Research Statement
S. M. Niaz Arifin

1 Overview

In today’s scientific world, computational science is considered the third pillar of scientific inquiry, along with the two traditional pillars of theory and experimentation. Although science is still carried out as an ongoing interplay between theory and experimentation, the increased scale and complexity of both have compelled computational science to be an integral aspect of almost every type of scientific research [1, 2].

Typically, computational science uses computer simulations to construct computational models and quantitative analysis techniques in order to analyze and solve scientific problems. Agent-based modeling and simulation (ABMS) is a class of computational modeling technique for simulating the actions and interactions of autonomous agents with a view to assessing their effects on the simulated system as a whole. ABMS is derived from artificial intelligence and computer science, and combines elements of game theory, complex systems, emergence, computational sociology, multi-agent systems, and evolutionary programming. Agent-based models (ABMs) have applications in diverse real-world problems and have become increasingly popular as a modeling approach in the social sciences because it enables one to build models where individual entities and their interactions are directly represented [3]. An ABM consists of a collection of autonomous decision-making entities called agents. Agents can be individual or collective entities (such as organizations or groups). Each agent represents an actor in the simulated (virtual) environment in which they interact with each other in space and time. Employing a set of rules, agents individually assess the environment and make decisions.

Modeling & simulation (M&S) paradigms such as ABMs are being increasingly used to model complex systems in public health science. In general, the following properties, which can be found in most complex systems, make ABMs especially suited as an efficient tool of the trade for modeling public health challenges: (1) heterogeneity: they comprise large number of heterogeneous elements, (2) dynamic interactions: system properties change in time as a result of the interactions of the actors within the system, (3) emergence: they exhibit emergent phenomena resulting from the interactions of individual entities, and (4) learning and adaptation: they exhibit other complex behavior such as learning and adaptation.

The study of epidemics and infectious disease dynamics can be termed as a signature success of ABMS [4]. Uses of M&S in public health include synthesizing knowledge from disparate disciplines, identifying gaps in existing knowledge, identifying cost-benefit tradeoffs, and generating hypotheses [5]. As such, an increasing number of U.S. universities are incorporating systems science and M&S into their curricula and research programs through the Schools of Public Health and other health-related academic departments [5].

My research focuses on ABMs for malaria epidemiology (and other public health problems), the spatial dimension of modeling, and related data management using dimensional modeling (DM) techniques. Some highlights of my past research are summarized in the next section.

2 Research Highlights

My Ph.D. dissertation focused on ABMs, spatial modeling, and data warehouses. All of these (except the last) are currently being synthesized into my first book (in collaboration with my supervisors) [6]. I also have considerable research experiences of working with the undergraduate students at the University of Notre Dame (ND).
2.1 Involvement in Undergraduate Research

Early undergraduate research involvement in the sciences, especially in the STEM fields, bears enormous importance as it promotes further interest in STEM areas and increase the number of undergraduates, particularly among underrepresented groups. Involving first- and second-year students in research is important for retention. Undergraduate research can capture a student’s interest before she or he decides to pursue other majors, and is a great way to enhance his or her understanding of a particular field or content area in a way that the traditional classroom setting would not fulfill.

The Research Experiences for Undergraduates (REU) Program at ND, which will celebrate 30 years in 2016, provides unique opportunities for undergraduate majors to experience hands-on participation in research in the STEM disciplines. The student, being a rising junior or senior, can come to ND and work with a professor for ten weeks. These programs offer the students a $7,000 stipend for travel and living expenses.

As a recent graduate student (graduated in 2013) and a current Research Assistant Professor, I have been enjoying the opportunity to closely work with some of our REU students at several projects. These projects include the Vector-Borne Disease Network (VecNet, https://www.vecnet.org/), the application of Dynamic Data Driven Application Systems (DDDAS) to the command and control of swarms of Unmanned Aerial Vehicles (UAVs), etc.

2.2 Grants and Contracts Awarded

I am a co-PI on the project An Enhanced Virtual Emergency Operation Center (EOC) for Research and Training in Disaster Management (Ensayo II) funded by the National Science Foundation (NSF). Award: #1405912, $577,474.00. Date: September 2014–August 2016.

2.3 Epidemiology Modeling using ABMs

I have developed ABMs for malaria epidemiology (in collaboration with colleagues). Malaria is one of the largest causes of global human mortality and morbidity. According to the World Health Organization (WHO), half of the world’s population (about 3.4 billion people) are currently at risk of malaria, with about 207 million cases and an estimated 627,000 deaths in 2012 [7]. The ABMs, developed by following a biological core model, simulate the life cycle of one of the most dangerous malaria-transmitting vector species Anopheles gambiae by tracking attributes relevant to the vector population dynamics for each individual mosquito. Evolution of the models have been guided by relevant biological concepts of the species. Details about the core model and the ABMs can be found in [8]. I have also developed a post-simulation analysis tool to analyze and graphically display statistics for an advanced epidemiological ABM named LiNK which models pathogen transmission among long-tailed macaques on the Indonesian island of Bali [9, 10].

2.4 Spatial Modeling

Since the spatial locations of pathogens, hosts, and vectors are fundamentally important to disease dynamics, the representation and treatment of the spatial dimension demand special attention in most epidemiological ABMs. Spatial heterogeneity is considered as one of the most important factors for an effective representation of the environment being modeled. Spatial epidemiology, medical geography and geographical epidemiology are all effectively synonymous terms for the study of the geographical distribution of disease spread or population at risk. A closely related research field, landscape epidemiology, studies the patterns, processes, and risk factors of diseases across time and space. The emergence and spread of infectious diseases in a changing environment
require the development of new methodologies and tools. As such, disease dynamics models on geographic scales ranging from village to continental levels are increasingly needed for research, surveillance, and control programs of vector-borne diseases (VBDs).

I have developed spatial extensions for the malaria ABM (as described in Section 2.3) [11, 12], and a landscape epidemiology modeling framework which incorporates the ABM with a geographic information system (GIS) (in collaboration with colleagues) [13, 14]. Three spatial output indices and five scenarios are created that represent two coverage levels of the two mosquito control interventions being modeled. Using spatial statistics tools, hot spot analysis is performed for all scenarios and two output indices in order to determine the statistical significance of the simulation results. The modeling framework can serve as a valuable tool for simulation modelers, epidemiologists, disease control managers and public health officials by assisting in refining research questions and surveillance needs, and in guiding control efforts and field studies. Although the landscape epidemiology modeling framework utilizes a malaria ABM, it is applicable to a wider range of other infectious VBDs including dengue, yellow fever, etc.

2.5 Verification, Validation, and Replication of Models

Verification & validation (V&V), accreditation, quality assurance (QA), certification, replication & reproducibility (R&R), acceptability assessment and uncertainty quantification are critical steps in M&S processes. In general, these activities primarily deal with the measurement and assessment of accuracy of ABMs. Verification involves transformational accuracy of the model artifacts in model development, in order to ensure that the implementation is a correct realization of the conceptual model. Validation, on the other hand, involves substantiation that a model within its domain of applicability possesses a satisfactory range of accuracy consistent with its intended application.

Replication and reproducibility (R&R) cover a wide spectrum of M&S issues. Replication is also known as model-to-model comparison, alignment, or cross-model validation. In most replication tasks, the primary goal of the modeler is to qualitatively (as opposed to quantitatively) replicate the results of previously published models. Replication is treated as the scientific gold standard to judge scientific claims. It allows independent researchers to address a scientific hypothesis and produce evidence for or against the hypothesis. Replication confirms reproducibility, which refers to the independent verification of prior findings, and is at the core of the spirit of science [15]. Reproducibility, as a fundamental principle of the scientific method, refers to the ability to independently replicate, reproduce and, if needed, extend previously published computational artefacts [16].

My V&V and R&R research is summarized in [17]. I performed a series of V&V experiments using ABMs (as described in Section 2.3) primarily using the V&V process of docking (in collaboration with colleagues) [18, 19]. The studies showed how docking helped in increasing confidence to the conceptual model, revealing conceptual and/or programming errors, and eliminating dubious assumptions from the models.

My R&R experiences involve replicating the results and extending some assumptions of several published malaria models (both mathematical and agent-based). Critical examination of some of these studies revealed that despite providing reasonably plausible results, the models adopted several dubious assumptions regarding, for example, the number of replicated simulation runs, and the boundary type of the landscapes. We emphasized the need for qualitative (as opposed to absolute or quantitative) matches between the respective models, and showed that performing sufficient number of replicated runs is important for model validation [20].

www3.nd.edu/~sarifin/
2.6 Data Warehouses & Dimensional Modeling

Epidemiological data management shares many common challenges typically faced in generic biological data management tasks. Traditional relational database management systems (RDBMS) cannot always cope with the complex requirements of advanced data analysis and large volume data aggregation. To overcome some of these challenges, data warehouses can offer a viable alternative. A data warehouse (DW) is a central repository of data created by integrating one or more disparate data sources. It is usually defined as a collection of subject-oriented, integrated, nonvolatile, and time-varying data to support management decisions. DWs use a special design technique known as dimensional modeling (DM), which views data as consisting of facts linked to several dimensions of interest [21]. Being different from the entity-relationship modeling technique (which is usually used to model traditional RDBMSs), dimensional modeling is at the heart of on-line analytical processing (OLAP) systems. In addition to eliminating some of the challenges, a well-designed DW may offer several other advantages, including a better response to the increasing demands of its decision-making users.

I contributed to the design and implementation of a multi-dimensional OLAP prototype DW to store, access, and analyze malaria-related data from heterogeneous sources. The prototype DW has been integrated into the open-access cyberinfrastructure (CI) framework of the Vector-Borne Disease Network (VecNet) [22] as the VecNet Data Warehouse (VecNet-DW). Details about the VecNet-DW can be found in [23].

3 Future Research

In the future, I plan to continue my research in the field of agent-based modeling and simulation (ABMS), landscape epidemiology, and data management systems, with a shared focus on public health and bioinformatics challenges. I also plan to sustain and increase my active collaboration with the undergraduate research programs. Some of my research plans are highlighted below:

3.1 Modeling & Simulation (M&S) in Public Health

Building on my experiences with M&S, I plan to expand my work to model other infectious diseases (such as dengue and Ebola). As natural companions of the research areas described above, I also plan to continue my work in verification & validation (V&V) and replication & reproducibility (R&R) research.

Dengue is a mosquito-borne viral disease which causes an estimated 50 – 100 million annual infections in over 100 endemic countries, putting almost half of the world’s population at risk [24]. The incidence of dengue has increased 30-fold over the last 50 years. Dengue flourishes in urban poor areas, suburbs, and the countryside, but also affects more affluent neighborhoods in tropical and subtropical countries. Dengue is transmitted by several species of mosquito within the genus *Aedes*, principally *Aedes aegypti*, which breeds in water containers in and around homes. Thus, dengue is closely affected by environmental factors linked to human behavior. An ABM of dengue may include, for example, an *epidemiological transmission component* which models the vector density, dispersal, biting, infection, etc., and a *vector control component* which models interventions like larval control, insecticide spraying, biological control, etc. In addition, a *vaccination strategies component* may be developed to explore the effectiveness and impact of vaccine against all four dengue serotypes.

A major focus of recent M&S research in public health is on building simulation tools to model and predict the emergence and spread of existing or new infectious disease outbreaks. For example, models can be valuable tools in analyzing the 2014 *Ebola epidemic*, which is the largest in history, and has affected multiple countries in Africa, including the Democratic Republic of the Congo, Guinea, Liberia, and Sierra Leone [25]. Ebola is a disease of humans and other primates caused by viruses. The disease has a high risk of death, killing between
25–90% percent of those who are infected [26]. The virus spreads by direct contact with body fluids (e.g., blood) of an infected human or other animals, or through contact with an item recently contaminated with bodily fluids.

### 3.2 Landscape Epidemiology Modeling

In my view, despite the recent proliferation of spatial disease risk models, geospatial data is still underutilized, and there is still a lack of and hence a serious need to develop efficient and useful tools to integrate spatial components for research, surveillance, and control programs of infectious diseases. As a logical extension of the landscape epidemiology modeling framework (see Section 2.4), I plan to integrate my future models of public health with high quality geospatial data covering geographic regions of communities, provinces, and countries, and to combine other expert knowledge bases (e.g., entomological, epidemiological, etc.) with them. Considering new infectious disease models and their heterogeneous geographical scales of transmission, the possibilities seem almost endless in this field of research.

### 3.3 Data Management Systems

The complex nature of infectious diseases and the integration of related data from heterogeneous, autonomous, distributed, and fast-changing data sources make the task of data management uniquely challenging. Instead of highly centralized and specialized traditional relational databases, decentralized, open-access computing and storage environments are gaining popularity in recent times.

Building on my past and current experiences in working with the open-access cyberinfrastructure framework of the Vector-Borne Disease Network (VecNet) [22], I plan to continue research on other data management systems and web-based user tools for online collaboration, storage, analysis, and retrieval of related data, possibly for hosting a multitude of models.

Given the highly interdisciplinary nature of my research, I plan to establish my own research program, to invite collaborations from colleagues in other academic departments across and outside the university, and to engage students from both undergraduate and graduate levels in my projects.

References


www3.nd.edu/~sarifin/


www3.nd.edu/~sarifin/


