and launch on the World Community Grid, the simulation model described here has been built and tested on a University of Virginia High Performance Computing cluster.
THE MODEL

Like the participatory simulation version, this model is focused on the health (quality) of the Chesapeake Bay as measured by the total nitrogen and phosphorus loads (nutrients) entering the Bay. The model calculates the amount of nitrogen and phosphorus applied to areas of the watershed and follows how those nutrients are either taken up in the soil or run off, eventually reaching the Bay. The water quality problem in the Chesapeake Bay Watershed, and similar regions, is caused by excess nutrients flowing into the watershed.

Nutrients, such as nitrogen and phosphorus, are extremely beneficial to the soil in enhancing crop growth and keeping lawns green and healthy. Too much of a good thing, however, is problematic. Excess nutrients flowing into the Bay lead to excessive algal growth. Algae both cloud the water preventing sunlight from reaching underwater grasses and, when they die, sink to the bottom consuming necessary oxygen. The lack of oxygen creates hypoxic and anoxic regions in the Bay (dead zones) that are inhospitable to aquatic species. In the case of the Chesapeake Bay, this is most challenging to the blue crab population, an industry important to the economic health and sustainability of the Bay.

While there are naturally occurring sources of nutrients in the environment and these do reach the Bay, the excess nutrient problem is a direct result of human action—agricultural fertilization, lack of proper waste removal on animal farms, land development, storm water runoff, increased area of impervious surfaces, septic system use, and atmospheric deposition (a large proportion of the nitrogen in the Bay is the result of air pollution). Such sources of excess nutrients are called non-point sources. They are not easily monitored or controlled. In contrast, point sources such as wastewater treatment facilities and industrial sites are quite effectively monitored and controlled through existing policy actions, such as the Clean Water Act and Clean Air Act.

In this model, all human action is simulated. But, through the ability to create and simulate a wide range of scenarios, the impacts of human action and a range of possible policy actions to mitigate the effects of those actions, can be efficiently and effectively examined.

Spatial dimension

Much of the data used to construct the model comes from diverse sources including the Chesapeake Bay Program Office’s Phase 5 Watershed Model, the Bureau of Labor and Statistics, the US Census Bureau, and the National Agricultural Statistics Service.

The Chesapeake Bay Watershed includes all, or a portion of, 197 distinct Federal Information Processing (FIPS) coded areas. These roughly correspond to counties and cities. Some are further subdivided based on variations in altitude (which affects rainfall measurements) giving a total 238 “parcels.” Each parcel is further characterized by 25 land uses (Table 1).

Lastly, using a cutoff flow rate of at least 100 gallons per minute, 952 river segments are identified. Combining parcels, land uses, and river segments, the spatial dimension of the model consists of approximately 35,000 “areas” consisting of a variable number of farms (about 33,000 in all) and a fixed number (2,259) of non-farm areas. The “area” then is the smallest spatial unit in the model. Farm acreage comprises about 10.2M acres of the
approximately 41.2M sizes in the Chesapeake Bay Watershed. Simulated farm sizes are
distributed according to USDA estimates.

<table>
<thead>
<tr>
<th>Animal Feeding Operations</th>
<th>Nutrient Management - Alfalfa</th>
</tr>
</thead>
<tbody>
<tr>
<td>Alfalfa</td>
<td>Nutrient Management - High Tillage with Manure</td>
</tr>
<tr>
<td>Bare-Construction</td>
<td>Nutrient Management - High Tillage without Manure</td>
</tr>
<tr>
<td>Extractive (Active/Abandoned Mines)</td>
<td>Nutrient Management - Hay</td>
</tr>
<tr>
<td>Forests, Woodlots and Wooded</td>
<td>Nutrient Management - Low Tillage</td>
</tr>
<tr>
<td>High Tillage without Manure</td>
<td>Nutrient Management - Pasture</td>
</tr>
<tr>
<td>Harvested Forest</td>
<td>Pasture</td>
</tr>
<tr>
<td>High Tillage with Manure</td>
<td>High Intensity Developed, Pervious</td>
</tr>
<tr>
<td>Hay without Nutrients</td>
<td>Low Intensity Developed, Pervious</td>
</tr>
<tr>
<td>Hay with Nutrients</td>
<td>Degraded Riparian Pasture</td>
</tr>
<tr>
<td>High Intensity Developed Impervious</td>
<td>Nursery</td>
</tr>
<tr>
<td>Low Intensity Developed Impervious</td>
<td>Open Water</td>
</tr>
<tr>
<td>Low Tillage with Manure</td>
<td></td>
</tr>
</tbody>
</table>

**Table 1: Land uses**

The simulation model distributes the population of the watershed over the approximately
35,000 areas in units of households. The number of persons per household is a parameter
of the model with a typical mean value of 2.4 and a standard deviation of 0.5 persons. Thus,
the model has approximately 7M households. All impacts of human action and decision-
making are captured at the level of the household, for example, whether it uses a sewer
connection versus a septic system.

The population of the Chesapeake Bay Watershed over the twenty-year simulated period
grows in accordance with U.S. Census Bureau projections at the level of each parcel. The
population increase in a parcel is then apportioned based on population density within each
area in the parcel.

**Modeling Nutrient Flow**

Having established the spatial and population distributions, it remains to model nutrient
ccontributions from each farm and non-farm area.¹

The process begins by with *nutrient introduction* to each area based on land use. Total
nitrogen (TN) is composed of NH₃-N (ammonia), NO₃-N (nitrate), and ORG-N (organic
nitrogen) and total phosphorus (TP) is composed of PO₄-P (phosphate) and ORG-P (organic
phosphorus). Next, *nutrient uptake* is calculated based on land use and month of the year
(capturing seasonality).

¹ Note that point sources of nutrient flow are not explicitly modeled because they have steadily
deprecated over recent years as a result of federal, state, and local regulatory actions.
Once nutrients have been introduced and some percentage has been taken up, the rest are presumed to leave the area in the form of nutrient runoff. Depending on the land use type, a percentage of the excess nutrients are transferred to the river segments adjacent to the area. River segments then transfer their nutrients to the next downstream river segment. This river-segment-to-river-segment transfer is limited by a user-configurable discount factor, which simulates the deposition of nutrients into the streambed.

Model execution

The computer code implementing this model is written in C++ and uses a relational database during execution. The actual code size is approximately 9,200 lines and requires about 300MB of memory when run. A single run (replicate) of the model with a given set of parameters takes between 3.5 and 5 hours to complete. When eventually run on the World Community Grid, a large number (on the order of 100) independent replicates for a given set of parameters will be executed. This is will provide the opportunity to assess not only point estimates but also the statistical significance of results.

The set of parameters that define and constitute an experimental model run is chosen from among the following:

- Population growth rate
- Population densities per land use
- Household mean size and standard deviation
- Percent of the population on septic/sewer systems
- Household lawn acreage and number of fertilizations per year
- Nutrient contributions of septic/sewer systems
- Land segment development rates
- Farm size statistics
- BMP efficiencies
- Ratio of degraded riparian buffers to farm size

Even with a small range of discrete values for each parameter, the potential number of experimental runs becomes combinatorially large. The opportunity to use the World Community Grid and its extraordinarily large number of computers makes this exploring this space of parameter settings feasible. However, as part of the model testing protocol, preliminary results over a small number of test cases are available. These are presented below.

Preliminary model results

A total of 144 test runs were completed on the University of Virginia High Performance Computing cluster. These include varying the following parameters and their levels:

- Mean household size (3 levels – 2.2, 2.4, and 2.6 persons)
- Fertilizer per acre (6 levels – 37.6, 39.6, 41.6, 43.6, 45.6, and 47.6 pounds)
- Percentage of land area in lawns (4 levels: 1.0%, 1.5%, 2.0%, and 2.5%)
- Percentage of degraded riparian buffer (3 levels: 0.0%, 0.5%, and 1.0%)

In all cases, point sources of nutrients are not included, neither is atmospheric deposition. It is also well known that non-point sources of nitrogen and phosphorus are highly correlated
to the volume of annual flow of water into the Bay from its river systems. The research questions to be addressed here are to determine the impact of levels of the parameters listed above. By including the highly variable flow volumes it would be impossible to isolate their effects in the presence of such flow volume. Consequently, flow volume is effectively held constant at an average rate experienced over the period 1990–2009.

Figures 2 and 3 show the annual Total Nitrogen and Total Phosphorus loads entering the Chesapeake Bay with Mean Household Sizes of 2.2, 2.4, and 2.6 persons per household. (The mean household size in 2010 was 2.4.) As expected, the annual loads remain relatively constant over the 20-year period, all other parameters being held constant. This simple result provides implicit verification of the underlying model.

**Figure 2: Total nitrogen for levels of Mean Household size**

**Figure 3: Total phosphorus for levels of Mean Household size**
Figures 5 and 6 show the annual Total Nitrogen and Total Phosphorus loads entering the Chesapeake Bay based on the Percentage of Degraded Riparian Buffers. A riparian buffer is an area at the edge-of-stream—run, creek, river, etc.—that protect the water body from the effects of runoff from adjacent land use.

In agricultural areas, creating and managing proper riparian buffers is considered a “Best Management Practice” and is often incentivized through direct payments to land owners to offset the cost of planting and maintaining these areas which are then not available for crop planting or as pasture for livestock.

**Figure 4: Riparian Buffers - badly degraded, degraded, and well-managed (Stream Notes, 1999)**

**Figure 5: Total nitrogen for levels of degraded riparian buffers**
Figure 6: Total phosphorus for levels of degraded riparian buffers

Figures 5 and 6 are the result of single replicate runs and display natural variability. However, the results generally indicate increases in nitrogen and phosphorus loads with larger areas of degraded riparian buffers. When this scenario is run with a large number of independent replicates on the World Community Grid—a Monte Carlo approach—and averages over the replicates are plotted, the levels will separate. It is of some note that regardless of the percentage of degraded riparian buffer space, the annual phosphorus load over the years 1990 – 2010 declines. This is a result of regulation—phosphates are no longer used in detergents and other cleansers—a victory of policy.

Figures 7 and 8 show the annual Total Nitrogen and Total Phosphorus loads entering the Chesapeake Bay based on the Percentage of Land Areas in Lawns. These represent single replicate runs of the model.

Figure 7: Total nitrogen for levels of Percentage Land Area in Lawns
While large-scale agricultural operations and even land development are commonly cited as the principal sources of nutrient flows into the Chesapeake Bay Watershed, an overlooked, though highly significant source of nutrients are household lawns (and public parks and golf courses). Here the direct impact of individual citizens’ activities is seen. As the percentage of lawn area increases, nitrogen loads expectedly increase. Phosphorus loads decline steadily and do not change significantly with percentage land area devoted to lawns. The decline is again due in large part to an overall reduction in phosphorus use. The lack of separation over percentages is a result of the fact that phosphorus is less of a factor in runoff from lawns than nitrogen.

CONCLUSION

The large-scale, high fidelity model described here is the latest project conducted at the University of Virginia to better understands the impact of human activities on the complex system that is the Chesapeake Bay Watershed. The UVa Bay Game® is a participatory simulation that enables individuals—students, policy-makers, business executives—to personally grasp the inherent challenges of balancing environmental and economic sustainability, especially in the presence of uncertain environmental conditions.

At a relatively high level of aggregation, the Bay Game provides realistic responses to player decisions and for those players to experiment with decisions in a responsive game context. Despite its richness, the Bay Game does not have the necessary fidelity to deeply understand and experiment with prospective policy choices. To address this need, a simulation-only model was built to achieve a larger scale and greater fidelity to both human and natural activities. The resulting model is about to be launched on the World Community Grid sponsored by IBM where a very large number of experiments over a set of selected parameters, individually and in combination, will be conducted. Preliminary results were presented here. Over the next twelve months, detailed simulation results will become available.
References


