

INTEGRATING AN AGENT-BASED MODEL OF MALARIA MOSQUITOES WITH A GEOGRAPHIC INFORMATION SYSTEM

S. M. Niaz Arifin^(a), Rumana Reaz Arifin^(b,c), Dilkushi de Alwis Pitts^(c), Gregory R. Madey^(d)

^(a)Department of Computer Science and Engineering, University of Notre Dame

^(b)Department of Civil & Environmental Engineering & Earth Sciences, University of Notre Dame

^(c)Center for Research Computing, University of Notre Dame

^(d)Department of Computer Science and Engineering, University of Notre Dame

^(a)sarifin@nd.edu, ^(b,c)rarifin@nd.edu, ^(c)dpitts@nd.edu, ^(d)gmadey@nd.edu

ABSTRACT

Agent-based models (ABMs) are used to model infectious diseases and disease-transmitting vectors. Malaria is a deadly infectious disease in humans, transmitted by *Anopheles* mosquito vectors. Although geographic information system (GIS) has been used before with ABMs, no ABM-based malaria study showed the usage of custom-built spatial outputs integrated within a modeling framework. In this paper, we show how to effectively integrate a malaria ABM with GIS-based, spatially derived parameters. For a specific study area, we process GIS data layers, create hypothetical scenarios, produce maps, and analyze biological insights. Results indicate that availability of resources and relative distances between them are crucial determinants for malaria transmission. The maps also reveal potential hotspots for the measured variables. We argue that such integrated approaches, which combine knowledge from entomological, epidemiological, simulation-based, and geo-spatial domains, are required for the identification of relationships between spatial variables, and may have important implications for malaria vector control.

Keywords: agent-based model, malaria, *Anopheles gambiae*, geographic information system

1. INTRODUCTION

Malaria is one of the oldest and deadliest infectious diseases in humans, transmitted by female mosquitoes of the genus *Anopheles*, which are regarded as the primary vector for transmission. Agent-based models (ABMs) can play important roles in malaria modeling, and answer interesting research questions. For example, ABMs can assist in selecting appropriate combinations of vector control interventions to interrupt malaria transmission, and in setting response timelines and expectations of impact.

Earlier, we developed a spatial ABM of the mosquito vector *Anopheles gambiae* for malaria epidemiology (Arifin, Davis, and Zhou 2011a; Arifin, Davis, and Zhou 2011b). Following a biological core model that describes the mosquito vector population

dynamics, the ABM simulates the life-cycle of mosquito agents by tracking attributes of each individual mosquito.

A geographic information system (GIS) is a system designed to capture, store, manipulate, analyze, manage, and present all types of geographical data. The idea of integrating GIS with ABMs is not new. Several studies, ranging across multiple domains, have shown such integration. For example, Brown *et al.* (2005) addressed the coupling of GIS-based data models with agent-based process models, and analyze different requirements for integrating ABM and GIS functionality. They illustrate the integration approach with four ABMs: urban land-use change, military mobile communications, dynamic landscape analysis and modeling system, and infrastructure simulations.

GIS has also been used in various epidemiological studies. For example, Gimnig, Hightower, and Hawley (2005) discussed the application of GIS to the study of mosquito ecology and mosquito-borne diseases, including malaria. Khormi and Kumar (2011) presented a review of mosquito-borne diseases, with examples of the use of spatial information technologies to visualize and analyze mosquito vector and epidemiological data.

However, no *model-based malaria study* has yet shown how to integrate an ABM with GIS, and thereby harness the full power of GIS, especially by utilizing custom-built spatial outputs. There is also a *vacuum of knowledge* in building robust integration frameworks that can guide the use of ecological, geo-spatial, environmental, and other types of features (related to malaria transmission) as model inputs, as opposed to simply use these features as cartographic outputs from the models (as done by most previous studies).

In this paper, we show how to effectively integrate a spatial ABM of malaria vector mosquitoes with a GIS. For a specific study area (Asembo, Kenya), we identify the relevant data layers, and collect, analyze, and prepare the data for the ABM. We rank different aquatic habitat types based on their characteristics. Then, we assign relative carrying capacities to the habitats, and build two hypothetical scenarios. Once the ABM is run with both scenarios, we analyze custom spatial variables

(outputs of the ABM), which include adult abundance by location, cumulative biomass per aquatic habitat, cumulative number of females oviposited per aquatic habitat, and cumulative number of bloodmeals per house. Lastly, we produce GIS maps by overlaying the spatial variables on top of the relevant data layers, and analyze important biological insights as discovered from the maps.

The organization of this paper is as follows: Section 2 describes some relevant studies. Section 3 briefly describes the ABM, the study area, and the GIS-ABM Workflow. In Section 4, we describe, in details, the processing steps of the GIS data layers. Section 5 describes two hypothetical scenarios created for the ABM. Section 6 describes the assumptions of the simulations, and defines the four custom spatial variables produced as outputs of the ABM. Section 7 describes our results, and Section 8 concludes.

2. LITERATURE REVIEW

In this section, we discuss several malaria-related studies that use GIS, Global Positioning System (GPS), spatial statistical methods, geo-spatial features, *etc.* In general, GIS has been extensively used in epidemiological studies. In particular, for malaria as a disease, GIS has been used for measuring the distribution of mosquito species, their habitats, the control and management of the disease, *etc.* GIS and spatial statistical methods are regarded as important tools in epidemiology to identify areas with increased risk of diseases, and determine spatial association between disease and risk factors (Mmbando *et al.* 2011).

Mbogo *et al.* (2003) studied the seasonal dynamics and spatial distributions of *Anopheles* mosquitoes and *Plasmodium falciparum* parasites along the coast of Kenya. Using hand-held GPS, they recorded latitude and longitude data at each site, and produced the spatial distribution maps for three *Anopheles* species. Li, Bian, and Yan (2006) presented a spatially distributed mosquito habitat modeling approach, integrating a Bayesian modeling method with Ecological Niche Factor Analysis (ENFA) using GIS. They used data for seven environmental variables to represent the environmental conditions of larval habitats in the Kenya highlands. The *Malaria Atlas Project (MAP)* developed the science of malaria cartography by modeling the global spatial distribution of *P. falciparum* malaria endemicity (Hay and Snow 2006). Focusing on the spatial heterogeneity of malaria transmission intensity, they effectively produced and used maps as essential tools for malaria control (Hay *et al.* 2009).

Zhou *et al.* (2007) used GIS layers of larval habitats, land use type, human population distribution, house structure, and hydrologic schemes, overlaid with adult mosquito abundance, to investigate the impact of environmental heterogeneity and larval habitats on the spatial distribution of adult *Anopheles* mosquitoes in western Kenya. Mmbando *et al.* (2011) conducted a study of four cross-sectional malaria surveys in 14

villages located in highland, lowland, and urban areas of northeastern Tanzania during the rainy seasons. Their results show a significant spatial variation of *P. falciparum* infection in the region, identifying altitude, socio-economic status, high bednet coverage, and urbanization as important factors associated with the spatial variability in malaria. Ndenga *et al.* (2011) used a GPS unit to classify aquatic habitats within highland sites in western Kenya. They recorded the latitude, longitude, and altitude of the habitats, and classified them as natural swamp, cultivated swamp, river fringe, puddle, open drain or burrow pit, and showed that the productivity of malaria vectors from different habitat types are highly heterogeneous.

3. ABM, STUDY AREA, AND WORKFLOW

In this section, we describe the study area, the GIS-ABM Workflow, and the selected GIS data layers, which can be broadly classified into two categories: aquatic habitats and houses.

3.1. The Agent-based Model (ABM)

The agent-based model (ABM) was described earlier in (Zhou, Arifin, Gentile, Kurtz, Davis, and Wendelberger 2010). It is derived from a core entomological model of the dominant malaria vector species *An. gambiae*. The core model, essentially conceptual in nature, is governed by the biology underlying *An. gambiae*, and describes the vector population dynamics. The verification and validation processes for the ABM were described in (Arifin, Davis, and Zhou 2010a; Arifin, Davis, Kurtz, Gentile, and Zhou 2010b). For this study, we use a spatial extension of the ABM, which was described in detail in (Arifin, Davis, and Zhou 2011a; Arifin, Davis, and Zhou 2011b).

In the spatial ABM, each aquatic habitat is associated with a finite *carrying capacity (CC)*, which is treated as a soft limit on the aquatic mosquito population that the aquatic habitat can sustain. The *combined carrying capacity (CCC)* for a given landscape (with one or more aquatic habitats) represents the sum of the *CCs* of all aquatic habitats.

3.2. The Study Area

For this study, a village cluster in Kenya's Rarieda Division in Nyanza Province, known locally as *Asembo*, is chosen as the study area (see Figure 1). *Asembo* is located within a subsection of the Siaya and Bondo Districts in western Kenya. According to estimates from the 1989 Kenya Government census statistics, it covers an area of 200 km² and had a population of approximately 60,000 persons (Phillips-Howard *et al.* 2003). *Asembo* includes a study site of 15 villages (with an area of approximately 70 km² near *Asembo Bay*, and experiences intense, perennial malaria transmission (Nahlen, Clark, and Alnwick 2003).

The primary reason for selecting *Asembo* as our study area is the availability of data from the *Asembo Bay Cohort Project* (McElroy *et al.* 2001) and the *Asembo ITN project* (Phillips-Howard *et al.* 2003),

which, in a series of 23 articles, report important public health findings from a successful trial of insecticide-treated bednets (ITNs) in western Kenya (Kazura 2003). Their research findings provide substantial evidence that high coverage of ITNs in the study area will result in significant health benefits for affected communities (Nahlen, Clark, and Alnwick 2003).

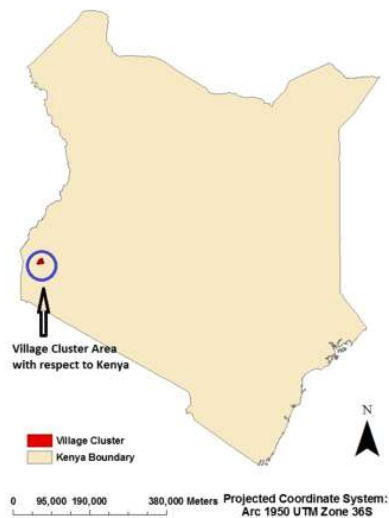


Figure 1: The Study Area of Village Clusters in Asembo, Kenya

3.3. The GIS-ABM Workflow

The GIS-ABM workflow is shown in Figure 2. The GIS module, using the ArcGIS software (ArcGIS Desktop 2011), produces, processes, and analyzes the relevant data layers, and converts them into plain-text ASCII format. The ASCII files are then converted into XML format by using a customized Java program (the Input Formatter), and fed as input to the spatial ABM. Once the ABM completes the simulations, the outputs are analyzed by using a custom-built Perl module (the Output Analyzer). Plots and other figures are then produced from the analyzed output. To perform the spatial analysis, we produce ASCII files from the analyzed output, and feed those into the GIS module. The GIS module then produces spatial maps with relevant information portrayed on top of the data layers.

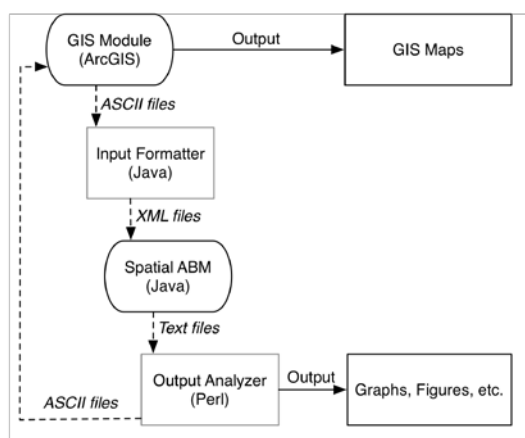


Figure 2: The GIS-ABM Workflow

4. GIS DATA LAYERS

The GIS data layers represent several thematic layers of the study area that are relevant to our spatial ABM. These layers fall into two categories: aquatic habitats and houses. The aquatic habitat types include two types of mosquito breeding sites (*type-1 breeding site* and *type-2 breeding site*), boreholes, pit latrines, and wetland. A type-2 breeding site is composed of a type-1 breeding site and several other data points (*e.g.*, compounds, boreholes *etc.*). Boreholes, also known as borrow pits, have great potentials as breeding sites in this area, and represent holes or pits made in the ground when local people use clay or soil for building houses, making pots, *etc.*, thereby leaving depressions in the ground that easily get filled with rain water. Pit latrines are very common to households in the study area. The wetland represents a stretch of waterbody lying to the northwest corner of the study area. Human houses serve as bloodmeal locations for the mosquitoes, and include houses, huts, *etc.*

4.1. Processing the Data Layers with GIS

We start with the feature identification and extraction process for the whole of Kenya; then, we describe the *scale down* process to the study area of Asembo, followed by the selection of a subset of villages within Asembo, and finally, to the selection of a polygon within the village clusters, which is used as input to our spatial ABM.

We first identify and extract different water features (rivers, wetlands, and several water-points) and villages (including human houses) for all over Kenya, as shown in Figure 3. Different water features (rivers, wetlands, *etc.*) and villages are projected to the projection system Arc 1960 UTM Zone 36S. In Figure 3, the figure on the left shows different water features (rivers, wetlands and several water-points) all over Kenya, and the figure on the right shows villages for Kenya.

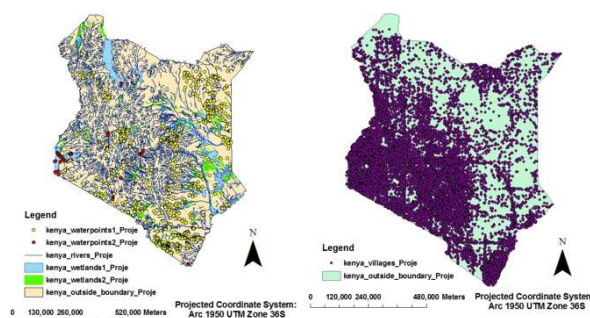


Figure 3: Water Sources and Villages Projections for Kenya

4.2. Selecting Aquatic Sites for the Study Area

We collect water features data for different types of aquatic sites. Each water source is assigned a unique ID (*ID data-feature*).

Once we thoroughly examine the water sources' data layers, we encounter some overlapping problems.

To overcome these, we provide precedence ranking for the data layers, by sub-grouping water source layers based on their attributes, and creating new shapefiles. In the process, we combine similar types of water features in the same data layer.

Figure 4 shows the selected data layers for different water features. We use the *Select By Attributes* tool (with SQL query) to select features, and export the selected features to create new shapefiles. We also assign new IDs for each water feature type by using the *Field Calculator* tool to calculate value for the $ID_{data-feature}$ field.

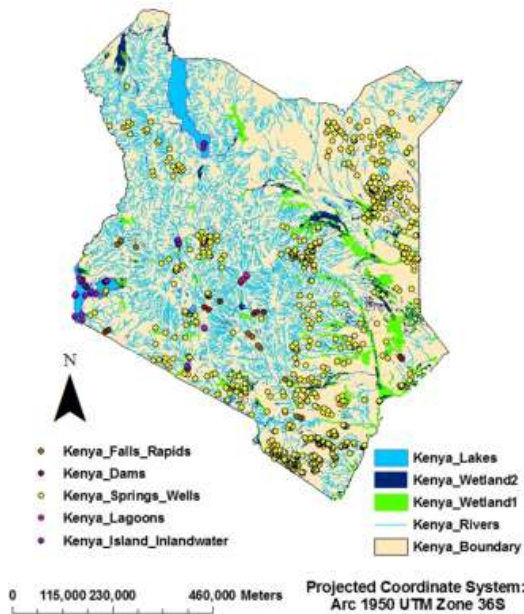


Figure 4: Selected Water Features for Kenya

4.3. Scaling Down to Village and Household Level

Since for the spatial ABM we need household level data that include water features available near the houses, we scale down the data to selected village cluster area from the entire Kenya boundary area.

We select a village cluster in Asembo, based on higher frequency of aquatic sites availability near households (than other clusters in Asembo). After analyzing the water features for all over Kenya, we also discover some wetlands and rivers features in the selected area. Figure 5 shows the selected village cluster area with houses and all water features located in the study area.

For reasons of performance and complexity (for example, large number of features), we further select a subset of villages from the village cluster area. The ABM, without explicit parallelization or multiple runs, can handle a landscape with maximum dimensions of $95\text{ columns} * 96\text{ rows}$. To reflect the available field data that points to limited flight ability and perceptual ranges of *Anopheles* mosquitoes, each cell in the landscape is set to $50\text{ m} * 50\text{ m}$, yielding a total area of $\approx 25\text{ km}^2$. Hence, we further scale down the area, and select a 25 km^2 polygon, as outlined in magenta in Figure 5.

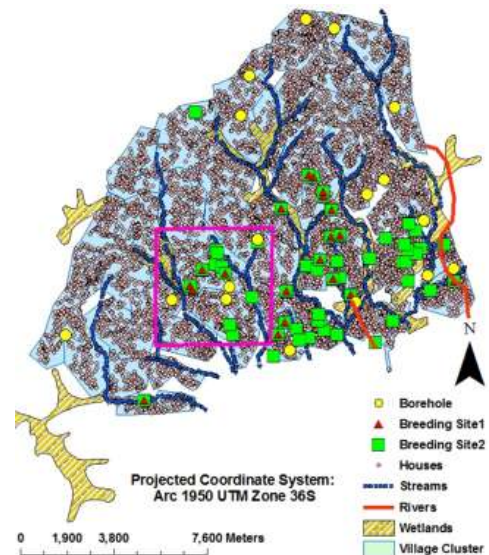


Figure 5: The Selected Polygon, Outlined in Magenta, within the Village Clusters in Asembo, Kenya

Next, we clip (crop) the aquatic habitats and houses within the outside boundary of the polygon. The clipped features include wetlands, streams, boreholes, breeding sites, and pit latrines. We eliminate the stream and river features, which, being moving (non-stagnant) water sources, are usually not considered as prospective breeding sites for *Anopheles* mosquitoes.

4.4. Conversion of Data Formats

Since the ABM needs data in the ASCII format, we first convert the selected layers to raster grid format. The cell size is set to $50\text{ m} * 50\text{ m}$, with the value field set to $ID_{data-feature}$. All point feature data layers for type-1 breeding sites, type-2 breeding sites, boreholes, pit latrines, and houses are converted using the *Point to Raster* tool. The data layer for pit latrines is created from the data layer for houses (since pit latrines are usually found inside household boundaries).

Due to the resolution (cell size), more than one feature type may fall in a single cell. In these cases, to calculate the number of features (of each type) in each cell, we set the *Cell Assignment Type* as *SUM*, since it sums the attributes of all points in the cell. Thus, it acquires the summation of the value fields (of $ID_{data-feature}$), and helps us to determine the number of features. Next, we set the extent for the conversion as the boundary coordinates of the polygon area shapefile (see Figure 5), and convert the raster files to ASCII format.

4.5. Generating the Study Area

Finally, we generate the study area for the ABM, which is shown in Figure 6: Map 1 shows the study area polygon for the ABM, outlined in magenta. The same polygon, within the village cluster area of Asembo, is shown in Map 2. In Map 3, the village cluster is marked in red circle within the map of Kenya. The figure clearly identifies the comparative scale down process of the area, as described previously in this section.

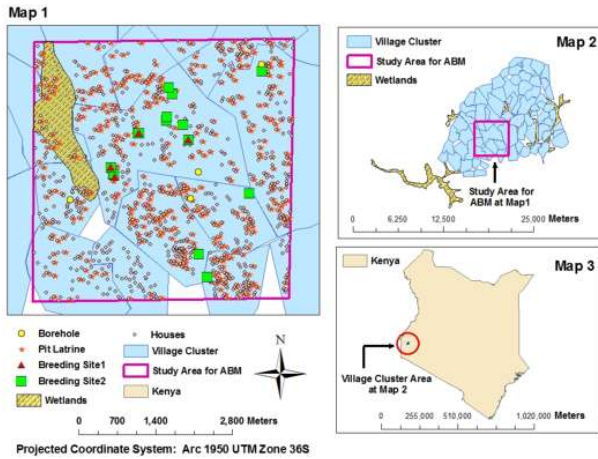


Figure 6: Study Area for the Spatial ABM

We reserve the use of shades of blue for all aquatic habitats, and square brown symbols for the houses. To aid in visualization, gridlines are also added to the map to every tenth point (starting from 1), using the Hawth's Tool (Hawth 2013), which we import into ArcGIS. The output map is shown in Figure 7.

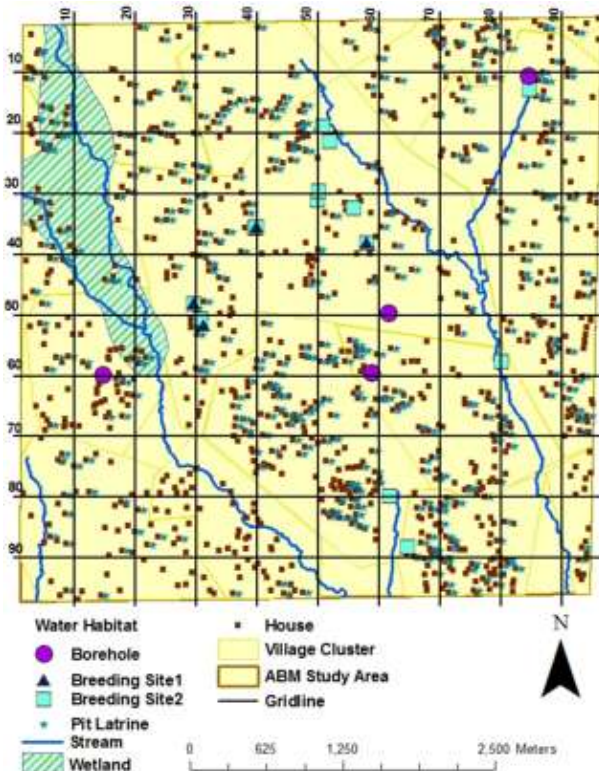


Figure 7: Study Area with Selected Data Layers and Gridlines for the Spatial ABM

4.6. Generating the Feature Counts for the ABM

From the GIS data layers described above (*i.e.*, houses and aquatic habitats), we generate the feature counts to use as inputs to our ABM. The feature counts, as shown in Table 1, appear as 1976 (982 aquatic habitats of different types, and 994 houses).

Table 1: GIS Feature Counts For The ABM

Feature Type	Feature Count
Type-1 breeding site	4
Type-2 breeding site	14
Borehole	4
Pit latrine	401
Wetland	559
House	994
Total	1976

5. CREATING SCENARIOS FOR THE ABM

To run simulations with the selected data layers, we create two hypothetical scenarios with different combined carrying capacities (*CCCs*, see Section 3.1).

We assign carrying capacities to the selected GIS layers that represent the aquatic habitats. However, since we cannot obtain absolute *CC* values for the habitats (due to the lack of habitat data), we assign relative *CC* values to the habitats based on the available spatial data, ensuring that the relative magnitudes of *CCs* are in accordance with: 1) the malaria vector productivity among distinct habitat types, and 2) the biological reality of the environment. For example, considering different cells in the spatial grid, a large breeding site cell would have higher *CC* than that of a wetland cell, although both cells represent the same surface area.

In terms of the magnitudes of the assigned *CCs*, we arbitrarily order the different aquatic habitat types in decreasing order of *CC* per cell: 1) type-1 breeding site, 2) type-2 breeding site, 3) borehole, 4) pit latrine, and 5) wetland. For wetland, which covers multiple cells in the northwest corner of the study area (see Figure 7), we assign the same *CC* value for each cell. In the future, when the data is available, the order, as well as the assigned *CC* values (to different aquatic habitat types), can be readily changed, and the ABM is ready to run with the newly assigned values.

To run the ABM with the selected data layers, we create two hypothetical scenarios with different *CCCs* by assigning relative *CCs* to the different aquatic habitat types, keeping the order of magnitudes intact. The *CCCs* for the scenarios appear as **21K** and **150K**, as shown in Tables 2 and 3.

Table 2: Dry Season Scenario 21K

Feature Type	Feature Count	Assigned <i>CC</i>	Total
Type-1 breeding site	4	1000	4000
Type-2 breeding site	14	500	7000
Borehole	4	100	400
Pit latrine	401	10	4010
Wetland (each cell)	559	10	5590
Total (CCC)			21000

Table 3: Rainy Season Scenario 150K

Feature Type	Feature Count	Assigned CC	Total
Type-1 breeding site	4	5000	20000
Type-2 breeding site	14	2000	28000
Borehole	4	1500	6000
Pit latrine	401	100	40100
Wetland (each cell)	559	100	55900
Total (CCC)			150000

The two scenarios, depicted in Tables 2 and 3, may effectively represent two ecological settings with *low* (21K) and *high* (150K) potentials for mosquito populations, resembling the *dry* and *rainy* weather seasons, respectively, for the study area.

6. SIMULATIONS

We assume the following for the spatial ABM: the model starts with 1000 initial female adult mosquito agents (no male agents). All new agents (entering as eggs) are female. For each initial female agent, the state is set to *Gravid*, the age is set to 120 hours (for being in the *Gravid* state), and the agent is assigned to an aquatic habitat chosen at random. Since available field data points to limited flight ability and perceptual ranges of mosquitoes, the speed and range of movement (of mosquitoes) in our spatial ABM are controlled by special agent-level variables. Unlike other traditional malaria transmission models, we assume senescence (biological aging) of the mosquitoes, and the ABM implements age-specific mortality rates for the adult mosquitoes and the larvae (*i.e.*, the probability of death for mosquito agents increases with their age).

In order to seek for resources (aquatic habitats or houses), and hence to complete the gonotrophic cycles, the adult female mosquito agents move only while they are in *Bloodmeal Seeking* and *Gravid* states. At any point in the resource-seeking process, a mosquito's neighborhood is modeled as an eight-directional *Moore* neighborhood. The landscape is assumed to have a non-absorbing boundary, modeled topologically as 2D torus spaces). For details, see (Arifin, Davis, and Zhou 2011a; Arifin, Davis, and Zhou 2011b).

6.1. Spatial Variables

For the two scenarios (21K and 150K), the output of our spatial ABM includes four custom spatial variables:

1. *Adult abundance by location*: shows the distribution of the adult mosquitoes over the entire landscape at the end of the simulation.
2. *Cumulative biomass per aquatic habitat*: overlaid on top of the aquatic habitats, it represents the sum of biomass (eggs, larvae, and pupae) present in an aquatic habitat.
3. *Cumulative number of females oviposited per aquatic habitat*: also overlaid on top of the

aquatic habitats, it represents the sum of the number of times female adults oviposited in the aquatic habitat.

4. *Cumulative number of bloodmeals per house*: overlaid on top of the houses, it represents the sum of the number of times female adults obtained bloodmeals in the house.

The last three spatial variables are sampled across all daily timesteps throughout the entire simulation. The output GIS maps, described in the next section, are produced by overlaying the above spatial variables on top of the relevant data layers. These variables allow us to analyze spatial correlations and find spatial patterns from the outputs of the ABM.

We use a special GIS map symbolizing technique known as *graduated symbols*. Graduated symbols, used to compare quantitative values, vary in size according to the relative magnitudes of the values. In all output maps, we use graduated symbols as hollow circles, where the relative radii of the circles are determined by the output values generated by the ABM.

7. RESULTS

7.1. Mosquito Abundance (Non-spatial)

In the ABM, mosquito abundance depends, among other factors, on the carrying capacities of the aquatic habitats. Hence, not surprisingly, the 150K scenario yields much higher abundance than the 21K scenario, as shown in Figure 8. This also validates the abundance patterns usually observed in the *dry* and *rainy* seasons.

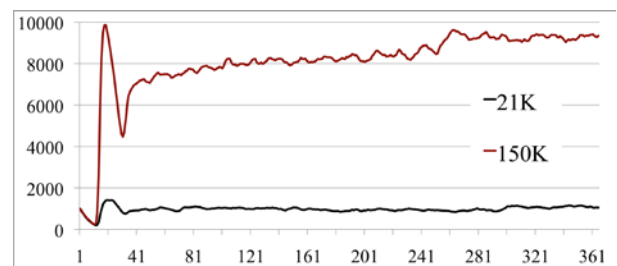


Figure 8: Mosquito Abundance (Non-spatial)

However, the output maps from all four spatial variables indicate that the *relative distances* between the aquatic habitats and the houses play a crucial role in determining the variables of interest, as shown in the following.

7.2. Adult Abundance by Location

Adult abundances by location are shown in Figures 9 and 10 for scenarios 21K and 150K, respectively. They indicate that higher abundances are associated with type-1 breeding sites, followed by type-2 breeding sites. Out of the four and 14 breeding sites (of type-1 and type-2, respectively), highest abundances are observed in locations where type-1 sites are in close proximity with type-2 sites, surrounded by human houses.

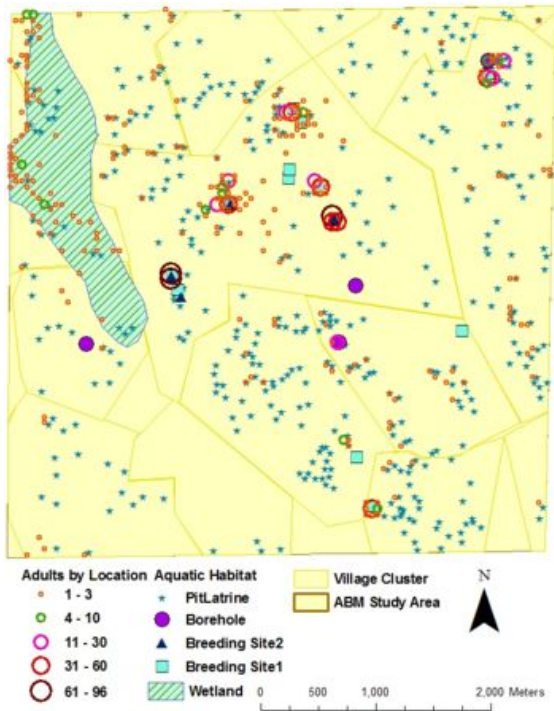


Figure 9: Adult Abundance by Location for 21K

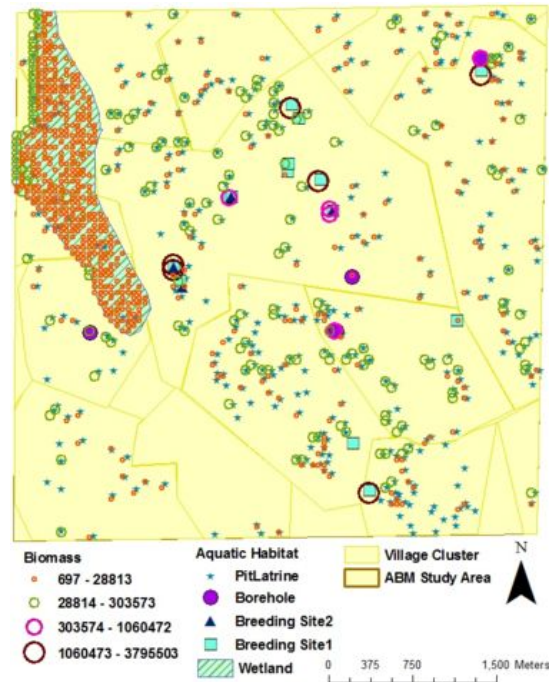


Figure 11: Cumulative Biomass for 21K

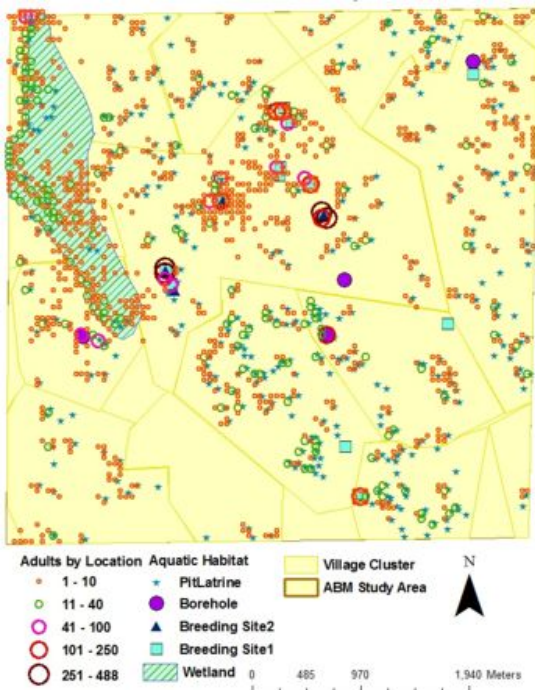


Figure 10: Adult Abundance by Location for 150K

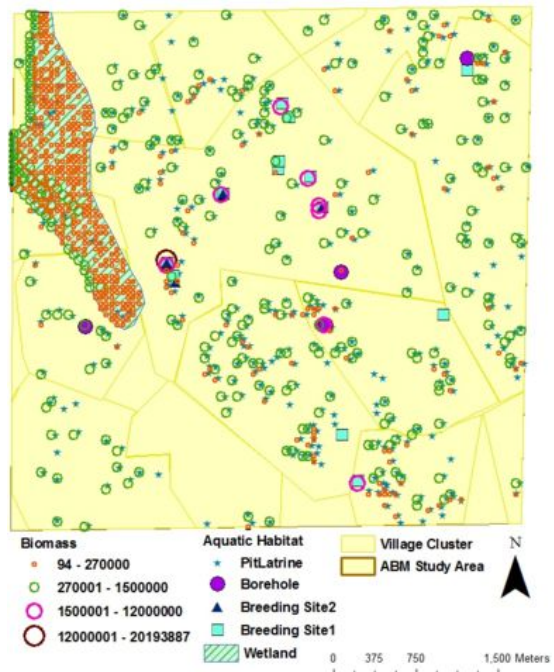


Figure 12: Cumulative Biomass for 150K

We also observe very low (1-3 mosquitoes per cell) abundance in the wetland, which may be attributed to reduced human habitation around the wetland, and low carrying capacities associated with the wetland cells.

7.3. Cumulative Biomass and Females Oviposited

Figures 11 and 12 show the cumulative biomass *per aquatic habitat* for scenarios 21K and 150K, respectively.

Figures 13 and 14 show the cumulative number of females oviposited *per aquatic habitat* for scenarios 21K and 150K, respectively.

Both of these metrics (Figures 11-14) show that higher abundances are associated with type-1 breeding sites, followed by type-2 sites, which are close to boreholes. However, an interesting insight reveals that two (out of 14) type-2 sites, suitable to yield high outputs (like other type-2 sites), yield only 0.07%-0.8% cumulative biomass, and only 0.005%-1% cumulative

number of females oviposited, when compared to other type-2 sites.

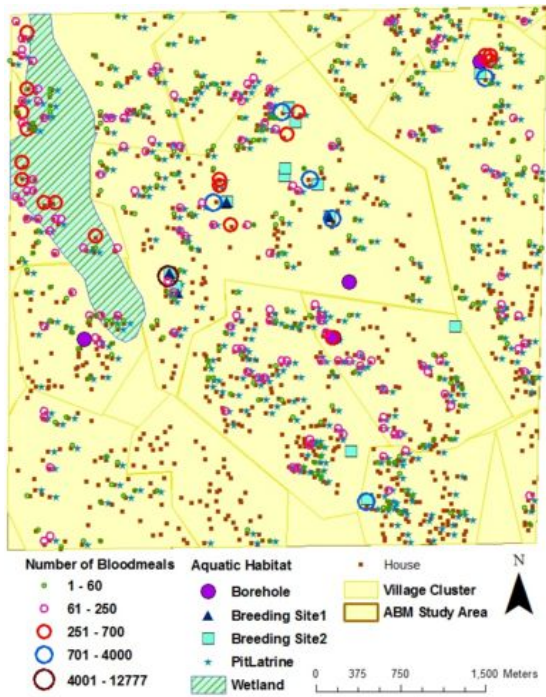


Figure 13: Cumulative Number of Females for 21K

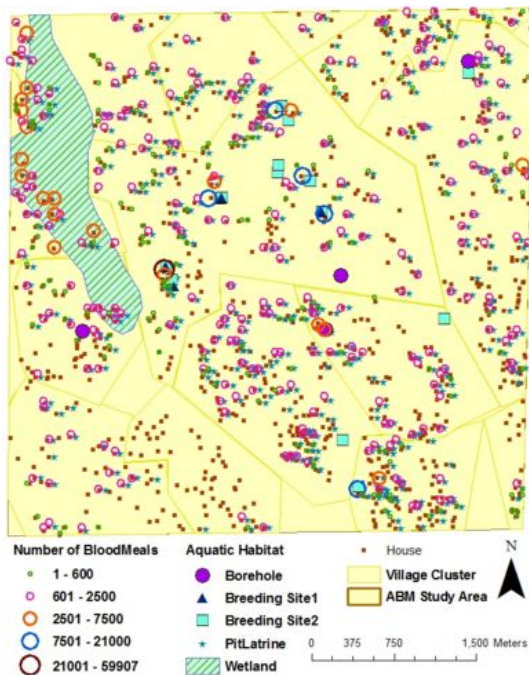


Figure 14: Cumulative Number of Females for 150K

Closer inspection of the corresponding output maps (Figures 11-14) reveals that the nearest human houses around these two type-2 breeding sites are situated much further than other type-2 sites. Since there are not enough houses in the close proximity, the female mosquitoes, ovipositing in these breeding sites, cannot find bloodmeals, and hence are forced to search longer distances. Since the mortality rate of mosquitoes

increases with their age (recall that the ABM implements age-specific mortality rates that incorporate senescence, or biological aging), the additional delays in obtaining bloodmeals actually reduce abundance around these sites, causing much lower cumulative biomass and cumulative number of females oviposited.

7.4. Cumulative Number of Bloodmeals

Lastly, Figures 15 and 16 show the cumulative number of bloodmeals *per house* for scenarios 21K and 150K, respectively.

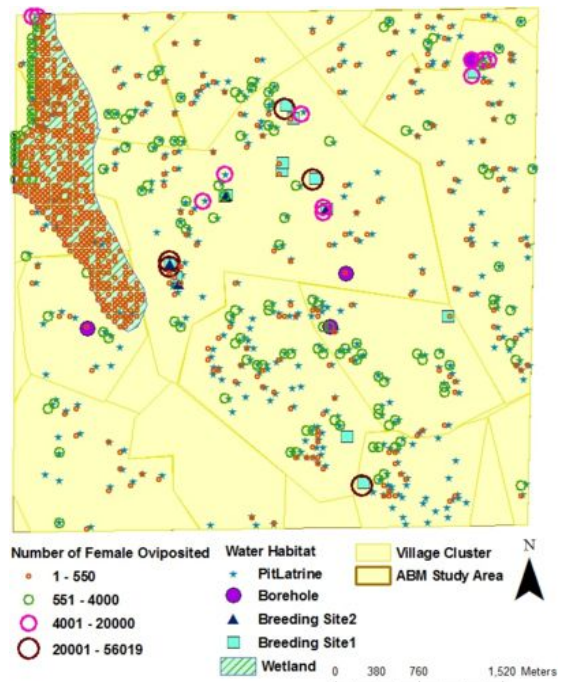


Figure 15: Cumulative Number of Bloodmeals for 21K

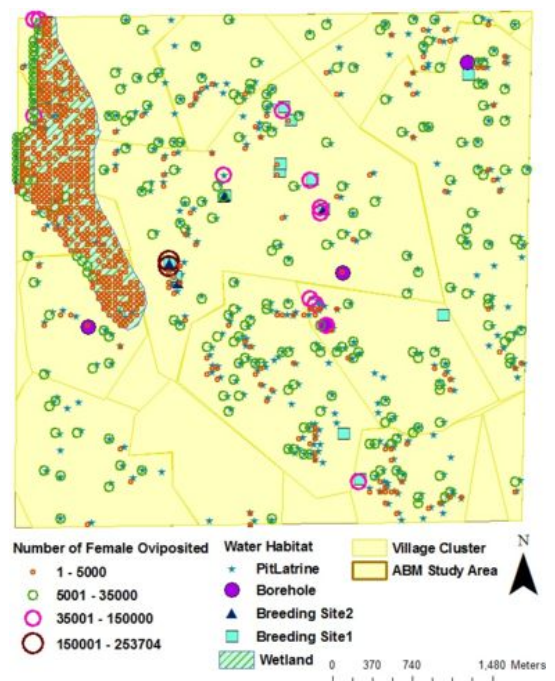


Figure 16: Cumulative Number of Bloodmeals for 150K

Figures 15 and 16 depict that higher values (number of bloodmeals) are associated with houses that have nearby type-1 and type-2 breeding sites, and moderate values are found in houses that have nearby aquatic habitats (of different types) with at least some carrying capacities. Interestingly, a large number of houses, located at the lower left quadrant of the study area, show no bloodmeals, due to the absence of aquatic habitats around them.

8. CONCLUSION

As our findings suggest, *availability* of the ecological resources, *i.e.*, the aquatic habitats and houses, and the *relative distances* between these distinct resource types, are two crucial determinants for the female mosquitoes to complete their gonotrophic cycles. From the viewpoint of mosquito agents, these resources directly define landscape features such as spatial heterogeneity, host availability, *etc.*, the importance of which for vector control have been demonstrated by several studies. Reduced availability of either type of these spatial resources would prolong the gonotrophic cycle of the female mosquito, and potentially affect malaria transmission.

In our study, spatial analysis of the output variables generated by the ABM reveals important biological insights. The use of maps and spatial statistical methods allows readily identifying and displaying interesting spatial patterns, which, without the maps, are difficult to detect. The output maps also reveal potential *hotspots* with higher rates for the measured variables of interest.

The proposed robust integration framework also allows easy parameterization of the model. For example, the arbitrary order of the different aquatic habitat types, and the assigned *CC* per habitat, can be readily changed to suit new scenarios and/or new areas of study. This will allow the ABM to produce site-specific outputs (without the need of modifying the ABM itself). The simplicity in the scenario-based approach allows to feed in different scenarios to the ABM by using different *CCCs* for various aquatic habitat types, without requiring to change the data layers, features, *etc.*, for future simulation runs.

Our results also indicate that disease-specific maps can play important roles in disease control activities, including monitoring the changes of malaria epidemiology, guiding resource allocation for malaria control, and identifying hotspots for further investigation. For example, the results highlight the importance of eliminating the aquatic habitats close to human habitations by means of environmental modifications and manipulations, supporting the arguments presented by several malaria control studies (*e.g.*, Fillinger and Lindsay 2011).

Although in this pilot study we handled a comparatively small study area of $\approx 25 \text{ km}^2$ (which transforms to a *95 columns* * *96 rows* landscape for the spatial ABM), the methodology described here can be readily extended to include larger areas (*e.g.*, the whole Asembo area). For the new regions to be modeled,

either real data can be used, or synthetic/predicted data can be interpolated from a few point regions (on which the described methodology is applied first).

We conclude that such integrated approaches, which combine knowledge from entomological, epidemiological, simulation-based, and geo-spatial domains, are required for the identification and analysis of relationships between various transmission variables, as demonstrated by our study. Eventually, such integration efforts may facilitate the *Integrated Vector Management (IVM)* agenda, promoted by the World Health Organization (WHO), to achieve improved efficacy, cost-effectiveness, ecological soundness, and sustainability of malaria vector control.

ACKNOWLEDGMENTS

Partial support for this work came from the Bill & Melinda Gates Foundation under the Vector Ecology and Control Network (VECNet) Project. We also wish to acknowledge helpful comments from Dr. Frank H. Collins and Dr. Neil F. Lobo at the University of Notre Dame.

REFERENCES

- ArcGIS Desktop (Release 9.3) [Computer software]. (2011). Environmental Systems Research Institute (ESRI). Available from: <http://www.esri.com/>
- Arifin, S. M. N., Davis, G. J., and Zhou, Y., 2010a. Verification & Validation by Docking: A Case Study of Agent-Based Models of *Anopheles gambiae*. *Summer Computer Simulation Conference (SCSC)*, 236-243, Jul. 2010, Ottawa, ON, Canada.
- Arifin, S. M. N., Davis, G. J., and Zhou, Y., 2011a. A Spatial Agent-Based Model of Malaria: Model Verification and Effects of Spatial Heterogeneity. *International Journal of Agent Technologies and Systems*, 3(3):17-34.
- Arifin, S. M. N., Davis, G. J., and Zhou, Y., 2011b. Modeling Space in an Agent-Based Model of Malaria: Comparison between Non-Spatial and Spatial Models. *Proceedings of the 2011 Workshop on Agent-Directed Simulation*, 92-99, Apr. 2011, Boston, Massachusetts.
- Arifin, S. M. N., Davis, G. J., Kurtz, S. J., Gentile, J. E., and Zhou, Y., 2010b. Divide and Conquer: A Four-fold Docking Experience of Agent-based Models. *Winter Simulation Conference (WSC)*, 575 - 586, Dec. 2010, Baltimore, Maryland, USA.
- Brown, D. G., Riolo, R., Robinson, D. T., North, M., and Rand, W., 2005. Spatial process and data models: Toward integration of agent-based models and GIS. *Journal of Geographical Systems*, 7:25-47.
- Fillinger, U. and Lindsay, S., 2011. Larval source management for malaria control in Africa: myths and reality. *Malaria Journal*, 10:353.
- Gimnig, J. E., Hightower, A. W., and Hawley, W. A., 2005. Application of geographic information systems to the study of the ecology of mosquitoes

- and mosquito-borne diseases. In: *Wageningen UR Frontis Series*, 9, 27-39.
- Hawth's Analysis Tools for ArcGIS [Website]. (2013). Available from: <http://www.spatial ecology.com/>
- Hay, S. I. and Snow, R. W., 2006. The Malaria Atlas Project: Developing Global Maps of Malaria Risk. *PLoS Med*, 3(12).
- Hay, S. I., Guerra, C. A., Gething, P. W., Patil, A. P., Tatem, A. J., Noor, A. M., ... Snow, R. W., 2009. A world malaria map: *Plasmodium falciparum* endemicity in 2007. *PLoS Med*, 6(3).
- Kazura, J. W., ed. *The American Journal of Tropical Medicine and Hygiene*, (2003). 68(4 suppl).
- Khormi, H. M. and Kumar, L. 2011. Examples of using spatial information technologies for mapping and modelling mosquito-borne diseases based on environmental, climatic and socio-economic factors and different spatial statistics, temporal risk indices and spatial analysis: A review. *Journal of Food, Agriculture & Environment*, 9(2):41-49.
- Li, L., Bian, L., and Yan, G. 2006. An Integrated Bayesian Modelling Approach for Predicting Mosquito Larval Habitats. In: *UCGIS 2006 Summer Assembly*.
- Mbogo, C. M., Mwangangi, J. M., Nzovu, J., Gu, W., Yan, G., Gunter, J. T., ... Beier, J. C., 2003. Spatial and temporal heterogeneity of *Anopheles* mosquitoes and *Plasmodium falciparum* transmission along the Kenyan coast. *The American Journal of Tropical Medicine and Hygiene*, 68(6):734-742.
- McElroy, P. D., ter Kuile, F. O., Hightower, A. W., Hawley, W. A., Phillips-Howard, P. A., Oloo, A. J., ... Nahlen, B. L., 2001. All-cause mortality among young children in western Kenya. VI: the Asembo Bay Cohort Project. *The American Journal of Tropical Medicine and Hygiene*, 64:18-27.
- Mmbando, B., Kamugisha, M., Lusingu, J., Francis, F., Ishengoma, D., Theander, T., ... Scheike, T., 2011. Spatial variation and socio-economic determinants of plasmodium falciparum infection in northeastern Tanzania. *Malaria Journal*, 10(1):145.
- Nahlen, B. L., Clark, J. P., and Alnwick, D., 2003. Insecticide-treated bed nets. *The American Journal of Tropical Medicine and Hygiene*, 68:1-2.
- Ndenga, B. A., Simbauni, J. A., Mbugi, J. P., Githeko, A. K., and Fillinger, U., 2011. Productivity of Malaria Vectors from Different Habitat Types in the Western Kenya Highlands. *PLoS ONE*, 6(4).
- Phillips-Howard, P. A., Nahlen, B. L., Alaii, J. A., ter Kuile, F. O., Gimnig, J. E., Terlouw, D. J., ... Hawley, W. A., 2003. The efficacy of permethrin-treated bed nets on child mortality and morbidity in western Kenya I. Development of infrastructure and description of study site. *The American Journal of Tropical Medicine and Hygiene*, 68:3-9.
- Vector Ecology and Control Network (VECNet) [Website]. (2013). Available from: <http://www.vecnet.org/>
- Zhou, G., Munga, S., Minakawa, N., Githeko, A. K., and Yan, G., 2007. Spatial Relationship Between Adult Malaria Vector Abundance and Environmental Factors in Western Kenya Highlands. *The American Journal of Tropical Medicine and Hygiene*, 77(1):29-35.
- Zhou, Y., Arifin, S. M. N., Gentile, J. E., Kurtz, S. J., Davis, G. J., and Wendelberger, B. A., 2010. An Agent-based Model of the *Anopheles gambiae* Mosquito Life Cycle. *Summer Computer Simulation Conference (SCSC)*, 201-208, Jul. 2010, Ottawa, ON, Canada.

AUTHORS BIOGRAPHY

S. M. Niaz Arifin is a PhD student in the Department of Computer Science and Engineering at the University of Notre Dame. His research interests include Data Warehousing, Agent-Based Modeling & Simulation, Geographic Information Systems, *etc.* He received his MS from the University of Texas at Dallas in 2006, and BS from Bangladesh University of Engineering and Technology (BUET) in 2004. He served as a software developer at Xcision Medical Systems, California and the Rails Online Database project at Sabre Holdings Corporation, Texas. **Rumana Reaz Arifin** is a PhD student in the Department of Civil & Environmental Engineering and Earth Sciences at the University of Notre Dame. Her research interests include Environmental Engineering and Geographic Information Science. She received her MSs in Civil Engineering (Notre Dame) and Geographic Information Science (University of Texas Dallas). She provides GIS support for Center for Research Computing (CRC) at Notre Dame. **Dilkushi de Alwis Pitts** is an Adjunct Research Assistant Professor at the University of Notre Dame. She earned a PhD in Environmental Engineering (specializing in Hydrology) at Cornell University, NY and a MS in Remote Sensing at the Rochester Institute of Technology, NY. She has nearly 17 years of experience assimilating spatial and attribute data at different spatial and temporal scales and up/down-scaling spatial data for integration into mathematical models and validation of model output. **Gregory R. Mady** is a Research Professor in the Department of Computer Science and Engineering at the University of Notre Dame. His research interests include Agent-Based Modeling & Simulation, Cyberinfrastructure, Open Source Software, Data Warehousing, Emergency Operations Management, Bioinformatics, Modeling Epidemiology, and Health Informatics. His recent research grants include a VECNet Cyberinfrastructure grant from the Bill & Melinda Gates Foundation, and Open Sourcing of the Civil Infrastructure Design grant from the US National Science Foundation.